PART I

STRUCTURE AND CATALYSIS

Biochemistry is nothing less than the chemistry of life, and, yes, life can be investigated, analyzed, and understood. To begin, every student of biochemistry needs both a language and some fundamentals; these are provided in Part I.

The chapters of Part I are devoted to the structure and function of the major classes of cellular constituents: water (Chapter 2), amino acids and proteins (Chapters 3 through 6), sugars and polysaccharides (Chapter 7), nucleotides and nucleic acids (Chapter 8), fatty acids and lipids (Chapter 10), and, finally, membranes and membrane signaling proteins (Chapters 11 and 12). We also discuss, in the context of structure and function, the technologies used to study each class of biomolecules. One whole chapter (Chapter 9) is devoted entirely to biotechnologies associated with cloning and genomics.

We begin, in Chapter 2, with water, because its properties affect the structure and function of all other cellular constituents. For each class of organic molecules, we first consider the covalent chemistry of the monomeric units (amino acids, monosaccharides, nucleotides, and fatty acids) and then describe the structure of the macromolecules and supramolecular complexes derived from them. An overarching theme is that the polymeric macromolecules in living systems, though large, are highly ordered chemical entities, with specific sequences of monomeric subunits giving rise to discrete structures and functions. This fundamental theme can be broken down into three interrelated principles: (1) the unique structure of each macromolecule determines its function; (2) noncovalent interactions play a critical role in the structure and thus the function of macromolecules; and (3) the monomeric subunits in polymeric macromolecules occur in specific sequences, representing a form of information on which the ordered living state depends.

The relationship between structure and function is especially evident in proteins, which exhibit an extraordinary diversity of functions. One particular polymeric sequence of amino acids produces a strong, fibrous structure found in hair and wool; another produces a protein that transports oxygen in the blood; a third binds other proteins and catalyzes the cleavage of the bonds between their amino acids. Similarly, the special functions of polysaccharides, nucleic acids, and lipids can be understood as resulting directly from their chemical structure, with their characteristic monomeric subunits precisely linked to form functional polymers.

Sugars linked together become energy stores, structural fibers, and points of specific molecular recognition; nucleotides strung together in DNA or RNA provide the blueprint for an entire organism; and aggregated lipids form membranes. Chapter 12 unifies the discussion of biomolecule function, describing how specific signaling systems regulate the activities of biomolecules—within a cell, within an organ, and among organs—to keep an organism in homeostasis.

As we move from monomeric units to larger and larger polymers, the chemical focus shifts from covalent bonds to noncovalent interactions. Covalent bonds, at the monomeric and macromolecular level, place constraints on the shapes assumed by large biomolecules. It is the numerous noncovalent interactions, however, that dictate the stable, native conformations of large molecules while permitting the flexibility necessary for their biological function. As we shall see, noncovalent interactions are essential to the catalytic power of enzymes, the critical interaction of complementary base pairs in nucleic acids, and the arrangement and properties of lipids in membranes. The principle that sequences of monomeric subunits are rich in information emerges most fully in the discussion of nucleic acids (Chapter 8). However, proteins and some short polymers of sugars (oligosaccharides) are also information-rich molecules. The amino acid sequence is a form of information that directs the folding of the protein into its unique three-dimensional structure and ultimately determines the function of the protein. Some oligosaccharides also have unique sequences and three-dimensional structures that are recognized by other macromolecules.

Each class of molecules has a similar structural hierarchy: subunits of fixed structure are connected by bonds of limited flexibility to form macromolecules with three-dimensional structures determined by noncovalent interactions. These macromolecules then interact to form the supramolecular structures and organelles that allow a cell to carry out its many metabolic functions. Together, the molecules described in Part I are the stuff of life.

2

Water

Water is the most abundant substance in living systems, making up 70% or more of the weight of most organisms. The first living organisms on Earth doubtless arose in an aqueous environment, and the course of evolution has been shaped by the properties of the aqueous medium in which life began.

This chapter begins with descriptions of the physical and chemical properties of water, to which all aspects of cell structure and function are adapted. The attractive forces between water molecules and the slight tendency of water to ionize are of crucial importance to the structure and function of biomolecules. We review the topic of ionization in terms of equilibrium constants, pH, and titration curves, and consider how aqueous solutions of weak acids or bases and their salts act as buffers against pH changes in biological systems. The water molecule and its ionization products, H^+ and OH^-, profoundly influence the structure, self-assembly, and properties of all cellular components, including proteins, nucleic acids, and lipids. The noncovalent interactions responsible for the strength and specificity of "recognition" among biomolecules are decisively influenced by water's properties as a solvent, including its ability to form hydrogen bonds with itself and with solutes.

2.1 Weak Interactions in Aqueous Systems

Hydrogen bonds between water molecules provide the cohesive forces that make water a liquid at room temperature and a crystalline solid (ice) with a highly ordered arrangement of molecules at cold temperatures. Polar biomolecules dissolve readily in water because they can replace water-water interactions with more energetically favorable water-solute interactions. In contrast, nonpolar biomolecules are poorly soluble in water because they interfere with water-water interactions but are unable to form water-solute interactions. In aqueous solutions, nonpolar molecules tend to cluster together. Hydrogen bonds and ionic, hydrophobic (Greek, "water-fearing"), and van der Waals interactions are individually weak, but collectively they have a very significant influence on the three-dimensional structures of proteins, nucleic acids, polysaccharides, and membrane lipids.

Hydrogen Bonding Gives Water Its Unusual Properties

Water has a higher melting point, boiling point, and heat of vaporization than most other common solvents (Table 2–1). These unusual properties are a consequence of attractions between adjacent water molecules that give liquid water great internal cohesion. A look at the electron structure of the H_2O molecule reveals the cause of these intermolecular attractions.

Each hydrogen atom of a water molecule shares an electron pair with the central oxygen atom. The geometry of the molecule is dictated by the shapes of the outer electron orbitals of the oxygen atom, which are similar to the sp^3 bonding orbitals of carbon (see Fig. 1–15). These orbitals describe a rough tetrahedron, with a hydrogen atom at each of two corners and unshared electron pairs at the other two corners **(Fig. 2–1a)**. The H—O—H bond angle is 104.5°, slightly less than the 109.5° of a perfect tetrahedron because of crowding by the nonbonding orbitals of the oxygen atom.

The oxygen nucleus attracts electrons more strongly than does the hydrogen nucleus (a proton); that is, oxygen is more electronegative. This means that the shared electrons are more often in the vicinity of the oxygen atom than of the hydrogen. The result of this unequal electron sharing is two electric dipoles in the water molecule, one along each of the H—O bonds;

TABLE 2–1 Melting Point, Boiling Point, and Heat of Vaporization of Some Common Solvents

	Melting point (°C)	Boiling point (°C)	Heat of vaporization (J/g)*
Water	0	100	2,260
Methanol (CH_3OH)	−98	65	1,100
Ethanol (CH_3CH_2OH)	−117	78	854
Propanol ($CH_3CH_2CH_2OH$)	−127	97	687
Butanol ($CH_3(CH_2)_2CH_2OH$)	−90	117	590
Acetone (CH_3COCH_3)	−95	56	523
Hexane ($CH_3(CH_2)_4CH_3$)	−98	69	423
Benzene (C_6H_6)	6	80	394
Butane ($CH_3(CH_2)_2CH_3$)	−135	−0.5	381
Chloroform ($CHCl_3$)	−63	61	247

*The heat energy required to convert 1.0 g of a liquid at its boiling point and at atmospheric pressure into its gaseous state at the same temperature. It is a direct measure of the energy required to overcome attractive forces between molecules in the liquid phase.

each hydrogen atom bears a partial positive charge ($\delta+$), and the oxygen atom bears a partial negative charge equal in magnitude to the sum of the two partial positives ($2\delta-$). As a result, there is an electrostatic attraction between the oxygen atom of one water molecule and the hydrogen of another (Fig. 2–1b), called a **hydrogen bond**. Throughout this book, we represent hydrogen bonds with three parallel blue lines, as in Figure 2–1b.

Hydrogen bonds are relatively weak. Those in liquid water have a **bond dissociation energy** (the energy required to break a bond) of about 23 kJ/mol, compared with 470 kJ/mol for the covalent O—H bond in water or 348 kJ/mol for a covalent C—C bond. The hydrogen bond is about 10% covalent, due to overlaps in the bonding orbitals, and about 90% electrostatic. At room temperature, the thermal energy of an aqueous solution (the kinetic energy of motion of the individual atoms and molecules) is of the same order of magnitude as that required to break hydrogen bonds. When water is heated, the increase in temperature reflects the faster motion of individual water molecules. At any given time, most of the molecules in liquid water are hydrogen bonded, but the lifetime of each hydrogen bond is just 1 to 20 picoseconds ($1\ ps = 10^{-12}$ s); when one hydrogen bond breaks, another hydrogen bond forms, with the same partner or a new one, within 0.1 ps. The apt phrase "flickering clusters" has been applied to the short-lived groups of water molecules interlinked by hydrogen bonds in liquid water. The sum of all the hydrogen bonds between H_2O molecules confers great internal cohesion on liquid water. Extended networks of hydrogen-bonded water molecules also form bridges between solutes (proteins and nucleic acids, for example) that allow the larger molecules to interact with each other over distances of several nanometers without physically touching.

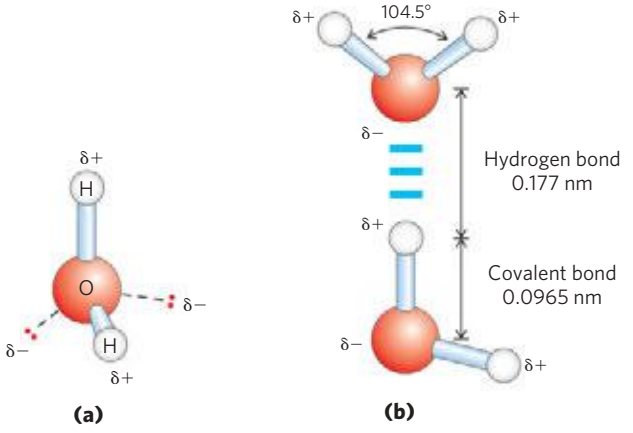


FIGURE 2–1 Structure of the water molecule. (a) The dipolar nature of the H_2O molecule is shown in a ball-and-stick model; the dashed lines represent the nonbonding orbitals. There is a nearly tetrahedral arrangement of the outer-shell electron pairs around the oxygen atom; the two hydrogen atoms have localized partial positive charges ($\delta+$) and the oxygen atom has a partial negative charge ($\delta-$). **(b)** Two H_2O molecules joined by a hydrogen bond (designated here, and throughout this book, by three blue lines) between the oxygen atom of the upper molecule and a hydrogen atom of the lower one. Hydrogen bonds are longer and weaker than covalent O—H bonds.

The nearly tetrahedral arrangement of the orbitals about the oxygen atom (Fig. 2–1a) allows each water molecule to form hydrogen bonds with as many as four neighboring water molecules. In liquid water at room temperature and atmospheric pressure, however, water molecules are disorganized and in continuous motion, so that each molecule forms hydrogen bonds with an average of only 3.4 other molecules. In ice, on the other hand, each water molecule is fixed in space and forms hydrogen bonds with a full complement of four other water molecules to yield a regular lattice structure **(Fig. 2–2)**. Hydrogen bonds account for the relatively high melting point of water, because much thermal energy is required to break a sufficient proportion of hydrogen bonds to destabilize the crystal lattice of ice

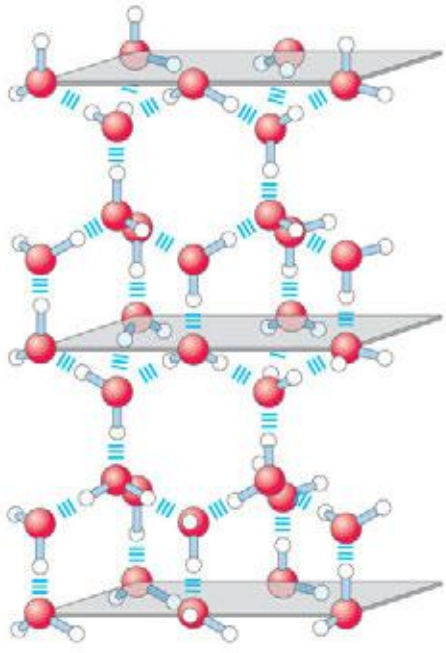


FIGURE 2–2 Hydrogen bonding in ice. In ice, each water molecule forms four hydrogen bonds, the maximum possible for a water molecule, creating a regular crystal lattice. By contrast, in liquid water at room temperature and atmospheric pressure, each water molecule hydrogen-bonds with an average of 3.4 other water molecules. This crystal lattice structure makes ice less dense than liquid water, and thus ice floats on liquid water.

(Table 2–1). When ice melts or water evaporates, heat is taken up by the system:

$$H_2O(\text{solid}) \longrightarrow H_2O(\text{liquid}) \qquad \Delta H = +5.9 \text{ kJ/mol}$$
$$H_2O(\text{liquid}) \longrightarrow H_2O(\text{gas}) \qquad \Delta H = +44.0 \text{ kJ/mol}$$

During melting or evaporation, the entropy of the aqueous system increases as the highly ordered arrays of water molecules in ice relax into the less orderly hydrogen-bonded arrays in liquid water or into the wholly disordered gaseous state. At room temperature, both the melting of ice and the evaporation of water occur spontaneously; the tendency of the water molecules to associate through hydrogen bonds is outweighed by the energetic push toward randomness. Recall that the free-energy change (ΔG) must have a negative value for a process to occur spontaneously: $\Delta G = \Delta H - T\Delta S$, where ΔG represents the driving force, ΔH the enthalpy change from making and breaking bonds, and ΔS the change in randomness. Because ΔH is positive for melting and evaporation, it is clearly the increase in entropy (ΔS) that makes ΔG negative and drives these changes.

Water Forms Hydrogen Bonds with Polar Solutes

Hydrogen bonds are not unique to water. They readily form between an electronegative atom (the hydrogen acceptor, usually oxygen or nitrogen) and a hydrogen atom covalently bonded to another electronegative atom (the hydrogen donor) in the same or another molecule **(Fig. 2–3)**. Hydrogen atoms covalently bonded to carbon atoms do not participate in hydrogen bonding, because carbon is only slightly more electronegative than hydrogen and thus the C—H bond is only very weakly polar. The distinction explains why butanol ($CH_3(CH_2)_2CH_2OH$) has a relatively high boiling point of 117 °C, whereas butane ($CH_3(CH_2)_2CH_3$) has a boiling point of only −0.5 °C. Butanol has a polar hydroxyl group and thus can form intermolecular hydrogen bonds. Uncharged but polar biomolecules such as sugars dissolve readily in water because of the stabilizing effect of hydrogen bonds between the hydroxyl groups or carbonyl oxygen of the sugar and the polar water molecules. Alcohols, aldehydes, ketones, and compounds containing N—H bonds all form hydrogen bonds with water molecules **(Fig. 2–4)** and tend to be soluble in water.

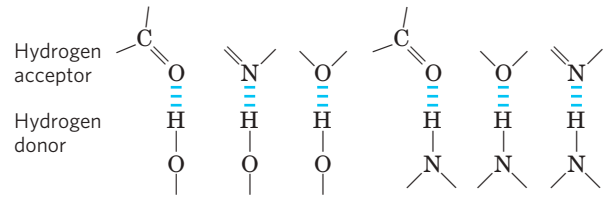


FIGURE 2–3 Common hydrogen bonds in biological systems. The hydrogen acceptor is usually oxygen or nitrogen; the hydrogen donor is another electronegative atom.

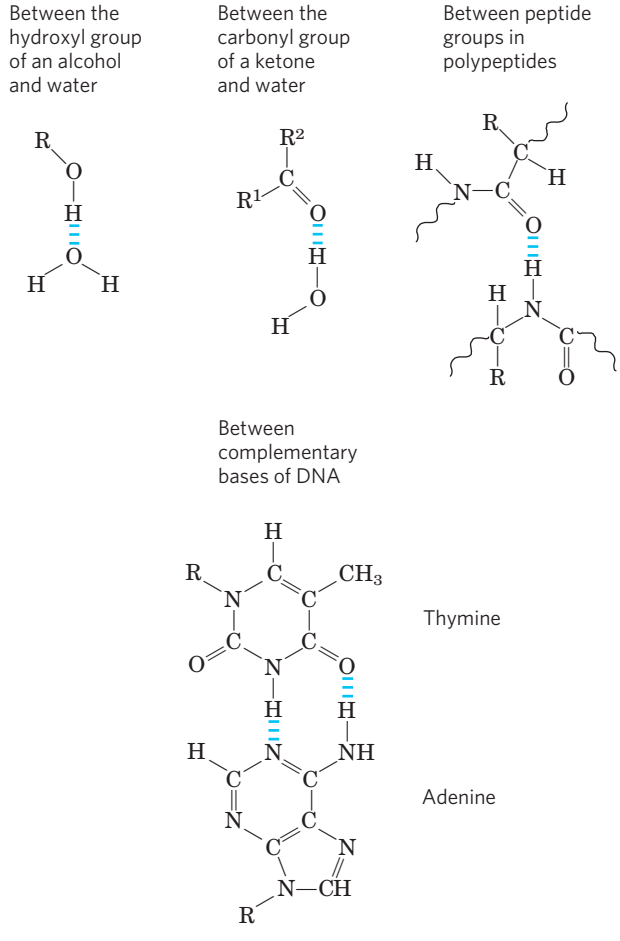


FIGURE 2–4 Some biologically important hydrogen bonds.

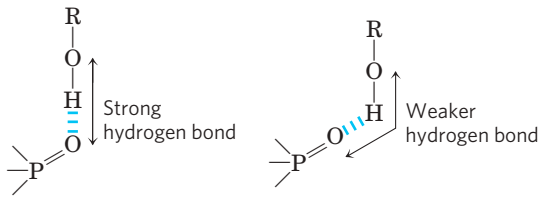


FIGURE 2–5 Directionality of the hydrogen bond. The attraction between the partial electric charges (see Fig. 2–1) is greatest when the three atoms involved in the bond (in this case O, H, and O) lie in a straight line. When the hydrogen-bonded moieties are structurally constrained (when they are parts of a single protein molecule, for example), this ideal geometry may not be possible and the resulting hydrogen bond is weaker.

Hydrogen bonds are strongest when the bonded molecules are oriented to maximize electrostatic interaction, which occurs when the hydrogen atom and the two atoms that share it are in a straight line—that is, when the acceptor atom is in line with the covalent bond between the donor atom and H **(Fig. 2–5)**. This arrangement puts the positive charge of the hydrogen ion directly between the two partial negative charges. Hydrogen bonds are thus highly directional and capable of holding two hydrogen-bonded molecules or groups in a specific geometric arrangement. As we shall see later, this property of hydrogen bonds confers very precise three-dimensional structures on protein and nucleic acid molecules, which have many intramolecular hydrogen bonds.

Water Interacts Electrostatically with Charged Solutes

Water is a polar solvent. It readily dissolves most biomolecules, which are generally charged or polar compounds (Table 2–2); compounds that dissolve easily in water are **hydrophilic** (Greek, "water-loving"). In contrast, nonpolar solvents such as chloroform and benzene are poor solvents for polar biomolecules but easily dissolve those that are **hydrophobic**—nonpolar molecules such as lipids and waxes.

Water dissolves salts such as NaCl by hydrating and stabilizing the Na^+ and Cl^- ions, weakening the electrostatic interactions between them and thus counteracting their tendency to associate in a crystalline lattice **(Fig. 2–6)**. Water also readily dissolves charged biomolecules, including compounds with functional groups such as ionized carboxylic acids (—COO^-), protonated amines ($-NH_3^+$), and phosphate esters or anhydrides. Water replaces the solute-solute hydrogen bonds linking these biomolecules to each other with solute-water hydrogen bonds, thus screening the electrostatic interactions between solute molecules.

Water is effective in screening the electrostatic interactions between dissolved ions because it has a high dielectric constant, a physical property that reflects the number of dipoles in a solvent. The strength, or force (F), of ionic interactions in a solution depends on the magnitude of the charges (Q), the distance between the charged groups (r), and the dielectric constant (ε, which is dimensionless) of the solvent in which the interactions occur:

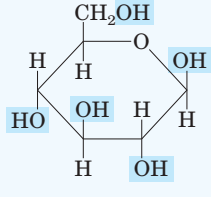
$$F = \frac{Q_1 Q_2}{\varepsilon r^2}$$

For water at 25 °C, ε is 78.5, and for the very nonpolar solvent benzene, ε is 4.6. Thus, ionic interactions between dissolved ions are much stronger in less polar environments. The dependence on r^2 is such that ionic attractions or repulsions operate only over short distances—in the range of 10 to 40 nm (depending on the electrolyte concentration) when the solvent is water.

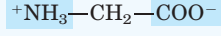
TABLE 2–2 Some Examples of Polar, Nonpolar, and Amphipathic Biomolecules (Shown as Ionic Forms at pH 7)

Polar

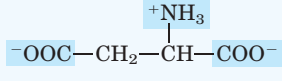
Glucose



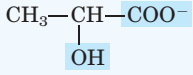
Glycine



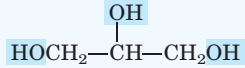
Aspartate



Lactate

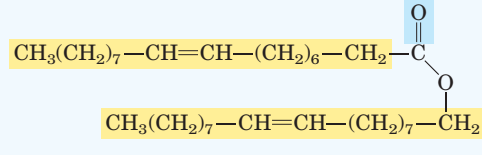


Glycerol



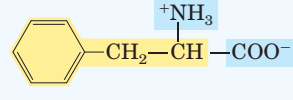
Nonpolar

Typical wax

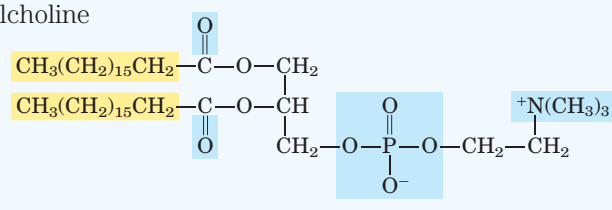


Amphipathic

Phenylalanine



Phosphatidylcholine



Polar groups Nonpolar groups

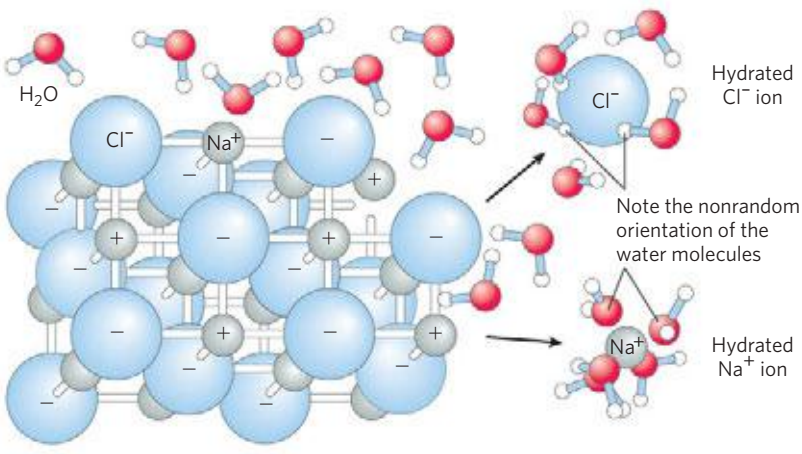


FIGURE 2–6 Water as solvent. Water dissolves many crystalline salts by hydrating their component ions. The NaCl crystal lattice is disrupted as water molecules cluster about the Cl^- and Na^+ ions. The ionic charges are partially neutralized, and the electrostatic attractions necessary for lattice formation are weakened.

Entropy Increases as Crystalline Substances Dissolve

As a salt such as NaCl dissolves, the Na^+ and Cl^- ions leaving the crystal lattice acquire far greater freedom of motion (Fig. 2–6). The resulting increase in entropy (randomness) of the system is largely responsible for the ease of dissolving salts such as NaCl in water. In thermodynamic terms, formation of the solution occurs with a favorable free-energy change: $\Delta G = \Delta H - T\Delta S$, where ΔH has a small positive value and $T\Delta S$ a large positive value; thus ΔG is negative.

Nonpolar Gases Are Poorly Soluble in Water

The molecules of the biologically important gases CO_2, O_2, and N_2 are nonpolar. In O_2 and N_2, electrons are shared equally by both atoms. In CO_2, each C═O bond is polar, but the two dipoles are oppositely directed and cancel each other (Table 2–3). The movement of molecules from the disordered gas phase into aqueous solution constrains their motion and the motion of water molecules and therefore represents a decrease in entropy. The nonpolar nature of these gases and the decrease in entropy when they enter solution combine to make them very poorly soluble in water (Table 2–3). Some organisms have water-soluble "carrier proteins" (hemoglobin and myoglobin, for example) that facilitate the transport of O_2. Carbon dioxide forms carbonic acid (H_2CO_3) in aqueous solution and is transported as the HCO_3^- (bicarbonate) ion, either free—bicarbonate is very soluble in water (~100 g/L at 25 °C)—or bound to hemoglobin. Three other gases, NH_3, NO, and H_2S, also have biological roles in some organisms; these gases are polar, dissolve readily in water, and ionize in aqueous solution.

Nonpolar Compounds Force Energetically Unfavorable Changes in the Structure of Water

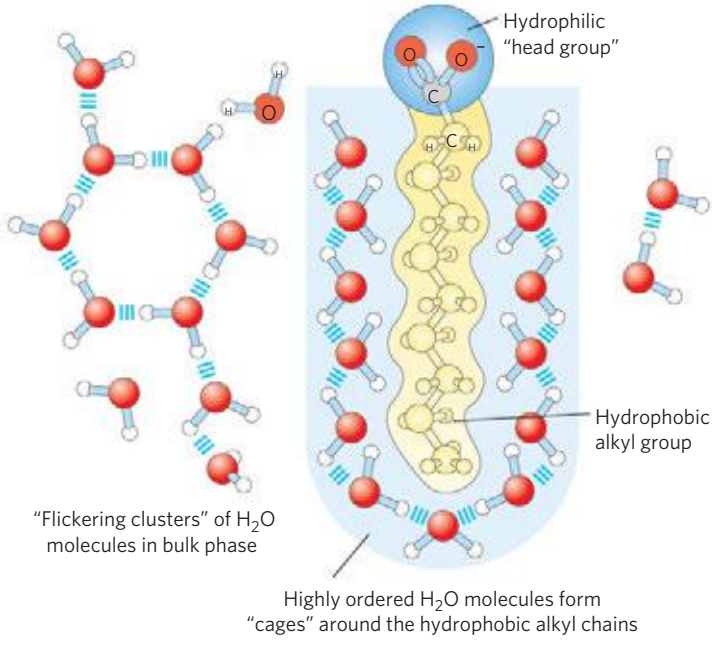
When water is mixed with benzene or hexane, two phases form; neither liquid is soluble in the other. Nonpolar compounds such as benzene and hexane are hydrophobic—they are unable to undergo energetically favorable interactions with water molecules, and they interfere with the hydrogen bonding among water molecules. All molecules or ions in aqueous solution interfere with the hydrogen bonding of some water

TABLE 2–3 Solubilities of Some Gases in Water

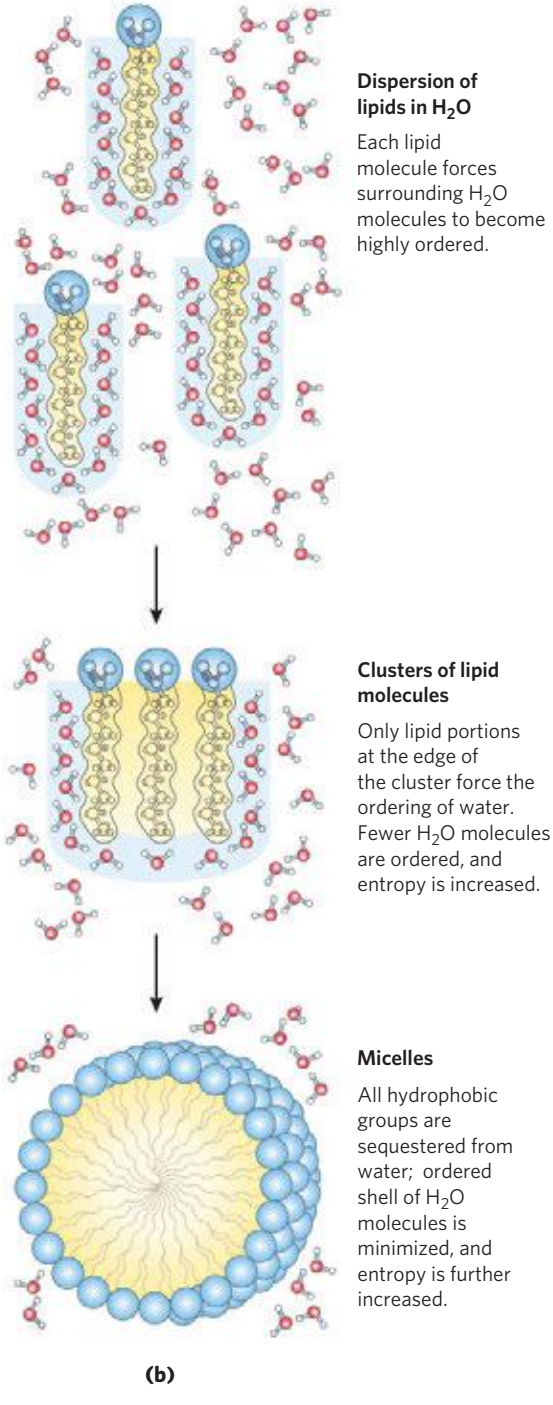
Gas	Structure*	Polarity	Solubility in water (g/L)†
Nitrogen	N≡N	Nonpolar	0.018 (40 °C)
Oxygen	O═O	Nonpolar	0.035 (50 °C)
Carbon dioxide	δ− ← O═C═O → δ−	Nonpolar	0.97 (45 °C)
Ammonia	H, H, H, N, δ−	Polar	900 (10 °C)
Hydrogen sulfide	H, H, S, δ−	Polar	1,860 (40 °C)

*The arrows represent electric dipoles; there is a partial negative charge (δ−) at the head of the arrow, a partial positive charge (δ+ ; not shown here) at the tail.

†Note that polar molecules dissolve far better even at low temperatures than do nonpolar molecules at relatively high temperatures.



(a)



(b)

FIGURE 2–7 Amphipathic compounds in aqueous solution. (a) Long-chain fatty acids have very hydrophobic alkyl chains, each of which is surrounded by a layer of highly ordered water molecules. **(b)** By clustering together in micelles, the fatty acid molecules expose the smallest possible hydrophobic surface area to the water, and fewer water molecules are required in the shell of ordered water. The energy gained by freeing immobilized water molecules stabilizes the micelle.

molecules in their immediate vicinity, but polar or charged solutes (such as NaCl) compensate for lost water-water hydrogen bonds by forming new solute-water interactions. The net change in enthalpy (ΔH) for dissolving these solutes is generally small. Hydrophobic solutes, however, offer no such compensation, and their addition to water may therefore result in a small gain of enthalpy; the breaking of hydrogen bonds between water molecules takes up energy from the system, requiring the input of energy from the surroundings. In addition to requiring this input of energy, dissolving hydrophobic compounds in water produces a measurable decrease in entropy. Water molecules in the immediate vicinity of a nonpolar solute are constrained in their possible orientations as they form a highly ordered cagelike shell around each solute molecule. These water molecules are not as highly oriented as those in **clathrates**, crystalline compounds of nonpolar solutes and water, but the effect is the same in both cases: the ordering of water molecules reduces entropy. The number of ordered water molecules, and therefore the magnitude of the entropy decrease, is proportional to the surface area of the hydrophobic solute enclosed within the cage of water molecules. The free-energy change for dissolving a nonpolar solute in water is thus unfavorable: $\Delta G = \Delta H - T\Delta S$, where ΔH has a positive value, ΔS has a negative value, and ΔG is positive.

Amphipathic compounds contain regions that are polar (or charged) and regions that are nonpolar (Table 2–2). When an amphipathic compound is mixed with water, the polar, hydrophilic region interacts favorably with the water and tends to dissolve, but the nonpolar, hydrophobic region tends to avoid contact with the water **(Fig. 2–7a)**. The nonpolar regions of the molecules cluster together to present the smallest hydrophobic area to the aqueous solvent, and the polar regions are arranged to maximize their interaction with the solvent (Fig. 2–7b). These stable structures of amphipathic compounds in water, called **micelles**, may contain hundreds or thousands of molecules. The forces that

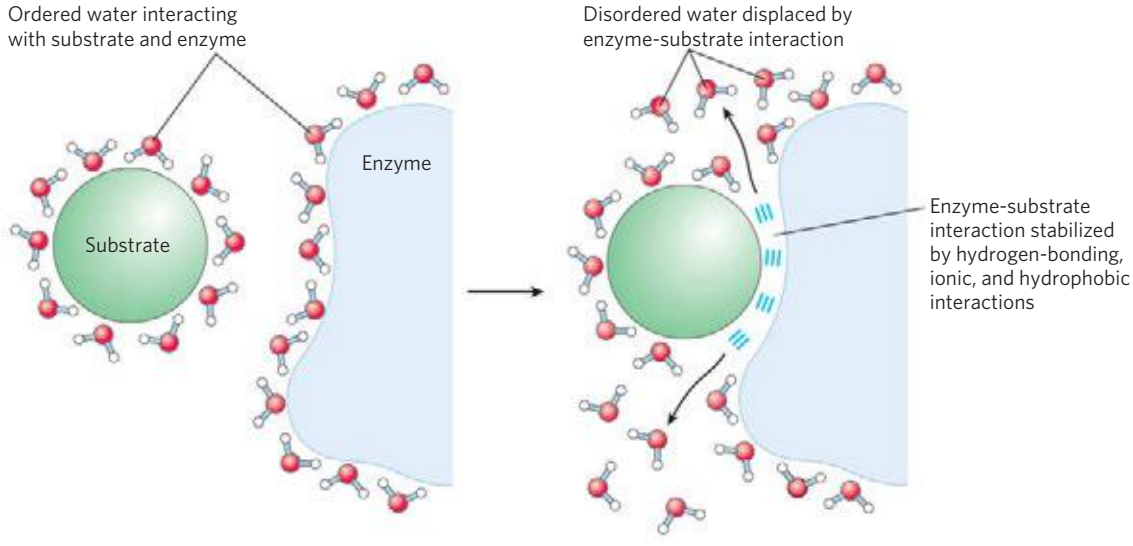


FIGURE 2–8 Release of ordered water favors formation of an enzyme-substrate complex. While separate, both enzyme and substrate force neighboring water molecules into an ordered shell. Binding of substrate to enzyme releases some of the ordered water, and the resulting increase in entropy provides a thermodynamic push toward formation of the enzyme-substrate complex (see p. 198).

hold the nonpolar regions of the molecules together are called **hydrophobic interactions.** The strength of hydrophobic interactions is not due to any intrinsic attraction between nonpolar moieties. Rather, it results from the system's achieving the greatest thermodynamic stability by minimizing the number of ordered water molecules required to surround hydrophobic portions of the solute molecules.

Many biomolecules are amphipathic; proteins, pigments, certain vitamins, and the sterols and phospholipids of membranes all have both polar and nonpolar surface regions. Structures composed of these molecules are stabilized by hydrophobic interactions among the nonpolar regions. Hydrophobic interactions among lipids, and between lipids and proteins, are the most important determinants of structure in biological membranes. Hydrophobic interactions between nonpolar amino acids also stabilize the three-dimensional structures of proteins.

Hydrogen bonding between water and polar solutes also causes an ordering of water molecules, but the energetic effect is less significant than with nonpolar solutes. Disruption of ordered water molecules is part of the driving force for binding of a polar substrate (reactant) to the complementary polar surface of an enzyme: entropy increases as the enzyme displaces ordered water from the substrate, and as the substrate displaces ordered water from the enzyme surface **(Fig. 2–8)**.

van der Waals Interactions Are Weak Interatomic Attractions

When two uncharged atoms are brought very close together, their surrounding electron clouds influence each other. Random variations in the positions of the electrons around one nucleus may create a transient electric dipole, which induces a transient, opposite electric dipole in the nearby atom. The two dipoles weakly attract each other, bringing the two nuclei closer. These weak attractions are called **van der Waals interactions** (also known as London forces). As the two nuclei draw closer together, their electron clouds begin to repel each other. At the point where the net attraction is maximal, the nuclei are said to be in van der Waals contact. Each atom has a characteristic **van der Waals radius**, a measure of how close that atom will allow another to approach (Table 2–4). In the "space-filling" molecular models shown throughout this book, the atoms are depicted in sizes proportional to their van der Waals radii.

TABLE 2–4 van der Waals Radii and Covalent (Single-Bond) Radii of Some Elements

Element	van der Waals radius (nm)	Covalent radius for single bond (nm)
H	0.11	0.030
O	0.15	0.066
N	0.15	0.070
C	0.17	0.077
S	0.18	0.104
P	0.19	0.110
I	0.21	0.133

Sources: For van der Waals radii, Chauvin, R. (1992) Explicit periodic trend of van der Waals radii. *J. Phys. Chem.* 96, 9194–9197. For covalent radii, Pauling, L. (1960) *Nature of the Chemical Bond,* 3rd edn, Cornell University Press, Ithaca, NY.

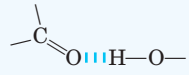
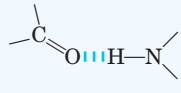
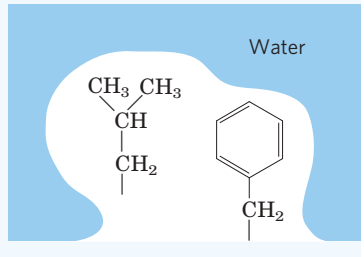
Note: van der Waals radii describe the space-filling dimensions of atoms. When two atoms are joined covalently, the atomic radii at the point of bonding are less than the van der Waals radii, because the joined atoms are pulled together by the shared electron pair. The distance between nuclei in a van der Waals interaction or a covalent bond is about equal to the sum of the van der Waals or covalent radii, respectively, for the two atoms. Thus the length of a carbon–carbon single bond is about 0.077 nm + 0.077 nm = 0.154 nm.

Weak Interactions Are Crucial to Macromolecular Structure and Function

> I believe that as the methods of structural chemistry are further applied to physiological problems, it will be found that the significance of the hydrogen bond for physiology is greater than that of any other single structural feature.
>
> —*Linus Pauling,*
> The Nature of the Chemical Bond, *1939*

The noncovalent interactions we have described—hydrogen bonds and ionic, hydrophobic, and van der Waals interactions (Table 2–5)—are much weaker than covalent bonds. An input of about 350 kJ of energy is required to break a mole of (6×10^{23}) C—C single bonds, and about 410 kJ to break a mole of C—H bonds, but as little as 4 kJ is sufficient to disrupt a mole of typical van der Waals interactions. Hydrophobic interactions are also much weaker than covalent bonds, although they are substantially strengthened by a highly polar solvent (a concentrated salt solution, for example). Ionic interactions and hydrogen bonds are variable in strength, depending on the polarity of the solvent and the alignment of the hydrogen-bonded atoms, but they are always significantly weaker than covalent bonds. In aqueous solvent at 25 °C, the available thermal energy can be of the same order of magnitude as the strength of these weak interactions, and the interaction between solute and solvent (water) molecules is nearly as favorable as solute-solute interactions. Consequently, hydrogen bonds and ionic, hydrophobic, and van der Waals interactions are continually forming and breaking.

TABLE 2–5 Four Types of Noncovalent ("Weak") Interactions among Biomolecules in Aqueous Solvent

Hydrogen bonds	
Between neutral groups	
Between peptide bonds	
Ionic interactions	
Attraction	
Repulsion	
Hydrophobic interactions	
van der Waals interactions	Any two atoms in close proximity

Although these four types of interactions are individually weak relative to covalent bonds, the cumulative effect of many such interactions can be very significant. For example, the noncovalent binding of an enzyme to its substrate may involve several hydrogen bonds and one or more ionic interactions, as well as hydrophobic and van der Waals interactions. The formation of each of these weak bonds contributes to a net decrease in the free energy of the system. We can calculate the stability of a noncovalent interaction, such as the hydrogen bonding of a small molecule to its macromolecular partner, from the binding energy, the reduction in the energy of the system when binding occurs. Stability, as measured by the equilibrium constant (see below) of the binding reaction, varies *exponentially* with binding energy. In order to dissociate two biomolecules (such as an enzyme and its bound substrate) that are associated noncovalently through multiple weak interactions, all these interactions must be disrupted at the same time. Because the interactions fluctuate randomly, such simultaneous disruptions are very unlikely. Therefore, 5 or 20 weak interactions bestow much greater molecular stability than would be expected intuitively from a simple summation of small binding energies.

Macromolecules such as proteins, DNA, and RNA contain so many sites of potential hydrogen bonding or ionic, van der Waals, or hydrophobic interactions that the cumulative effect of the many small binding forces can be enormous. For macromolecules, the most stable (that is, the native) structure is usually that in which weak interactions are maximized. The folding of a single polypeptide or polynucleotide chain into its three-dimensional shape is determined by this principle. The binding of an antigen to a specific antibody depends on the cumulative effects of many weak interactions. As noted earlier, the energy released when an enzyme binds noncovalently to its substrate is the main source of the enzyme's catalytic power. The binding of a hormone or a neurotransmitter to its cellular receptor protein is the result of multiple weak interactions. One consequence of the large size of enzymes and receptors (relative to their substrates or ligands) is that their extensive surfaces provide many opportunities for weak interactions. At the molecular level, the complementarity between interacting biomolecules reflects the complementarity and weak interactions between polar, charged, and hydrophobic groups on the surfaces of the molecules.

When the structure of a protein such as hemoglobin **(Fig. 2–9)** is determined by x-ray crystallography (see Box 4–5), water molecules are often found to be bound so tightly that they are part of the crystal structure; the same is true for water in crystals of RNA or DNA. These bound water molecules, which can also be detected in aqueous solutions by nuclear magnetic resonance, have distinctly different properties from

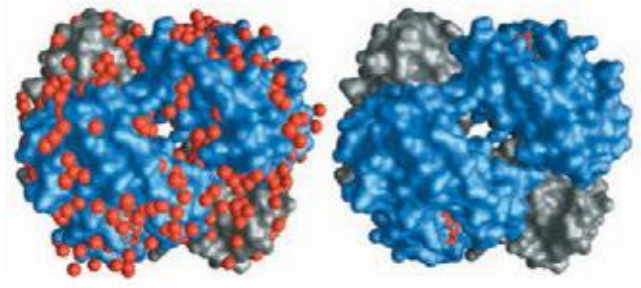


FIGURE 2–9 Water binding in hemoglobin. (PDB ID 1A3N) The crystal structure of hemoglobin, shown **(a)** with bound water molecules (red spheres) and **(b)** without the water molecules. The water molecules are so firmly bound to the protein that they affect the x-ray diffraction pattern as though they were fixed parts of the crystal. The two α subunits of hemoglobin are shown in gray, the two β subunits in blue. Each subunit has a bound heme group (red stick structure), visible only in the β subunits in this view. The structure and function of hemoglobin are discussed in detail in Chapter 5.

those of the "bulk" water of the solvent. They are, for example, not osmotically active (see below). For many proteins, tightly bound water molecules are essential to their function. In a key reaction in photosynthesis, for example, protons flow across a biological membrane as light drives the flow of electrons through a series of electron-carrying proteins (see Fig. 19–62). One of these proteins, cytochrome *f*, has a chain of five bound water molecules **(Fig. 2–10)** that may provide a path for protons to move through the membrane by a process known as "proton hopping" (described below). Another such light-driven proton pump, bacteriorhodopsin, almost certainly uses a chain of precisely oriented bound water molecules in the transmembrane movement of protons (see Fig. 19–69b). Tightly bound water molecules can also form an essential part of the binding site of a protein for its ligand. In a bacterial arabinose-binding protein, for example, five water molecules form hydrogen bonds that provide critical cross-links between the sugar (arabinose) and the amino acid residues in the sugar-binding site **(Fig. 2–11)**.

Solutes Affect the Colligative Properties of Aqueous Solutions

Solutes of all kinds alter certain physical properties of the solvent, water: its vapor pressure, boiling point, melting point (freezing point), and osmotic pressure. These are called **colligative properties** (colligative meaning "tied together"), because the effect of solutes on all four properties has the same basis: the concentration of water is lower in solutions than in pure water. The effect of solute concentration on the colligative properties of water is independent of the chemical properties of the solute; it depends only on the *number* of solute particles (molecules or ions) in a given amount of water. For example, a compound such as NaCl, which dissociates in solution, has an effect on osmotic pressure

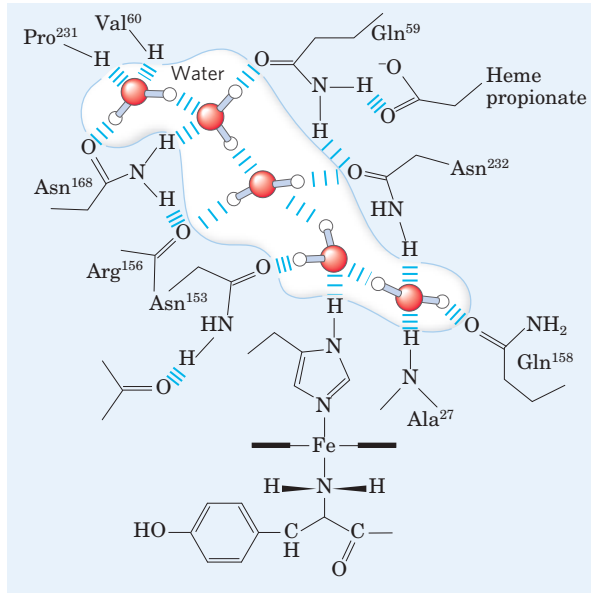


FIGURE 2–10 Water chain in cytochrome *f*. Water is bound in a proton channel of the membrane protein cytochrome *f*, which is part of the energy-trapping machinery of photosynthesis in chloroplasts (see Fig. 19–61). Five water molecules are hydrogen-bonded to each other and to functional groups of the protein: the peptide backbone atoms of valine, proline, arginine, and alanine residues, and the side chains of three asparagine and two glutamine residues. The protein has a bound heme (see Fig. 5–1), its iron ion facilitating electron flow during photosynthesis. Electron flow is coupled to the movement of protons across the membrane, which probably involves "proton hopping" (see Fig. 2–14) through this chain of bound water molecules.

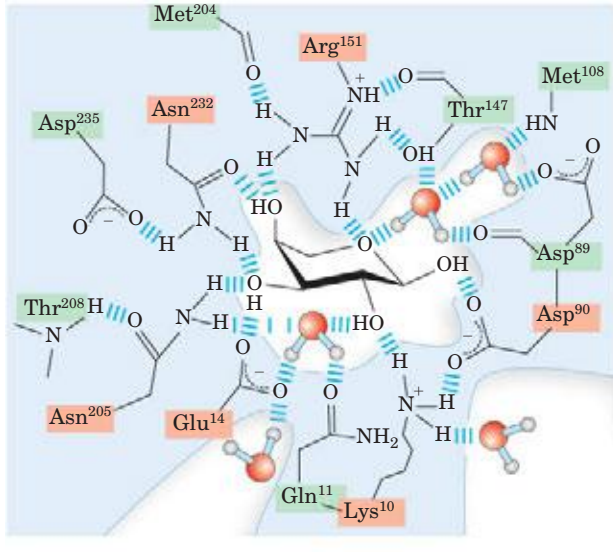


FIGURE 2–11 Hydrogen-bonded water as part of a protein's sugar-binding site. In the L-arabinose-binding protein of the bacterium *E. coli*, five water molecules are essential components of the hydrogen-bonded network of interactions between the sugar arabinose (center) and at least 13 amino acid residues in the sugar-binding site. Viewed in three dimensions, these interacting groups constitute two layers of binding moieties; amino acid residues in the first layer are screened in red, those in the second layer in green. Some of the hydrogen bonds are drawn longer than others for clarity; they are not actually longer than the others.

that is twice that of an equal number of moles of a nondissociating solute such as glucose.

Water molecules tend to move from a region of higher water concentration to one of lower water concentration, in accordance with the tendency in nature for a system to become disordered. When two different aqueous solutions are separated by a semipermeable membrane (one that allows the passage of water but not solute molecules), water molecules diffusing from the region of higher water concentration to the region of lower water concentration produce osmotic pressure **(Fig. 2–12)**. Osmotic pressure, Π, measured as the force necessary to resist water movement (Fig. 2–12c), is approximated by the van't Hoff equation:

$$\Pi = icRT$$

in which R is the gas constant and T is the absolute temperature. The symbol i is the van't Hoff factor, which is a measure of the extent to which the solute dissociates into two or more ionic species. The term ic is the **osmolarity** of the solution, the product of the van't Hoff factor i and the solute's molar concentration c. In dilute NaCl solutions, the solute completely dissociates into Na^+ and Cl^-, doubling the number of solute particles, and thus $i = 2$. For all nonionizing solutes, $i = 1$. For solutions of several (n) solutes, Π is the sum of the contributions of each species:

$$\Pi = RT(i_1c_1 + i_2c_2 + \cdots + i_nc_n)$$

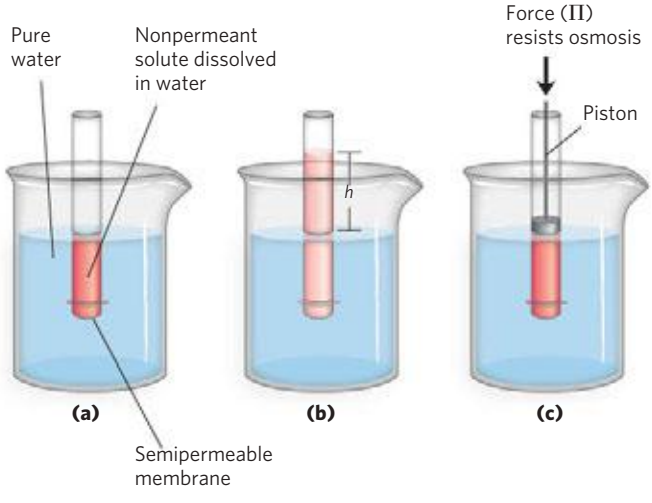


FIGURE 2–12 Osmosis and the measurement of osmotic pressure. **(a)** The initial state. The tube contains an aqueous solution, the beaker contains pure water, and the semipermeable membrane allows the passage of water but not solute. Water flows from the beaker into the tube to equalize its concentration across the membrane. **(b)** The final state. Water has moved into the solution of the nonpermeant compound, diluting it and raising the column of solution within the tube. At equilibrium, the force of gravity operating on the solution in the tube exactly balances the tendency of water to move into the tube, where its concentration is lower. **(c)** Osmotic pressure (Π) is measured as the force that must be applied to return the solution in the tube to the level of the water in the beaker. This force is proportional to the height, *h*, of the column in (b).

Osmosis, water movement across a semipermeable membrane driven by differences in osmotic pressure, is an important factor in the life of most cells. Plasma membranes are more permeable to water than to most other small molecules, ions, and macromolecules because protein channels (aquaporins; see Fig. 11–45) in the membrane selectively permit the passage of water. Solutions of osmolarity equal to that of a cell's cytosol are said to be **isotonic** relative to that cell. Surrounded by an isotonic solution, a cell neither gains nor loses water **(Fig. 2–13)**. In a **hypertonic** solution, one with higher osmolarity than that of the cytosol, the cell shrinks as water moves out. In a **hypotonic** solution, one with a lower osmolarity than the cytosol, the cell swells as water enters. In their natural environments, cells generally contain higher concentrations of biomolecules and ions than their surroundings, so osmotic pressure tends to drive water into cells. If not somehow counterbalanced, this inward movement of water would distend the plasma membrane and eventually cause bursting of the cell (osmotic lysis).

Several mechanisms have evolved to prevent this catastrophe. In bacteria and plants, the plasma membrane is surrounded by a nonexpandable cell wall of sufficient rigidity and strength to resist osmotic pressure

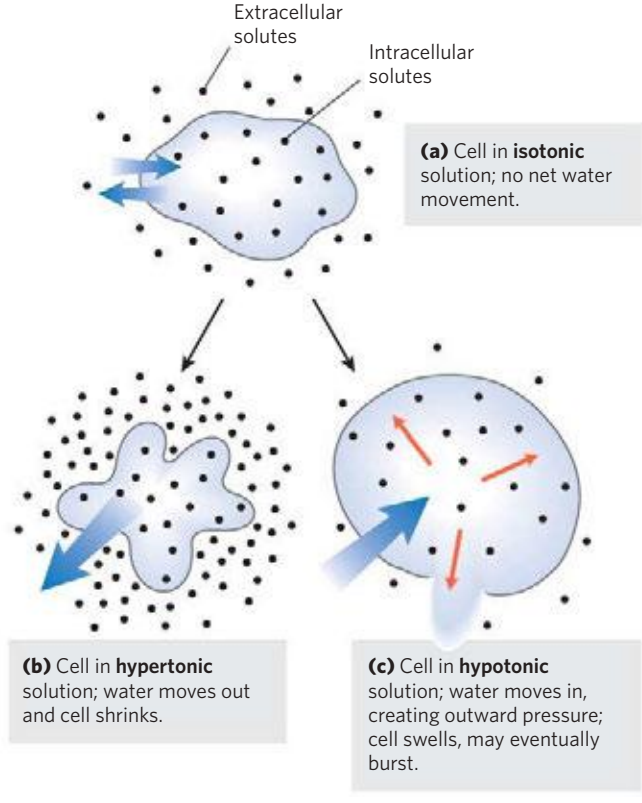


FIGURE 2–13 Effect of extracellular osmolarity on water movement across a plasma membrane. When a cell in osmotic balance with its surrounding medium—that is, a cell in **(a)** an isotonic medium—is transferred into **(b)** a hypertonic solution or **(c)** a hypotonic solution, water moves across the plasma membrane in the direction that tends to equalize osmolarity outside and inside the cell.

and prevent osmotic lysis. Certain freshwater protists that live in a highly hypotonic medium have an organelle (contractile vacuole) that pumps water out of the cell. In multicellular animals, blood plasma and interstitial fluid (the extracellular fluid of tissues) are maintained at an osmolarity close to that of the cytosol. The high concentration of albumin and other proteins in blood plasma contributes to its osmolarity. Cells also actively pump out Na^+ and other ions into the interstitial fluid to stay in osmotic balance with their surroundings.

Because the effect of solutes on osmolarity depends on the *number* of dissolved particles, not their *mass*, macromolecules (proteins, nucleic acids, polysaccharides) have far less effect on the osmolarity of a solution than would an equal mass of their monomeric components. For example, a *gram* of a polysaccharide composed of 1,000 glucose units has the same effect on osmolarity as a *milligram* of glucose. Storing fuel as polysaccharides (starch or glycogen) rather than as glucose or other simple sugars avoids an enormous increase in osmotic pressure in the storage cell.

Plants use osmotic pressure to achieve mechanical rigidity. The very high solute concentration in the plant cell vacuole draws water into the cell (Fig. 2–13), but the nonexpandable cell wall prevents swelling; instead, the pressure exerted against the cell wall (turgor pressure) increases, stiffening the cell, the tissue, and the plant body. When the lettuce in your salad wilts, it is because loss of water has reduced turgor pressure. Osmosis also has consequences for laboratory protocols. Mitochondria, chloroplasts, and lysosomes, for example, are enclosed by semipermeable membranes. In isolating these organelles from broken cells, biochemists must perform the fractionations in isotonic solutions (see Fig. 1–8) to prevent excessive entry of water into the organelles and the swelling and bursting that would follow. Buffers used in cellular fractionations commonly contain sufficient concentrations of sucrose or some other inert solute to protect the organelles from osmotic lysis.

WORKED EXAMPLE 2–1 Osmotic Strength of an Organelle I

Suppose the major solutes in intact lysosomes are KCl (~0.1 M) and NaCl (~0.03 M). When isolating lysosomes, what concentration of sucrose is required in the extracting solution at room temperature (25 °C) to prevent swelling and lysis?

Solution: We want to find a concentration of sucrose that gives an osmotic strength equal to that produced by the KCl and NaCl in the lysosomes. The equation for calculating osmotic strength (the van't Hoff equation) is

$$\Pi = RT(i_1c_1 + i_2c_2 + i_3c_3 + \cdots + i_nc_n)$$

where R is the gas constant 8.315 J/mol · K, T is the absolute temperature (Kelvin), c_1, c_2, and c_3 are the molar concentrations of each solute, and i_1, i_2, and i_3 are the numbers of particles each solute yields in solution ($i = 2$ for KCl and NaCl).

The osmotic strength of the lysosomal contents is

$$\begin{aligned}\Pi_{\text{lysosome}} &= RT(i_{\text{KCl}}c_{\text{KCl}} + i_{\text{NaCl}}c_{\text{NaCl}})\\ &= RT[(2)(0.1\text{ mol/L}) + (2)(0.03\text{ mol/L})]\\ &= RT(0.26\text{ mol/L})\end{aligned}$$

The osmotic strength of a sucrose solution is given by

$$\Pi_{\text{sucrose}} = RT(i_{\text{sucrose}}c_{\text{sucrose}})$$

In this case, $i_{\text{sucrose}} = 1$, because sucrose does not ionize. Thus,

$$\Pi_{\text{sucrose}} = RT(c_{\text{sucrose}})$$

The osmotic strength of the lysosomal contents equals that of the sucrose solution when

$$\begin{aligned}\Pi_{\text{sucrose}} &= \Pi_{\text{lysosome}}\\ RT(c_{\text{sucrose}}) &= RT(0.26\text{ mol/L})\\ c_{\text{sucrose}} &= 0.26\text{ mol/L}\end{aligned}$$

So the required concentration of sucrose (FW 342) is (0.26 mol/L)(342 g/mol) = 88.92 g/L. Because the solute concentrations are only accurate to one significant figure, $c_{\text{sucrose}} = 0.09$ kg/L.

WORKED EXAMPLE 2–2 Osmotic Strength of an Organelle II

Suppose we decided to use a solution of a polysaccharide, say glycogen (p. 255), to balance the osmotic strength of the lysosomes (described in Worked Example 2–1). Assuming a linear polymer of 100 glucose units, calculate the amount of this polymer needed to achieve the same osmotic strength as the sucrose solution in Worked Example 2–1. The M_r of the glucose polymer is ~18,000, and, like sucrose, it does not ionize in solution.

Solution: As derived in Worked Example 2–1,

$$\Pi_{\text{sucrose}} = RT(0.26\text{ mol/L})$$

Similarly,

$$\Pi_{\text{glycogen}} = RT(i_{\text{glycogen}}c_{\text{glycogen}}) = RT(c_{\text{glycogen}})$$

For a glycogen solution with the same osmotic strength as the sucrose solution,

$$\begin{aligned}\Pi_{\text{glycogen}} &= \Pi_{\text{sucrose}}\\ RT(c_{\text{glycogen}}) &= RT(0.26\text{ mol/L})\\ c_{\text{glycogen}} &= 0.26\text{ mol/L} = (0.26\text{ mol/L})(18{,}000\text{ g/mol})\\ &= 4.68\text{ kg/L}\end{aligned}$$

Or, when significant figures are taken into account, $c_{\text{glycogen}} = 5$ kg/L, an absurdly high concentration.

As we'll see later (p. 256), cells of liver and muscle store carbohydrate not as low molecular weight sugars such as glucose or sucrose but as the high molecular weight polymer glycogen. This allows the cell to contain a large mass of glycogen with a minimal effect on the osmolarity of the cytosol.

SUMMARY 2.1 Weak Interactions in Aqueous Systems

- The very different electronegativities of H and O make water a highly polar molecule, capable of forming hydrogen bonds with itself and with solutes. Hydrogen bonds are fleeting, primarily electrostatic, and weaker than covalent bonds. Water is a good solvent for polar (hydrophilic) solutes, with which it forms hydrogen bonds, and for charged solutes, with which it interacts electrostatically.
- Nonpolar (hydrophobic) compounds dissolve poorly in water; they cannot hydrogen-bond with the solvent, and their presence forces an energetically unfavorable ordering of water molecules at their hydrophobic surfaces. To minimize the surface exposed to water, nonpolar compounds such as lipids form aggregates (micelles) in which the hydrophobic moieties are sequestered in the interior, associating through hydrophobic interactions, and only the more polar moieties interact with water.
- Weak, noncovalent interactions, in large numbers, decisively influence the folding of macromolecules such as proteins and nucleic acids. The most stable macromolecular conformations are those in which hydrogen bonding is maximized within the molecule and between the molecule and the solvent, and in which hydrophobic moieties cluster in the interior of the molecule away from the aqueous solvent.
- The physical properties of aqueous solutions are strongly influenced by the concentrations of solutes. When two aqueous compartments are separated by a semipermeable membrane (such as the plasma membrane separating a cell from its surroundings), water moves across that membrane to equalize the osmolarity in the two compartments. This tendency for water to move across a semipermeable membrane produces the osmotic pressure.

2.2 Ionization of Water, Weak Acids, and Weak Bases

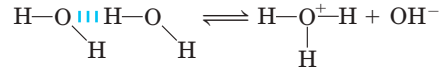
Although many of the solvent properties of water can be explained in terms of the uncharged H_2O molecule, the small degree of ionization of water to hydrogen ions (H^+) and hydroxide ions (OH^-) must also be taken into account. Like all reversible reactions, the ionization of water can be described by an equilibrium constant. When weak acids are dissolved in water, they contribute H^+ by ionizing; weak bases consume H^+ by becoming protonated. These processes are also governed by equilibrium constants. The total hydrogen ion concentration from all sources is experimentally measurable and is expressed as the pH of the solution. To predict the state of ionization of solutes in water, we must take into account the relevant equilibrium constants for each ionization reaction. We therefore turn now to a brief discussion of the ionization of water and of weak acids and bases dissolved in water.

Pure Water Is Slightly Ionized

Water molecules have a slight tendency to undergo reversible ionization to yield a hydrogen ion (a proton) and a hydroxide ion, giving the equilibrium

$$H_2O \rightleftharpoons H^+ + OH^- \qquad (2\text{–}1)$$

Although we commonly show the dissociation product of water as H^+, free protons do not exist in solution; hydrogen ions formed in water are immediately hydrated to form **hydronium ions** (H_3O^+). Hydrogen bonding between water molecules makes the hydration of dissociating protons virtually instantaneous:



The ionization of water can be measured by its electrical conductivity; pure water carries electrical current as H_3O^+ migrates toward the cathode and OH^- toward the anode. The movement of hydronium and hydroxide ions in the electric field is extremely fast compared with that of other ions such as Na^+, K^+, and Cl^-. This high ionic mobility results from the kind of "proton hopping" shown in **Figure 2–14**. No individual proton moves very

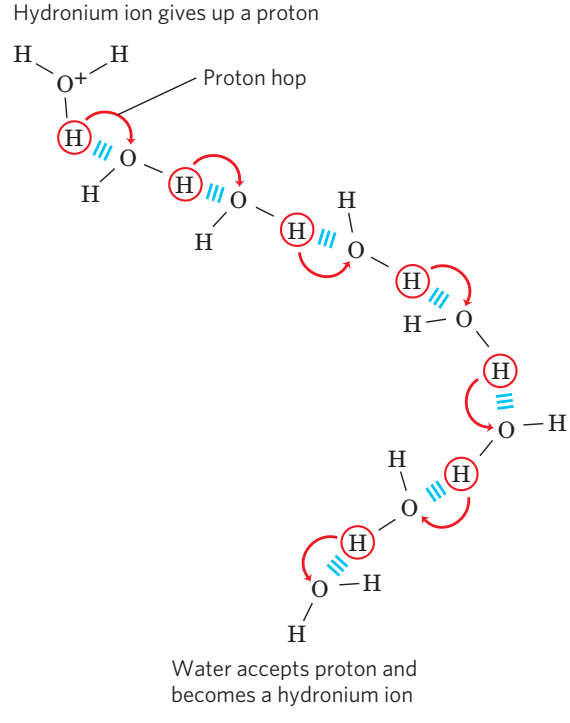


FIGURE 2–14 Proton hopping. Short "hops" of protons between a series of hydrogen-bonded water molecules result in an extremely rapid net movement of a proton over a long distance. As a hydronium ion (upper left) gives up a proton, a water molecule some distance away (bottom) acquires one, becoming a hydronium ion. Proton hopping is much faster than true diffusion and explains the remarkably high ionic mobility of H^+ ions compared with other monovalent cations such as Na^+ and K^+.

far through the bulk solution, but a series of proton hops between hydrogen-bonded water molecules causes the *net* movement of a proton over a long distance in a remarkably short time. (OH^- also moves rapidly by proton hopping, but in the opposite direction.) As a result of the high ionic mobility of H^+, acid-base reactions in aqueous solutions are exceptionally fast. As noted above, proton hopping very likely also plays a role in biological proton-transfer reactions (Fig. 2–10; see also Fig. 19–69b).

Because reversible ionization is crucial to the role of water in cellular function, we must have a means of expressing the extent of ionization of water in quantitative terms. A brief review of some properties of reversible chemical reactions shows how this can be done.

The position of equilibrium of any chemical reaction is given by its **equilibrium constant, K_{eq}** (sometimes expressed simply as K). For the generalized reaction

$$A + B \rightleftharpoons C + D \qquad (2–2)$$

the equilibrium constant K_{eq} can be defined in terms of the concentrations of reactants (A and B) and products (C and D) at equilibrium:

$$K_{eq} = \frac{[C]_{eq}[D]_{eq}}{[A]_{eq}[B]_{eq}}$$

Strictly speaking, the concentration terms should be the *activities*, or effective concentrations in nonideal solutions, of each species. Except in very accurate work, however, the equilibrium constant may be approximated by measuring the *concentrations* at equilibrium. For reasons beyond the scope of this discussion, equilibrium constants are dimensionless. Nonetheless, we have generally retained the concentration units (M) in the equilibrium expressions used in this book to remind you that molarity is the unit of concentration used in calculating K_{eq}.

The equilibrium constant is fixed and characteristic for any given chemical reaction at a specified temperature. It defines the composition of the final equilibrium mixture, regardless of the starting amounts of reactants and products. Conversely, we can calculate the equilibrium constant for a given reaction at a given temperature if the equilibrium concentrations of all its reactants and products are known. As we showed in Chapter 1 (p. 26), the standard free-energy change ($\Delta G°$) is directly related to $\ln K_{eq}$.

The Ionization of Water Is Expressed by an Equilibrium Constant

The degree of ionization of water at equilibrium (Eqn 2–1) is small; at 25 °C only about two of every 10^9 molecules in pure water are ionized at any instant. The equilibrium constant for the reversible ionization of water is

$$K_{eq} = \frac{[H^+][OH^-]}{[H_2O]} \qquad (2–3)$$

In pure water at 25 °C, the concentration of water is 55.5 M—grams of H_2O in 1 L divided by its gram molecular weight: (1,000 g/L)/(18.015 g/mol)—and is essentially constant in relation to the very low concentrations of H^+ and OH^-, namely 1×10^{-7} M. Accordingly, we can substitute 55.5 M in the equilibrium constant expression (Eqn 2–3) to yield

$$K_{eq} = \frac{[H^+][OH^-]}{[55.5\ \text{M}]}$$

On rearranging, this becomes

$$(55.5\ \text{M})(K_{eq}) = [H^+][OH^-] = K_w \qquad (2–4)$$

where K_w designates the product $(55.5\ \text{M})(K_{eq})$, the **ion product of water** at 25 °C.

The value for K_{eq}, determined by electrical-conductivity measurements of pure water, is 1.8×10^{-16} M at 25 °C. Substituting this value for K_{eq} in Equation 2–4 gives the value of the ion product of water:

$$K_w = [H^+][OH^-] = (55.5\ \text{M})(1.8 \times 10^{-16}\ \text{M})$$
$$= 1.0 \times 10^{-14}\ \text{M}^2$$

Thus the product $[H^+][OH^-]$ in aqueous solutions at 25 °C always equals 1×10^{-14} M^2. When there are exactly equal concentrations of H^+ and OH^-, as in pure water, the solution is said to be at **neutral pH**. At this pH, the concentration of H^+ and OH^- can be calculated from the ion product of water as follows:

$$K_w = [H^+][OH^-] = [H^+]^2 = [OH^-]^2$$

Solving for $[H^+]$ gives

$$[H^+] = \sqrt{K_w} = \sqrt{1 \times 10^{-14}\ \text{M}^2}$$
$$[H^+] = [OH^-] = 10^{-7}\ \text{M}$$

As the ion product of water is constant, whenever $[H^+]$ is greater than 1×10^{-7} M, $[OH^-]$ must be less than 1×10^{-7} M, and vice versa. When $[H^+]$ is very high, as in a solution of hydrochloric acid, $[OH^-]$ must be very low. From the ion product of water we can calculate $[H^+]$ if we know $[OH^-]$, and vice versa.

WORKED EXAMPLE 2–3 Calculation of $[H^+]$

What is the concentration of H^+ in a solution of 0.1 M NaOH?

Solution: We begin with the equation for the ion product of water:

$$K_w = [H^+][OH^-]$$

With $[OH^-] = 0.1$ M, solving for $[H^+]$ gives

$$[H^+] = \frac{K_w}{[OH^-]} = \frac{1 \times 10^{-14}\ \text{M}^2}{0.1\ \text{M}} = \frac{10^{-14}\ \text{M}^2}{10^{-1}\ \text{M}}$$
$$= 10^{-13}\ \text{M}$$

WORKED EXAMPLE 2–4 Calculation of $[OH^-]$

What is the concentration of OH^- in a solution with an H^+ concentration of 1.3×10^{-4} M?

Solution: We begin with the equation for the ion product of water:

$$K_w = [H^+][OH^-]$$

With $[H^+] = 1.3 \times 10^{-4}$ M, solving for $[OH^-]$ gives

$$[OH^-] = \frac{K_w}{[H^+]} = \frac{1 \times 10^{-14}\ M^2}{0.00013\ M} = \frac{10^{-14}\ M^2}{1.3 \times 10^{-4}\ M}$$
$$= 7.7 \times 10^{-11}\ M$$

In all calculations be sure to round your answer to the correct number of significant figures, as here.

The pH Scale Designates the H^+ and OH^- Concentrations

The ion product of water, K_w, is the basis for the **pH scale** (Table 2–6). It is a convenient means of designating the concentration of H^+ (and thus of OH^-) in any aqueous solution in the range between 1.0 M H^+ and 1.0 M OH^-. The term **pH** is defined by the expression

$$pH = \log \frac{1}{[H^+]} = -\log [H^+]$$

The symbol p denotes "negative logarithm of." For a precisely neutral solution at 25 °C, in which the concentration of hydrogen ions is 1.0×10^{-7} M, the pH can be calculated as follows:

$$pH = \log \frac{1}{1.0 \times 10^{-7}} = 7.0$$

Note that the concentration of H^+ must be expressed in molar (M) terms.

The value of 7 for the pH of a precisely neutral solution is not an arbitrarily chosen figure; it is derived from the absolute value of the ion product of water at 25 °C, which by convenient coincidence is a round number. Solutions having a pH greater than 7 are alkaline or basic; the concentration of OH^- is greater than that of H^+. Conversely, solutions having a pH less than 7 are acidic.

Keep in mind that the pH scale is logarithmic, not arithmetic. To say that two solutions differ in pH by 1 pH unit means that one solution has ten times the H^+ concentration of the other, but it does not tell us the absolute magnitude of the difference. **Figure 2–15** gives the pH values of some common aqueous fluids. A cola drink (pH 3.0) or red wine (pH 3.7) has an H^+ concentration approximately 10,000 times that of blood (pH 7.4).

The pH of an aqueous solution can be approximately measured with various indicator dyes, including litmus, phenolphthalein, and phenol red. These dyes undergo color changes as a proton dissociates from the dye

TABLE 2–6 The pH Scale

$[H^+]$ (M)	pH	$[OH^-]$ (M)	pOH*
10^0 (1)	0	10^{-14}	14
10^{-1}	1	10^{-13}	13
10^{-2}	2	10^{-12}	12
10^{-3}	3	10^{-11}	11
10^{-4}	4	10^{-10}	10
10^{-5}	5	10^{-9}	9
10^{-6}	6	10^{-8}	8
10^{-7}	7	10^{-7}	7
10^{-8}	8	10^{-6}	6
10^{-9}	9	10^{-5}	5
10^{-10}	10	10^{-4}	4
10^{-11}	11	10^{-3}	3
10^{-12}	12	10^{-2}	2
10^{-13}	13	10^{-1}	1
10^{-14}	14	10^0 (1)	0

*The expression pOH is sometimes used to describe the basicity, or OH^- concentration, of a solution; pOH is defined by the expression $pOH = -\log[OH^-]$, which is analogous to the expression for pH. Note that in all cases, pH + pOH = 14.

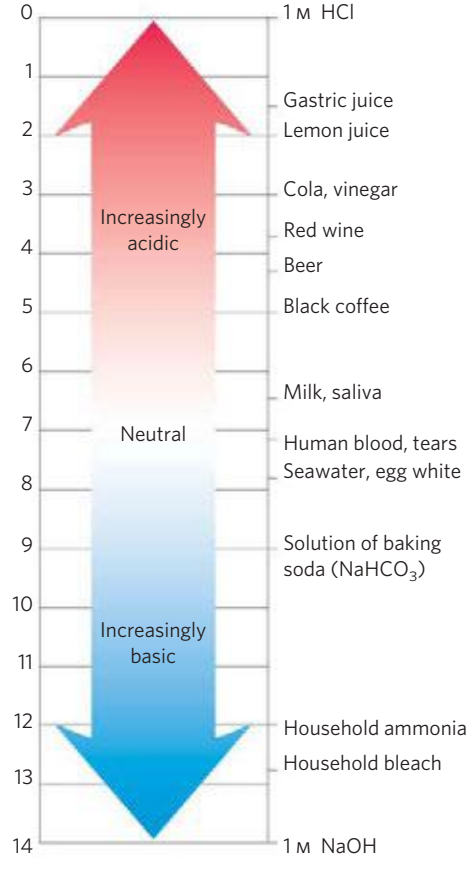


FIGURE 2–15 The pH of some aqueous fluids.

molecule. Accurate determinations of pH in the chemical or clinical laboratory are made with a glass electrode that is selectively sensitive to H^+ concentration but insensitive to Na^+, K^+, and other cations. In a pH meter, the signal from the glass electrode placed in a test solution is amplified and compared with the signal generated by a solution of accurately known pH.

Measurement of pH is one of the most important and frequently used procedures in biochemistry. The pH affects the structure and activity of biological macromolecules; for example, the catalytic activity of enzymes is strongly dependent on pH (see Fig. 2–22). Measurements of the pH of blood and urine are commonly used in medical diagnoses. The pH of the blood plasma of people with severe, uncontrolled diabetes, for example, is often below the normal value of 7.4; this condition is called **acidosis** (described in more detail below). In certain other diseases the pH of the blood is higher than normal, a condition known as **alkalosis**. Extreme acidosis or alkalosis can be life-threatening. ■

Weak Acids and Bases Have Characteristic Acid Dissociation Constants

Hydrochloric, sulfuric, and nitric acids, commonly called strong acids, are completely ionized in dilute aqueous solutions; the strong bases NaOH and KOH are also completely ionized. Of more interest to biochemists is the behavior of weak acids and bases—those not completely ionized when dissolved in water. These are ubiquitous in biological systems and play important roles in metabolism and its regulation. The behavior of aqueous solutions of weak acids and bases is best understood if we first define some terms.

Acids may be defined as proton donors and bases as proton acceptors. When a proton donor such as acetic acid (CH_3COOH) loses a proton, it becomes the corresponding proton acceptor, in this case the acetate anion (CH_3COO^-). A proton donor and its corresponding proton acceptor make up a **conjugate acid-base pair (Fig. 2–16)**, related by the reversible reaction

$$CH_3COOH \rightleftharpoons CH_3COO^- + H^+$$

Each acid has a characteristic tendency to lose its proton in an aqueous solution. The stronger the acid, the greater its tendency to lose its proton. The tendency of any acid (HA) to lose a proton and form its conjugate base (A^-) is defined by the equilibrium constant (K_{eq}) for the reversible reaction

$$HA \rightleftharpoons H^+ + A^-$$

for which

$$K_{eq} = \frac{[H^+][A^-]}{[HA]} = K_a$$

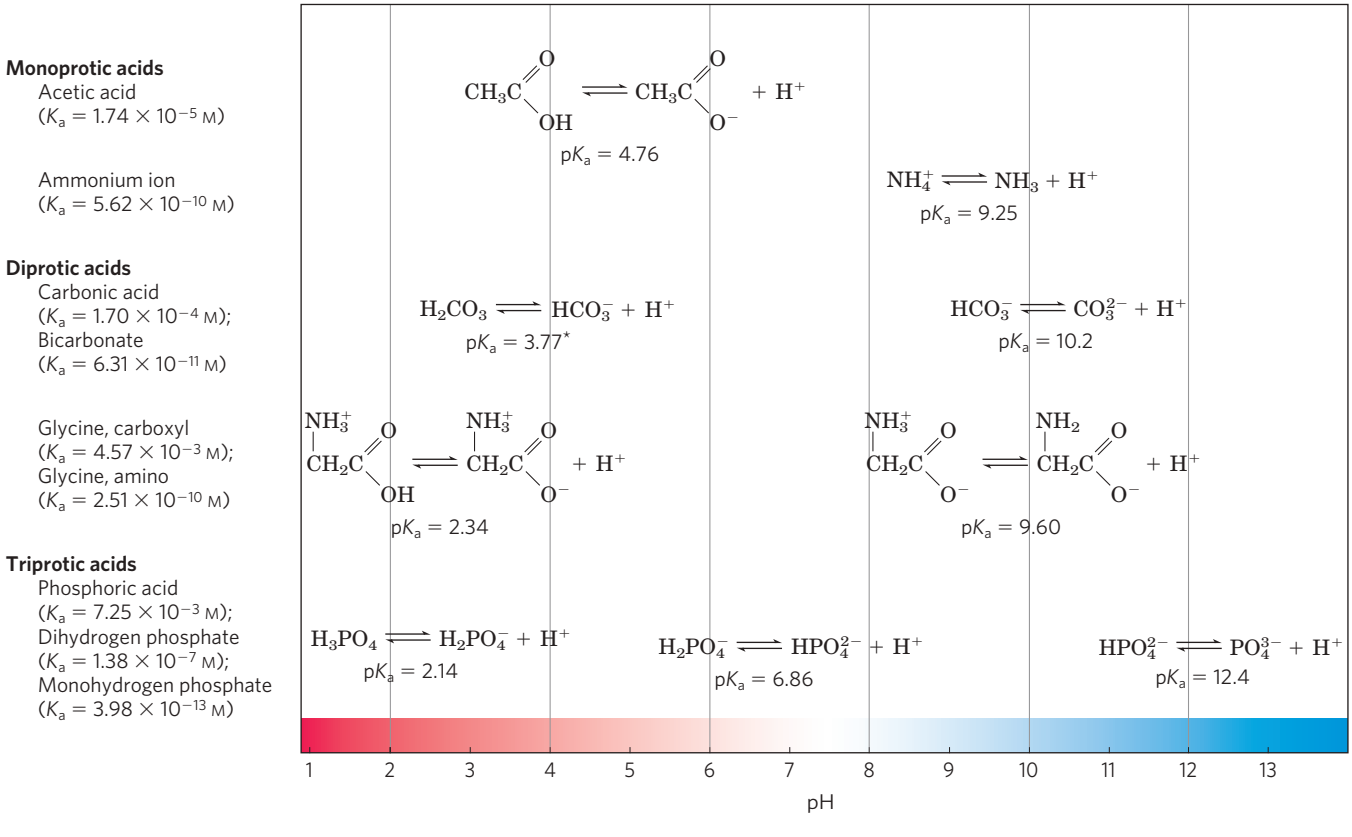


FIGURE 2–16 Conjugate acid-base pairs consist of a proton donor and a proton acceptor. Some compounds, such as acetic acid and ammonium ion, are monoprotic; they can give up only one proton. Others are diprotic (carbonic acid and glycine) or triprotic (phosphoric acid). The dissociation reactions for each pair are shown where they occur along a pH gradient. The equilibrium or dissociation constant (K_a) and its negative logarithm, the pK_a, are shown for each reaction. *For an explanation of apparent discrepancies in pK_a values for carbonic acid (H_2CO_3), see p. 67.

Equilibrium constants for ionization reactions are usually called **ionization constants** or **acid dissociation constants**, often designated K_a. The dissociation constants of some acids are given in Figure 2–16. Stronger acids, such as phosphoric and carbonic acids, have larger ionization constants; weaker acids, such as monohydrogen phosphate (HPO_4^{2-}), have smaller ionization constants.

Also included in Figure 2–16 are values of $\mathbf{pK_a}$, which is analogous to pH and is defined by the equation

$$pK_a = \log \frac{1}{K_a} = -\log K_a$$

The stronger the tendency to dissociate a proton, the stronger is the acid and the lower its pK_a. As we shall now see, the pK_a of any weak acid can be determined quite easily.

Titration Curves Reveal the pK_a of Weak Acids

Titration is used to determine the amount of an acid in a given solution. A measured volume of the acid is titrated with a solution of a strong base, usually sodium hydroxide (NaOH), of known concentration. The NaOH is added in small increments until the acid is consumed (neutralized), as determined with an indicator dye or a pH meter. The concentration of the acid in the original solution can be calculated from the volume and concentration of NaOH added. The amounts of acid and base in titrations are often expressed in terms of equivalents, where one equivalent is the amount of a substance that will react with, or supply, one mole of hydrogen ions in an acid-base reaction.

A plot of pH against the amount of NaOH added (a **titration curve**), reveals the pK_a of the weak acid. Consider the titration of a 0.1 M solution of acetic acid with 0.1 M NaOH at 25 °C **(Fig. 2–17)**. Two reversible equilibria are involved in the process (here, for simplicity, acetic acid is denoted HAc):

$$H_2O \rightleftharpoons H^+ + OH^- \qquad (2–5)$$

$$HAc \rightleftharpoons H^+ + Ac^- \qquad (2–6)$$

The equilibria must simultaneously conform to their characteristic equilibrium constants, which are, respectively,

$$K_w = [H^+][OH^-] = 1 \times 10^{-14}\ M^2 \qquad (2–7)$$

$$K_a = \frac{[H^+][Ac^-]}{[HAc]} = 1.74 \times 10^{-5}\ M \qquad (2–8)$$

At the beginning of the titration, before any NaOH is added, the acetic acid is already slightly ionized, to an extent that can be calculated from its ionization constant (Eqn 2–8).

As NaOH is gradually introduced, the added OH^- combines with the free H^+ in the solution to form H_2O, to an extent that satisfies the equilibrium relationship in Equation 2–7. As free H^+ is removed, HAc dissociates further to satisfy its own equilibrium constant (Eqn 2–8). The net result as the titration proceeds is that more and more HAc ionizes, forming Ac^-, as the NaOH is added. At the midpoint of the titration, at which exactly 0.5 equivalent of NaOH has been added per equivalent of the acid, one-half of the original acetic acid has undergone dissociation, so that the concentration of the proton donor, [HAc], now equals that of the proton acceptor, [Ac^-]. At this midpoint a very important relationship holds: the pH of the equimolar solution of acetic acid and acetate is exactly equal to the pK_a of acetic acid (pK_a = 4.76; Figs 2–16, 2–17). The basis for this relationship, which holds for all weak acids, will soon become clear.

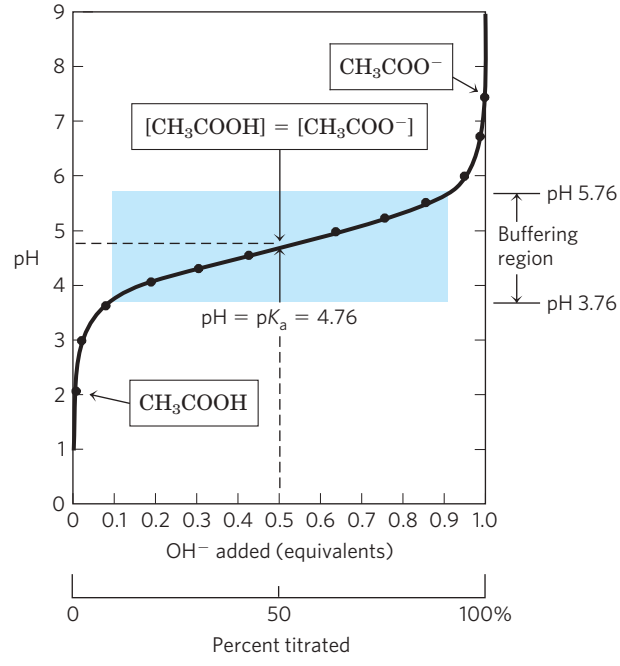


FIGURE 2–17 The titration curve of acetic acid. After addition of each increment of NaOH to the acetic acid solution, the pH of the mixture is measured. This value is plotted against the amount of NaOH added, expressed as a fraction of the total NaOH required to convert all the acetic acid (CH_3COOH) to its deprotonated form, acetate (CH_3COO^-). The points so obtained yield the titration curve. Shown in the boxes are the predominant ionic forms at the points designated. At the midpoint of the titration, the concentrations of the proton donor and proton acceptor are equal, and the pH is numerically equal to the pK_a. The shaded zone is the useful region of buffering power, generally between 10% and 90% titration of the weak acid.

As the titration is continued by adding further increments of NaOH, the remaining nondissociated acetic acid is gradually converted into acetate. The end point of the titration occurs at about pH 7.0: all the acetic acid has lost its protons to OH^-, to form H_2O and acetate. Throughout the titration the two equilibria (Eqns 2–5, 2–6) coexist, each always conforming to its equilibrium constant.

Figure 2–18 compares the titration curves of three weak acids with very different ionization constants: acetic acid (pK_a = 4.76); dihydrogen phosphate,

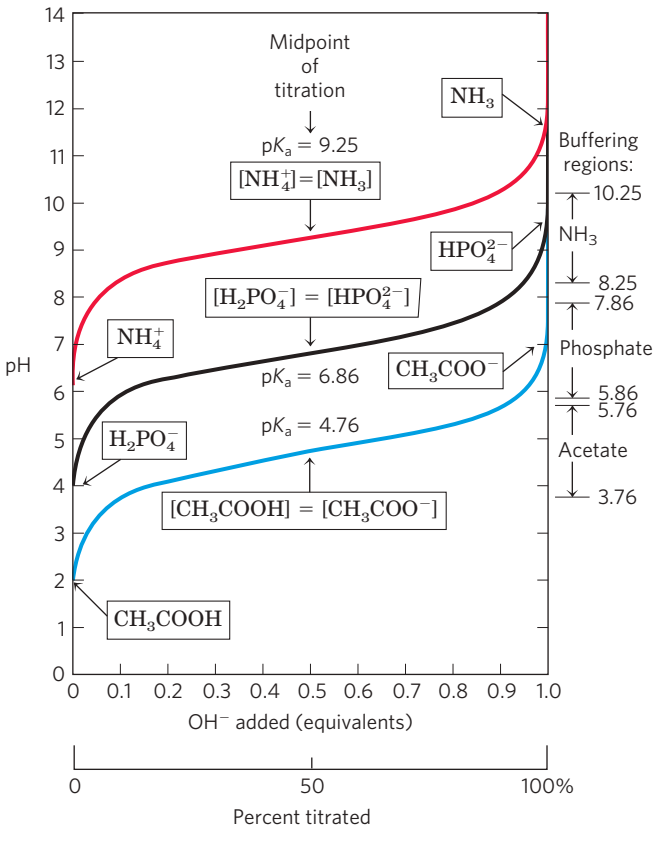


FIGURE 2–18 Comparison of the titration curves of three weak acids. Shown here are the titration curves for CH_3COOH, $H_2PO_4^-$, and NH_4^+. The predominant ionic forms at designated points in the titration are given in boxes. The regions of buffering capacity are indicated at the right. Conjugate acid-base pairs are effective buffers between approximately 10% and 90% neutralization of the proton-donor species.

$H_2PO_4^-$ ($pK_a = 6.86$); and ammonium ion, NH_4^+ ($pK_a = 9.25$). Although the titration curves of these acids have the same shape, they are displaced along the pH axis because the three acids have different strengths. Acetic acid, with the highest K_a (lowest pK_a) of the three, is the strongest of the three weak acids (loses its proton most readily); it is already half dissociated at pH 4.76. Dihydrogen phosphate loses a proton less readily, being half dissociated at pH 6.86. Ammonium ion is the weakest acid of the three and does not become half dissociated until pH 9.25.

The titration curve of a weak acid shows graphically that a weak acid and its anion—a conjugate acid-base pair—can act as a buffer, as we describe in the next section.

SUMMARY 2.2 Ionization of Water, Weak Acids, and Weak Bases

- Pure water ionizes slightly, forming equal numbers of hydrogen ions (hydronium ions, H_3O^+) and hydroxide ions. The extent of ionization is described by an equilibrium constant, $K_{eq} = \dfrac{[H^+][OH^-]}{[H_2O]}$, from which the ion product of water, K_w, is derived. At 25 °C, $K_w = [H^+][OH^-] = (55.5\ \text{M})(K_{eq}) = 10^{-14}\ \text{M}^2$.
- The pH of an aqueous solution reflects, on a logarithmic scale, the concentration of hydrogen ions:

$$pH = \log \frac{1}{[H^+]} = -\log [H^+].$$

- The greater the acidity of a solution, the lower its pH. Weak acids partially ionize to release a hydrogen ion, thus lowering the pH of the aqueous solution. Weak bases accept a hydrogen ion, increasing the pH. The extent of these processes is characteristic of each particular weak acid or base and is expressed as an acid dissociation constant:

$$K_{eq} = \frac{[H^+][A^-]}{[HA]} = K_a.$$

- The pK_a expresses, on a logarithmic scale, the relative strength of a weak acid or base:

$$pK_a = \log \frac{1}{K_a} = -\log K_a.$$

- The stronger the acid, the smaller its pK_a; the stronger the base, the larger its pK_a. The pK_a can be determined experimentally; it is the pH at the midpoint of the titration curve for the acid or base.

2.3 Buffering against pH Changes in Biological Systems

Almost every biological process is pH-dependent; a small change in pH produces a large change in the rate of the process. This is true not only for the many reactions in which the H^+ ion is a direct participant, but also for those reactions in which there is no apparent role for H^+ ions. The enzymes that catalyze cellular reactions, and many of the molecules on which they act, contain ionizable groups with characteristic pK_a values. The protonated amino and carboxyl groups of amino acids and the phosphate groups of nucleotides, for example, function as weak acids; their ionic state is determined by the pH of the surrounding medium. (When an ionizable group is sequestered in the middle of a protein, away from the aqueous solvent, its pK_a, or apparent pK_a, can be significantly different from its pK_a in water.) As we noted above, ionic interactions are among the forces that stabilize a protein molecule and allow an enzyme to recognize and bind its substrate.

Cells and organisms maintain a specific and constant cytosolic pH, usually near pH 7, keeping biomolecules in their optimal ionic state. In multicellular organisms, the pH of extracellular fluids is also tightly regulated. Constancy of pH is achieved primarily by biological buffers: mixtures of weak acids and their conjugate bases.

Buffers Are Mixtures of Weak Acids and Their Conjugate Bases

Buffers are aqueous systems that tend to resist changes in pH when small amounts of acid (H^+) or base (OH^-) are added. A buffer system consists of a weak acid (the proton donor) and its conjugate base (the proton acceptor). As an example, a mixture of equal concentrations of acetic acid and acetate ion, found at the midpoint of the titration curve in Figure 2–17, is a buffer system. Notice that the titration curve of acetic acid has a relatively flat zone extending about 1 pH unit on either side of its midpoint pH of 4.76. In this zone, a given amount of H^+ or OH^- added to the system has much less effect on pH than the same amount added outside the zone. This relatively flat zone is the **buffering region** of the acetic acid–acetate buffer pair. At the midpoint of the buffering region, where the concentration of the proton donor (acetic acid) exactly equals that of the proton acceptor (acetate), the buffering power of the system is maximal; that is, its pH changes least on addition of H^+ or OH^-. The pH at this point in the titration curve of acetic acid is equal to its pK_a. The pH of the acetate buffer system does change slightly when a small amount of H^+ or OH^- is added, but this change is very small compared with the pH change that would result if the same amount of H^+ or OH^- were added to pure water or to a solution of the salt of a strong acid and strong base, such as NaCl, which has no buffering power.

Buffering results from two reversible reaction equilibria occurring in a solution of nearly equal concentrations of a proton donor and its conjugate proton acceptor. **Figure 2–19** explains how a buffer system works. Whenever H^+ or OH^- is added to a buffer, the result is a small change in the ratio of the relative concentrations of the weak acid and its anion and thus a small change in pH. The decrease in concentration of one component of the system is balanced exactly by an increase in the other. The sum of the buffer components does not change, only their ratio.

Each conjugate acid-base pair has a characteristic pH zone in which it is an effective buffer (Fig. 2–18). The $H_2PO_4^-/HPO_4^{2-}$ pair has a pK_a of 6.86 and thus can serve as an effective buffer system between approximately pH 5.9 and pH 7.9; the NH_4^+/NH_3 pair, with a pK_a of 9.25, can act as a buffer between approximately pH 8.3 and pH 10.3.

The Henderson-Hasselbalch Equation Relates pH, pK_a, and Buffer Concentration

The titration curves of acetic acid, $H_2PO_4^-$, and NH_4^+ (Fig. 2–18) have nearly identical shapes, suggesting that these curves reflect a fundamental law or relationship. This is indeed the case. The shape of the titration curve of any weak acid is described by the Henderson-Hasselbalch equation, which is important for understanding buffer action and acid-base balance in the blood and tissues of vertebrates. This equation is simply a useful way of restating the expression for the ionization constant of an acid. For the ionization of a weak acid HA, the Henderson-Hasselbalch equation can be derived as follows:

$$K_a = \frac{[H^+][A^-]}{[HA]}$$

First solve for $[H^+]$:

$$[H^+] = K_a \frac{[HA]}{[A^-]}$$

Then take the negative logarithm of both sides:

$$-\log[H^+] = -\log K_a - \log \frac{[HA]}{[A^-]}$$

Substitute pH for $-\log [H^+]$ and pK_a for $-\log K_a$:

$$pH = pK_a - \log \frac{[HA]}{[A^-]}$$

Now invert $-\log [HA]/[A^-]$, which involves changing its sign, to obtain the **Henderson-Hasselbalch equation**:

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \qquad (2\text{–}9)$$

This equation fits the titration curve of all weak acids and enables us to deduce some important quantitative relationships. For example, it shows why the pK_a of a

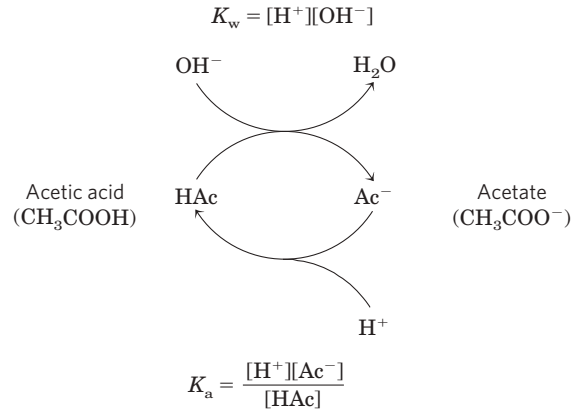


FIGURE 2–19 The acetic acid–acetate pair as a buffer system. The system is capable of absorbing either H^+ or OH^- through the reversibility of the dissociation of acetic acid. The proton donor, acetic acid (HAc), contains a reserve of bound H^+, which can be released to neutralize an addition of OH^- to the system, forming H_2O. This happens because the product $[H^+][OH^-]$ transiently exceeds K_w (1×10^{-14} M^2). The equilibrium quickly adjusts to restore the product to 1×10^{-14} M^2 (at 25 °C), thus transiently reducing the concentration of H^+. But now the quotient $[H^+][Ac^-]/[HAc]$ is less than K_a, so HAc dissociates further to restore equilibrium. Similarly, the conjugate base, Ac^-, can react with H^+ ions added to the system; again, the two ionization reactions simultaneously come to equilibrium. Thus a conjugate acid-base pair, such as acetic acid and acetate ion, tends to resist a change in pH when small amounts of acid or base are added. Buffering action is simply the consequence of two reversible reactions taking place simultaneously and reaching their points of equilibrium as governed by their equilibrium constants, K_w and K_a.

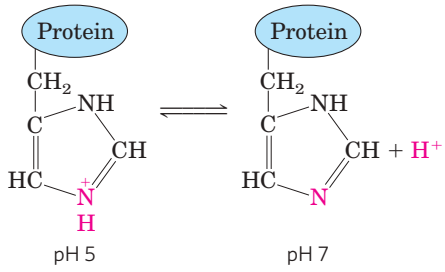


FIGURE 2–20 Ionization of histidine. The amino acid histidine, a component of proteins, is a weak acid. The pK_a of the protonated nitrogen of the side chain is 6.0.

weak acid is equal to the pH of the solution at the midpoint of its titration. At that point, [HA] = [A^-], and

$$\text{pH} = \text{p}K_a + \log 1 = \text{p}K_a + 0 = \text{p}K_a$$

The Henderson-Hasselbalch equation also allows us to (1) calculate pK_a, given pH and the molar ratio of proton donor and acceptor; (2) calculate pH, given pK_a and the molar ratio of proton donor and acceptor; and (3) calculate the molar ratio of proton donor and acceptor, given pH and pK_a.

Weak Acids or Bases Buffer Cells and Tissues against pH Changes

The intracellular and extracellular fluids of multicellular organisms have a characteristic and nearly constant pH. The organism's first line of defense against changes in internal pH is provided by buffer systems. The cytoplasm of most cells contains high concentrations of proteins, and these proteins contain many amino acids with functional groups that are weak acids or weak bases. For example, the side chain of histidine **(Fig. 2–20)** has a pK_a of 6.0 and thus can exist in either the protonated or unprotonated form near neutral pH. Proteins containing histidine residues therefore buffer effectively near neutral pH.

WORKED EXAMPLE 2–5 Ionization of Histidine

Calculate the fraction of histidine that has its imidazole side chain protonated at pH 7.3. The pK_a values for histidine are pK_1 = 1.8, pK_2 (imidazole) = 6.0, and pK_3 = 9.2 (see Fig. 3–12b).

Solution: The three ionizable groups in histidine have sufficiently different pK_a values that the first acid (—COOH) is completely ionized before the second (protonated imidazole) begins to dissociate a proton, and the second ionizes completely before the third ($—NH_3^+$) begins to dissociate its proton. (With the Henderson-Hasselbalch equation, we can easily show that a weak acid goes from 1% ionized at 2 pH units below its pK_a to 99% ionized at 2 pH units above its pK_a; see also Fig. 3–12b.) At pH 7.3, the carboxyl group of histidine is entirely deprotonated ($—COO^-$) and the α-amino group is fully protonated ($—NH_3^+$). We can therefore assume that at pH 7.3, the only group that is partially dissociated is the imidazole group, which can be protonated (we'll abbreviate as $HisH^+$) or not (His).

We use the Henderson-Hasselbalch equation:

$$\text{pH} = \text{p}K_a + \log \frac{[A^-]}{[HA]}$$

Substituting pK_2 = 6.0 and pH = 7.3:

$$7.3 = 6.0 + \log \frac{[\text{His}]}{[\text{HisH}^+]}$$

$$1.3 = \log \frac{[\text{His}]}{[\text{HisH}^+]}$$

$$\text{antilog } 1.3 = \frac{[\text{His}]}{[\text{HisH}^+]} = 2.0 \times 10^1$$

This gives us the *ratio* of [His] to [$HisH^+$] (20 to 1 in this case). We want to convert this ratio to the *fraction* of total histidine that is in the unprotonated form His at pH 7.3. That fraction is 20/21 (20 parts His per 1 part $HisH^+$, *in a total of 21 parts histidine* in either form), or about 95.2%; the remainder (100% minus 95.2%) is protonated—about 5%.

Nucleotides such as ATP, as well as many metabolites of low molecular weight, contain ionizable groups that can contribute buffering power to the cytoplasm. Some highly specialized organelles and extracellular compartments have high concentrations of compounds that contribute buffering capacity: organic acids buffer the vacuoles of plant cells; ammonia buffers urine.

Two especially important biological buffers are the phosphate and bicarbonate systems. The phosphate buffer system, which acts in the cytoplasm of all cells, consists of $H_2PO_4^-$ as proton donor and HPO_4^{2-} as proton acceptor:

$$H_2PO_4^- \rightleftharpoons H^+ + HPO_4^{2-}$$

The phosphate buffer system is maximally effective at a pH close to its pK_a of 6.86 (Figs 2–16, 2–18) and thus tends to resist pH changes in the range between about 5.9 and 7.9. It is therefore an effective buffer in biological fluids; in mammals, for example, extracellular fluids and most cytoplasmic compartments have a pH in the range of 6.9 to 7.4.

WORKED EXAMPLE 2–6 Phosphate Buffers

(a) What is the pH of a mixture of 0.042 M NaH_2PO_4 and 0.058 M Na_2HPO_4?

Solution: We use the Henderson-Hasselbalch equation, which we'll express here as

$$\text{pH} = \text{p}K_a + \log \frac{[\text{conjugate base}]}{[\text{acid}]}$$

In this case, the acid (the species that gives up a proton) is $H_2PO_4^-$, and the conjugate base (the species that gains a proton) is HPO_4^{2-}. Substituting the given concentrations of acid and conjugate base and the pK_a (6.86),

$$\text{pH} = 6.86 + \log \frac{0.058}{0.042} = 6.86 + 0.14 = 7.0$$

We can roughly check this answer. When more conjugate base than acid is present, the acid is more than 50% titrated and thus the pH is above the pK_a (6.86), where the acid is exactly 50% titrated.

(b) If 1.0 mL of 10.0 M NaOH is added to a liter of the buffer prepared in (a), how much will the pH change?

Solution: A liter of the buffer contains 0.042 mol of NaH_2PO_4. Adding 1.0 mL of 10.0 M NaOH (0.010 mol) would titrate an equivalent amount (0.010 mol) of NaH_2PO_4 to Na_2HPO_4, resulting in 0.032 mol of NaH_2PO_4 and 0.068 mol of Na_2HPO_4. The new pH is

$$\text{pH} = \text{p}K_a + \log\frac{[HPO_4^{2-}]}{[H_2PO_4^-]}$$
$$= 6.86 + \log\frac{0.068}{0.032} = 6.86 + 0.33 = 7.2$$

(c) If 1.0 mL of 10.0 M NaOH is added to a liter of pure water at pH 7.0, what is the final pH? Compare this with the answer in (b).

Solution: The NaOH dissociates completely into Na^+ and OH^-, giving $[OH^-] = 0.010$ mol/L $= 1.0 \times 10^{-2}$ M. The pOH is the negative logarithm of $[OH^-]$, so pOH = 2.0. Given that in all solutions, pH + pOH = 14, the pH of the solution is 12.

So, an amount of NaOH that increases the pH of water from 7 to 12 increases the pH of a buffered solution, as in (b), from 7.0 to just 7.2. Such is the power of buffering!

Blood plasma is buffered in part by the bicarbonate system, consisting of carbonic acid (H_2CO_3) as proton donor and bicarbonate (HCO_3^-) as proton acceptor (K_1 is the first of several equilibrium constants in the bicarbonate buffering system):

$$H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$
$$K_1 = \frac{[H^+][HCO_3^-]}{[H_2CO_3]}$$

This buffer system is more complex than other conjugate acid-base pairs because one of its components, carbonic acid (H_2CO_3), is formed from dissolved (d) carbon dioxide and water, in a reversible reaction:

$$CO_2(d) + H_2O \rightleftharpoons H_2CO_3$$
$$K_2 = \frac{[H_2CO_3]}{[CO_2(d)][H_2O]}$$

Carbon dioxide is a gas under normal conditions, and CO_2 dissolved in an aqueous solution is in equilibrium with CO_2 in the gas (g) phase:

$$CO_2(g) \rightleftharpoons CO_2(d)$$
$$K_3 = \frac{[CO_2(d)]}{[CO_2(g)]}$$

The pH of a bicarbonate buffer system depends on the concentration of H_2CO_3 and HCO_3^-, the proton donor and acceptor components. The concentration of H_2CO_3 in turn depends on the concentration of dissolved CO_2, which in turn depends on the concentration of CO_2 in the gas phase, or the **partial pressure** of CO_2, denoted pCO_2. Thus the pH of a bicarbonate buffer exposed to a gas phase is ultimately determined by the concentration of HCO_3^- in the aqueous phase and by pCO_2 in the gas phase.

The bicarbonate buffer system is an effective physiological buffer near pH 7.4, because the H_2CO_3 of blood plasma is in equilibrium with a large reserve capacity of CO_2(g) in the air space of the lungs. As noted above, this buffer system involves three reversible equilibria, in this case between gaseous CO_2 in the lungs and bicarbonate (HCO_3^-) in the blood plasma **(Fig. 2–21)**.

Blood can pick up H^+, such as from the lactic acid produced in muscle tissue during vigorous exercise. Alternatively, it can lose H^+, such as by protonation of the NH_3 produced during protein catabolism. When H^+ is added to blood as it passes through the tissues, reaction 1 in Figure 2–21 proceeds toward a new equilibrium, in which $[H_2CO_3]$ is increased. This in turn increases $[CO_2(d)]$ in the blood (reaction 2) and thus increases the partial pressure of CO_2(g) in the air space of the lungs (reaction 3); the extra CO_2 is exhaled. Conversely, when H^+ is lost from the blood, the opposite events occur: more H_2CO_3 dissociates into H^+ and HCO_3^- and thus more CO_2(g) from the lungs dissolves in blood plasma. The rate of respiration—that is, the rate of inhaling and exhaling—can quickly adjust these equilibria to keep the blood pH nearly constant. The rate of respiration is controlled by the brain stem, where detection of an increased blood pCO_2 or decreased blood pH triggers deeper and more frequent breathing.

At the pH of blood plasma (7.4) very little H_2CO_3 is present in comparison with HCO_3^-, and the addition of a small amount of base (NH_3 or OH^-) would titrate this H_2CO_3, exhausting the buffering capacity. The important

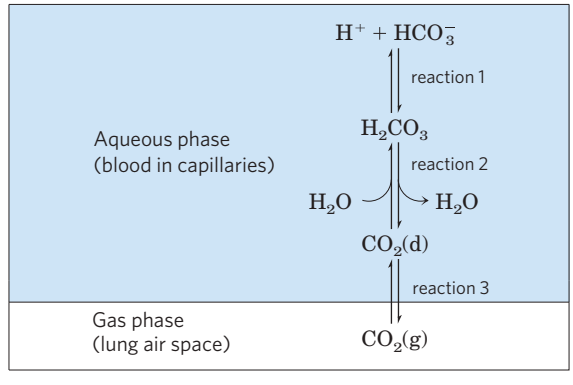


FIGURE 2–21 **The bicarbonate buffer system.** CO_2 in the air space of the lungs is in equilibrium with the bicarbonate buffer in the blood plasma passing through the lung capillaries. Because the concentration of dissolved CO_2 can be adjusted rapidly through changes in the rate of breathing, the bicarbonate buffer system of the blood is in near-equilibrium with a large potential reservoir of CO_2.

role for H_2CO_3 (pK_a = 3.57 at 37 °C) in buffering blood plasma (~pH 7.4) seems inconsistent with our earlier statement that a buffer is most effective in the range of 1 pH unit above and below its pK_a. The explanation for this paradox is the large reservoir of CO_2(d) in blood. Its rapid equilibration with H_2CO_3 results in the formation of additional H_2CO_3:

$$CO_2(d) + H_2O \rightleftharpoons H_2CO_3$$

It is useful in clinical medicine to have a simple expression for blood pH in terms of dissolved CO_2, which is commonly monitored along with other blood gases. We can define a constant, K_h, which is the equilibrium constant for the hydration of CO_2 to form H_2CO_3:

$$K_h = \frac{[H_2CO_3]}{[CO_2(d)]}$$

Then, to take the CO_2(d) reservoir into account, we can express $[H_2CO_3]$ as $K_h[CO_2(d)]$, and substitute this expression for $[H_2CO_3]$ in the equation for the acid dissociation of H_2CO_3:

$$K_a = \frac{[H^+][HCO_3^-]}{[H_2CO_3]} = \frac{[H^+][HCO_3^-]}{K_h[CO_2(d)]}$$

Now, the overall equilibrium for dissociation of H_2CO_3 can be expressed in these terms:

$$K_hK_a = K_{combined} = \frac{[H^+][HCO_3^-]}{[CO_2(d)]}$$

We can calculate the value of the new constant, $K_{combined}$, and the corresponding apparent pK, or p$K_{combined}$, from the experimentally determined values of K_h (3.0×10^{-3} M) and K_a (2.7×10^{-4} M) at 37 °C:

$$K_{combined} = (3.0 \times 10^{-3}\ \text{M})(2.7 \times 10^{-4}\ \text{M})$$
$$= 8.1 \times 10^{-7}\ \text{M}^2$$
$$pK_{combined} = 6.1$$

In clinical medicine, it is common to refer to CO_2(d) as the conjugate acid and to use the apparent, or combined, pK_a of 6.1 to simplify calculation of pH from $[CO_2(d)]$. In this convention,

$$pH = 6.1 + \log \frac{[HCO_3^-]}{(0.23 \times pCO_2)}$$

where pCO_2 is expressed in kilopascals (kPa; typically, pCO_2 is 4.6 to 6.7 kPa) and 0.23 is the corresponding solubility coefficient for CO_2 in water; thus the term $0.23 \times pCO_2 \approx 1.2$ kPa. Plasma $[HCO_3^-]$ is normally about 24 mM. ■

Untreated Diabetes Produces Life-Threatening Acidosis

Human blood plasma normally has a pH between 7.35 and 7.45, and many of the enzymes that function in the blood have evolved to have maximal activity in that pH range. Enzymes typically show maximal catalytic activity at a characteristic pH, called the **pH optimum (Fig. 2–22)**. On either side of this optimum pH, catalytic activity often declines sharply. Thus, a small change in pH can make a large difference in the rate of some crucial enzyme-catalyzed reactions. Biological control of the pH of cells and body fluids is therefore of central importance in all aspects of metabolism and cellular activities, and changes in blood pH have marked physiological consequences (described with gusto in Box 2–1!).

In individuals with untreated diabetes mellitus, the lack of insulin, or insensitivity to insulin (depending on the type of diabetes), disrupts the uptake of glucose from blood into the tissues and forces the tissues to use stored fatty acids as their primary fuel. For reasons we will describe in detail later (see Fig. 24–30), this dependence on fatty acids results in the accumulation of high concentrations of two carboxylic acids, β-hydroxybutyric acid and acetoacetic acid (blood plasma level of 90 mg/100 mL, compared with <3 mg/100 mL in control (healthy) individuals; urinary excretion of 5 g/24 hr, compared with <125 mg/24 hr in controls). Dissociation of these acids lowers the pH of blood plasma to less than 7.35, causing acidosis. Severe acidosis leads to headache, drowsiness, nausea, vomiting, and diarrhea, followed by stupor, coma, and convulsions, presumably because at the lower pH, some enzyme(s) do not function optimally. When a patient is found to have high blood glucose, low plasma pH, and high levels of β-hydroxybutyric acid and acetoacetic acid in blood and urine, diabetes mellitus is the likely diagnosis.

Other conditions can also produce acidosis. Fasting and starvation force the use of stored fatty acids as fuel,

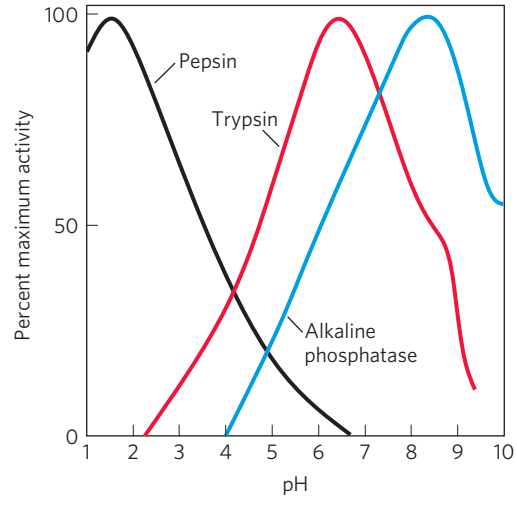


FIGURE 2–22 The pH optima of some enzymes. Pepsin is a digestive enzyme secreted into gastric juice, which has a pH of ~1.5, allowing pepsin to act optimally. Trypsin, a digestive enzyme that acts in the small intestine, has a pH optimum that matches the neutral pH in the lumen of the small intestine. Alkaline phosphatase of bone tissue is a hydrolytic enzyme thought to aid in bone mineralization.

BOX 2–1 MEDICINE On Being One's Own Rabbit (Don't Try This at Home!)

This is an account by J.B.S. Haldane of physiological experiments on controlling blood pH, from his book *Possible Worlds* (Harper and Brothers, 1928).

"I wanted to find out what happened to a man when one made him more acid or more alkaline . . . One might, of course, have tried experiments on a rabbit first, and some work had been done along these lines; but it is difficult to be sure how a rabbit feels at any time. Indeed, some rabbits make no serious attempt to cooperate with one.

". . . A human colleague and I therefore began experiments on one another . . . My colleague Dr. H.W. Davies and I made ourselves alkaline by over-breathing and by eating anything up to three ounces of bicarbonate of soda. We made ourselves acid by sitting in an airtight room with between six and seven per cent of carbon dioxide in the air. This makes one breathe as if one had just completed a boat-race, and also gives one a rather violent headache . . . Two hours was as long as any one wanted to stay in the carbon dioxide, even if the gas chamber at our disposal had not retained an ineradicable odour of 'yellow cross gas' from some wartime experiments, which made one weep gently every time one entered it. The most obvious thing to try was drinking hydrochloric acid. If one takes it strong it dissolves one's teeth and burns one's throat, whereas I wanted to let it diffuse gently all through my body. The strongest I ever cared to drink was about one part of the commercial strong acid in a hundred of water, but a pint of that was enough for me, as it irritated my throat and stomach, while my calculations showed that I needed a gallon and a half to get the effect I wanted . . . I argued that if one ate ammonium chloride, it would partly break up in the body, liberating hydrochloric acid. This proved to be correct . . . the liver turns ammonia into a harmless substance called urea before it reaches the heart and brain on absorption from the gut. The hydrochloric acid is left behind and combines with sodium bicarbonate, which exists in all the tissues, producing sodium chloride and carbon dioxide. I have had this gas produced in me in this way at the rate of six quarts an hour (though not for an hour on end at that rate) . . .

"I was quite satisfied to have reproduced in myself the type of shortness of breath which occurs in the terminal stages of kidney disease and diabetes. This had long been known to be due to acid poisoning, but in each case the acid poisoning is complicated by other chemical abnormalities, and it had been rather uncertain which of the symptoms were due to the acid as such.

"The scene now shifts to Heidelberg, where Freudenberg and György were studying tetany in babies . . . it occurred to them that it would be well worth trying the effect of making the body unusually acid. For tetany had occasionally been observed in patients who had been treated for other complaints by very large doses of sodium bicarbonate, or had lost large amounts of hydrochloric acid by constant vomiting; and if alkalinity of the tissues will produce tetany, acidity may be expected to cure it. Unfortunately, one could hardly try to cure a dying baby by shutting it up in a room full of carbonic acid, and still less would one give it hydrochloric acid to drink; so nothing had come of their idea, and they were using lime salts, which are not very easily absorbed, and which upset the digestion, but certainly benefit many cases of tetany.

"However, the moment they read my paper on the effects of ammonium chloride, they began giving it to babies, and were delighted to find that the tetany cleared up in a few hours. Since then it has been used with effect both in England and America, both on children and adults. It does not remove the cause, but it brings the patient into a condition from which he has a very fair chance of recovering."

with the same consequences as for diabetes. Very heavy exertion, such as a sprint by runners or cyclists, leads to temporary accumulation of lactic acid in the blood. Kidney failure results in a diminished capacity to regulate bicarbonate levels. Lung diseases (such as emphysema, pneumonia, and asthma) reduce the capacity to dispose of the CO_2 produced by fuel oxidation in the tissues, with the resulting accumulation of H_2CO_3. Acidosis is treated by dealing with the underlying condition—administering insulin to people with diabetes, and steroids or antibiotics to people with lung disease. Severe acidosis can be reversed by administering bicarbonate solution intravenously. ■

WORKED EXAMPLE 2–7 Treatment of Acidosis with Bicarbonate

Why does intravenous administration of a bicarbonate solution raise the plasma pH?

Solution: The ratio of $[HCO_3^-]$ to $[CO_2(d)]$ determines the pH of the bicarbonate buffer, according to the equation

$$pH = 6.1 + \log \frac{[HCO_3^-]}{(0.23 \times pCO_2)}$$

If $[HCO_3^-]$ is increased with no change in pCO_2, the pH will rise.

SUMMARY 2.3 Buffering against pH Changes in Biological Systems

- A mixture of a weak acid (or base) and its salt resists changes in pH caused by the addition of H^+ or OH^-. The mixture thus functions as a buffer.
- The pH of a solution of a weak acid (or base) and its salt is given by the Henderson-Hasselbalch equation: $pH = pK_a + \log \frac{[A^-]}{[HA]}$.
- In cells and tissues, phosphate and bicarbonate buffer systems maintain intracellular and extracellular fluids at their optimum (physiological) pH, which is usually close to pH 7. Enzymes generally work optimally at this pH.
- Medical conditions that lower the pH of blood, causing acidosis, or raise it, causing alkalosis, can be life threatening.

2.4 Water as a Reactant

Water is not just the solvent in which the chemical reactions of living cells occur; it is very often a direct participant in those reactions. The formation of ATP from ADP and inorganic phosphate is an example of a **condensation reaction** in which the elements of water are eliminated **(Fig. 2–23)**. The reverse of this reaction—cleavage accompanied by the addition of the elements of water—is a **hydrolysis reaction**. Hydrolysis reactions are also responsible for the enzymatic depolymerization of proteins, carbohydrates, and nucleic acids. Hydrolysis reactions, catalyzed by enzymes called **hydrolases**, are almost invariably exergonic; by producing two molecules from one, they lead to an increase in the randomness of the system. The formation of cellular polymers from their subunits by simple reversal of hydrolysis (that is, by condensation reactions) would be endergonic and therefore does not occur. As we shall see, cells circumvent this thermodynamic obstacle by coupling endergonic condensation reactions to exergonic processes, such as breakage of the anhydride bond in ATP.

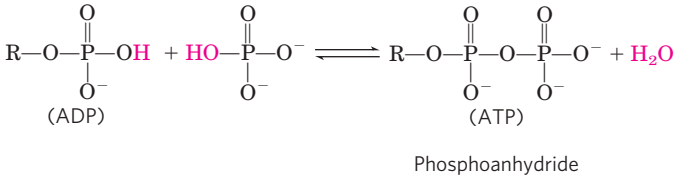


FIGURE 2–23 Participation of water in biological reactions. ATP is a phosphoanhydride formed by a condensation reaction (loss of the elements of water) between ADP and phosphate. R represents adenosine monophosphate (AMP). This condensation reaction requires energy. The hydrolysis of (addition of the elements of water to) ATP to form ADP and phosphate releases an equivalent amount of energy. These condensation and hydrolysis reactions of ATP are just one example of the role of water as a reactant in biological processes.

You are (we hope!) consuming oxygen as you read. Water and carbon dioxide are the end products of the oxidation of fuels such as glucose. The overall reaction can be summarized as

$$\underset{\text{Glucose}}{C_6H_{12}O_6} + 6O_2 \longrightarrow 6CO_2 + 6H_2O$$

The "metabolic water" formed by oxidation of foods and stored fats is actually enough to allow some animals in very dry habitats (gerbils, kangaroo rats, camels) to survive for extended periods without drinking water.

The CO_2 produced by glucose oxidation is converted in erythrocytes to the more soluble HCO_3^-, in a reaction catalyzed by the enzyme carbonic anhydrase:

$$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$$

In this reaction, water not only is a substrate but also functions in proton transfer by forming a network of hydrogen-bonded water molecules through which proton hopping occurs (Fig. 2–14).

Green plants and algae use the energy of sunlight to split water in the process of photosynthesis:

$$2H_2O + 2A \xrightarrow{\text{light}} O_2 + 2AH_2$$

In this reaction, A is an electron-accepting species, which varies with the type of photosynthetic organism, and water serves as the electron donor in an oxidation-reduction sequence (see Fig. 19–59) that is fundamental to all life.

SUMMARY 2.4 Water as a Reactant

- Water is both the solvent in which metabolic reactions occur and a reactant in many biochemical processes, including hydrolysis, condensation, and oxidation-reduction reactions.

2.5 The Fitness of the Aqueous Environment for Living Organisms

Organisms have effectively adapted to their aqueous environment and have evolved means of exploiting the unusual properties of water. The high specific heat of water (the heat energy required to raise the temperature of 1 g of water by 1 °C) is useful to cells and organisms because it allows water to act as a "heat buffer," keeping the temperature of an organism relatively constant as the temperature of the surroundings fluctuates and as heat is generated as a byproduct of metabolism. Furthermore, some vertebrates exploit the high heat of vaporization of water (Table 2–1) by using (thus losing) excess body heat to evaporate sweat. The high degree of internal cohesion of liquid water, due to hydrogen bonding, is exploited by plants as a means of transporting dissolved nutrients from the roots to the leaves during the process of transpiration. Even the density of ice, lower than that of liquid water, has important biological consequences in the life cycles of aquatic organisms. Ponds freeze from

the top down, and the layer of ice at the top insulates the water below from frigid air, preventing the pond (and the organisms in it) from freezing solid. Most fundamental to all living organisms is the fact that many physical and biological properties of cell macromolecules, particularly the proteins and nucleic acids, derive from their interactions with water molecules of the surrounding medium. The influence of water on the course of biological evolution has been profound and determinative. If life forms have evolved elsewhere in the universe, they are unlikely to resemble those of Earth unless liquid water is plentiful in their planet of origin.

Key Terms

Terms in bold are defined in the glossary.

hydrogen bond 48
bond energy 48
hydrophilic 50
hydrophobic 50
amphipathic 52
micelle 52
hydrophobic interactions 53
van der Waals interactions 53
osmolarity 56
osmosis 56
isotonic 56
hypertonic 56
hypotonic 56
equilibrium constant (K_{eq}) 59
ion product of water (K_w) 59
pH 60
acidosis 61
alkalosis 61
conjugate acid-base pair 61
acid dissociation constant (K_a) 62
pK_a 62
titration curve 62
buffer 64
buffering region 64
Henderson-Hasselbalch equation 64
condensation 69
hydrolysis 69

Further Reading

General

Ball, P. (2001) *Life's Matrix: A Biography of Water*, University of California Press, Berkeley, CA.

A highly accessible and entertaining description of water, from the Big Bang to its many roles in the chemistry of life.

Denny, M.W. (1993) *Air and Water: The Biology and Physics of Life's Media*, Princeton University Press, Princeton, NJ.

A wonderful investigation of the biological relevance of the properties of water.

Eisenberg, D. & Kauzmann, W. (1969) *The Structure and Properties of Water*, Oxford University Press, New York.

An advanced, classic treatment of the physical chemistry of water and hydrophobic interactions.

Franks, F. & Mathias, S.F. (eds). (1982) *Biophysics of Water*, John Wiley & Sons, Inc., New York.

A large collection of papers on the structure of pure water and of the cytoplasm.

Gerstein, M. & Levitt, M. (1998) Simulating water and the molecules of life. *Sci. Am.* **279** (November), 100–105.

A well-illustrated description of the use of computer simulation to study the biologically important association of water with proteins and nucleic acids.

Kandori, H. (2000) Role of internal water molecules in bacteriorhodopsin. *Biochim. Biophys. Acta* **1460**, 177–191.

Intermediate-level review of the role of an internal chain of water molecules in proton movement through this protein.

Kornblatt, J. & Kornblatt, J. (1997) The role of water in recognition and catalysis by enzymes. *Biochemist* **19** (3), 14–17.

A short, useful summary of the ways in which bound water influences the structure and activity of proteins.

Lemieux, R.U. (1996) How water provides the impetus for molecular recognition in aqueous solution. *Acc. Chem. Res.* **29**, 373–380.

A study of the role of water in the binding of a sugar to its binding protein.

Luecke, H. (2000) Atomic resolution structures of bacteriorhodopsin photocycle intermediates: the role of discrete water molecules in the function of this light-driven ion pump. *Biochim. Biophys. Acta* **1460**, 133–156.

Advanced review of a proton pump that employs an internal chain of water molecules.

Nicolls, P. (2000) Introduction: the biology of the water molecule. *Cell. Mol. Life Sci.* **57**, 987–992.

A short review of the properties of water, introducing several excellent advanced reviews published in the same issue (see especially Pocker, 2000, and Rand et al., 2000, below).

Symons, M.C. (2000) Spectroscopy of aqueous solutions: protein and DNA interactions with water. *Cell. Mol. Life Sci.* **57**, 999–1007.

Wiggins, P.M. (1990) Role of water in some biological processes. *Microbiol. Rev.* **54**, 432–449.

A review of water in biology, including discussion of the physical structure of liquid water, its interaction with biomolecules, and the state of water in living cells.

Osmosis

Cayley, D.S., Guttman, H.J., & Record, M.T., Jr. (2000) Biophysical characterization of changes in amounts and activity of *Escherichia coli* cell and compartment water and turgor pressure in response to osmotic stress. *Biophys. J.* **78**, 1748–1764.

An advanced physical investigation of the cytoplasmic water fraction of the bacterium *Escherichia coli* grown in media of different osmolarities. (See also Record et al., 1998, below.)

Rand, R.P., Parsegian, V.A., & Rau, D.C. (2000) Intracellular osmotic action. *Cell. Mol. Life Sci.* **57**, 1018–1032.

Review of the roles of water in enzyme catalysis as revealed by studies in water-poor solutes.

Record, M.T., Jr., Courtenay, E.S., Cayley, D.S., & Guttman, H.J. (1998) Responses of *E. coli* to osmotic stress: large changes in amounts of cytoplasmic solutes and water. *Trends Biochem. Sci.* **23**, 143–148.

Intermediate-level review of the ways in which a bacterial cell counters changes in the osmolarity of its surroundings. (See also Cayley et al., 2000, above.)

Zonia, L. & Munnik, T. (2007) Life under pressure: hydrostatic pressure in cell growth and function. *Trends Plant Sci.* **12**, 90–97.

Weak Interactions in Aqueous Systems

Baldwin, R.L. (2007) Energetics of protein folding. *J. Mol. Biol.* **371**, 283–301.

Advanced discussion of the thermodynamic factors, including weak interactions, that determine the course of protein folding.

Ball, P. (2008) Water as an active constituent in cell biology. *Chem. Rev.* **108**, 74–108.

An advanced discussion of the role of water in biological structure and function.

Blokzijl, W. & Engberts, J.B.F.N. (1993) Hydrophobic effects. Opinions and facts. *Angew. Chem. Int. Ed. Engl.* **32**, 1545–1579.

Advanced, monumental, and critical review.

Chaplin, M. (2006) Do we underestimate the importance of water in cell biology? *Nat. Rev. Mol. Cell Biol.* **7**, 861–866.

Fersht, A.R. (1987) The hydrogen bond in molecular recognition. *Trends Biochem. Sci.* **12**, 301–304.
A clear, brief, quantitative discussion of the contribution of hydrogen bonding to molecular recognition and enzyme catalysis.

Jeffrey, G.A. (1997) *An Introduction to Hydrogen Bonding*, Oxford University Press, New York.
A detailed, advanced discussion of the structure and properties of hydrogen bonds, including those in water and biomolecules.

Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1–63.
Remains the classic statement of the importance of hydrophobic interactions in the stability of proteins.

Ladbury, J. (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* **3**, 973–980.

Levy, Y. & Onuchic, J.N. (2006) Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389–415.
An advanced discussion of the role of water in protein structure.

Martin, T.W. & Derewenda, Z.S. (1999) The name is bond—H bond. *Nat. Struct. Biol.* **6**, 403–406.
Brief review of the evidence that hydrogen bonds have some covalent character.

Pace, C.N. (2009) Energetics of protein hydrogen bonds. *Nat. Struct. Mol. Biol*. **16**, 681–682.
Brief account of the historical contributions to understanding the strength of hydrogen bonds in proteins.

Pocker, Y. (2000) Water in enzyme reactions: biophysical aspects of hydration-dehydration processes. *Cell. Mol. Life Sci.* **57**, 1008–1017.
Review of the role of water in enzyme catalysis, with carbonic anhydrase as the featured example.

Schwabe, J.W.R. (1997) The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* **7**, 126–134.
An examination of the important role of water in both the specificity and the affinity of protein-DNA interactions.

Stillinger, F.H. (1980) Water revisited. *Science* **209**, 451–457.
A short review of the physical structure of water, including the importance of hydrogen bonding and the nature of hydrophobic interactions.

Tanford, C. (1978) The hydrophobic effect and the organization of living matter. *Science* **200**, 1012–1018.
A classic review of the chemical and energetic bases for hydrophobic interactions between biomolecules in aqueous solutions.

Weak Acids, Weak Bases, and Buffers: Problems for Practice

Segel, I.H. (1976) *Biochemical Calculations*, 2nd edn, John Wiley & Sons, Inc., New York.

Problems

1. Solubility of Ethanol in Water Explain why ethanol (CH_3CH_2OH) is more soluble in water than is ethane (CH_3CH_3).

2. Calculation of pH from Hydrogen Ion Concentration What is the pH of a solution that has an H^+ concentration of (a) 1.75×10^{-5} mol/L; (b) 6.50×10^{-10} mol/L; (c) 1.0×10^{-4} mol/L; (d) 1.50×10^{-5} mol/L?

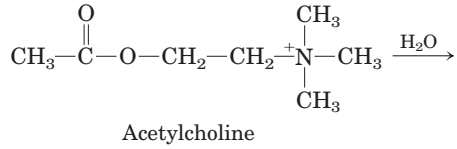
3. Calculation of Hydrogen Ion Concentration from pH What is the H^+ concentration of a solution with pH of (a) 3.82; (b) 6.52; (c) 11.11?

4. Acidity of Gastric HCl In a hospital laboratory, a 10.0 mL sample of gastric juice, obtained several hours after a meal, was titrated with 0.1 M NaOH to neutrality; 7.2 mL of NaOH was required. The patient's stomach contained no ingested food or drink; thus assume that no buffers were present. What was the pH of the gastric juice?

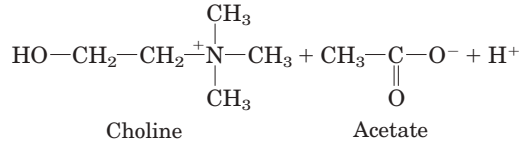
5. Calculation of the pH of a Strong Acid or Base (a) Write out the acid dissociation reaction for hydrochloric acid. (b) Calculate the pH of a solution of 5.0×10^{-4} M HCl. (c) Write out the acid dissociation reaction for sodium hydroxide. (d) Calculate the pH of a solution of 7.0×10^{-5} M NaOH.

6. Calculation of pH from Concentration of Strong Acid Calculate the pH of a solution prepared by diluting 3.0 mL of 2.5 M HCl to a final volume of 100 mL with H_2O.

7. Measurement of Acetylcholine Levels by pH Changes The concentration of acetylcholine (a neurotransmitter) in a sample can be determined from the pH changes that accompany its hydrolysis. When the sample is incubated with the enzyme acetylcholinesterase, acetylcholine is converted to choline and acetic acid, which dissociates to yield acetate and a hydrogen ion:



Acetylcholine



Choline Acetate

In a typical analysis, 15 mL of an aqueous solution containing an unknown amount of acetylcholine had a pH of 7.65. When incubated with acetylcholinesterase, the pH of the solution decreased to 6.87. Assuming there was no buffer in the assay mixture, determine the number of moles of acetylcholine in the 15 mL sample.

8. Physical Meaning of pK_a Which of the following aqueous solutions has the lowest pH: 0.1 M HCl; 0.1 M acetic acid ($pK_a = 4.86$); 0.1 M formic acid ($pK_a = 3.75$)?

9. Meanings of K_a and pK_a (a) Does a strong acid have a greater or lesser tendency to lose its proton than a weak acid? (b) Does the strong acid have a higher or lower K_a than the weak acid? (c) Does the strong acid have a higher or lower pK_a than the weak acid?

10. Simulated Vinegar One way to make vinegar (*not* the preferred way) is to prepare a solution of acetic acid, the sole acid component of vinegar, at the proper pH (see Fig. 2–15)

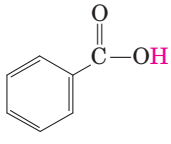
and add appropriate flavoring agents. Acetic acid (M_r 60) is a liquid at 25 °C, with a density of 1.049 g/mL. Calculate the volume that must be added to distilled water to make 1 L of simulated vinegar (see Fig. 2–16).

11. Identifying the Conjugate Base Which is the conjugate base in each of the pairs below?

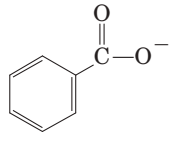
(a) RCOOH, $RCOO^-$
(b) RNH_2, RNH_3^+
(c) $H_2PO_4^-$, H_3PO_4
(d) H_2CO_3, HCO_3^-

12. Calculation of the pH of a Mixture of a Weak Acid and Its Conjugate Base Calculate the pH of a dilute solution that contains a molar ratio of potassium acetate to acetic acid ($pK_a = 4.76$) of (a) 2:1; (b) 1:3; (c) 5:1; (d) 1:1; (e) 1:10.

13. Effect of pH on Solubility The strongly polar, hydrogen-bonding properties of water make it an excellent solvent for ionic (charged) species. By contrast, nonionized, nonpolar organic molecules, such as benzene, are relatively insoluble in water. In principle, the aqueous solubility of any organic acid or base can be increased by converting the molecules to charged species. For example, the solubility of benzoic acid in water is low. The addition of sodium bicarbonate to a mixture of water and benzoic acid raises the pH and deprotonates the benzoic acid to form benzoate ion, which is quite soluble in water.

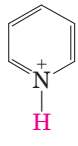


Benzoic acid
$pK_a \approx 5$



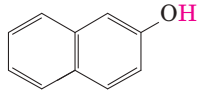
Benzoate ion

Are the following compounds more soluble in an aqueous solution of 0.1 M NaOH or 0.1 M HCl? (The dissociable protons are shown in red.)



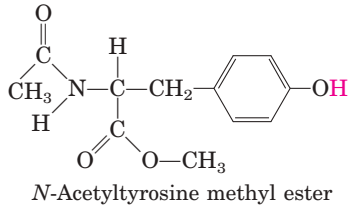
Pyridine ion
$pK_a \approx 5$

(a)



β-Naphthol
$pK_a \approx 10$

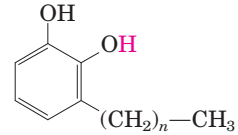
(b)



N-Acetyltyrosine methyl ester
$pK_a \approx 10$

(c)

14. Treatment of Poison Ivy Rash The components of poison ivy and poison oak that produce the characteristic itchy rash are catechols substituted with long-chain alkyl groups.

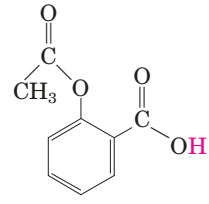


$pK_a \approx 8$

If you were exposed to poison ivy, which of the treatments below would you apply to the affected area? Justify your choice.

(a) Wash the area with cold water.
(b) Wash the area with dilute vinegar or lemon juice.
(c) Wash the area with soap and water.
(d) Wash the area with soap, water, and baking soda (sodium bicarbonate).

15. pH and Drug Absorption Aspirin is a weak acid with a pK_a of 3.5 (the ionizable H is shown in red):



It is absorbed into the blood through the cells lining the stomach and the small intestine. Absorption requires passage through the plasma membrane, the rate of which is determined by the polarity of the molecule: charged and highly polar molecules pass slowly, whereas neutral hydrophobic ones pass rapidly. The pH of the stomach contents is about 1.5, and the pH of the contents of the small intestine is about 6. Is more aspirin absorbed into the bloodstream from the stomach or from the small intestine? Clearly justify your choice.

16. Calculation of pH from Molar Concentrations What is the pH of a solution containing 0.12 mol/L of NH_4Cl and 0.03 mol/L of NaOH (pK_a of NH_4^+/NH_3 is 9.25)?

17. Calculation of pH after Titration of Weak Acid A compound has a pK_a of 7.4. To 100 mL of a 1.0 M solution of this compound at pH 8.0 is added 30 mL of 1.0 M hydrochloric acid. What is the pH of the resulting solution?

18. Properties of a Buffer The amino acid glycine is often used as the main ingredient of a buffer in biochemical experiments. The amino group of glycine, which has a pK_a of 9.6, can exist either in the protonated form ($—NH_3^+$) or as the free base ($—NH_2$), because of the reversible equilibrium

$$R—NH_3^+ \rightleftharpoons R—NH_2 + H^+$$

(a) In what pH range can glycine be used as an effective buffer due to its amino group?

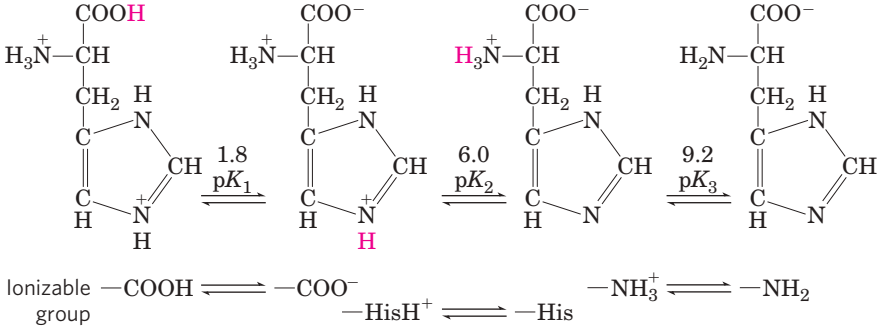
(b) In a 0.1 M solution of glycine at pH 9.0, what fraction of glycine has its amino group in the $—NH_3^+$ form?

(c) How much 5 M KOH must be added to 1.0 L of 0.1 M glycine at pH 9.0 to bring its pH to exactly 10.0?

(d) When 99% of the glycine is in its $—NH_3^+$ form, what is the numerical relation between the pH of the solution and the pK_a of the amino group?

19. Calculation of the pK_a of an Ionizable Group by Titration The pK_a values of a compound with two ionizable groups are pK_1 = 4.10 and pK_2 between 7 and 10. A biochemist has 10 mL of a 1.0 M solution of this compound at a pH of 8.00. She adds 10.0 mL of 1.00 M HCl, which changes the pH to 3.20. What is pK_2?

20. Calculation of the pH of a Solution of a Polyprotic Acid Histidine has ionizable groups with pK_a values of 1.8, 6.0, and 9.2, as shown below (His = imidazole group). A biochemist makes up 100 mL of a 0.100 M solution of histidine at a pH of 5.40. She then adds 40 mL of 0.10 M HCl. What is the pH of the resulting solution?



21. Calculation of the Original pH from the Final pH after Titration A biochemist has 100 mL of a 0.10 M solution of a weak acid with a pK_a of 6.3. She adds 6.0 mL of 1.0 M HCl, which changes the pH to 5.7. What was the pH of the original solution?

22. Preparation of a Phosphate Buffer What molar ratio of HPO_4^{2-} to $H_2PO_4^-$ in solution would produce a pH of 7.0? Phosphoric acid (H_3PO_4), a triprotic acid, has three pK_a values: 2.14, 6.86, and 12.4. Hint: Only one of the pK_a values is relevant here.

23. Preparation of Standard Buffer for Calibration of a pH Meter The glass electrode used in commercial pH meters gives an electrical response proportional to the concentration of hydrogen ion. To convert these responses to a pH reading, the electrode must be calibrated against standard solutions of known H^+ concentration. Determine the weight in grams of sodium dihydrogen phosphate ($NaH_2PO_4 \cdot H_2O$; FW 138) and disodium hydrogen phosphate (Na_2HPO_4; FW 142) needed to prepare 1 L of a standard buffer at pH 7.00 with a total phosphate concentration of 0.100 M (see Fig. 2–16). See problem 22 for the pK_a values of phosphoric acid.

24. Calculation of Molar Ratios of Conjugate Base to Weak Acid from pH For a weak acid with a pK_a of 6.0, calculate the ratio of conjugate base to acid at a pH of 5.0.

25. Preparation of Buffer of Known pH and Strength Given 0.10 M solutions of acetic acid (pK_a = 4.76) and sodium acetate, describe how you would go about preparing 1.0 L of 0.10 M acetate buffer of pH 4.00.

26. Choice of Weak Acid for a Buffer Which of these compounds would be the best buffer at pH 5.0: formic acid (pK_a = 3.8), acetic acid (pK_a = 4.76), or ethylamine (pK_a = 9.0)? Briefly justify your answer.

27. Working with Buffers A buffer contains 0.010 mol of lactic acid (pK_a = 3.86) and 0.050 mol of sodium lactate per liter. (a) Calculate the pH of the buffer. (b) Calculate the change in pH when 5 mL of 0.5 M HCl is added to 1 L of the buffer. (c) What pH change would you expect if you added the same quantity of HCl to 1 L of pure water?

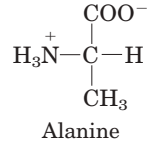
28. Use of Molar Concentrations to Calculate pH What is the pH of a solution that contains 0.20 M sodium acetate and 0.60 M acetic acid (pK_a = 4.76)?

29. Preparation of an Acetate Buffer Calculate the concentrations of acetic acid (pK_a = 4.76) and sodium acetate necessary to prepare a 0.2 M buffer solution at pH 5.0.

30. pH of Insect Defensive Secretion You have been observing an insect that defends itself from enemies by secreting a caustic liquid. Analysis of the liquid shows it to have a total concentration of formate plus formic acid ($K_a = 1.8 \times 10^{-4}$) of 1.45 M; the concentration of formate ion is 0.015 M. What is the pH of the secretion?

31. Calculation of pK_a An unknown compound, X, is thought to have a carboxyl group with a pK_a of 2.0 and another ionizable group with a pK_a between 5 and 8. When 75 mL of 0.1 M NaOH is added to 100 mL of a 0.1 M solution of X at pH 2.0, the pH increases to 6.72. Calculate the pK_a of the second ionizable group of X.

32. Ionic Forms of Alanine Alanine is a diprotic acid that can undergo two dissociation reactions (see Table 3–1 for pK_a values). (a) Given the structure of the partially protonated form (or zwitterion; see Fig. 3–9) below, draw the chemical structures of the other two forms of alanine that predominate in aqueous solution: the fully protonated form and the fully deprotonated form.



Alanine

Of the three possible forms of alanine, which would be present at the highest concentration in solutions of the following pH: (b) 1.0; (c) 6.2; (d) 8.02; (e) 11.9. Explain your answers in terms of pH relative to the two pK_a values.

33. Control of Blood pH by Respiratory Rate

(a) The partial pressure of CO_2 in the lungs can be varied rapidly by the rate and depth of breathing. For example, a common remedy to alleviate hiccups is to increase the concentration of CO_2 in the lungs. This can be achieved by holding one's breath, by very slow and shallow breathing (hypoventilation), or by breathing in and out of a paper bag. Under such conditions, pCO_2 in the air space of the lungs rises above normal.

Qualitatively explain the effect of these procedures on the blood pH.

(b) A common practice of competitive short-distance runners is to breathe rapidly and deeply (hyperventilate) for about half a minute to remove CO_2 from their lungs just before the race begins. Blood pH may rise to 7.60. Explain why the blood pH increases.

(c) During a short-distance run, the muscles produce a large amount of lactic acid ($CH_3CH(OH)COOH$; $K_a = 1.38 \times 10^{-4}$ M) from their glucose stores. In view of this fact, why might hyperventilation before a dash be useful?

34. Calculation of Blood pH from CO_2 and Bicarbonate Levels Calculate the pH of a blood plasma sample with a total CO_2 concentration of 26.9 mM and bicarbonate concentration of 25.6 mM. Recall from page 67 that the relevant pK_a of carbonic acid is 6.1.

35. Effect of Holding One's Breath on Blood pH The pH of the extracellular fluid is buffered by the bicarbonate/carbonic acid system. Holding your breath can increase the concentration of $CO_2(g)$ in the blood. What effect might this have on the pH of the extracellular fluid? Explain by showing the relevant equilibrium equation(s) for this buffer system.

Data Analysis Problem

36. "Switchable" Surfactants Hydrophobic molecules do not dissolve well in water. Given that water is a very commonly used solvent, this makes certain processes very difficult: washing oily food residue off dishes, cleaning up spilled oil, keeping the oil and water phases of salad dressings well mixed, and carrying out chemical reactions that involve both hydrophobic and hydrophilic components.

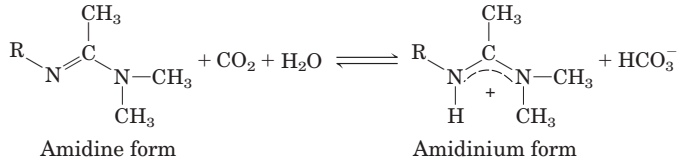
Surfactants are a class of amphipathic compounds that includes soaps, detergents, and emulsifiers. With the use of surfactants, hydrophobic compounds can be suspended in aqueous solution by forming micelles (see Fig. 2–7). A micelle has a hydrophobic core consisting of the hydrophobic compound and the hydrophobic "tails" of the surfactant; the hydrophilic "heads" of the surfactant cover the surface of the micelle. A suspension of micelles is called an emulsion. The more hydrophilic the head group of the surfactant, the more powerful it is—that is, the greater its capacity to emulsify hydrophobic material.

When you use soap to remove grease from dirty dishes, the soap forms an emulsion with the grease that is easily removed by water through interaction with the hydrophilic head of the soap molecules. Likewise, a detergent can be used to emulsify spilled oil for removal by water. And emulsifiers in commercial salad dressings keep the oil suspended evenly throughout the water-based mixture.

There are some situations in which it would be very useful to have a "switchable" surfactant: a molecule that could be reversibly converted between a surfactant and a nonsurfactant.

(a) Imagine such a "switchable" surfactant existed. How would you use it to clean up and then recover the oil from an oil spill?

Liu et al. describe a prototypical switchable surfactant in their 2006 article "Switchable Surfactants." The switching is based on the following reaction:

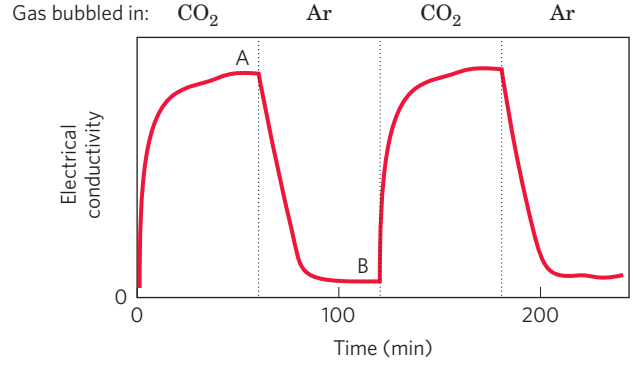


(b) Given that the pK_a of a typical amidinium ion is 12.4, in which direction (left or right) would you expect the equilibrium of the above reaction to lie? (See Fig. 2–16 for relevant pK_a values.) Justify your answer. Hint: Remember the reaction $H_2O + CO_2 \rightleftharpoons H_2CO_3$.

Liu and colleagues produced a switchable surfactant for which R = $C_{16}H_{33}$. They do not name the molecule in their article; for brevity, we'll call it s-surf.

(c) The amidinium form of s-surf is a powerful surfactant; the amidine form is not. Explain this observation.

Liu and colleagues found that they could switch between the two forms of s-surf by changing the gas that they bubbled through a solution of the surfactant. They demonstrated this switch by measuring the electrical conductivity of the s-surf solution; aqueous solutions of ionic compounds have higher conductivity than solutions of nonionic compounds. They started with a solution of the amidine form of s-surf in water. Their results are shown below; dotted lines indicate the switch from one gas to another.



(d) In which form is the majority of s-surf at point A? At point B?

(e) Why does the electrical conductivity rise from time 0 to point A?

(f) Why does the electrical conductivity fall from point A to point B?

(g) Explain how you would use s-surf to clean up and recover the oil from an oil spill.

Reference

Liu, Y., Jessop, P.G., Cunningham, M., Eckert, C.A., & Liotta, C.L. (2006) Switchable surfactants. *Science* **313**, 958–960.

3

Amino Acids, Peptides, and Proteins

Proteins mediate virtually every process that takes place in a cell, exhibiting an almost endless diversity of functions. To explore the molecular mechanism of a biological process, a biochemist almost inevitably studies one or more proteins. Proteins are the most abundant biological macromolecules, occurring in all cells and all parts of cells. Proteins also occur in great variety; thousands of different kinds may be found in a single cell. As the arbiters of molecular function, proteins are the most important final products of the information pathways discussed in Part III of this book. Proteins are the molecular instruments through which genetic information is expressed.

Relatively simple monomeric subunits provide the key to the structure of the thousands of different proteins. The proteins of every organism, from the simplest of bacteria to human beings, are constructed from the same ubiquitous set of 20 amino acids. Because each of these amino acids has a side chain with distinctive chemical properties, this group of 20 precursor molecules may be regarded as the alphabet in which the language of protein structure is written.

To generate a particular protein, amino acids are covalently linked in a characteristic linear sequence. What is most remarkable is that cells can produce proteins with strikingly different properties and activities by joining the same 20 amino acids in many different combinations and sequences. From these building blocks different organisms can make such widely diverse products as enzymes, hormones, antibodies, transporters, muscle fibers, the lens protein of the eye, feathers, spider webs, rhinoceros horn, milk proteins, antibiotics, mushroom poisons, and myriad other substances having distinct biological activities **(Fig. 3–1)**. Among these protein products, the enzymes are the most varied and specialized. As the catalysts of virtually all cellular reactions, enzymes are one of the keys to understanding the chemistry of life and thus provide a focal point for any course in biochemistry.

Protein structure and function are the topics of this and the next three chapters. Here, we begin with a description of the fundamental chemical properties of amino acids, peptides, and proteins. We also consider how a biochemist works with proteins.

FIGURE 3–1 Some functions of proteins. (a) The light produced by fireflies is the result of a reaction involving the protein luciferin and ATP, catalyzed by the enzyme luciferase (see Box 13–1). **(b)** Erythrocytes contain large amounts of the oxygen-transporting protein hemoglobin. **(c)** The protein keratin, formed by all vertebrates, is the chief structural component of hair, scales, horn, wool, nails, and feathers. The black rhinoceros is nearing extinction in the wild because of the belief prevalent in some parts of the world that a powder derived from its horn has aphrodisiac properties. In reality, the chemical properties of powdered rhinoceros horn are no different from those of powdered bovine hooves or human fingernails.

3.1 Amino Acids

Protein Architecture—Amino Acids Proteins are polymers of amino acids, with each **amino acid residue** joined to its neighbor by a specific type of covalent bond. (The term "residue" reflects the loss of the elements of water when one amino acid is joined to another.) Proteins can be broken down (hydrolyzed) to their constituent amino acids by a variety of methods, and the earliest studies of proteins naturally focused on the free amino acids derived from them. Twenty different amino acids are commonly found in proteins. The first to be discovered was asparagine, in 1806. The last of the 20 to be found, threonine, was not identified until 1938. All the amino acids have trivial or common names, in some cases derived from the source from which they were first isolated. Asparagine was first found in asparagus, and glutamate in wheat gluten; tyrosine was first isolated from cheese (its name is derived from the Greek *tyros*, "cheese"); and glycine (Greek *glykos*, "sweet") was so named because of its sweet taste.

Amino Acids Share Common Structural Features

All 20 of the common amino acids are α-amino acids. They have a carboxyl group and an amino group bonded to the same carbon atom (the α carbon) **(Fig. 3–2)**. They differ from each other in their side chains, or **R groups**, which vary in structure, size, and electric charge, and which influence the solubility of the amino acids in water. In addition to these 20 amino acids there are many less common ones. Some are residues modified after a protein has been synthesized; others are amino acids present in living organisms but not as constituents of proteins. The common amino acids of proteins have been assigned three-letter abbreviations and one-letter symbols (Table 3–1), which are used as shorthand to indicate the composition and sequence of amino acids polymerized in proteins.

KEY CONVENTION: The three-letter code is transparent, the abbreviations generally consisting of the first three letters of the amino acid name. The one-letter code was devised by Margaret Oakley Dayhoff, considered by many to be the founder of the field of bioinformatics. The one-letter code reflects an attempt to reduce the size of the data files (in an era of punch-card computing) used to describe amino acid sequences. It was designed to be easily memorized, and understanding its origin can help students do just that. For six amino acids (CHIMSV), the first letter of the amino acid name is unique and thus is used as the symbol. For five others (AGLPT), the first letter is not unique but is assigned to the amino acid that is most common in proteins (for example, leucine is more common than lysine). For another four, the letter used is phonetically suggestive (RFYW: aRginine, Fenyl-alanine, tYrosine, tWiptophan). The rest were harder to assign. Four (DNEQ) were assigned letters found within or suggested by their names (asparDic, asparagiNe, glutamEke, Q-tamine). That left lysine. Only a few letters were left in the alphabet, and K was chosen because it was the closest to L. ■

Margaret Oakley Dayhoff, 1925–1983

For all the common amino acids except glycine, the α carbon is bonded to four different groups: a carboxyl group, an amino group, an R group, and a hydrogen atom (Fig. 3–2; in glycine, the R group is another hydrogen atom). The α-carbon atom is thus a **chiral center** (p. 17). Because of the tetrahedral arrangement of the bonding orbitals around the α-carbon atom, the four different groups can occupy two unique spatial arrangements, and thus amino acids have two possible stereoisomers. Since they are nonsuperposable mirror images of each other **(Fig. 3–3)**, the two forms

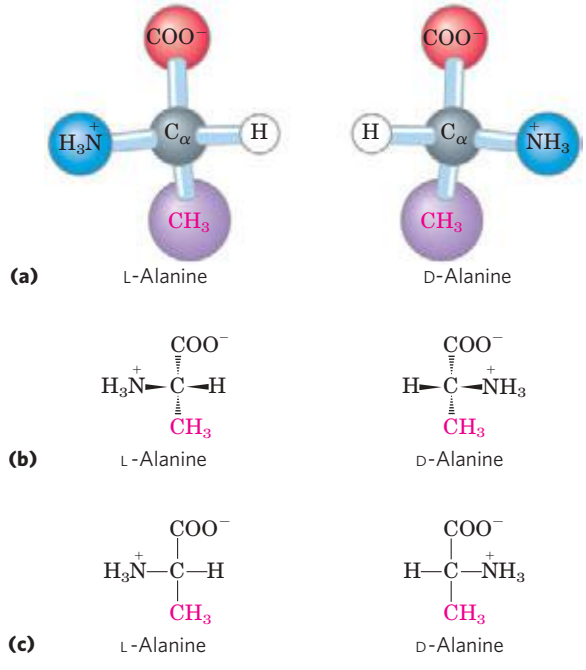


FIGURE 3–3 Stereoisomerism in α-amino acids. (a) The two stereoisomers of alanine, L- and D-alanine, are nonsuperposable mirror images of each other (enantiomers). **(b, c)** Two different conventions for showing the configurations in space of stereoisomers. In perspective formulas **(b)**, the solid wedge-shaped bonds project out of the plane of the paper, the dashed bonds behind it. In projection formulas **(c)**, the horizontal bonds are assumed to project out of the plane of the paper, the vertical bonds behind. However, projection formulas are often used casually and are not always intended to portray a specific stereochemical configuration.

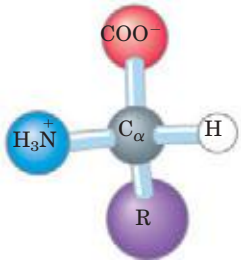


FIGURE 3–2 General structure of an amino acid. This structure is common to all but one of the α-amino acids. (Proline, a cyclic amino acid, is the exception.) The R group, or side chain (purple), attached to the α carbon (gray) is different in each amino acid.

TABLE 3–1 Properties and Conventions Associated with the Common Amino Acids Found in Proteins

Amino acid	Abbreviation/ symbol	M_r*	pK_a values pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R (R group)	pI	Hydropathy index†	Occurrence in proteins (%)‡
Nonpolar, aliphatic R groups								
Glycine	Gly G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala A	89	2.34	9.69		6.01	1.8	7.8
Proline	Pro P	115	1.99	10.96		6.48	−1.6	5.2
Valine	Val V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups								
Phenylalanine	Phe F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups								
Serine	Ser S	105	2.21	9.15		5.68	−0.8	6.8
Threonine	Thr T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine¶	Cys C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups								
Lysine	Lys K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups								
Aspartate	Asp D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu E	147	2.19	9.67	4.25	3.22	−3.5	6.3

*M_r values reflect the structures as shown in Figure 3–5. The elements of water (M_r 18) are deleted when the amino acid is incorporated into a polypeptide.

†A scale combining hydrophobicity and hydrophilicity of R groups. The values reflect the free energy (ΔG) of transfer of the amino acid side chain from a hydrophobic solvent to water. This transfer is favorable ($\Delta G < 0$; negative value in the index) for charged or polar amino acid side chains, and unfavorable ($\Delta G > 0$; positive value in the index) for amino acids with nonpolar or more hydrophobic side chains. See Chapter 11. From Kyte, J. & Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105–132.

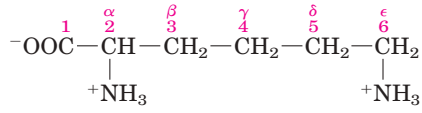
‡Average occurrence in more than 1,150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed.), pp. 599–623, Plenum Press, New York.

¶Cysteine is generally classified as polar despite having a positive hydropathy index. This reflects the ability of the sulfhydryl group to act as a weak acid and to form a weak hydrogen bond with oxygen or nitrogen.

represent a class of stereoisomers called **enantiomers** (see Fig. 1–20). All molecules with a chiral center are also **optically active**—that is, they rotate plane-polarized light (see Box 1–2).

KEY CONVENTION: Two conventions are used to identify the carbons in an amino acid—a practice that can be confusing. The additional carbons in an R group are commonly designated β, γ, δ, ε, and so forth, proceeding out from the α carbon. For most other organic molecules, carbon atoms are simply numbered from one end, giving highest priority (C-1) to the carbon with the substituent containing the atom of highest atomic number. Within this latter convention, the carboxyl

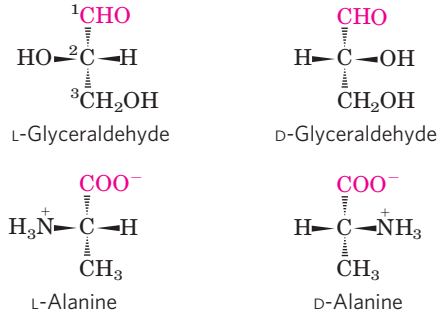
carbon of an amino acid would be C-1 and the α carbon would be C-2.



Lysine

In some cases, such as amino acids with heterocyclic R groups (such as histidine), the Greek lettering system is ambiguous and the numbering convention is therefore used. For branched amino acid side chains, equivalent carbons are given numbers after the Greek letters. Leucine thus has $\delta 1$ and $\delta 2$ carbons (see the structure in Fig. 3–5). ■

Special nomenclature has been developed to specify the **absolute configuration** of the four substituents of asymmetric carbon atoms. The absolute configurations of simple sugars and amino acids are specified by the **D, L system (Fig. 3–4)**, based on the absolute configuration of the three-carbon sugar glyceraldehyde, a convention proposed by Emil Fischer in 1891. (Fischer knew what groups surrounded the asymmetric carbon of glyceraldehyde but had to guess at their absolute configuration; he guessed right, as was later confirmed by x-ray diffraction analysis.) For all chiral compounds, stereoisomers having a configuration related to that of L-glyceraldehyde are designated L, and stereoisomers related to D-glyceraldehyde are designated D. The functional groups of L-alanine are matched with those of L-glyceraldehyde by aligning those that can be interconverted by simple, one-step chemical reactions. Thus the carboxyl group of L-alanine occupies the same position about the chiral carbon as does the aldehyde group of L-glyceraldehyde, because an aldehyde is readily converted to a carboxyl group via a one-step oxidation. Historically, the similar L and D designations were used for levorotatory (rotating plane-polarized light to the left) and dextrorotatory (rotating light to the right). However, not all L-amino acids are levorotatory, and the convention shown in Figure 3–4 was needed to avoid potential ambiguities about absolute configuration. By Fischer's convention, L and D refer *only* to the absolute configuration of the four substituents around the chiral carbon, not to optical properties of the molecule.



L-Glyceraldehyde D-Glyceraldehyde

L-Alanine D-Alanine

FIGURE 3–4 Steric relationship of the stereoisomers of alanine to the absolute configuration of L- and D-glyceraldehyde. In these perspective formulas, the carbons are lined up vertically, with the chiral atom in the center. The carbons in these molecules are numbered beginning with the terminal aldehyde or carboxyl carbon (red), 1 to 3 from top to bottom as shown. When presented in this way, the R group of the amino acid (in this case the methyl group of alanine) is always below the α carbon. L-Amino acids are those with the α-amino group on the left, and D-amino acids have the α-amino group on the right.

Another system of specifying configuration around a chiral center is the **RS system**, which is used in the systematic nomenclature of organic chemistry and describes more precisely the configuration of molecules with more than one chiral center (p. 18).

The Amino Acid Residues in Proteins Are L Stereoisomers

Nearly all biological compounds with a chiral center occur naturally in only one stereoisomeric form, either D or L. The amino acid residues in protein molecules are exclusively L stereoisomers. D-Amino acid residues have been found in only a few, generally small peptides, including some peptides of bacterial cell walls and certain peptide antibiotics.

It is remarkable that virtually all amino acid residues in proteins are L stereoisomers. When chiral compounds are formed by ordinary chemical reactions, the result is a racemic mixture of D and L isomers, which are difficult for a chemist to distinguish and separate. But to a living system, D and L isomers are as different as the right hand and the left. The formation of stable, repeating substructures in proteins (Chapter 4) generally requires that their constituent amino acids be of one stereochemical series. Cells are able to specifically synthesize the L isomers of amino acids because the active sites of enzymes are asymmetric, causing the reactions they catalyze to be stereospecific.

Amino Acids Can Be Classified by R Group

Knowledge of the chemical properties of the common amino acids is central to an understanding of biochemistry. The topic can be simplified by grouping the amino acids into five main classes based on the properties of their R groups (Table 3–1), particularly their **polarity**, or tendency to interact with water at biological pH (near pH 7.0). The polarity of the R groups varies widely, from nonpolar and hydrophobic (water-insoluble) to highly polar and hydrophilic (water-soluble). A few amino acids are somewhat difficult to characterize or do not fit perfectly in any one group, particularly glycine, histidine, and cysteine. Their assignments to particular groupings are the results of considered judgments rather than absolutes.

The structures of the 20 common amino acids are shown in **Figure 3–5**, and some of their properties are listed in Table 3–1. Within each class there are gradations of polarity, size, and shape of the R groups.

Nonpolar, Aliphatic R Groups The R groups in this class of amino acids are nonpolar and hydrophobic. The side

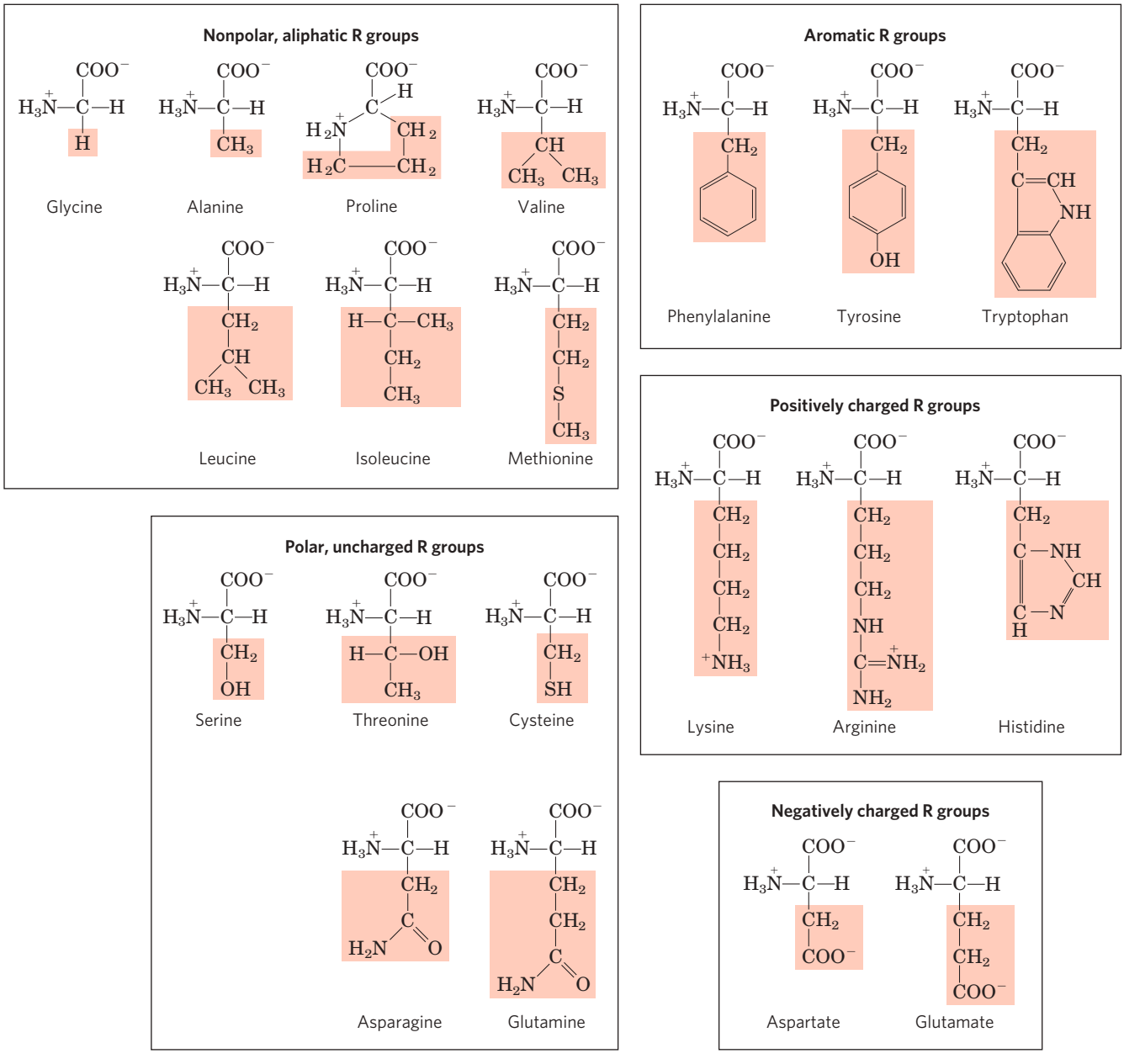


FIGURE 3–5 The 20 common amino acids of proteins. The structural formulas show the state of ionization that would predominate at pH 7.0. The unshaded portions are those common to all the amino acids; the shaded portions are the R groups. Although the R group of histidine is shown uncharged, its pK_a (see Table 3–1) is such that a small but significant fraction of these groups are positively charged at pH 7.0. The protonated form of histidine is shown above the graph in Figure 3–12b.

chains of **alanine**, **valine**, **leucine**, and **isoleucine** tend to cluster together within proteins, stabilizing protein structure by means of hydrophobic interactions. **Glycine** has the simplest structure. Although it is most easily grouped with the nonpolar amino acids, its very small side chain makes no real contribution to hydrophobic interactions. **Methionine**, one of the two sulfur-containing amino acids, has a slightly nonpolar thioether group in its side chain. **Proline** has an aliphatic side chain with a distinctive cyclic structure. The secondary amino (imino) group of proline residues is held in a rigid conformation that reduces the structural flexibility of polypeptide regions containing proline.

Aromatic R Groups **Phenylalanine**, **tyrosine**, and **tryptophan**, with their aromatic side chains, are relatively nonpolar (hydrophobic). All can participate in hydrophobic interactions. The hydroxyl group of tyrosine can form hydrogen bonds, and it is an important functional group in some enzymes. Tyrosine and tryptophan are significantly more polar than phenylalanine, because of the tyrosine hydroxyl group and the nitrogen of the tryptophan indole ring.

Tryptophan and tyrosine, and to a much lesser extent phenylalanine, absorb ultraviolet light (**Fig. 3–6**; see also Box 3–1). This accounts for the characteristic strong absorbance of light by most proteins at a wavelength of

280 nm, a property exploited by researchers in the characterization of proteins.

Polar, Uncharged R Groups The R groups of these amino acids are more soluble in water, or more hydrophilic, than those of the nonpolar amino acids, because they contain functional groups that form hydrogen bonds

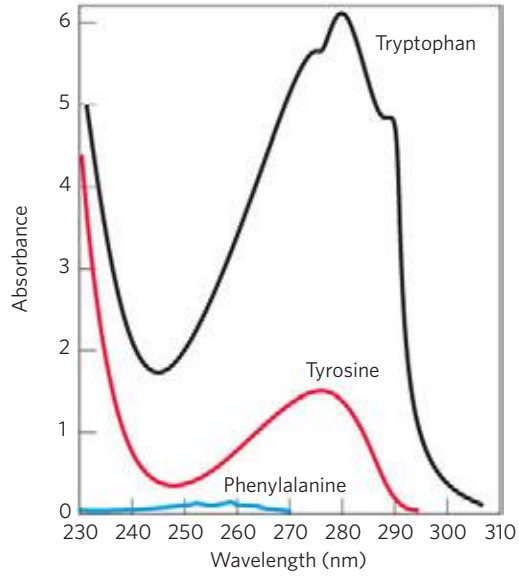


FIGURE 3–6 Absorption of ultraviolet light by aromatic amino acids. Comparison of the light absorption spectra of the aromatic amino acids tryptophan, tyrosine, and phenylalanine at pH 6.0. The amino acids are present in equimolar amounts (10^{-3} M) under identical conditions. The measured absorbance of tryptophan is more than four times that of tyrosine at a wavelength of 280 nm. Note that the maximum light absorption for both tryptophan and tyrosine occurs near 280 nm. Light absorption by phenylalanine generally contributes little to the spectroscopic properties of proteins.

BOX 3–1 METHODS Absorption of Light by Molecules: The Lambert-Beer Law

A wide range of biomolecules absorb light at characteristic wavelengths, just as tryptophan absorbs light at 280 nm (see Fig. 3–6). Measurement of light absorption by a spectrophotometer is used to detect and identify molecules and to measure their concentration in solution. The fraction of the incident light absorbed by a solution at a given wavelength is related to the thickness of the absorbing layer (path length) and the concentration of the absorbing species (Fig. 1). These two relationships are combined into the Lambert-Beer law,

$$\log \frac{I_0}{I} = \varepsilon c l$$

where I_0 is the intensity of the incident light, I is the intensity of the transmitted light, the ratio I/I_0 (the inverse of the ratio in the equation) is the transmittance, ε is the molar extinction coefficient (in units of liters per mole-centimeter), c is the concentration of the absorbing species (in moles per liter), and l is the path length of the light-absorbing sample (in centimeters). The Lambert-Beer law assumes that the incident light is parallel and monochromatic (of a single wavelength) and that the solvent and solute molecules are randomly oriented. The expression log (I_0/I) is called the **absorbance**, designated A.

It is important to note that each successive millimeter of path length of absorbing solution in a 1.0 cm cell absorbs not a constant amount but a constant fraction of the light that is incident upon it. However, with an absorbing layer of fixed path length, *the absorbance, A, is directly proportional to the concentration of the absorbing solute*.

The molar extinction coefficient varies with the nature of the absorbing compound, the solvent, and the wavelength, and also with pH if the light-absorbing species is in equilibrium with an ionization state that has different absorbance properties.

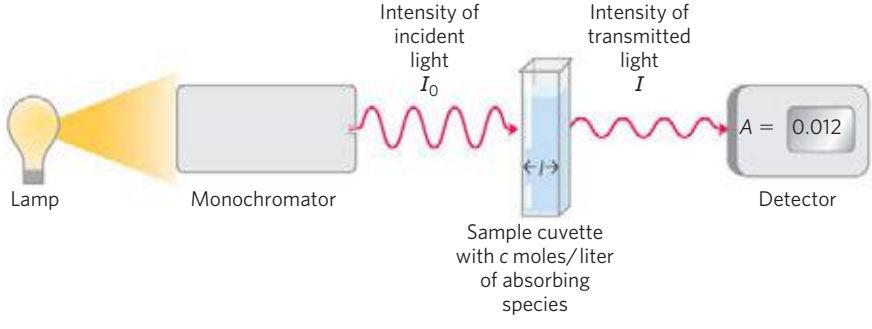


FIGURE 1 The principal components of a spectrophotometer. A light source emits light along a broad spectrum, then the monochromator selects and transmits light of a particular wavelength. The monochromatic light passes through the sample in a cuvette of path length l. The absorbance of the sample, log (I_0/I), is proportional to the concentration of the absorbing species. The transmitted light is measured by a detector.

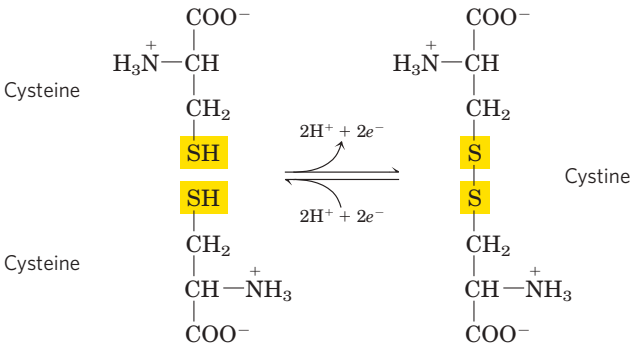


FIGURE 3–7 Reversible formation of a disulfide bond by the oxidation of two molecules of cysteine. Disulfide bonds between Cys residues stabilize the structures of many proteins.

with water. This class of amino acids includes **serine**, **threonine**, **cysteine**, **asparagine**, and **glutamine**. The polarity of serine and threonine is contributed by their hydroxyl groups, and that of asparagine and glutamine by their amide groups. Cysteine is an outlier here because its polarity, contributed by its sulfhydryl group, is quite modest. Cysteine is a weak acid and can make weak hydrogen bonds with oxygen or nitrogen.

Asparagine and glutamine are the amides of two other amino acids also found in proteins—aspartate and glutamate, respectively—to which asparagine and glutamine are easily hydrolyzed by acid or base. Cysteine is readily oxidized to form a covalently linked dimeric amino acid called **cystine**, in which two cysteine molecules or residues are joined by a disulfide bond **(Fig. 3–7)**. The disulfide-linked residues are strongly hydrophobic (nonpolar). Disulfide bonds play a special role in the structures of many proteins by forming covalent links between parts of a polypeptide molecule or between two different polypeptide chains.

Positively Charged (Basic) R Groups The most hydrophilic R groups are those that are either positively or negatively charged. The amino acids in which the R groups have significant positive charge at pH 7.0 are **lysine**, which has a second primary amino group at the ε position on its aliphatic chain; **arginine**, which has a positively charged guanidinium group; and **histidine**, which has an aromatic imidazole group. As the only common amino acid having an ionizable side chain with pK_a near neutrality, histidine may be positively charged (protonated form) or uncharged at pH 7.0. His residues facilitate many enzyme-catalyzed reactions by serving as proton donors/acceptors.

Negatively Charged (Acidic) R Groups The two amino acids having R groups with a net negative charge at pH 7.0 are **aspartate** and **glutamate**, each of which has a second carboxyl group.

Uncommon Amino Acids Also Have Important Functions

In addition to the 20 common amino acids, proteins may contain residues created by modification of common residues already incorporated into a polypeptide **(Fig. 3–8a)**. Among these uncommon amino acids are **4-hydroxyproline**, a derivative of proline, and **5-hydroxylysine**, derived from lysine. The former is found in plant cell wall proteins, and both are found in collagen, a fibrous protein of connective tissues. **6-*N*-Methyllysine** is a constituent of myosin, a contractile protein of muscle. Another important uncommon amino acid is **γ-carboxyglutamate**, found in the blood-clotting protein prothrombin and in certain other proteins that bind Ca^{2+} as part of their biological function. More complex is **desmosine**, a derivative of four Lys residues, which is found in the fibrous protein elastin.

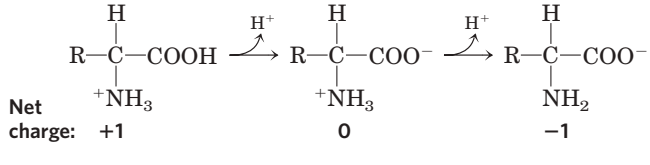
Selenocysteine is a special case. This rare amino acid residue is introduced during protein synthesis rather than created through a postsynthetic modification. It contains selenium rather than the sulfur of cysteine. Actually derived from serine, selenocysteine is a constituent of just a few known proteins.

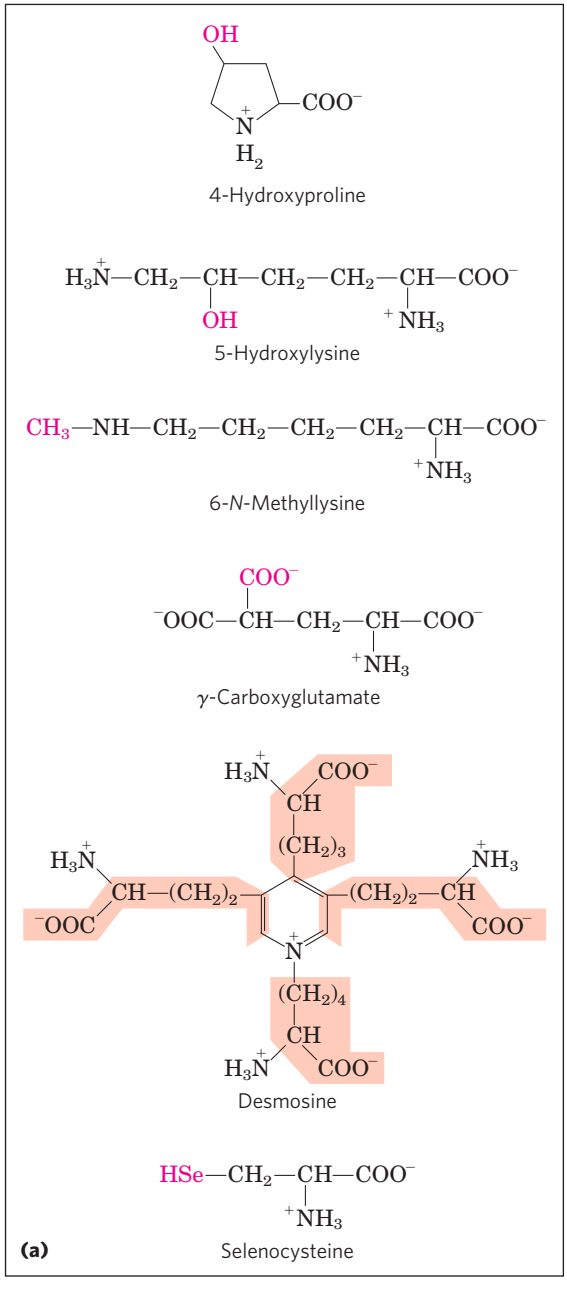
Some amino acid residues in a protein may be modified transiently to alter the protein's function. The addition of phosphoryl, methyl, acetyl, adenylyl, ADP-ribosyl, or other groups to particular amino acid residues can increase or decrease a protein's activity (Fig. 3–8b). Phosphorylation is a particularly common regulatory modification. Covalent modification as a protein regulatory strategy is discussed in more detail in Chapter 6.

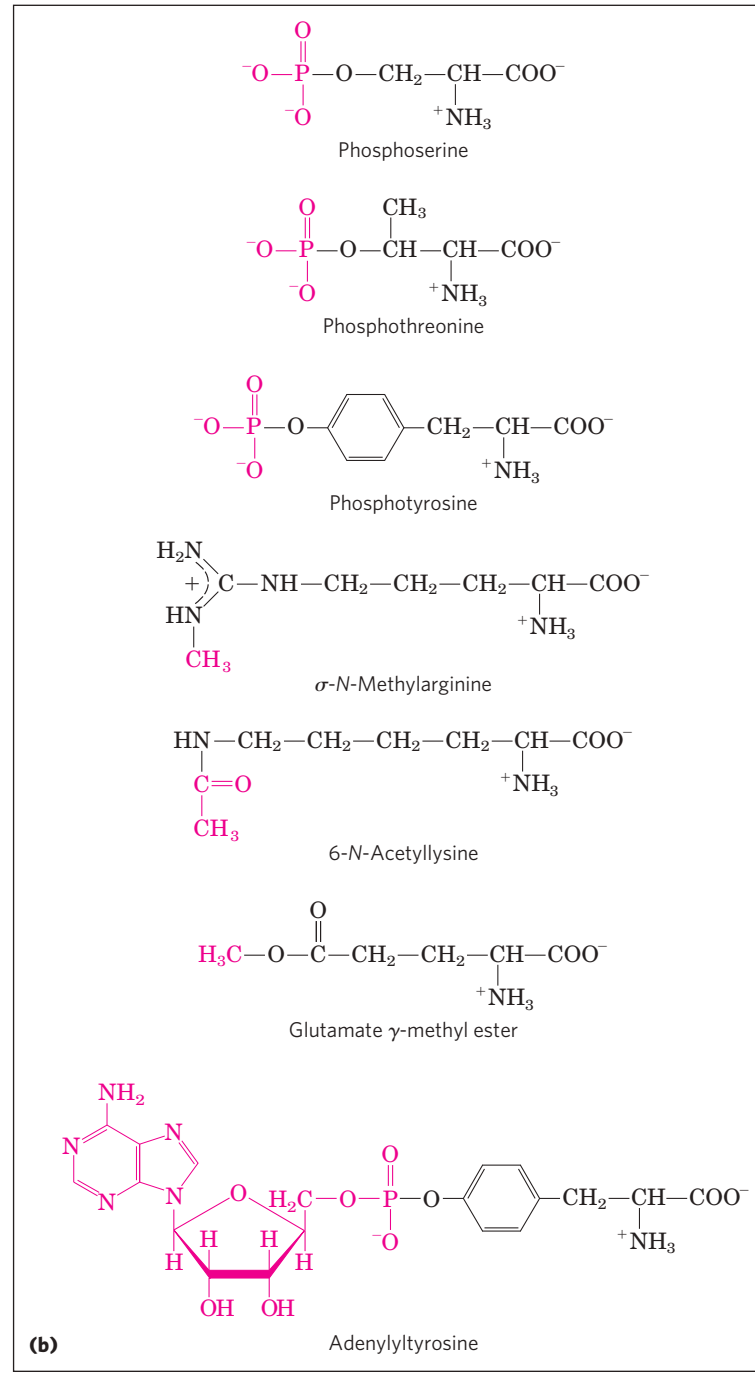
Some 300 additional amino acids have been found in cells. They have a variety of functions but are not all constituents of proteins. **Ornithine** and **citrulline** (Fig. 3–8c) deserve special note because they are key intermediates (metabolites) in the biosynthesis of arginine (Chapter 22) and in the urea cycle (Chapter 18).

Amino Acids Can Act as Acids and Bases

The amino and carboxyl groups of amino acids, along with the ionizable R groups of some amino acids, function as weak acids and bases. When an amino acid lacking an ionizable R group is dissolved in water at neutral pH, it exists in solution as the dipolar ion, or **zwitterion** (German for "hybrid ion"), which can act as either an acid or a base **(Fig. 3–9)**. Substances having this dual (acid-base) nature are **amphoteric** and are often called **ampholytes** (from "amphoteric electrolytes"). A simple monoamino monocarboxylic α-amino acid, such as alanine, is a diprotic acid when fully protonated; it has two groups, the —COOH group and the —NH_3^+ group, that can yield protons:







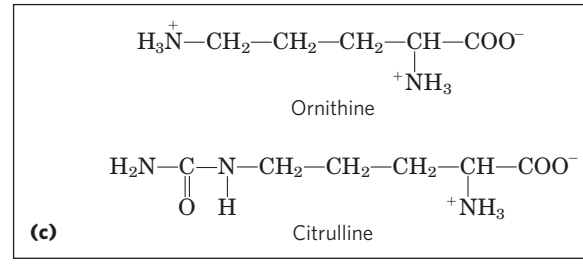


FIGURE 3–8 Uncommon amino acids. (a) Some uncommon amino acids found in proteins. All are derived from common amino acids. Extra functional groups added by modification reactions are shown in red. Desmosine is formed from four Lys residues (the carbon backbones are shaded in light red). Note the use of either numbers or Greek letters in the names of these structures to identify the altered carbon atoms. **(b)** Reversible amino acid modifications involved in regulation of protein activity. Phosphorylation is the most common type of regulatory modification. **(c)** Ornithine and citrulline, which are not found in proteins, are intermediates in the biosynthesis of arginine and in the urea cycle.

Amino Acids Have Characteristic Titration Curves

Acid-base titration involves the gradual addition or removal of protons (Chapter 2). **Figure 3–10** shows the titration curve of the diprotic form of glycine. The two ionizable groups of glycine, the carboxyl group and the amino group, are titrated with a strong base such as NaOH. The plot has two distinct stages, corresponding to deprotonation of two different groups on glycine. Each of the two stages resembles in shape the titration

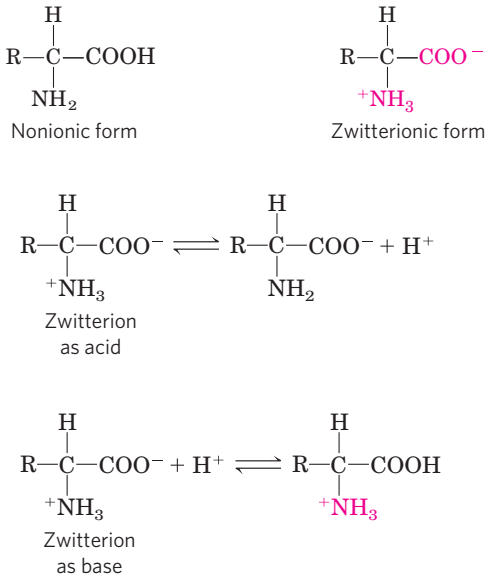


FIGURE 3–9 **Nonionic and zwitterionic forms of amino acids.** The nonionic form does not occur in significant amounts in aqueous solutions. The zwitterion predominates at neutral pH. A zwitterion can act as either an acid (proton donor) or a base (proton acceptor).

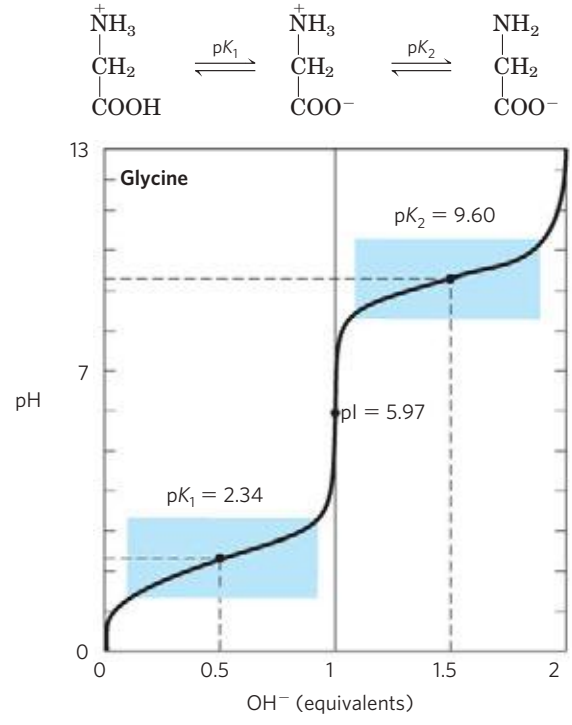


FIGURE 3–10 **Titration of an amino acid.** Shown here is the titration curve of 0.1 M glycine at 25°C. The ionic species predominating at key points in the titration are shown above the graph. The shaded boxes, centered at about $pK_1 = 2.34$ and $pK_2 = 9.60$, indicate the regions of greatest buffering power. Note that 1 equivalent of OH^- = 0.1 M NaOH added.

curve of a monoprotic acid, such as acetic acid (see Fig. 2–17), and can be analyzed in the same way. At very low pH, the predominant ionic species of glycine is the fully protonated form, $^+H_3N—CH_2—COOH$. In the first stage of the titration, the —COOH group of glycine loses its proton. At the midpoint of this stage, equimolar concentrations of the proton-donor ($^+H_3N—CH_2—COOH$) and proton-acceptor ($^+H_3N—CH_2—COO^-$) species are present. As in the titration of any weak acid, a point of inflection is reached at this midpoint where the pH is equal to the pK_a of the protonated group being titrated (see Fig. 2–18). For glycine, the pH at the midpoint is 2.34, thus its —COOH group has a pK_a (labeled pK_1 in Fig. 3–10) of 2.34. (Recall from Chapter 2 that pH and pK_a are simply convenient notations for proton concentration and the equilibrium constant for ionization, respectively. The pK_a is a measure of the tendency of a group to give up a proton, with that tendency decreasing tenfold as the pK_a increases by one unit.) As the titration of glycine proceeds, another important point is reached at pH 5.97. Here there is another point of inflection, at which removal of the first proton is essentially complete and removal of the second has just begun. At this pH glycine is present largely as the dipolar ion (zwitterion) $^+H_3N—CH_2—COO^-$. We shall return to the significance of this inflection point in the titration curve (labeled pI in Fig. 3–10) shortly.

The second stage of the titration corresponds to the removal of a proton from the $—NH_3^+$ group of glycine. The pH at the midpoint of this stage is 9.60, equal to the pK_a (labeled pK_2 in Fig. 3–10) for the $—NH_3^+$ group. The titration is essentially complete at a pH of about 12, at which point the predominant form of glycine is $H_2N—CH_2—COO^-$.

From the titration curve of glycine we can derive several important pieces of information. First, it gives a quantitative measure of the pK_a of each of the two ionizing groups: 2.34 for the —COOH group and 9.60 for the $—NH_3^+$ group. Note that the carboxyl group of glycine is over 100 times more acidic (more easily ionized) than the carboxyl group of acetic acid, which, as we saw in Chapter 2, has a pK_a of 4.76—about average for a carboxyl group attached to an otherwise unsubstituted aliphatic hydrocarbon. The perturbed pK_a of glycine is caused by repulsion between the departing proton and the nearby positively charged amino group on the α-carbon atom, as described in **Figure 3–11**. The opposite charges on the resulting zwitterion are stabilizing. Similarly, the pK_a of the amino group in glycine is perturbed downward relative to the average pK_a of an amino group. This effect is due partly to the electronegative oxygen atoms in the carboxyl groups, which tend to pull electrons toward them, increasing the tendency of the amino group to give up a proton. Hence, the α-amino group has a pK_a that is lower than that of an aliphatic amine such as methylamine (Fig. 3–11). In short, the pK_a of any functional group is greatly affected by its chemical environment, a phenomenon sometimes exploited in the active sites of enzymes to promote exquisitely adapted reaction mechanisms that depend

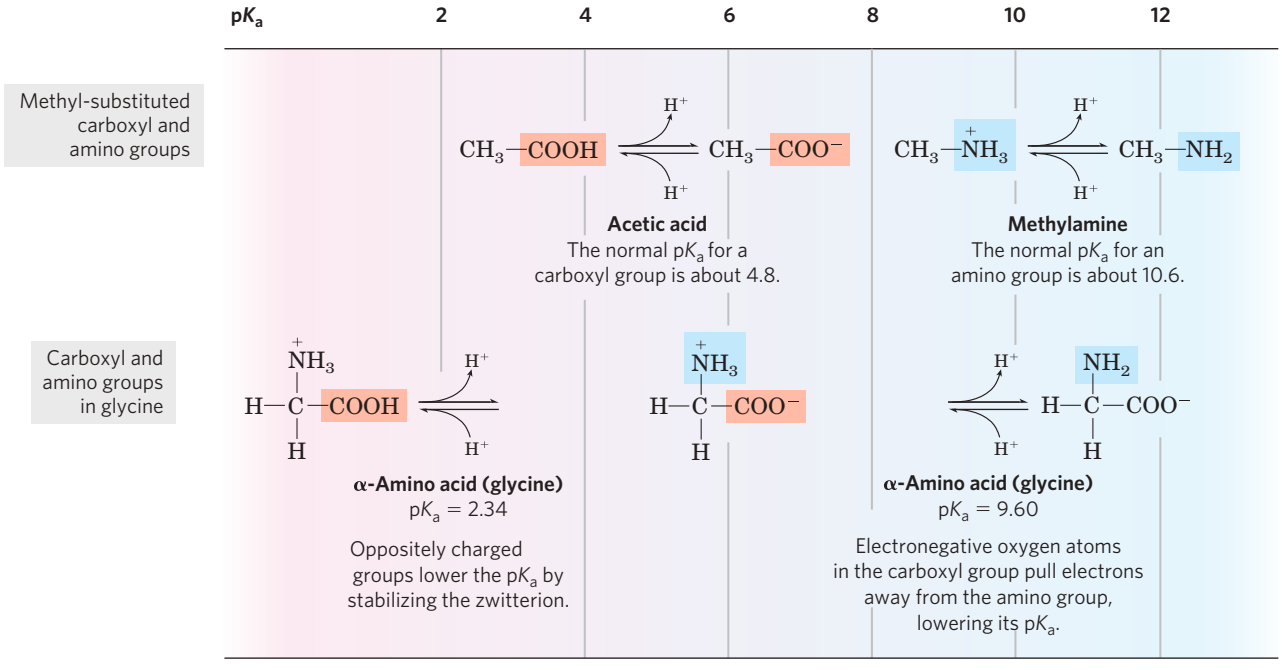


FIGURE 3–11 Effect of the chemical environment on pK_a. The pK_a values for the ionizable groups in glycine are lower than those for simple, methyl-substituted amino and carboxyl groups. These downward perturbations of pK_a are due to intramolecular interactions. Similar effects can be caused by chemical groups that happen to be positioned nearby—for example, in the active site of an enzyme.

on the perturbed pK_a values of proton donor/acceptor groups of specific residues.

The second piece of information provided by the titration curve of glycine is that this amino acid has two regions of buffering power. One of these is the relatively flat portion of the curve, extending for approximately 1 pH unit on either side of the first pK_a of 2.34, indicating that glycine is a good buffer near this pH. The other buffering zone is centered around pH 9.60. (Note that glycine is not a good buffer at the pH of intracellular fluid or blood, about 7.4.) Within the buffering ranges of glycine, the Henderson-Hasselbalch equation (p. 64) can be used to calculate the proportions of proton-donor and proton-acceptor species of glycine required to make a buffer at a given pH.

Titration Curves Predict the Electric Charge of Amino Acids

Another important piece of information derived from the titration curve of an amino acid is the relationship between its net charge and the pH of the solution. At pH 5.97, the point of inflection between the two stages in its titration curve, glycine is present predominantly as its dipolar form, fully ionized but with no *net* electric charge (Fig. 3–10). The characteristic pH at which the *net* electric charge is zero is called the **isoelectric point** or **isoelectric pH**, designated **pI**. For glycine, which has no ionizable group in its side chain, the isoelectric point is simply the arithmetic mean of the two pK_a values:

$$\text{pI} = \frac{1}{2}(\text{p}K_1 + \text{p}K_2) = \frac{1}{2}(2.34 + 9.60) = 5.97$$

As is evident in Figure 3–10, glycine has a net negative charge at any pH above its pI and will thus move toward the positive electrode (the anode) when placed in an electric field. At any pH below its pI, glycine has a net positive charge and will move toward the negative electrode (the cathode). The farther the pH of a glycine solution is from its isoelectric point, the greater the net electric charge of the population of glycine molecules. At pH 1.0, for example, glycine exists almost entirely as the form $^{+}H_3N—CH_2—COOH$ with a net positive charge of 1.0. At pH 2.34, where there is an equal mixture of $^{+}H_3N—CH_2—COOH$ and $^{+}H_3N—CH_2—COO^{-}$, the average or net positive charge is 0.5. The sign and the magnitude of the net charge of any amino acid at any pH can be predicted in the same way.

Amino Acids Differ in Their Acid-Base Properties

The shared properties of many amino acids permit some simplifying generalizations about their acid-base behaviors. First, all amino acids with a single α-amino group, a single α-carboxyl group, and an R group that does not ionize have titration curves resembling that of glycine (Fig. 3–10). These amino acids have very similar, although not identical, pK_a values: pK_a of the —COOH group in the range of 1.8 to 2.4, and pK_a of the $—NH_3^+$

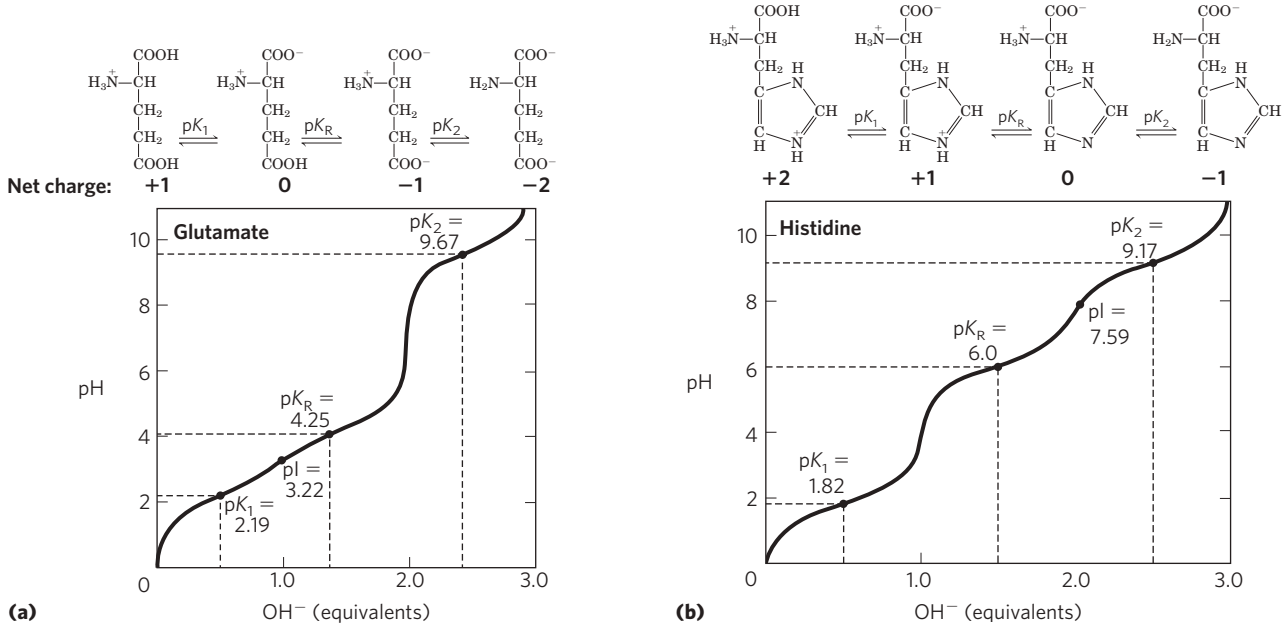


FIGURE 3–12 Titration curves for **(a)** glutamate and **(b)** histidine. The pK_a of the R group is designated here as pK_R.

group in the range of 8.8 to 11.0 (Table 3–1). The differences in these pK_a values reflect the chemical environments imposed by their R groups. Second, amino acids with an ionizable R group have more complex titration curves, with *three* stages corresponding to the three possible ionization steps; thus they have three pK_a values. The additional stage for the titration of the ionizable R group merges to some extent with that for the titration of the α-carboxyl group, the titration of the α-amino group, or both. The titration curves for two amino acids of this type, glutamate and histidine, are shown in **Figure 3–12**. The isoelectric points reflect the nature of the ionizing R groups present. For example, glutamate has a pI of 3.22, considerably lower than that of glycine. This is due to the presence of two carboxyl groups, which, at the average of their pK_a values (3.22), contribute a net charge of −1 that balances the +1 contributed by the amino group. Similarly, the pI of histidine, with two groups that are positively charged when protonated, is 7.59 (the average of the pK_a values of the amino and imidazole groups), much higher than that of glycine.

Finally, as pointed out earlier, under the general condition of free and open exposure to the aqueous environment, only histidine has an R group (pK_a = 6.0) providing significant buffering power near the neutral pH usually found in the intracellular and extracellular fluids of most animals and bacteria (Table 3–1).

SUMMARY 3.1 Amino Acids

- The 20 amino acids commonly found as residues in proteins contain an α-carboxyl group, an α-amino group, and a distinctive R group substituted on the α-carbon atom. The α-carbon atom of all amino acids except glycine is asymmetric, and thus amino acids can exist in at least two stereoisomeric forms. Only the L stereoisomers, with a configuration related to the absolute configuration of the reference molecule L-glyceraldehyde, are found in proteins.
- Other, less common amino acids also occur, either as constituents of proteins (through modification of common amino acid residues after protein synthesis) or as free metabolites.
- Amino acids can be classified into five types on the basis of the polarity and charge (at pH 7) of their R groups.
- Amino acids vary in their acid-base properties and have characteristic titration curves. Monoamino monocarboxylic amino acids (with nonionizable R groups) are diprotic acids ($^+H_3NCH(R)COOH$) at low pH and exist in several different ionic forms as the pH is increased. Amino acids with ionizable R groups have additional ionic species, depending on the pH of the medium and the pK_a of the R group.

3.2 Peptides and Proteins

We now turn to polymers of amino acids, the **peptides** and **proteins**. Biologically occurring polypeptides range in size from small to very large, consisting of two or three to thousands of linked amino acid residues. Our focus is on the fundamental chemical properties of these polymers.

Peptides Are Chains of Amino Acids

Two amino acid molecules can be covalently joined through a substituted amide linkage, termed a **peptide**

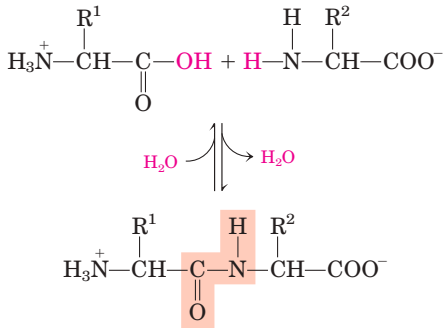


FIGURE 3–13 Formation of a peptide bond by condensation. The α-amino group of one amino acid (with R^2 group) acts as a nucleophile to displace the hydroxyl group of another amino acid (with R^1 group), forming a peptide bond (shaded in light red). Amino groups are good nucleophiles, but the hydroxyl group is a poor leaving group and is not readily displaced. At physiological pH, the reaction shown here does not occur to any appreciable extent.

bond, to yield a dipeptide. Such a linkage is formed by removal of the elements of water (dehydration) from the α-carboxyl group of one amino acid and the α-amino group of another **(Fig. 3–13)**. Peptide bond formation is an example of a condensation reaction, a common class of reactions in living cells. Under standard biochemical conditions, the equilibrium for the reaction shown in Figure 3–13 favors the amino acids over the dipeptide. To make the reaction thermodynamically more favorable, the carboxyl group must be chemically modified or activated so that the hydroxyl group can be more readily eliminated. A chemical approach to this problem is outlined later in this chapter. The biological approach to peptide bond formation is a major topic of Chapter 27.

Three amino acids can be joined by two peptide bonds to form a tripeptide; similarly, four amino acids can be linked to form a tetrapeptide, five to form a pentapeptide, and so forth. When a few amino acids are joined in this fashion, the structure is called an **oligopeptide**. When many amino acids are joined, the product is called a **polypeptide**. Proteins may have thousands of amino acid residues. Although the terms "protein" and "polypeptide" are sometimes used interchangeably, molecules referred to as polypeptides generally have molecular weights below 10,000, and those called proteins have higher molecular weights.

Figure 3–14 shows the structure of a pentapeptide. As already noted, an amino acid unit in a peptide is often called a residue (the part left over after losing the elements of water—a hydrogen atom from its amino group and the hydroxyl moiety from its carboxyl group). In a peptide, the amino acid residue at the end with a free α-amino group is the **amino-terminal** (or *N*-terminal) residue; the residue at the other end, which has a free carboxyl group, is the **carboxyl-terminal** (*C*-terminal) residue.

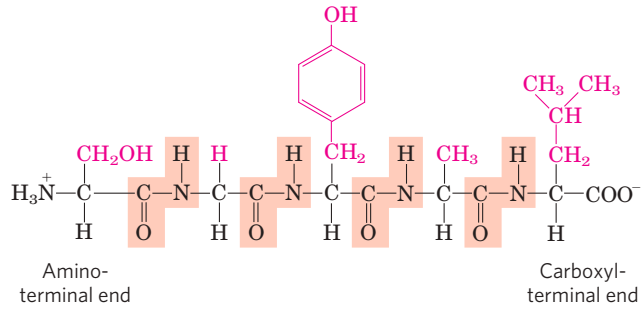


FIGURE 3–14 The pentapeptide serylglycyltyrosylalanylleucine, Ser-Gly-Tyr-Ala-Leu, or SGYAL. Peptides are named beginning with the amino-terminal residue, which by convention is placed at the left. The peptide bonds are shaded in light red; the R groups are in red.

KEY CONVENTION: When an amino acid sequence of a peptide, polypeptide, or protein is displayed, the amino-terminal end is placed on the left, the carboxyl-terminal end on the right. The sequence is read left to right, beginning with the amino-terminal end. ■

Although hydrolysis of a peptide bond is an exergonic reaction, it occurs only slowly because it has a high activation energy (p. 27). As a result, the peptide bonds in proteins are quite stable, with an average half-life ($t_{1/2}$) of about 7 years under most intracellular conditions.

Peptides Can Be Distinguished by Their Ionization Behavior

Peptides contain only one free α-amino group and one free α-carboxyl group, at opposite ends of the chain **(Fig. 3–15)**. These groups ionize as they do in free amino acids, although the ionization constants are different because an oppositely charged group is no longer

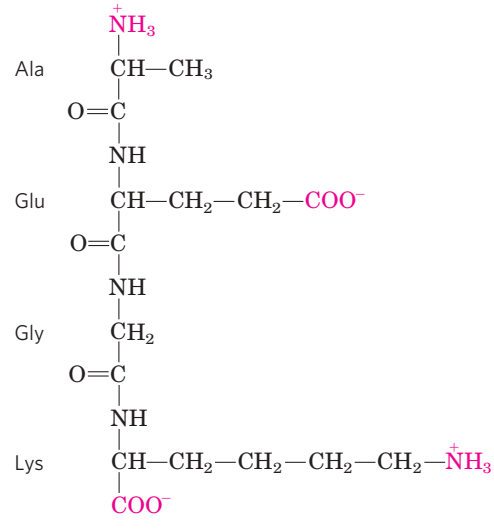


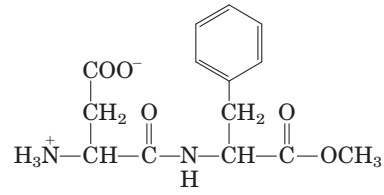
FIGURE 3–15 Alanylglutamylglycyllysine. This tetrapeptide has one free α-amino group, one free α-carboxyl group, and two ionizable R groups. The groups ionized at pH 7.0 are in red.

linked to the α carbon. The α-amino and α-carboxyl groups of all nonterminal amino acids are covalently joined in the peptide bonds, which do not ionize and thus do not contribute to the total acid-base behavior of peptides. However, the R groups of some amino acids can ionize (Table 3–1), and in a peptide these contribute to the overall acid-base properties of the molecule (Fig. 3–15). Thus the acid-base behavior of a peptide can be predicted from its free α-amino and α-carboxyl groups combined with the nature and number of its ionizable R groups.

Like free amino acids, peptides have characteristic titration curves and a characteristic isoelectric pH (pI) at which they do not move in an electric field. These properties are exploited in some of the techniques used to separate peptides and proteins, as we shall see later in the chapter. It should be emphasized that the pK_a value for an ionizable R group can change somewhat when an amino acid becomes a residue in a peptide. The loss of charge in the α-carboxyl and α-amino groups, the interactions with other peptide R groups, and other environmental factors can affect the pK_a. The pK_a values for R groups listed in Table 3–1 can be a useful guide to the pH range in which a given group will ionize, but they cannot be strictly applied to peptides.

Biologically Active Peptides and Polypeptides Occur in a Vast Range of Sizes and Compositions

No generalizations can be made about the molecular weights of biologically active peptides and proteins in relation to their functions. Naturally occurring peptides range in length from two to many thousands of amino acid residues. Even the smallest peptides can have biologically important effects. Consider the commercially synthesized dipeptide L-aspartyl-L-phenylalanine methyl ester, the artificial sweetener better known as aspartame or NutraSweet.



L-Aspartyl-L-phenylalanine methyl ester (aspartame)

Many small peptides exert their effects at very low concentrations. For example, a number of vertebrate hormones (Chapter 23) are small peptides. These include oxytocin (nine amino acid residues), which is secreted by the posterior pituitary gland and stimulates uterine contractions, and thyrotropin-releasing factor (three residues), which is formed in the hypothalamus and stimulates the release of another hormone, thyrotropin, from the anterior pituitary gland. Some extremely toxic mushroom poisons, such as amanitin, are also small peptides, as are many antibiotics.

How long are the polypeptide chains in proteins? As Table 3–2 shows, lengths vary considerably. Human cytochrome *c* has 104 amino acid residues linked in a single chain; bovine chymotrypsinogen has 245 residues. At the extreme is titin, a constituent of vertebrate muscle, which has nearly 27,000 amino acid residues and a molecular weight of about 3,000,000. The vast majority of naturally occurring proteins are much smaller than this, containing fewer than 2,000 amino acid residues.

Some proteins consist of a single polypeptide chain, but others, called **multisubunit** proteins, have two or more polypeptides associated noncovalently (Table 3–2). The individual polypeptide chains in a multisubunit

TABLE 3–2 Molecular Data on Some Proteins

	Molecular weight	Number of residues	Number of polypeptide chains
Cytochrome *c* (human)	12,400	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (chicken egg white)	14,300	129	1
Myoglobin (equine heart)	16,700	153	1
Chymotrypsin (bovine pancreas)	25,200	241	3
Chymotrypsinogen (bovine)	25,700	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	66,000	609	1
Hexokinase (yeast)	107,900	972	2
RNA polymerase (*E. coli*)	450,000	4,158	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamine synthetase (*E. coli*)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

protein may be identical or different. If at least two are identical the protein is said to be **oligomeric**, and the identical units (consisting of one or more polypeptide chains) are referred to as **protomers**. Hemoglobin, for example, has four polypeptide subunits: two identical α chains and two identical β chains, all four held together by noncovalent interactions. Each α subunit is paired in an identical way with a β subunit within the structure of this multisubunit protein, so that hemoglobin can be considered either a tetramer of four polypeptide subunits or a dimer of $\alpha\beta$ protomers.

A few proteins contain two or more polypeptide chains linked covalently. For example, the two polypeptide chains of insulin are linked by disulfide bonds. In such cases, the individual polypeptides are not considered subunits but are commonly referred to simply as chains.

The amino acid composition of proteins is also highly variable. The 20 common amino acids almost never occur in equal amounts in a protein. Some amino acids may occur only once or not at all in a given type of protein; others may occur in large numbers. Table 3–3 shows the amino acid composition of bovine cytochrome *c* and chymotrypsinogen, the inactive precursor of the digestive enzyme chymotrypsin. These two proteins, with very different functions, also differ significantly in the relative numbers of each kind of amino acid residue.

We can calculate the approximate number of amino acid residues in a simple protein containing no other chemical constituents by dividing its molecular weight by 110. Although the average molecular weight of the 20 common amino acids is about 138, the smaller amino acids predominate in most proteins. If we take into account the proportions in which the various amino acids occur in an average protein (Table 3–1; the averages are determined by surveying the amino acid compositions of more than 1,000 different proteins), the average molecular weight of protein amino acids is nearer to 128. Because a molecule of water (M_r 18) is removed to create each peptide bond, the average molecular weight of an amino acid residue in a protein is about 128 − 18 = 110.

TABLE 3–3 Amino Acid Composition of Two Proteins

	Bovine cytochrome *c*		Bovine chymotrypsinogen	
Amino acid	Number of residues per molecule	Percentage of total*	Number of residues per molecule	Percentage of total*
Ala	6	6	22	9
Arg	2	2	4	1.6
Asn	5	5	14	5.7
Asp	3	3	9	3.7
Cys	2	2	10	4
Gln	3	3	10	4
Glu	9	9	5	2
Gly	14	13	23	9.4
His	3	3	2	0.8
Ile	6	6	10	4
Leu	6	6	19	7.8
Lys	18	17	14	5.7
Met	2	2	2	0.8
Phe	4	4	6	2.4
Pro	4	4	9	3.7
Ser	1	1	28	11.4
Thr	8	8	23	9.4
Trp	1	1	8	3.3
Tyr	4	4	4	1.6
Val	3	3	23	9.4
Total	104	102	245	99.7

Note: In some common analyses, such as acid hydrolysis, Asp and Asn are not readily distinguished from each other and are together designated Asx (or B). Similarly, when Glu and Gln cannot be distinguished, they are together designated Glx (or Z). In addition, Trp is destroyed by acid hydrolysis. Additional procedures must be employed to obtain an accurate assessment of complete amino acid content.

*Percentages do not total to 100%, due to rounding.

TABLE 3–4 Conjugated Proteins

Class	Prosthetic group	Example
Lipoproteins	Lipids	β_1-Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin

Some Proteins Contain Chemical Groups Other Than Amino Acids

Many proteins, for example the enzymes ribonuclease A and chymotrypsin, contain only amino acid residues and no other chemical constituents; these are considered simple proteins. However, some proteins contain permanently associated chemical components in addition to amino acids; these are called **conjugated proteins**. The non–amino acid part of a conjugated protein is usually called its **prosthetic group**. Conjugated proteins are classified on the basis of the chemical nature of their prosthetic groups (Table 3–4); for example, **lipoproteins** contain lipids, **glycoproteins** contain sugar groups, and **metalloproteins** contain a specific metal. Some proteins contain more than one prosthetic group. Usually the prosthetic group plays an important role in the protein's biological function.

SUMMARY 3.2 Peptides and Proteins

- Amino acids can be joined covalently through peptide bonds to form peptides and proteins. Cells generally contain thousands of different proteins, each with a different biological activity.
- Proteins can be very long polypeptide chains of 100 to several thousand amino acid residues. However, some naturally occurring peptides have only a few amino acid residues. Some proteins are composed of several noncovalently associated polypeptide chains, called subunits.
- Simple proteins yield only amino acids on hydrolysis; conjugated proteins contain in addition some other component, such as a metal or organic prosthetic group.

3.3 Working with Proteins

Biochemists' understanding of protein structure and function has been derived from the study of many individual proteins. To study a protein in detail, the researcher must be able to separate it from other proteins in pure form and must have the techniques to determine its properties. The necessary methods come from protein chemistry, a discipline as old as biochemistry itself and one that retains a central position in biochemical research.

Proteins Can Be Separated and Purified

A pure preparation is essential before a protein's properties and activities can be determined. Given that cells contain thousands of different kinds of proteins, how can one protein be purified? Classical methods for separating proteins take advantage of properties that vary from one protein to the next, including size, charge, and binding properties. These have been supplemented in recent decades by methods involving DNA cloning and genome sequencing that can simplify the process of protein purification. The newer methods, presented in Chapter 9, often artificially modify the protein being purified, adding a few or many amino acid residues to one or both ends. Convenience thus comes at the price of potentially altering the activity of the purified protein. The purification of proteins in their native state (the form in which they function in the cell) usually relies on methods described here.

The source of a protein is generally tissue or microbial cells. The first step in any protein purification procedure is to break open these cells, releasing their proteins into a solution called a **crude extract**. If necessary, differential centrifugation can be used to prepare subcellular fractions or to isolate specific organelles (see Fig. 1–8).

Once the extract or organelle preparation is ready, various methods are available for purifying one or more of the proteins it contains. Commonly, the extract is subjected to treatments that separate the proteins into different **fractions** based on a property such as size or charge, a process referred to as **fractionation**. Early fractionation steps in a purification utilize differences in protein solubility, which is a complex function of pH, temperature, salt concentration, and other factors.

The solubility of proteins is lowered in the presence of some salts, an effect called "salting out." The addition of certain salts in the right amount can selectively precipitate some proteins, while others remain in solution. Ammonium sulfate ($(NH_4)_2SO_4$) is particularly effective and is often used to salt out proteins. The proteins thus precipitated are removed from those remaining in solution by low-speed centrifugation.

A solution containing the protein of interest usually must be further altered before subsequent purification steps are possible. For example, **dialysis** is a procedure that separates proteins from small solutes by taking advantage of the proteins' larger size. The partially purified extract is placed in a bag or tube made of a semipermeable membrane. When this is suspended in a much larger volume of buffered solution of appropriate ionic strength, the membrane allows the exchange of salt and buffer but not proteins. Thus dialysis retains large proteins within the membranous bag or tube while allowing the concentration of other solutes in the protein preparation to change until they come into equilibrium with the solution outside the membrane. Dialysis might be used, for example, to remove ammonium sulfate from the protein preparation.

The most powerful methods for fractionating proteins make use of **column chromatography**, which takes advantage of differences in protein charge, size, binding affinity, and other properties **(Fig. 3–16)**. A porous solid material with appropriate chemical properties (the stationary phase) is held in a column, and a buffered solution (the mobile phase) migrates through it. The protein, dissolved in the same buffered solution that was used to establish the mobile phase, is layered on the top of the column. The protein then percolates through the solid matrix as an ever-expanding band within the larger mobile phase. Individual proteins migrate faster or more slowly through the column depending on their properties.

Ion-exchange chromatography exploits differences in the sign and magnitude of the net electric charge of proteins at a given pH **(Fig. 3–17a)**. The column matrix is a synthetic polymer (resin) containing bound charged groups; those with bound anionic groups are called **cation exchangers**, and those with bound cationic groups are called **anion exchangers**. The affinity of each protein for the charged groups on the column is affected by the pH (which determines the ionization state of the molecule) and the concentration of competing free salt ions in the surrounding solution. Separation can be optimized by gradually changing the pH and/or salt concentration of the mobile phase so as to create a pH or salt gradient. In **cation-exchange chromatography**, the solid matrix has negatively charged groups. In the mobile phase, proteins with a net positive charge migrate through the matrix more slowly than those with a net negative charge, because the migration of the former is retarded more by interaction with the stationary phase.

In ion-exchange columns, the expansion of the protein band in the mobile phase (the protein solution) is caused both by separation of proteins with different properties and by diffusional spreading. As the length of the column increases, the resolution of two types of protein with different net charges generally improves. However, the rate at which the protein solution can flow through

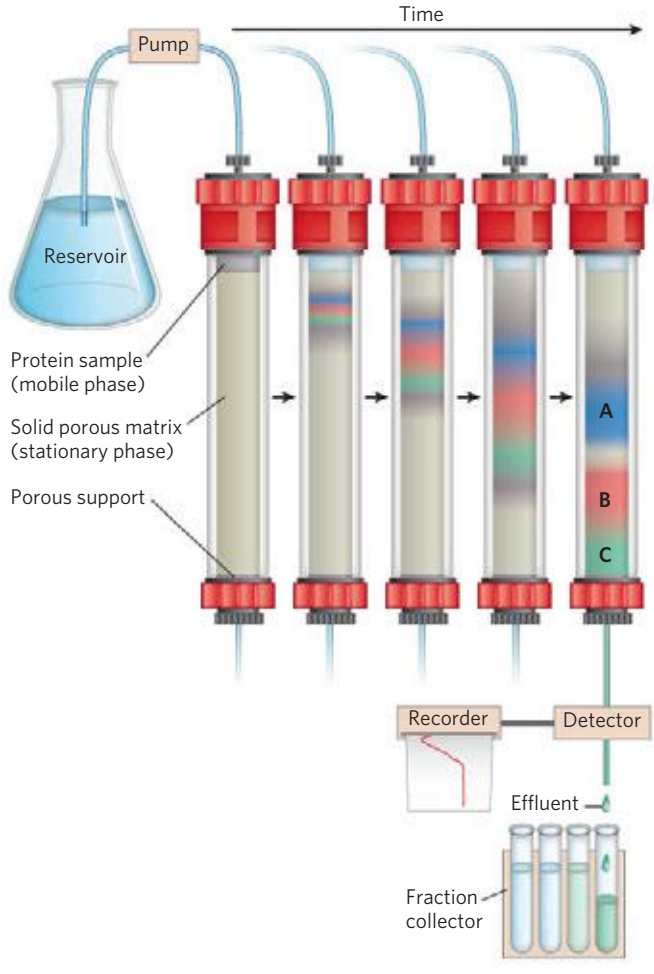


FIGURE 3–16 Column chromatography. The standard elements of a chromatographic column include a solid, porous material (matrix) supported inside a column, generally made of plastic or glass. A solution, the mobile phase, flows through the matrix, the stationary phase. The solution that passes out of the column at the bottom (the effluent) is constantly replaced by solution supplied from a reservoir at the top. The protein solution to be separated is layered on top of the column and allowed to percolate into the solid matrix. Additional solution is added on top. The protein solution forms a band within the mobile phase that is initially the depth of the protein solution applied to the column. As proteins migrate through the column (shown here at five different times), they are retarded to different degrees by their different interactions with the matrix material. The overall protein band thus widens as it moves through the column. Individual types of proteins (such as A, B, and C, shown in blue, red, and green) gradually separate from each other, forming bands within the broader protein band. Separation improves (i.e., resolution increases) as the length of the column increases. However, each individual protein band also broadens with time due to diffusional spreading, a process that decreases resolution. In this example, protein A is well separated from B and C, but diffusional spreading prevents complete separation of B and C under these conditions.

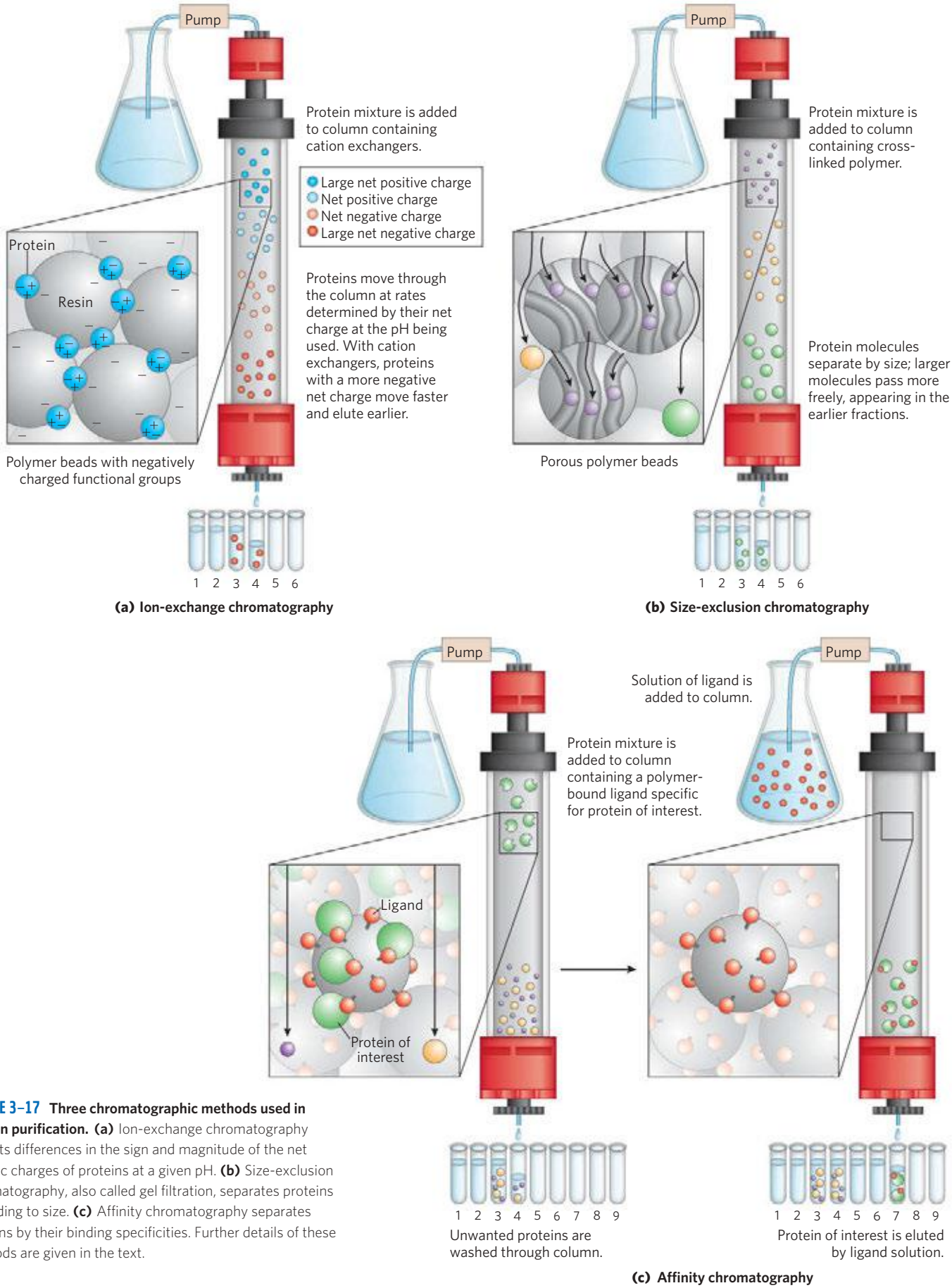


FIGURE 3–17 Three chromatographic methods used in protein purification. (a) Ion-exchange chromatography exploits differences in the sign and magnitude of the net electric charges of proteins at a given pH. **(b)** Size-exclusion chromatography, also called gel filtration, separates proteins according to size. **(c)** Affinity chromatography separates proteins by their binding specificities. Further details of these methods are given in the text.

the column usually decreases with column length. And as the length of time spent on the column increases, the resolution can decline as a result of diffusional spreading within each protein band. As the protein-containing solution exits a column, successive portions (fractions) of this effluent are collected in test tubes. Each fraction can be tested for the presence of the protein of interest as well as other properties, such as ionic strength or total protein concentration. All fractions positive for the protein of interest can be combined as the product of this chromatographic step of the protein purification.

WORKED EXAMPLE 3–1 Ion Exchange of Peptides

A biochemist wants to separate two peptides by ion-exchange chromatography. At the pH of the mobile phase to be used on the column, one peptide (A) has a net charge of −3, due to the presence of more Glu and Asp residues than Arg, Lys, and His residues. Peptide B has a net charge of +1. Which peptide would elute first from a cation-exchange resin? Which would elute first from an anion-exchange resin?

Solution: A cation-exchange resin has negative charges and binds positively charged molecules, retarding their progress through the column. Peptide B, with its net positive charge, will interact more strongly than peptide A with the cation-exchange resin, and thus peptide A will elute first. On the anion-exchange resin, peptide B will elute first. Peptide A, being negatively charged, will be retarded by its interaction with the positively charged resin.

Figure 3–17 shows two other variations of column chromatography in addition to ion exchange. **Size-exclusion chromatography**, also called gel filtration (Fig. 3–17b), separates proteins according to size. In this method, large proteins emerge from the column sooner than small ones—a somewhat counterintuitive result. The solid phase consists of cross-linked polymer beads with engineered pores or cavities of a particular size. Large proteins cannot enter the cavities and so take a shorter (and more rapid) path through the column, around the beads. Small proteins enter the cavities and are slowed by their more labyrinthine path through the column. Size-exclusion chromatography can also be used to approximate the size of a protein being purified, using methods similar to those described in Figure 3–19.

Affinity chromatography is based on binding affinity (Fig. 3–17c). The beads in the column have a covalently attached chemical group called a ligand—a group or molecule that binds to a macromolecule such as a protein. When a protein mixture is added to the column, any protein with affinity for this ligand binds to the beads, and its migration through the matrix is retarded. For example, if the biological function of a protein involves binding to ATP, then attaching a molecule that resembles ATP to the beads in the column creates an affinity matrix that can help purify the protein. As the protein solution moves through the column, ATP-binding proteins (including the protein of interest) bind to the matrix. After proteins that do not bind are washed through the column, the bound protein is eluted by a solution containing either a high concentration of salt or free ligand—in this case, ATP or an analog of ATP. Salt weakens the binding of the protein to the immobilized ligand, interfering with ionic interactions. Free ligand competes with the ligand attached to the beads, releasing the protein from the matrix; the protein product that elutes from the column is often bound to the ligand used to elute it.

Chromatographic methods are typically enhanced by the use of **HPLC**, or **high-performance liquid chromatography**. HPLC makes use of high-pressure pumps that speed the movement of the protein molecules down the column, as well as higher-quality chromatographic materials that can withstand the crushing force of the pressurized flow. By reducing the transit time on the column, HPLC can limit diffusional spreading of protein bands and thus greatly improve resolution.

The approach to purification of a protein that has not previously been isolated is guided both by established precedents and by common sense. In most cases, several different methods must be used sequentially to purify a protein completely, each method separating proteins on the basis of different properties. For example, if one step separates ATP-binding proteins from those that do not bind ATP, then the next step must separate the various ATP-binding proteins on the basis of size or charge to isolate the particular protein that is wanted. The choice of methods is somewhat empirical, and many strategies may be tried before the most effective one is found. Trial and error can often be minimized by basing the new procedure on purification techniques developed for similar proteins. Published purification protocols are available for many thousands of proteins. Common sense dictates that inexpensive procedures such as salting out be used first, when the total volume and the number of contaminants are greatest. Chromatographic methods are often impractical at early stages, because the amount of chromatographic medium needed increases with sample size. As each purification step is completed, the sample size generally becomes smaller (Table 3–5), making it feasible to use more sophisticated (and expensive) chromatographic procedures at later stages.

Proteins Can Be Separated and Characterized by Electrophoresis

Another important technique for the separation of proteins is based on the migration of charged proteins in an electric field, a process called **electrophoresis**. These procedures are not generally used to purify proteins, because simpler alternatives are usually available and electrophoretic methods often adversely affect the

TABLE 3–5 A Purification Table for a Hypothetical Enzyme

Procedure or step	Fraction volume (mL)	Total protein (mg)	Activity (units)	Specific activity (units/mg)
1. Crude cellular extract	1,400	10,000	100,000	10
2. Precipitation with ammonium sulfate	280	3,000	96,000	32
3. Ion-exchange chromatography	90	400	80,000	200
4. Size-exclusion chromatography	80	100	60,000	600
5. Affinity chromatography	6	3	45,000	15,000

Note: All data represent the status of the sample *after* the designated procedure has been carried out. Activity and specific activity are defined on page 95.

structure and thus the function of proteins. However, as an analytical method, electrophoresis is extremely important. Its advantage is that proteins can be visualized as well as separated, permitting a researcher to estimate quickly the number of different proteins in a mixture or the degree of purity of a particular protein preparation. Also, electrophoresis can be used to determine crucial properties of a protein such as its isoelectric point and approximate molecular weight.

Electrophoresis of proteins is generally carried out in gels made up of the cross-linked polymer polyacrylamide **(Fig. 3–18)**. The polyacrylamide gel acts as a molecular sieve, slowing the migration of proteins approximately in proportion to their charge-to-mass ratio. Migration may also be affected by protein shape. In electrophoresis, the force moving the macromolecule is the electrical potential, E. The electrophoretic mobility, μ, of a molecule is the ratio of its velocity, V,

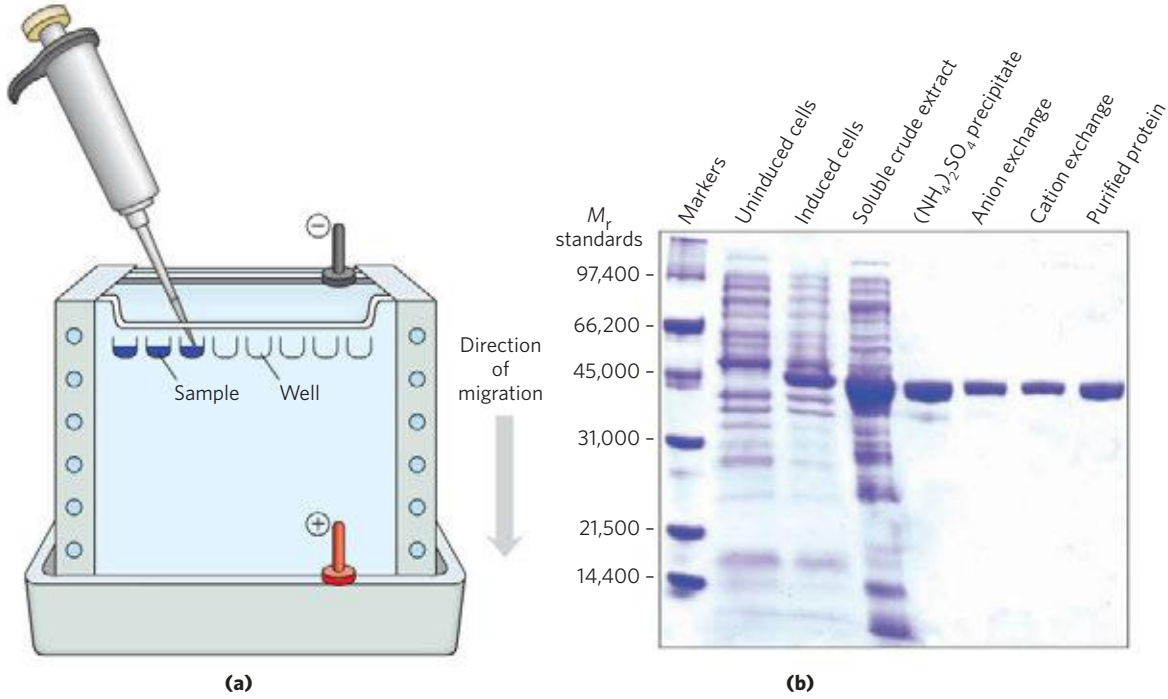


FIGURE 3–18 Electrophoresis. (a) Different samples are loaded in wells or depressions at the top of the SDS-polyacrylamide gel. The proteins move into the gel when an electric field is applied. The gel minimizes convection currents caused by small temperature gradients, as well as protein movements other than those induced by the electric field. **(b)** Proteins can be visualized after electrophoresis by treating the gel with a stain such as Coomassie blue, which binds to the proteins but not to the gel itself. Each band on the gel represents a different protein (or protein subunit); smaller proteins move through the gel more rapidly than larger proteins and therefore are found nearer the bottom of the gel. This gel illustrates purification of the RecA protein of *Escherichia coli* (described in Chapter 25). The gene for the RecA protein was cloned (Chapter 9) so that its expression (synthesis of the protein) could be controlled. The first lane shows a set of standard proteins (of known M_r), serving as molecular weight markers. The next two lanes show proteins from *E. coli* cells before and after synthesis of RecA protein was induced. The fourth lane shows the proteins in a crude cellular extract. Subsequent lanes (left to right) show the proteins present after successive purification steps. The purified protein is a single polypeptide chain (M_r ~38,000), as seen in the rightmost lane.

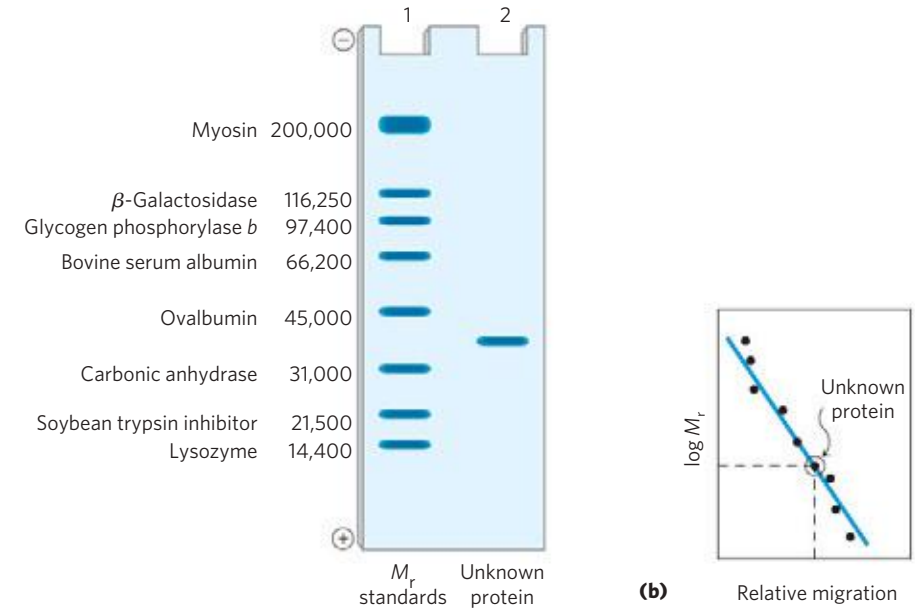


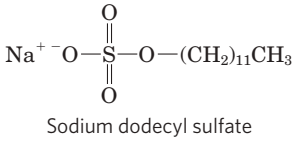
FIGURE 3–19 Estimating the molecular weight of a protein. The electrophoretic mobility of a protein on an SDS polyacrylamide gel is related to its molecular weight, M_r. **(a)** Standard proteins of known molecular weight are subjected to electrophoresis (lane 1). These marker proteins can be used to estimate the molecular weight of an unknown protein (lane 2). **(b)** A plot of log M_r of the marker proteins versus relative migration during electrophoresis is linear, which allows the molecular weight of the unknown protein to be read from the graph. (In similar fashion, a set of standard proteins with reproducible retention times on a size-exclusion column can be used to create a standard curve of retention time versus log M_r. The retention time of an unknown substance on the column can be compared with this standard curve to obtain an approximate M_r.)

to the electrical potential. Electrophoretic mobility is also equal to the net charge, Z, of the molecule divided by the frictional coefficient, f, which reflects in part a protein's shape. Thus:

$$\mu = \frac{V}{E} = \frac{Z}{f}$$

The migration of a protein in a gel during electrophoresis is therefore a function of its size and its shape.

An electrophoretic method commonly employed for estimation of purity and molecular weight makes use of the detergent **sodium dodecyl sulfate (SDS)** ("dodecyl" denoting a 12-carbon chain).



Sodium dodecyl sulfate (SDS)

A protein will bind about 1.4 times its weight of SDS, nearly one molecule of SDS for each amino acid residue. The bound SDS contributes a large net negative charge, rendering the intrinsic charge of the protein insignificant and conferring on each protein a similar charge-to-mass ratio. In addition, SDS binding partially unfolds proteins, such that most SDS-bound proteins assume a similar rodlike shape. Electrophoresis in the presence of SDS therefore separates proteins almost exclusively on the basis of mass (molecular weight), with smaller polypeptides migrating more rapidly. After electrophoresis, the proteins are visualized by adding a dye such as Coomassie blue, which binds to proteins but not to the gel itself (Fig. 3–18b). Thus, a researcher can monitor the progress of a protein purification procedure as the number of protein bands visible on the gel decreases after each new fractionation step. When compared with the positions to which proteins of known molecular weight migrate in the gel, the position of an unidentified protein can provide a good approximation of its molecular weight **(Fig. 3–19)**. If the protein has two or more different subunits, the subunits are generally separated by the SDS treatment, and a separate band appears for each. SDS Gel Electrophoresis

Isoelectric focusing is a procedure used to determine the isoelectric point (pI) of a protein **(Fig. 3–20)**. A pH gradient is established by allowing a mixture of low molecular weight organic acids and bases (ampholytes; p. 81) to distribute themselves in an electric field generated across the gel. When a protein mixture is applied, each protein migrates until it reaches the pH that matches its pI. Proteins with different isoelectric points are thus distributed differently throughout the gel.

Combining isoelectric focusing and SDS electrophoresis sequentially in a process called **two-dimensional electrophoresis** permits the resolution of complex mixtures of proteins **(Fig. 3–21)**. This is a more sensitive analytical method than either electrophoretic method alone. Two-dimensional electrophoresis separates proteins of identical molecular weight that differ

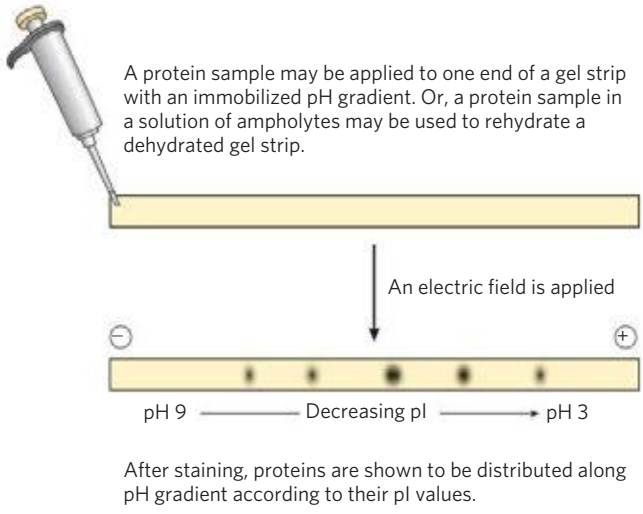


FIGURE 3–20 Isoelectric focusing. This technique separates proteins according to their isoelectric points. A protein mixture is placed on a gel strip containing an immobilized pH gradient. With an applied electric field, proteins enter the gel and migrate until each reaches a pH equivalent to its pI. Remember that when pH = pI, the net charge of a protein is zero.

in pI, or proteins with similar pI values but different molecular weights.

Unseparated Proteins Can Be Quantified

To purify a protein, it is essential to have a way of detecting and quantifying that protein in the presence of many other proteins at each stage of the procedure. Often, purification must proceed in the absence of any information about the size and physical properties of the protein or about the fraction of the total protein mass it represents in the extract. For proteins that are enzymes, the amount in a given solution or tissue extract can be measured, or assayed, in terms of the catalytic effect the enzyme produces—that is, the *increase* in the rate at which its substrate is converted to reaction products when the enzyme is present. For this purpose the researcher must know (1) the overall equation of the reaction catalyzed, (2) an analytical procedure for determining the disappearance of the substrate or the appearance of a reaction product, (3) whether the enzyme requires cofactors such as metal ions or coenzymes, (4) the dependence of the enzyme activity on substrate concentration, (5) the optimum pH, and (6) a temperature zone in which the enzyme is stable and has high activity. Enzymes are usually assayed at their optimum pH and at some convenient temperature within the range 25 to 38°C. Also, very high substrate concentrations are generally used so that the initial reaction rate, measured experimentally, is proportional to enzyme concentration (Chapter 6).

By international agreement, 1.0 unit of enzyme activity for most enzymes is defined as the amount of enzyme causing transformation of 1.0 μmol of substrate to product per minute at 25°C under optimal conditions of measurement (for some enzymes, this definition is

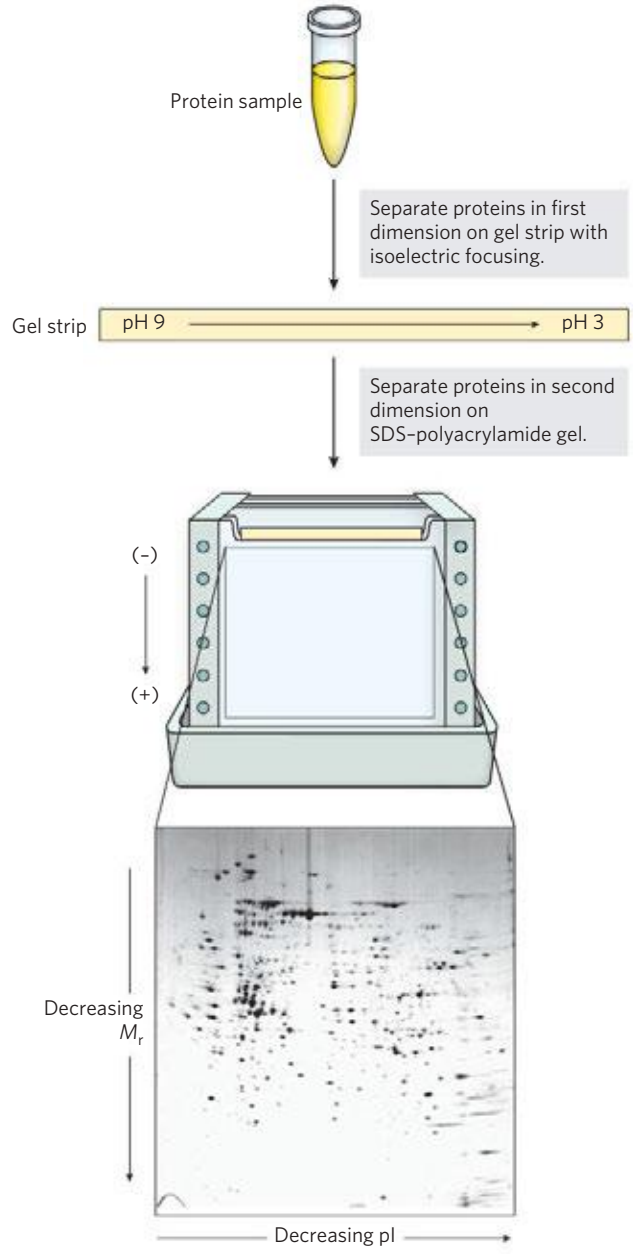


FIGURE 3–21 Two-dimensional electrophoresis. Proteins are first separated by isoelectric focusing in a thin strip gel. The gel is then laid horizontally on a second, slab-shaped gel, and the proteins are separated by SDS polyacrylamide gel electrophoresis. Horizontal separation reflects differences in pI; vertical separation reflects differences in molecular weight. The original protein complement is thus spread in two dimensions. Thousands of cellular proteins can be resolved using this technique. Individual protein spots can be cut out of the gel and identified by mass spectrometry (see Figs 3–30 and 3–31).

inconvenient, and a unit may be defined differently). The term **activity** refers to the total units of enzyme in a solution. The **specific activity** is the number of enzyme units per milligram of total protein **(Fig. 3–22)**. The specific activity is a measure of enzyme purity: it increases during purification of an enzyme and becomes maximal and constant when the enzyme is pure (Table 3–5, p. 93).

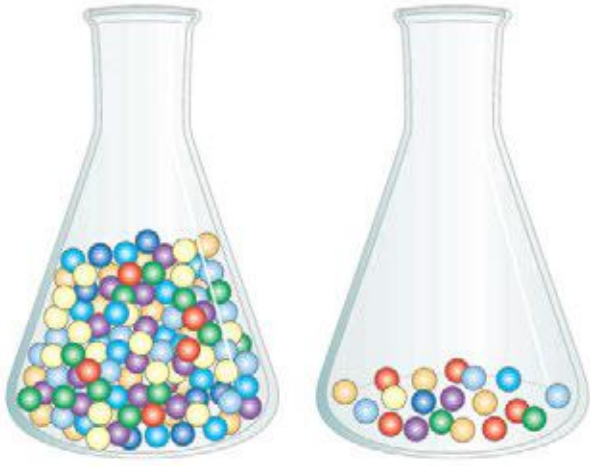


FIGURE 3–22 Activity versus specific activity. The difference between these terms can be illustrated by considering two flasks containing marbles. The flasks contain the same number of red marbles, but different numbers of marbles of other colors. If the marbles represent proteins, both flasks contain the same *activity* of the protein represented by the red marbles. The second flask, however, has the higher *specific activity* because red marbles represent a higher fraction of the total.

After each purification step, the activity of the preparation (in units of enzyme activity) is assayed, the total amount of protein is determined independently, and the ratio of the two gives the specific activity. Activity and total protein generally decrease with each step. Activity decreases because there is always some loss due to inactivation or nonideal interactions with chromatographic materials or other molecules in the solution. Total protein decreases because the objective is to remove as much unwanted or nonspecific protein as possible. In a successful step, the loss of nonspecific protein is much greater than the loss of activity; therefore, specific activity increases even as total activity falls. The data are assembled in a purification table similar to Table 3–5. A protein is generally considered pure when further purification steps fail to increase specific activity and when only a single protein species can be detected (for example, by electrophoresis).

For proteins that are not enzymes, other quantification methods are required. Transport proteins can be assayed by their binding to the molecule they transport, and hormones and toxins by the biological effect they produce; for example, growth hormones will stimulate the growth of certain cultured cells. Some structural proteins represent such a large fraction of a tissue mass that they can be readily extracted and purified without a functional assay. The approaches are as varied as the proteins themselves.

SUMMARY 3.3 Working with Proteins

- Proteins are separated and purified on the basis of differences in their properties. Proteins can be selectively precipitated by changes in pH or temperature, and particularly by the addition of certain salts. A wide range of chromatographic procedures makes use of differences in size, binding affinities, charge, and other properties. These include ion-exchange, size-exclusion, affinity, and high-performance liquid chromatography.
- Electrophoresis separates proteins on the basis of mass or charge. SDS gel electrophoresis and isoelectric focusing can be used separately or in combination for higher resolution.
- All purification procedures require a method for quantifying or assaying the protein of interest in the presence of other proteins. Purification can be monitored by assaying specific activity.

3.4 The Structure of Proteins: Primary Structure

Purification of a protein is usually only a prelude to a detailed biochemical dissection of its structure and function. What is it that makes one protein an enzyme, another a hormone, another a structural protein, and still another an antibody? How do they differ chemically? The most obvious distinctions are structural, and to protein structure we now turn.

The structure of large molecules such as proteins can be described at several levels of complexity, arranged in a kind of conceptual hierarchy. Four levels of protein structure are commonly defined **(Fig. 3–23)**. A description

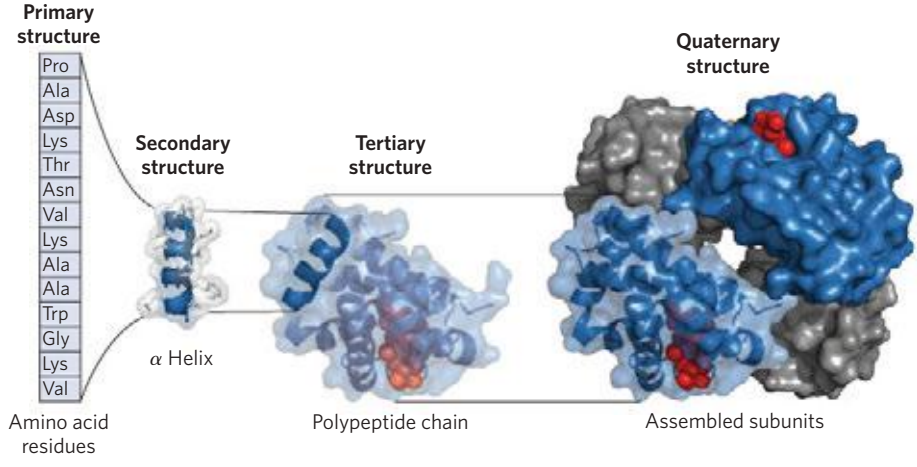


FIGURE 3–23 Levels of structure in proteins. The *primary structure* consists of a sequence of amino acids linked together by peptide bonds and includes any disulfide bonds. The resulting polypeptide can be arranged into units of *secondary structure*, such as an α helix. The helix is a part of the *tertiary structure* of the folded polypeptide, which is itself one of the subunits that make up the *quaternary structure* of the multisubunit protein, in this case hemoglobin.

of all covalent bonds (mainly peptide bonds and disulfide bonds) linking amino acid residues in a polypeptide chain is its **primary structure**. The most important element of primary structure is the *sequence* of amino acid residues. **Secondary structure** refers to particularly stable arrangements of amino acid residues giving rise to recurring structural patterns. **Tertiary structure** describes all aspects of the three-dimensional folding of a polypeptide. When a protein has two or more polypeptide subunits, their arrangement in space is referred to as **quaternary structure**. Our exploration of proteins will eventually include complex protein machines consisting of dozens to thousands of subunits. Primary structure is the focus of the remainder of this chapter; the higher levels of structure are discussed in Chapter 4.

Differences in primary structure can be especially informative. Each protein has a distinctive number and sequence of amino acid residues. As we shall see in Chapter 4, the primary structure of a protein determines how it folds up into its unique three-dimensional structure, and this in turn determines the function of the protein. We first consider empirical clues that amino acid sequence and protein function are closely linked, then describe how amino acid sequence is determined; finally, we outline the many uses to which this information can be put.

The Function of a Protein Depends on Its Amino Acid Sequence

The bacterium *Escherichia coli* produces more than 3,000 different proteins; a human has ~25,000 genes encoding a much larger number of proteins (through genetic processes discussed in Part III of this text). In both cases, each type of protein has a unique amino acid sequence that confers a particular three-dimensional structure. This structure in turn confers a unique function.

Some simple observations illustrate the importance of primary structure, or the amino acid sequence of a protein. First, as we have already noted, proteins with different functions always have different amino acid sequences. Second, thousands of human genetic diseases have been traced to the production of defective proteins. The defect can range from a single change in the amino acid sequence (as in sickle cell anemia, described in Chapter 5) to deletion of a larger portion of the polypeptide chain (as in most cases of Duchenne muscular dystrophy: a large deletion in the gene encoding the protein dystrophin leads to production of a shortened, inactive protein). Finally, on comparing functionally similar proteins from different species, we find that these proteins often have similar amino acid sequences. Thus, a close link between protein primary structure and function is evident.

Is the amino acid sequence absolutely fixed, or invariant, for a particular protein? No; some flexibility is possible. An estimated 20% to 30% of the proteins in humans are **polymorphic**, having amino acid sequence variants in the human population. Many of these variations in sequence have little or no effect on the function of the protein. Furthermore, proteins that carry out a broadly similar function in distantly related species can differ greatly in overall size and amino acid sequence.

Although the amino acid sequence in some regions of the primary structure might vary considerably without affecting biological function, most proteins contain crucial regions that are essential to their function and whose sequence is therefore conserved. The fraction of the overall sequence that is critical varies from protein to protein, complicating the task of relating sequence to three-dimensional structure, and structure to function. Before we can consider this problem further, however, we must examine how sequence information is obtained.

The Amino Acid Sequences of Millions of Proteins Have Been Determined

Two major discoveries in 1953 were of crucial importance in the history of biochemistry. In that year, James D. Watson and Francis Crick deduced the double-helical structure of DNA and proposed a structural basis for its precise replication (Chapter 8). Their proposal illuminated the molecular reality behind the idea of a gene. In the same year, Frederick Sanger worked out the sequence of amino acid residues in the polypeptide chains of the hormone insulin **(Fig. 3–24)**, surprising many researchers who had long thought that determining the amino acid sequence of a polypeptide would be a hopelessly difficult task. It quickly became evident that the nucleotide sequence in DNA and the amino acid sequence in proteins were somehow related. Barely a decade after these discoveries, the genetic code relating the nucleotide sequence of DNA to the amino acid sequence of protein molecules was elucidated (Chapter 27). The amino acid sequences of proteins are now most often derived indirectly from the DNA sequences in genome databases. However, an array of techniques derived from traditional methods of polypeptide sequencing still command an important place in protein chemistry. Below, we summarize the traditional method and mention a few of the techniques derived from it.

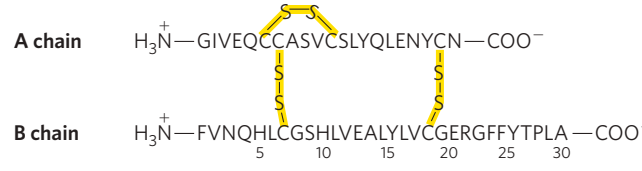
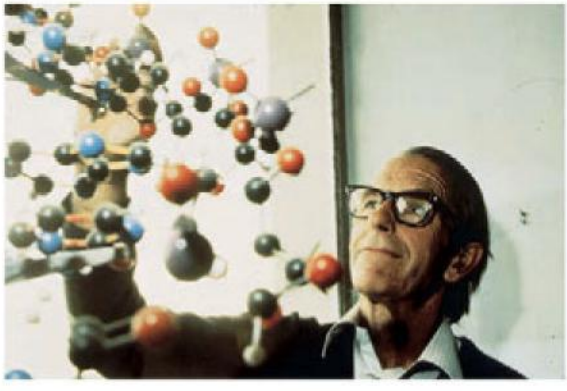


FIGURE 3–24 Amino acid sequence of bovine insulin. The two polypeptide chains are joined by disulfide cross-linkages (yellow). The A chain of insulin is identical in human, pig, dog, rabbit, and sperm whale insulins. The B chains of the cow, pig, dog, goat, and horse are identical.

Protein Chemistry Is Enriched by Methods Derived from Classical Polypeptide Sequencing

The methods used in the 1950s by Fred Sanger to determine the sequence of the protein insulin are summarized, in their modern form, in **Figure 3–25**. Few proteins are sequenced in this way now, at least in their entirety. However, these traditional sequencing protocols have provided a rich array of tools for biochemists, and almost every step in Figure 3–25 makes use of methods that are widely used, sometimes in quite different contexts.



Frederick Sanger

In the traditional scheme for sequencing large proteins, the amino-terminal amino acid residue was first labeled and its identity determined. The amino-terminal α-amino group can be labeled with 1-fluoro-2,4-dinitrobenzene (FDNB), dansyl chloride, or dabsyl chloride **(Fig. 3–26)**.

The chemical sequencing process itself is based on a two-step process developed by Pehr Edman **(Fig. 3–27)**. The **Edman degradation** procedure labels and removes only the amino-terminal residue from a peptide, leaving all other peptide bonds intact. The peptide is reacted with phenylisothiocyanate under mildly alkaline conditions, which converts the amino-terminal amino acid to a phenylthiocarbamoyl (PTC) adduct. The peptide bond next to the PTC adduct is then cleaved in a step carried out in anhydrous trifluoroacetic acid, with removal of the amino-terminal amino acid as an anilinothiazolinone derivative. The derivatized amino acid is extracted with organic solvents, converted to the more stable phenylthiohydantoin derivative by treatment with aqueous acid, and then identified. The use of sequential reactions carried out under first basic and then acidic conditions provides a means of controlling the entire process. Each reaction with the amino-terminal amino acid can go essentially to completion without affecting any of the other peptide bonds in the peptide. The process is repeated until, typically, as many as 40 sequential amino acid residues are identified. The reactions of the Edman degradation have been automated.

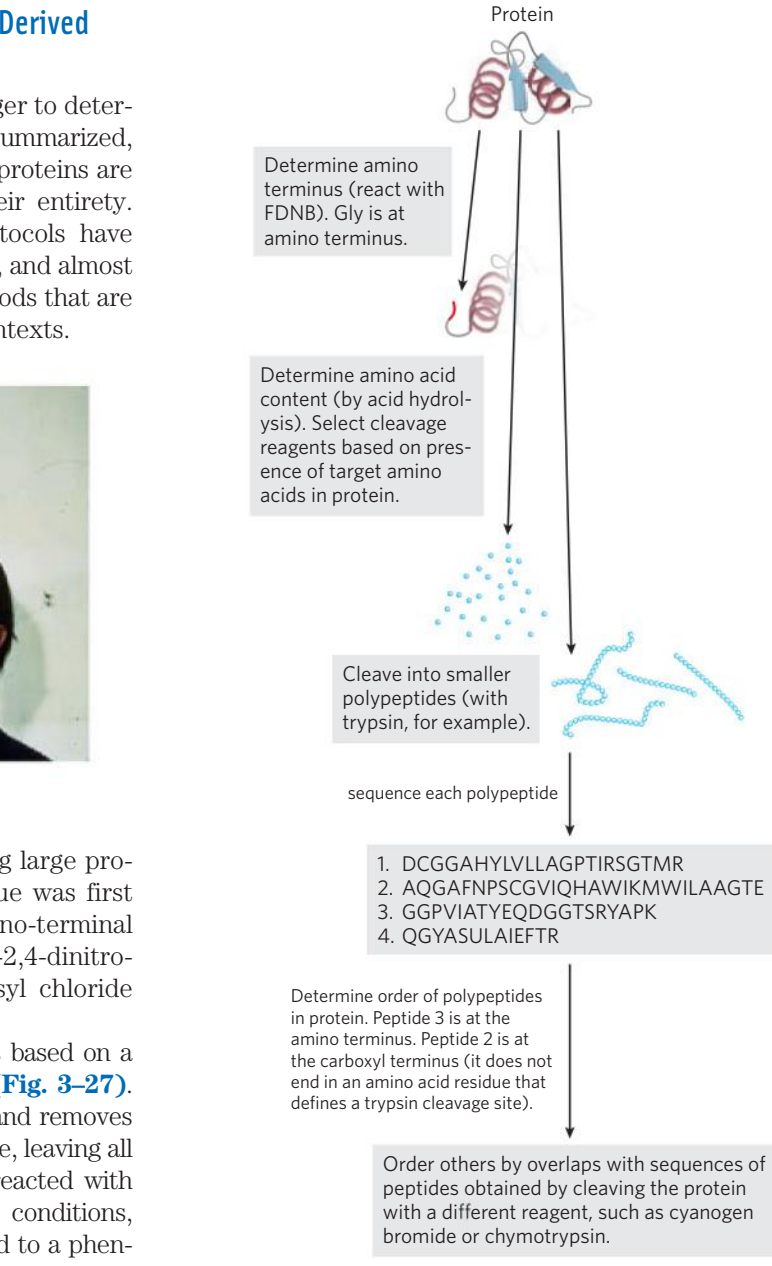
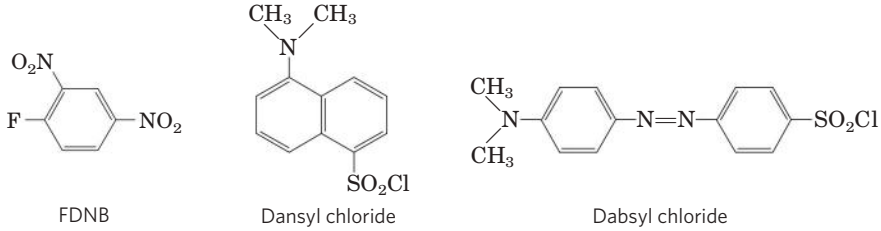


FIGURE 3–25 Direct protein sequencing. The procedures shown here are those developed by Fred Sanger to sequence insulin, and they have been used subsequently for many additional proteins. FDNB is 1-fluoro-2,4-dinitrobenzene (see text and Fig. 3–26).



FDNB Dansyl chloride Dabsyl chloride

FIGURE 3–26 Reagents used to modify the α-amino group at the amino terminus.

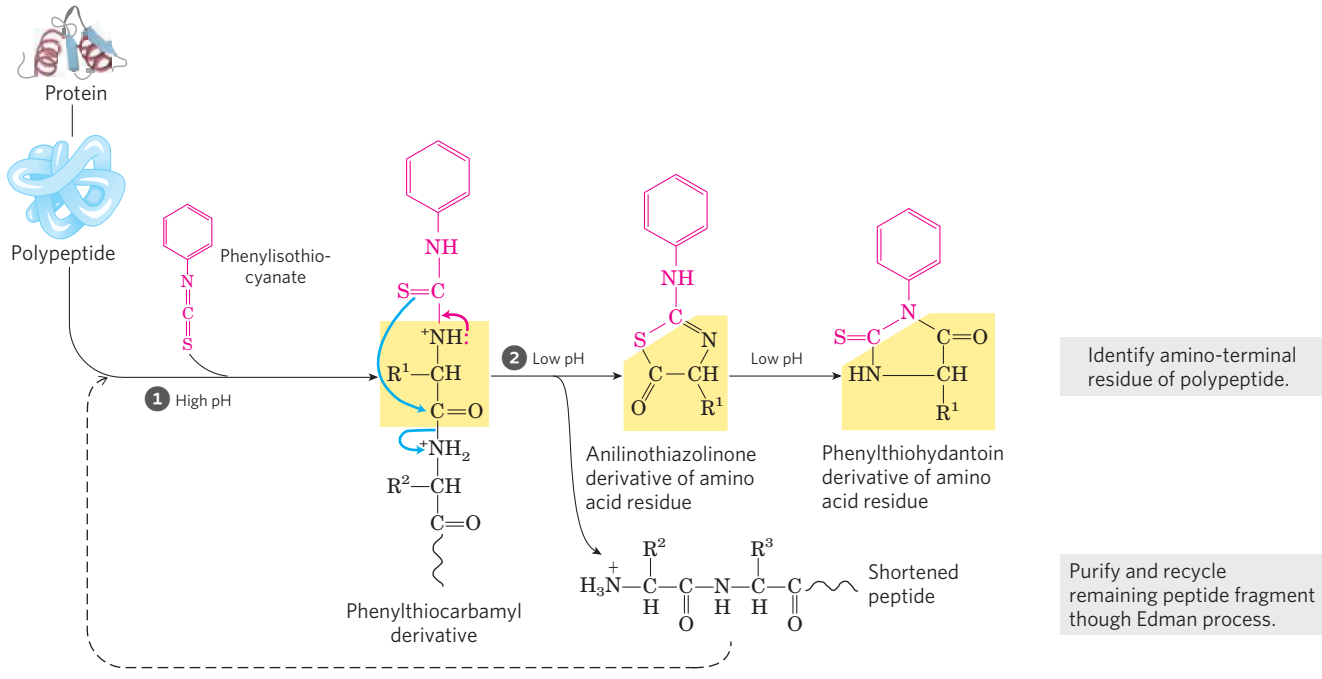


FIGURE 3–27 The protein sequencing chemistry devised by Pehr Edman. The peptide bond nearest to the amino terminus of the protein or polypeptide is cleaved in two steps. The two steps are carried out under very different reaction conditions (basic conditions in step 1, acidic in step 2), allowing one step to proceed to completion before the second is initiated.

To determine the sequence of large proteins, early developers of sequencing protocols had to devise methods to eliminate disulfide bonds and to cleave proteins precisely into smaller polypeptides. Two approaches to irreversible breakage of disulfide bonds are outlined in **Figure 3–28**. Enzymes called **proteases** catalyze the hydrolytic cleavage of peptide bonds. Some proteases cleave only the peptide bond adjacent to particular amino acid residues (Table 3–6) and thus fragment a polypeptide chain in a predictable and reproducible way. A few chemical reagents also cleave the peptide bond adjacent to specific residues. Among proteases, the digestive enzyme trypsin catalyzes the hydrolysis of only those peptide bonds in

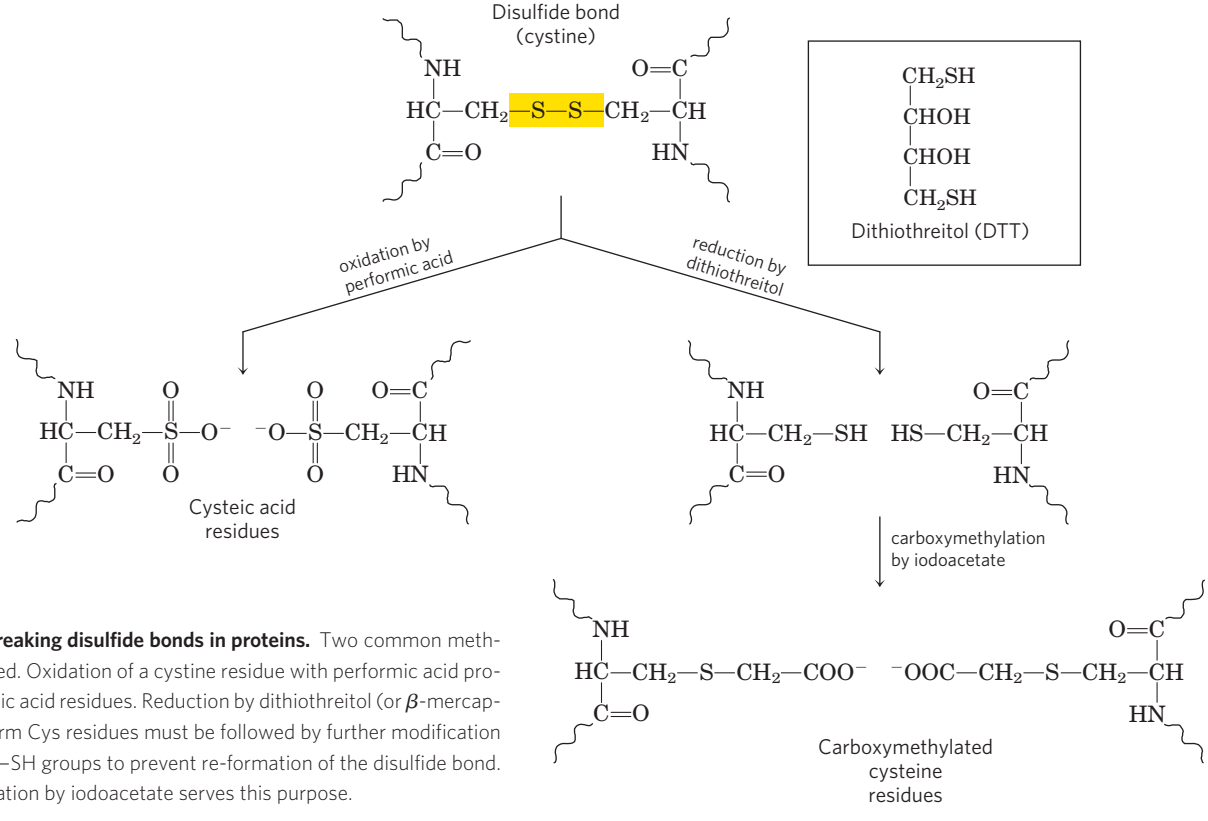


FIGURE 3–28 Breaking disulfide bonds in proteins. Two common methods are illustrated. Oxidation of a cystine residue with performic acid produces two cysteic acid residues. Reduction by dithiothreitol (or β-mercaptoethanol) to form Cys residues must be followed by further modification of the reactive —SH groups to prevent re-formation of the disulfide bond. Carboxymethylation by iodoacetate serves this purpose.

TABLE 3–6 The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

Reagent (biological source)*	Cleavage points†
Trypsin (bovine pancreas)	Lys, Arg (C)
Submaxillary protease (mouse submaxillary gland)	Arg (C)
Chymotrypsin (bovine pancreas)	Phe, Trp, Tyr (C)
Staphylococcus aureus V8 protease (bacterium *S. aureus*)	Asp, Glu (C)
Asp-*N*-protease (bacterium *Pseudomonas fragi*)	Asp, Glu (N)
Pepsin (porcine stomach)	Leu, Phe, Trp, Tyr (N)
Endoproteinase Lys C (bacterium *Lysobacter enzymogenes*)	Lys (C)
Cyanogen bromide	Met (C)

*All reagents except cyanogen bromide are proteases. All are available from commercial sources.

†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

which the carbonyl group is contributed by either a Lys or an Arg residue, regardless of the length or amino acid sequence of the chain. A polypeptide with three Lys and/or Arg residues will usually yield four smaller peptides on cleavage with trypsin. Moreover, all except one of these will have a carboxyl-terminal Lys or Arg. The choice of a reagent to cleave the protein into smaller peptides can be aided by first determining the amino acid content of the entire protein, employing acid to reduce the protein to its constituent amino acids. Trypsin would be used only on proteins that have an appropriate number of Lys or Arg residues.

In classical sequencing, a large protein would be cleaved into fragments twice, using a different protease or cleavage reagent each time so that the fragment endpoints would be different. Both sets of fragments would be purified and sequenced. The order in which the fragments appeared in the original protein could then be determined by examining the overlaps in sequence between the two sets of fragments.

Even if no longer used to sequence entire proteins, the traditional sequencing methods are still valuable in the lab. The sequencing of some amino acids from the amino terminus using the Edman chemistry is often sufficient to confirm the identity of a known protein that has just been purified, or to identify an unknown protein purified on the basis of an unusual activity. Techniques employed in individual steps of the traditional sequencing method are also useful for other purposes. For example, the methods used to break disulfide bonds can also be used to denature proteins when that is required. Furthermore, the effort to label the amino-terminal amino acid residue led eventually to the development of an array of reagents that could react with specific groups on a protein. The same reagents used to label the amino-terminal α-amino group can be used to label the primary–amines of Lys residues (Fig. 3–26). The sulfhydryl group on Cys residues can be modified with iodoacetamides, maleimides, benzyl halides, and bromomethyl ketones **(Fig. 3–29)**. Other amino acid residues can be modified by reagents linked to a dye or other molecule to aid in protein detection or functional studies.

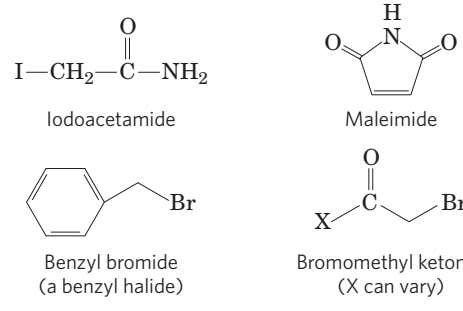


FIGURE 3–29 Reagents used to modify the sulfhydryl groups of Cys residues. (See also Fig. 3–28.)

Mass Spectrometry Offers an Alternative Method to Determine Amino Acid Sequences

Modern adaptations of **mass spectrometry** provide an important alternative to the sequencing methods described above. Mass spectrometry can provide a highly accurate measure of the molecular weight of a protein, but can also do much more. In particular, some variants of mass spectrometry can provide the sequences of multiple short polypeptide segments (20 to 30 amino acid residues) in a protein sample quite rapidly.

The mass spectrometer has long been an indispensable tool in chemistry. Molecules to be analyzed, referred to as **analytes**, are first ionized in a vacuum. When the newly charged molecules are introduced into an electric and/or magnetic field, their paths through the field are a function of their mass-to-charge ratio, m/z. This measured property of the ionized species can be used to deduce the mass (m) of the analyte with very high precision.

Although mass spectrometry has been in use for many years, it could not be applied to macromolecules such as proteins and nucleic acids. The m/z measurements are made on molecules in the gas phase, and the heating or other treatment needed to transfer a macromolecule to the gas phase usually caused its rapid

decomposition. In 1988, two different techniques were developed to overcome this problem. In one, proteins are placed in a light-absorbing matrix. With a short pulse of laser light, the proteins are ionized and then desorbed from the matrix into the vacuum system. This process, known as **matrix-assisted laser desorption/ionization mass spectrometry**, or **MALDI MS**, has been successfully used to measure the mass of a wide range of macromolecules. In a second and equally successful method, macromolecules in solution are forced directly from the liquid to gas phase. A solution of analytes is passed through a charged needle that is kept at a high electrical potential, dispersing the solution into a fine mist of charged microdroplets. The solvent surrounding the macromolecules rapidly evaporates, leaving multiply charged macromolecular ions in the gas phase. This technique is called **electrospray ionization mass spectrometry**, or **ESI MS**. Protons added during passage through the needle give additional charge to the macromolecule. The *m/z* of the molecule can be analyzed in the vacuum chamber.

Mass spectrometry provides a wealth of information for proteomics research, enzymology, and protein chemistry in general. The techniques require only miniscule amounts of sample, so they can be readily applied to the small amounts of protein that can be extracted from a two-dimensional electrophoretic gel. The accurately measured molecular mass of a protein is critical to its identification. Once the mass of a protein is accurately known, mass spectrometry is a convenient and accurate method for detecting changes in mass due to the presence of bound cofactors, bound metal ions, covalent modifications, and so on.

The process for determining the molecular mass of a protein with ESI MS is illustrated in **Figure 3–30**. As it is injected into the gas phase, a protein acquires a variable number of protons, and thus positive charges, from the solvent. The variable addition of these charges creates a spectrum of species with different mass-to-charge ratios. Each successive peak corresponds to a species that differs from that of its neighboring peak by a charge difference of 1 and a mass difference of 1 (1 proton). The mass of the protein can be determined from any two neighboring peaks.

Mass spectrometry can also be used to sequence short stretches of polypeptide, an application that has emerged as an invaluable tool for quickly identifying unknown proteins. Sequence information is extracted using a technique called **tandem MS**, or **MS/MS**. A solution containing the protein under investigation is first treated with a protease or chemical reagent to hydrolyze it to a mixture of shorter peptides. The mixture is then injected into a device that is essentially two mass spectrometers in tandem (**Fig. 3–31a**, top). In the first, the peptide mixture is sorted so that only one of the several types of peptides produced by cleavage emerges at the other end. The sample of the selected peptide, each molecule of which has a charge some-

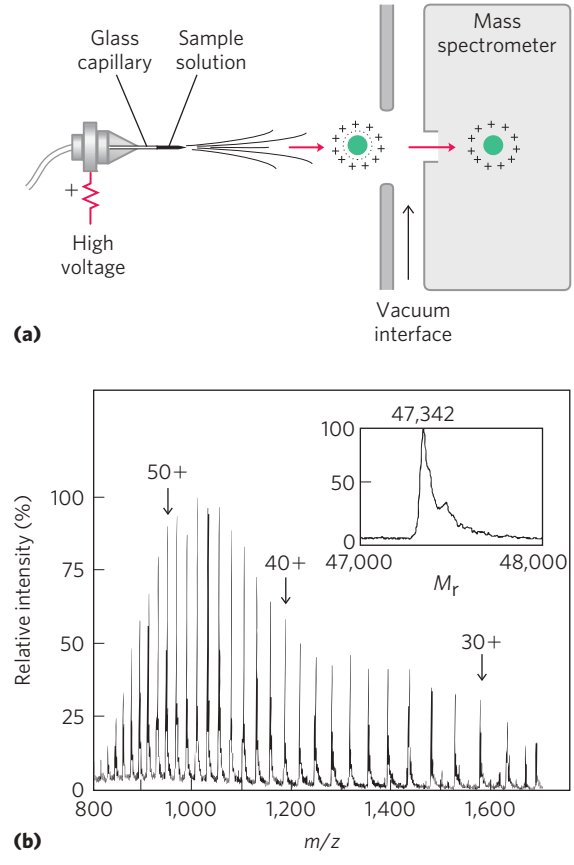


FIGURE 3–30 Electrospray ionization mass spectrometry of a protein. **(a)** A protein solution is dispersed into highly charged droplets by passage through a needle under the influence of a high-voltage electric field. The droplets evaporate, and the ions (with added protons in this case) enter the mass spectrometer for *m/z* measurement. The spectrum generated **(b)** is a family of peaks, with each successive peak (from right to left) corresponding to a charged species increased by 1 in both mass and charge. The inset shows a computer-generated transformation of this spectrum.

where along its length, then travels through a vacuum chamber between the two mass spectrometers. In this collision cell, the peptide is further fragmented by high-energy impact with a "collision gas" such as helium or argon that is bled into the vacuum chamber. Each individual peptide is broken in only one place, on average. Although the breaks are not hydrolytic, most occur at the peptide bonds.

The second mass spectrometer then measures the *m/z* ratios of all the charged fragments. This process generates one or more sets of peaks. A given set of peaks (Fig. 3–31b) consists of all the charged fragments that were generated by breaking the same type of bond (but at different points in the peptide). One set of peaks includes only the fragments in which the charge was retained on the amino-terminal side of the broken bonds; another includes only the fragments in which the charge was retained on the carboxyl-terminal side of the broken bonds. Each successive peak in a given set has one less amino acid than the peak before. The difference in mass from peak to peak identifies the amino

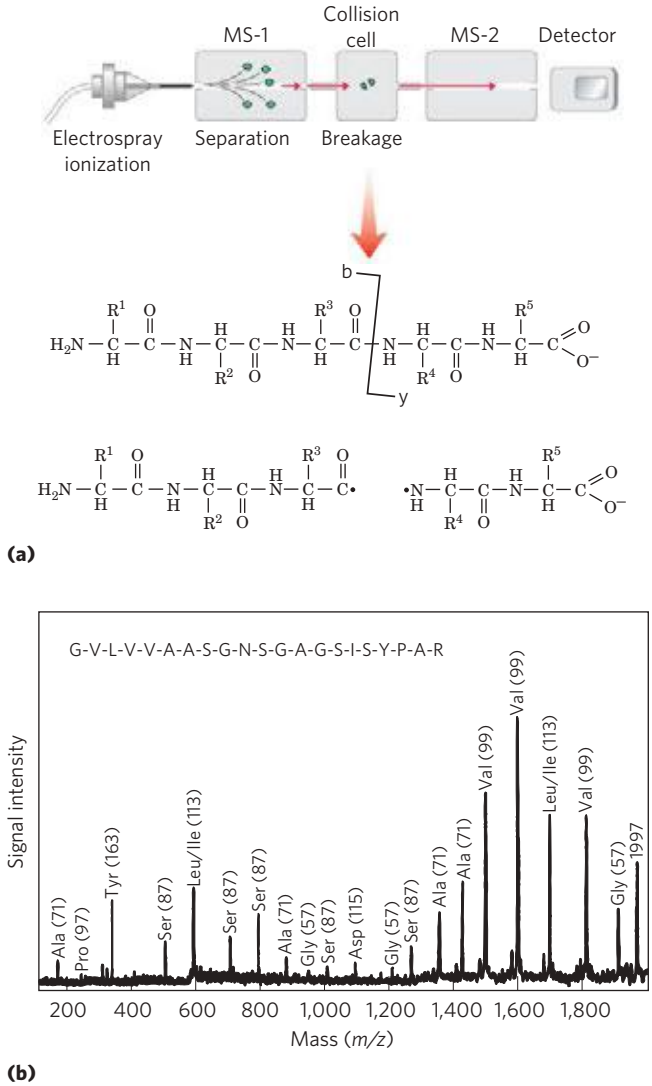


FIGURE 3–31 Obtaining protein sequence information with tandem MS. (a) After proteolytic hydrolysis, a protein solution is injected into a mass spectrometer (MS-1). The different peptides are sorted so that only one type is selected for further analysis. The selected peptide is further fragmented in a chamber between the two mass spectrometers, and *m/z* for each fragment is measured in the second mass spectrometer (MS-2). Many of the ions generated during this second fragmentation result from breakage of the peptide bond, as shown. These are called b-type or y-type ions, depending on whether the charge is retained on the amino- or carboxyl-terminal side, respectively. **(b)** A typical spectrum with peaks representing the peptide fragments generated from a sample of one small peptide (21 residues). The labeled peaks are y-type ions derived from amino acid residues. The number in parentheses over each peak is the molecular weight of the amino acid ion. The successive peaks differ by the mass of a particular amino acid in the original peptide. The deduced sequence is shown at the top.

acid that was lost in each case, thus revealing the sequence of the peptide. The only ambiguities involve leucine and isoleucine, which have the same mass. Although multiple sets of peaks are usually generated, the two most prominent sets generally consist of charged fragments derived from breakage of the peptide bonds. The amino acid sequence derived from one set can be confirmed by the other, improving the confidence in the sequence information obtained.

The various methods for obtaining protein sequence information complement one another. The Edman degradation procedure is sometimes convenient to get sequence information uniquely from the amino terminus of a protein or peptide. However, it is relatively slow and requires a larger sample than does mass spectrometry. Mass spectrometry can be used for small amounts of sample and for mixed samples. It provides sequence information, but the fragmentation processes can leave unpredictable sequence gaps. Although most protein sequences are now extracted from genomic DNA sequences (Chapter 9) by employing our understanding of the genetic code (Chapter 27), direct protein sequencing is often necessary to identify unknown protein samples. Both protein sequencing methods permit the unambiguous identification of newly purified proteins. Mass spectrometry is the method of choice to identify proteins that are present in small amounts. For example, the technique is sensitive enough to analyze the few hundred nanograms of protein that might be extracted from a single protein band on a polyacrylamide gel. Direct sequencing by mass spectrometry also can reveal the addition of phosphoryl groups or other modifications (Chapter 6). Sequencing by either method can reveal changes in protein sequence that result from the editing of messenger RNA in eukaryotes (Chapter 26). Thus, these methods are all part of a robust toolbox used to investigate proteins and their functions.

Small Peptides and Proteins Can Be Chemically Synthesized

Many peptides are potentially useful as pharmacologic agents, and their production is of considerable commercial importance. There are three ways to obtain a peptide: (1) purification from tissue, a task often made difficult by the vanishingly low concentrations of some peptides; (2) genetic engineering (Chapter 9); or (3) direct chemical synthesis. Powerful techniques now make direct chemical synthesis an attractive option in many cases. In addition to commercial applications, the synthesis of specific peptide portions of larger proteins is an increasingly important tool for the study of protein structure and function.

The complexity of proteins makes the traditional synthetic approaches of organic chemistry impractical for peptides with more than four or five amino acid residues. One problem is the difficulty of purifying the product after each step.

The major breakthrough in this technology was provided by R. Bruce Merrifield in 1962. His innovation was to synthesize a peptide while keeping it attached at one end to a solid support. The support is an insoluble polymer (resin) contained within a column, similar to that used for chromatographic procedures. The peptide is built up on this support one amino acid at a time,

through a standard set of reactions in a repeating cycle **(Fig. 3–32)**. At each successive step in the cycle, protective chemical groups block unwanted reactions.

The technology for chemical peptide synthesis is now automated. An important limitation of the process (a limitation shared by the Edman degradation sequencing process) is the efficiency of each chemical cycle, as can be seen by calculating the overall yields of peptides of various lengths when the yield for addition of each new amino acid is 96.0% versus 99.8% (Table 3–7).

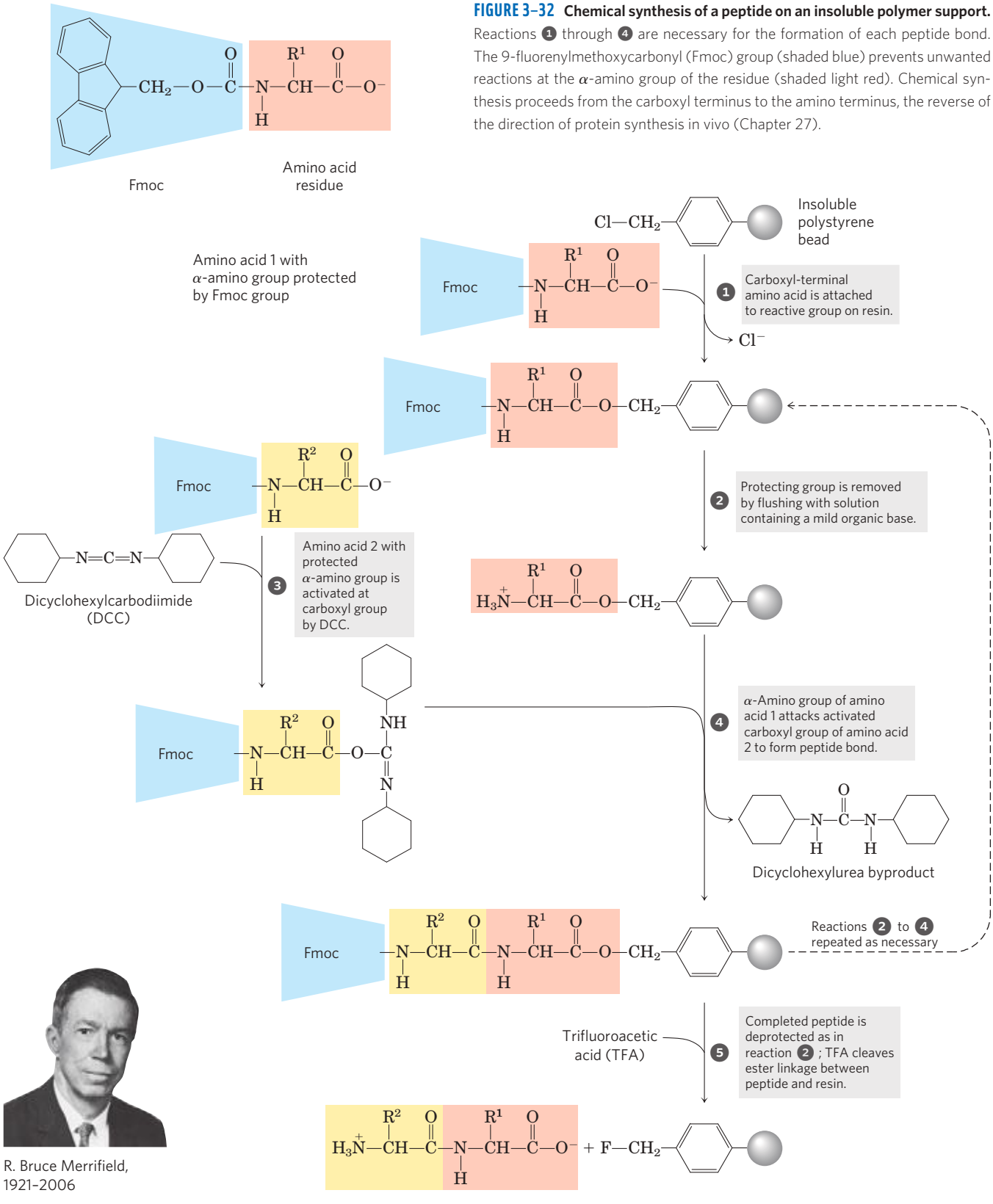


FIGURE 3–32 Chemical synthesis of a peptide on an insoluble polymer support. Reactions 1 through 4 are necessary for the formation of each peptide bond. The 9-fluorenylmethoxycarbonyl (Fmoc) group (shaded blue) prevents unwanted reactions at the α-amino group of the residue (shaded light red). Chemical synthesis proceeds from the carboxyl terminus to the amino terminus, the reverse of the direction of protein synthesis in vivo (Chapter 27).

R. Bruce Merrifield, 1921–2006

TABLE 3–7 Effect of Stepwise Yield on Overall Yield in Peptide Synthesis

Number of residues in the final polypeptide	Overall yield of final peptide (%) when the yield of each step is:	
	96.0%	99.8%
11	66	98
21	44	96
31	29	94
51	13	90
100	1.8	82

Incomplete reaction at one stage can lead to formation of an impurity (in the form of a shorter peptide) in the next. The chemistry has been optimized to permit the synthesis of proteins of 100 amino acid residues in a few days in reasonable yield. A very similar approach is used to synthesize nucleic acids (see Fig. 8–35). It is worth noting that this technology, impressive as it is, still pales when compared with biological processes. The same 100-residue protein would be synthesized with exquisite fidelity in about 5 seconds in a bacterial cell.

A variety of new methods for the efficient ligation (joining together) of peptides has made possible the assembly of synthetic peptides into larger polypeptides and proteins. With these methods, novel forms of proteins can be created with precisely positioned chemical groups, including those that might not normally be found in a cellular protein. These novel forms provide new ways to test theories of enzyme catalysis, to create proteins with new chemical properties, and to design protein sequences that will fold into particular structures. This last application provides the ultimate test of our increasing ability to relate the primary structure of a peptide to the three-dimensional structure that it takes up in solution.

Amino Acid Sequences Provide Important Biochemical Information

Knowledge of the sequence of amino acids in a protein can offer insights into its three-dimensional structure and its function, cellular location, and evolution. Most of these insights are derived by searching for similarities between a protein of interest and previously studied proteins. Thousands of sequences are known and available in databases accessible through the Internet. A comparison of a newly obtained sequence with this large bank of stored sequences often reveals relationships both surprising and enlightening.

Exactly how the amino acid sequence determines three-dimensional structure is not understood in detail, nor can we always predict function from sequence. However, protein families that have some shared structural or functional features can be readily identified on the basis of amino acid sequence similarities. Individual proteins are assigned to families based on the degree of similarity in amino acid sequence. Members of a family are usually identical across 25% or more of their sequences, and proteins in these families generally share at least some structural and functional characteristics. Some families are defined, however, by identities involving only a few amino acid residues that are critical to a certain function. A number of similar substructures, or "domains" (to be defined more fully in Chapter 4), occur in many functionally unrelated proteins. These domains often fold into structural configurations that have an unusual degree of stability or that are specialized for a certain environment. Evolutionary relationships can also be inferred from the structural and functional similarities within protein families.

Certain amino acid sequences serve as signals that determine the cellular location, chemical modification, and half-life of a protein. Special signal sequences, usually at the amino terminus, are used to target certain proteins for export from the cell; other proteins are targeted for distribution to the nucleus, the cell surface, the cytosol, or other cellular locations. Other sequences act as attachment sites for prosthetic groups, such as sugar groups in glycoproteins and lipids in lipoproteins. Some of these signals are well characterized and are easily recognized in the sequence of a newly characterized protein (Chapter 27).

KEY CONVENTION: Much of the functional information encapsulated in protein sequences comes in the form of **consensus sequences**. This term is applied to such sequences in DNA, RNA, or protein. When a series of related nucleic acid or protein sequences are compared, a consensus sequence is the one that reflects the most common base or amino acid at each position. Parts of the sequence that have particularly good agreement often represent evolutionarily conserved functional domains. A range of mathematical tools available on the Internet can be used to generate consensus sequences or identify them in sequence databases. Box 3–2 illustrates common conventions for displaying consensus sequences. ■

Protein Sequences Can Elucidate the History of Life on Earth

The simple string of letters denoting the amino acid sequence of a protein holds a surprising wealth of information. As more protein sequences have become available, the development of more powerful methods for extracting information from them has become a major biochemical enterprise. Analysis of the information available in the many, ever-expanding biological databases, including gene and protein sequences and macromolecular structures, has given rise to the new field of **bioinformatics**. One outcome of this discipline is a growing suite of computer programs, many readily available on the Internet, that can be used by any scientist,

BOX 3–2 Consensus Sequences and Sequence Logos

Consensus sequences can be represented in several ways. To illustrate two types of conventions, we use two examples of consensus sequences, shown in Figure 1: (a) an ATP-binding structure called a P loop (see Box 12–2) and (b) a Ca^{2+}-binding structure called an EF hand (see Fig. 12–11). The rules described here are adapted from those used by the sequence comparison website PROSITE (expasy.org/prosite); they use the standard one-letter codes for the amino acids.

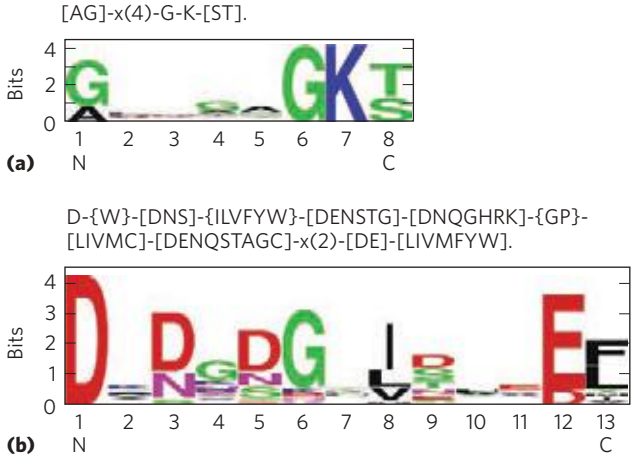


FIGURE 1 **Pepresentations of two consensus sequences. (a)** P loop, an ATP-binding structure; **(b)** EF hand, a Ca^{2+}-binding structure.

In one type of consensus sequence designation (shown at the top of (a) and (b)), each position is separated from its neighbor by a hyphen. A position where any amino acid is allowed is designated x. Ambiguities are indicated by listing the acceptable amino acids for a given position between square brackets. For example, in (a) [AG] means Ala or Gly. If all but a few amino acids are allowed at one position, the amino acids that are *not* allowed are listed between curly brackets. For example, in (b) {W} means any amino acid except Trp. Repetition of an element of the pattern is indicated by following that element with a number or range of numbers between parentheses. In (a), for example, x(4) means x-x-x-x; x(2,4) would mean x-x, or x-x-x, or x-x-x-x. When a pattern is restricted to either the amino or carboxyl terminus of a sequence, that pattern starts with < or ends with >, respectively (not so for either example here). A period ends the pattern. Applying these rules to the consensus sequence in (a), either A or G can be found at the first position. Any amino acid can occupy the next four positions, followed by an invariant G and an invariant K. The last position is either S or T.

Sequence logos provide a more informative and graphic representation of an amino acid (or nucleic acid) multiple sequence alignment. Each logo consists of a stack of symbols for each position in the sequence. The overall height of the stack (in bits) indicates the degree of sequence conservation at that position, while the height of each symbol in the stack indicates the relative frequency of that amino acid (or nucleotide). For amino acid sequences, the colors denote the characteristics of the amino acid: polar (G, S, T, Y, C, Q, N) green; basic (K, R, H) blue; acidic (D, E) red; and hydrophobic (A, V, L, I, P, W, F, M) black. The classification of amino acids in this scheme is somewhat different from that in Table 3–1 and Figure 3–5. The amino acids with aromatic side chains are subsumed into the nonpolar (F, W) and polar (Y) classifications. Glycine, always hard to group, is assigned to the polar group. Note that when multiple amino acids are acceptable at a particular position, they rarely occur with equal probability. One or a few usually predominate. The logo representation makes the predominance clear, and a conserved sequence in a protein is made obvious. However, the logo obscures some amino acid residues that may be allowed at a position, such as the Cys that occasionally occurs at position 8 of the EF hand in (b).

student, or knowledgeable layperson. Each protein's function relies on its three-dimensional structure, which in turn is determined largely by its primary structure. Thus, the biochemical information conveyed by a protein sequence is limited only by our own understanding of structural and functional principles. The constantly evolving tools of bioinformatics make it possible to identify functional segments in new proteins and help establish both their sequence and their structural relationships to proteins already in the databases. On a different level of inquiry, protein sequences are beginning to tell us how the proteins evolved and, ultimately, how life evolved on this planet.

The field of molecular evolution is often traced to Emile Zuckerkandl and Linus Pauling, whose work in the mid-1960s advanced the use of nucleotide and protein sequences to explore evolution. The premise is deceptively straightforward. If two organisms are closely related, the sequences of their genes and proteins should be similar. The sequences increasingly diverge as the evolutionary distance between two organisms increases. The promise of this approach began to be realized in the 1970s, when Carl Woese used ribosomal RNA sequences to define the Archaea as a group of living organisms distinct from the Bacteria and Eukarya (see Fig. 1–4). Protein sequences offer an opportunity to greatly refine the available information. With the advent of genome projects investigating organisms from bacteria to humans, the number of available sequences is growing at an enormous rate. This information can be

used to trace biological history. The challenge is in learning to read the genetic hieroglyphics.

Evolution has not taken a simple linear path. Complexities abound in any attempt to mine the evolutionary information stored in protein sequences. For a given protein, the amino acid residues essential for the activity of the protein are conserved over evolutionary time. The residues that are less important to function may vary over time—that is, one amino acid may substitute for another—and these variable residues can provide the information to trace evolution. Amino acid substitutions are not always random, however. At some positions in the primary structure, the need to maintain protein function may mean that only particular amino acid substitutions can be tolerated. Some proteins have more variable amino acid residues than others. For these and other reasons, different proteins can evolve at different rates.

Another complicating factor in tracing evolutionary history is the rare transfer of a gene or group of genes from one organism to another, a process called **horizontal gene transfer**. The transferred genes may be quite similar to the genes they were derived from in the original organism, whereas most other genes in the same two organisms may be quite distantly related. An example of horizontal gene transfer is the recent rapid spread of antibiotic-resistance genes in bacterial populations. The proteins derived from these transferred genes would not be good candidates for the study of bacterial evolution, because they share only a very limited evolutionary history with their "host" organisms.

The study of molecular evolution generally focuses on families of closely related proteins. In most cases, the families chosen for analysis have essential functions in cellular metabolism that must have been present in the earliest viable cells, thus greatly reducing the chance that they were introduced relatively recently by horizontal gene transfer. For example, a protein called EF-1α (elongation factor 1α) is involved in the synthesis of proteins in all eukaryotes. A similar protein, EF-Tu, with the same function, is found in bacteria. Similarities in sequence and function indicate that EF-1α and EF-Tu are members of a family of proteins that share a common ancestor. The members of protein families are called **homologous proteins**, or **homologs**. The concept of a homolog can be further refined. If two proteins in a family (that is, two homologs) are present in the same species, they are referred to as **paralogs**. Homologs from different species are called **orthologs**. The process of tracing evolution involves first identifying suitable families of homologous proteins and then using them to reconstruct evolutionary paths.

Homologs are identified through the use of increasingly powerful computer programs that can directly compare two or more chosen protein sequences, or can search vast databases to find the evolutionary relatives of one selected protein sequence. The electronic search process can be thought of as sliding one sequence past the other until a section with a good match is found. Within this sequence alignment, a positive score is assigned for each position where the amino acid residues in the two sequences are identical—the value of the score varying from one program to the next—to provide a measure of the quality of the alignment. The process has some complications. Sometimes the proteins being compared match well at, say, two sequence segments, and these segments are connected by less related sequences of different lengths. Thus the two matching segments cannot be aligned at the same time. To handle this, the computer program introduces "gaps" in one of the sequences to bring the matching segments into register **(Fig. 3–33)**. Of course, if a sufficient number of gaps are introduced, almost any two sequences could be brought into some sort of alignment. To avoid uninformative alignments, the programs include penalties for each gap introduced, thus lowering the overall alignment score. With electronic trial and error, the program selects the alignment with the optimal score that maximizes identical amino acid residues while minimizing the introduction of gaps.

Finding identical amino acids is often inadequate to identify related proteins or, more importantly, to determine how closely related the proteins are on an evolutionary time scale. A more useful analysis also considers the chemical properties of substituted amino acids. Many of the amino acid differences within a protein family may be conservative—that is, an amino acid residue is replaced by a residue having similar chemical properties. For example, a Glu residue may substitute in one family member for the Asp residue found in another; both amino acids are negatively charged. Such a conservative substitution should logically receive a higher score in a sequence alignment than does a nonconservative substitution, such as the replacement of the Asp residue with a hydrophobic Phe residue.

For most efforts to find homologies and explore evolutionary relationships, protein sequences (derived either directly from protein sequencing or from the

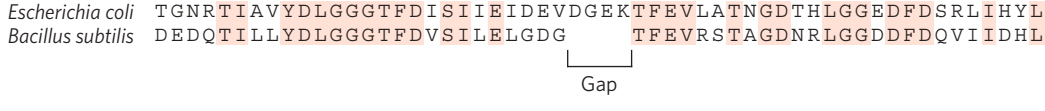


FIGURE 3–33 Aligning protein sequences with the use of gaps. Shown here is the sequence alignment of a short section of the Hsp70 proteins (a widespread class of protein-folding chaperones) from two well-studied bacterial species, *E. coli* and *Bacillus subtilis*. Introduction of a gap in the *B. subtilis* sequence allows a better alignment of amino acid residues on either side of the gap. Identical amino acid residues are shaded.

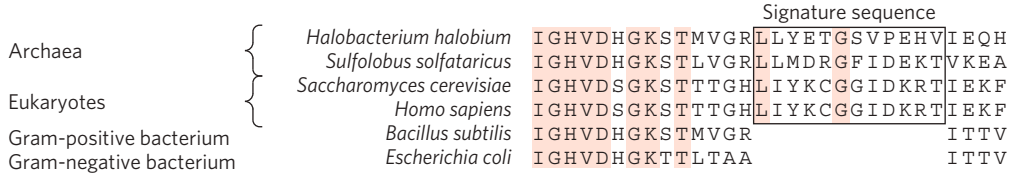


FIGURE 3–34 A signature sequence in the EF-1α/EF-Tu protein family. The signature sequence (boxed) is a 12-residue insertion near the amino terminus of the sequence. Residues that align in all species are shaded. Both archaea and eukaryotes have the signature, although the sequences of the insertions are quite distinct for the two groups. The variation in the signature sequence reflects the significant evolutionary divergence that has occurred at this site since it first appeared in a common ancestor of both groups.

sequencing of the DNA encoding the protein) are superior to nongenic nucleic acid sequences (those that do not encode a protein or functional RNA). For a nucleic acid, with its four different types of residues, random alignment of nonhomologous sequences will generally yield matches for at least 25% of the positions. Introduction of a few gaps can often increase the fraction of matched residues to 40% or more, and the probability of chance alignment of unrelated sequences becomes quite high. The 20 different amino acid residues in proteins greatly lower the probability of uninformative chance alignments of this type.

The programs used to generate a sequence alignment are complemented by methods that test the reliability of the alignments. A common computerized test is to shuffle the amino acid sequence of one of the proteins being compared to produce a random sequence, then to instruct the program to align the shuffled sequence with the other, unshuffled one. Scores are assigned to the new alignment, and the shuffling and alignment process is repeated many times. The original alignment, before shuffling, should have a score significantly higher than any of those within the distribution of scores generated by the random alignments; this increases the confidence that the sequence alignment has identified a pair of homologs. Note that the absence of a significant alignment score does not necessarily mean that no evolutionary relationship exists between two proteins. As we shall see in Chapter 4, three-dimensional structural similarities sometimes reveal evolutionary relationships where sequence homology has been wiped away by time.

To use a protein family to explore evolution, researchers identify family members with similar molecular functions in the widest possible range of organisms. Information from the family can then be used to trace the evolution of those organisms. By analyzing the sequence divergence in selected protein families, investigators can segregate organisms into classes based on their evolutionary relationships. This information must be reconciled with more classical examinations of the physiology and biochemistry of the organisms.

Certain segments of a protein sequence may be found in the organisms of one taxonomic group but not in other groups; these segments can be used as **signature sequences** for the group in which they are found. An example of a signature sequence is an insertion of 12 amino acids near the amino terminus of the EF-1α/EF-Tu proteins in all archaea and eukaryotes but not in bacteria **(Fig. 3–34)**. This particular signature is one of many biochemical clues that can help establish the evolutionary relatedness of eukaryotes and archaea. Signature sequences have been used to establish evolutionary relationships among groups of organisms at many different taxonomic levels.

By considering the entire sequence of a protein, researchers can now construct more elaborate evolutionary trees with many species in each taxonomic group. **Figure 3–35** presents one such tree for bacteria, based on sequence divergence in the protein GroEL (a protein present in all bacteria that assists in the proper folding of proteins). The tree can be refined by basing it on the sequences of multiple proteins and by supplementing the sequence information with data on the unique biochemical and physiological properties of each species. There are many methods for generating trees, each method with its own advantages and shortcomings, and many ways to represent the resulting evolutionary relationships. In Figure 3–35, the free end points of lines are called “external nodes”; each represents an extant species, and each is so labeled. The points where two lines come together, the “internal nodes,” represent extinct ancestor species. In most representations (including Fig. 3–35), the lengths of the lines connecting the nodes are proportional to the number of amino acid substitutions separating one species from another. If we trace two extant species to a common internal node (representing the common ancestor of the two species), the length of the branch connecting each external node to the internal node represents the number of amino acid substitutions separating one extant species from this ancestor. The sum of the lengths of all the line segments that connect an extant species to another extant species through a common ancestor reflects the number of substitutions separating the two extant species. To determine how much time was needed for the various species to diverge, the tree must be calibrated by comparing it with information from the fossil record and other sources.

As more sequence information is made available in databases, we can generate evolutionary trees based on multiple proteins. And we can refine these trees as

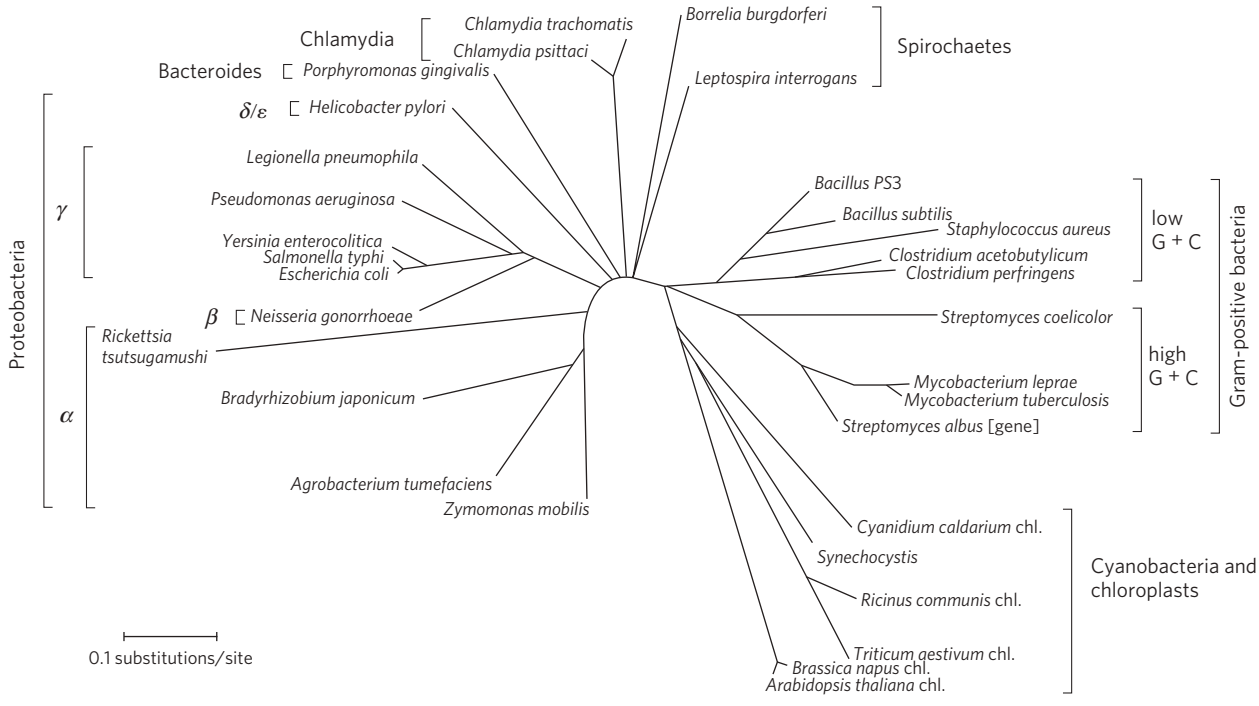


FIGURE 3–35 Evolutionary tree derived from amino acid sequence comparisons. A bacterial evolutionary tree, based on the sequence divergence observed in the GroEL family of proteins. Also included in this tree (lower right) are the chloroplasts (chl.) of some nonbacterial species.

additional genomic information emerges from increasingly sophisticated methods of analysis. All of this work moves us toward the goal of creating a detailed tree of life that describes the evolution and relationship of every organism on Earth. The story is a work in progress, of course **(Fig. 3–36)**. The questions being asked and answered are fundamental to how humans view themselves and the world around them. The field of molecular evolution promises to be among the most vibrant of the scientific frontiers in the twenty-first century.

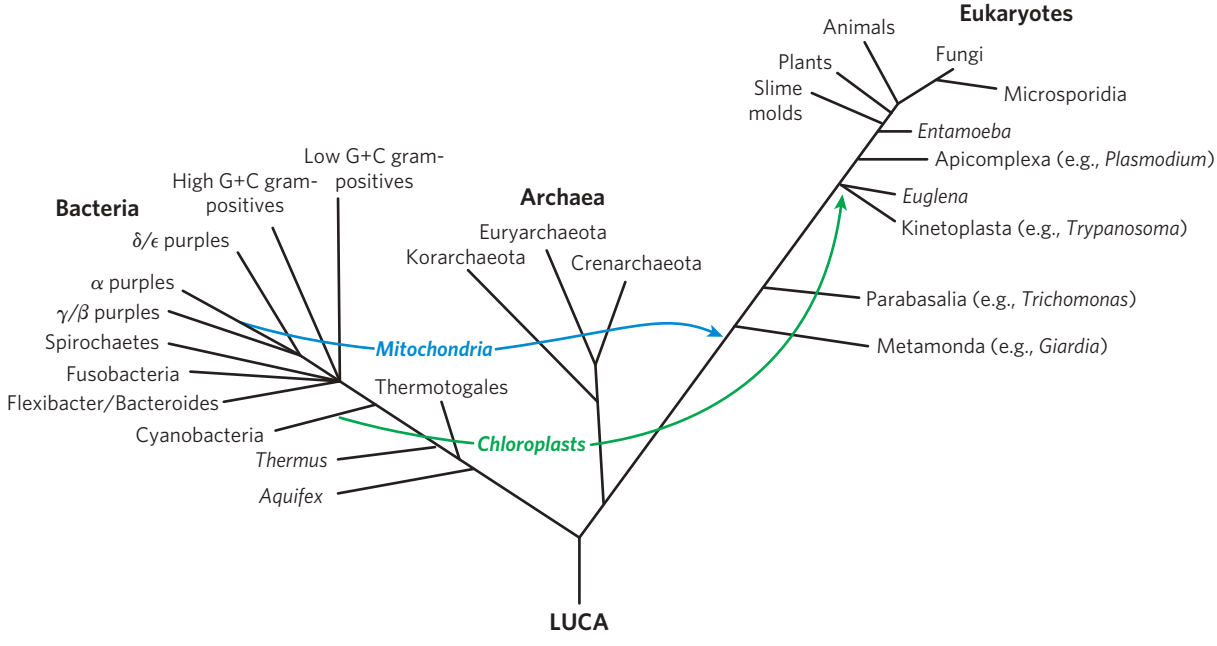


FIGURE 3–36 A consensus tree of life. The tree shown here is based on analyses of many different protein sequences and additional genomic features. The tree presents only a fraction of the available information, as well as only a fraction of the issues remaining to be resolved. Each extant group shown is a complex evolutionary story unto itself. LUCA is the last universal common ancestor from which all other life forms evolved. The blue and green arrows indicate the endosymbiotic assimilation of particular types of bacteria into eukaryotic cells to become mitochondria and chloroplasts, respectively (see Fig. 1–38).

SUMMARY 3.4 The Structure of Proteins: Primary Structure

- Differences in protein function result from differences in amino acid composition and sequence. Some variations in sequence may occur in a particular protein, with little or no effect on its function.
- Amino acid sequences are deduced by fragmenting polypeptides into smaller peptides with reagents known to cleave specific peptide bonds; determining the amino acid sequence of each fragment by the automated Edman degradation procedure; then ordering the peptide fragments by finding sequence overlaps between fragments generated by different reagents. A protein sequence can also be deduced from the nucleotide sequence of its corresponding gene in DNA, or by mass spectrometry.
- Short proteins and peptides (up to about 100 residues) can be chemically synthesized. The peptide is built up, one amino acid residue at a time, while tethered to a solid support.
- Protein sequences are a rich source of information about protein structure and function, as well as the evolution of life on Earth. Sophisticated methods are being developed to trace evolution by analyzing the resultant slow changes in amino acid sequences of homologous proteins.

Key Terms

Terms in bold are defined in the glossary.

amino acids 76
residue 76
R group 76
chiral center 76
enantiomers 76
absolute configuration 78
D, L system 78
polarity 78
absorbance, *A* 80
zwitterion 81
isoelectric pH (isoelectric point, pI) 84
peptide 85
protein 85
peptide bond 85
oligopeptide 86
polypeptide 86
oligomeric protein 88
protomer 88
conjugated protein 89
prosthetic group 89
crude extract 89
fraction 89
fractionation 89
dialysis 90
column chromatography 90
ion-exchange chromatography 90
size-exclusion chromatography 92
affinity chromatography 92
high-performance liquid chromatography (HPLC) 92
electrophoresis 92
sodium dodecyl sulfate (SDS) 94
isoelectric focusing 94
specific activity 95
primary structure 97
secondary structure 97
tertiary structure 97
quaternary structure 97
Edman degradation 98
proteases 99
MALDI MS 101
ESI MS 101
consensus sequence 104
bioinformatics 104
horizontal gene transfer 106
homologous proteins 106
homologs 106
paralogs 106
orthologs 106
signature sequence 107

Further Reading

Amino Acids

Dougherty, D.A. (2000) Unnatural amino acids as probes of protein structure and function. *Curr. Opin. Chem. Biol.* **4**, 645–652.

Kreil, G. (1997) D-Amino acids in animal peptides. *Annu. Rev. Biochem.* **66**, 337–345.
Details the occurrence of these unusual stereoisomers of amino acids.

Meister, A. (1965) *Biochemistry of the Amino Acids*, 2nd edn, Vols 1 and 2, Academic Press, Inc., New York.
Encyclopedic treatment of the properties, occurrence, and metabolism of amino acids.

Peptides and Proteins

Creighton, T.E. (1992) *Proteins: Structures and Molecular Properties*, 2nd edn, W. H. Freeman and Company, New York.
Very useful general source.

Working with Proteins

Dunn, M.J. & Corbett, J.M. (1996) Two-dimensional polyacrylamide gel electrophoresis. *Methods Enzymol.* **271**, 177–203.
A detailed description of the technology.

Kornberg, A. (1990) Why purify enzymes? *Methods Enzymol.* **182**, 1–5.
The critical role of classical biochemical methods in a new age.

Scopes, R.K. (1994) *Protein Purification: Principles and Practice*, 3rd edn, Springer-Verlag, New York.
A good source for more complete descriptions of the principles underlying chromatography and other methods.

Protein Primary Structure and Evolution

Andersson, L., Blomberg, L., Flegel, M., Lepsa, L., Nilsson, B., & Verlander, M. (2000) Large-scale synthesis of peptides. *Biopolymers* **55**, 227–250.
A discussion of approaches to manufacturing peptides as pharmaceuticals.

Dell, A. & Morris, H.R. (2001) Glycoprotein structure determination by mass spectrometry. *Science* **291**, 2351–2356.
Glycoproteins can be complex; mass spectrometry is a preferred method for sorting things out.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005) Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375.

Gogarten, J.P. & Townsend, J.P. (2005) Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3**, 679–687.

Gygi, S.P. & Aebersold, R. (2000) Mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.* **4**, 489–494.
Uses of mass spectrometry to identify and study cellular proteins.

Koonin, E.V., Tatusov, R.L., & Galperin, M.Y. (1998) Beyond complete genomes: from sequence to structure and function. *Curr. Opin. Struct. Biol.* **8**, 355–363.
A good discussion about the possible uses of the increasing amount of information on protein sequences.

Li, W.-H. & Graur, D. (2000) *Fundamentals of Molecular Evolution*, 2nd edn, Sinauer Associates, Inc., Sunderland, MA.
A very readable text describing methods used to analyze protein and nucleic acid sequences. Chapter 5 provides one of the best

available descriptions of how evolutionary trees are constructed from sequence data.

Mayo, K.H. (2000) Recent advances in the design and construction of synthetic peptides: for the love of basics or just for the technology of it. *Trends Biotechnol.* **18**, 212–217.

Miranda, L.P. & Alewood, P.F. (2000) Challenges for protein chemical synthesis in the 21st century: bridging genomics and proteomics. *Biopolymers* **55**, 217–226.

This and the article by Mayo (above) describe how to make peptides and splice them together to address a wide range of problems in protein biochemistry.

Ramisetty, S.R. & Washburn, M.P. (2011) Unraveling the dynamics of protein interactions with quantitative mass spectrometry. *Crit. Rev. Biochem. Mol. Biol.* **46**, 216–228.

Rokas, A., Williams, B.L., King, N., & Carroll, S.B. (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804.

How sequence comparisons of multiple proteins can yield accurate evolutionary information.

Sanger, F. (1988) Sequences, sequences, sequences. *Annu. Rev. Biochem.* **57**, 1–28.

A nice historical account of the development of sequencing methods.

Snel, B., Huynen, M.A., & Dutilh, B.E. (2005) Genome trees and the nature of genome evolution. *Annu. Rev. Microbiol.* **59**, 191–209.

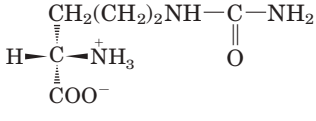
Steen, H. & Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5**, 699–711.

Zuckerkandl, E. & Pauling, L. (1965) Molecules as documents of evolutionary history. *J. Theor. Biol.* **8**, 357–366.

Many consider this the founding paper in the field of molecular evolution.

Problems

1. Absolute Configuration of Citrulline The citrulline isolated from watermelons has the structure shown below. Is it a D- or L-amino acid? Explain.



2. Relationship between the Titration Curve and the Acid-Base Properties of Glycine A 100 mL solution of 0.1 M glycine at pH 1.72 was titrated with 2 M NaOH solution. The pH was monitored and the results were plotted as shown in the graph. The key points in the titration are designated I to V. For each of the statements (a) to (o), *identify* the appropriate key point in the titration and *justify* your choice.

(a) Glycine is present predominantly as the species $^{+}H_3N—CH_2—COOH$.

(b) The *average* net charge of glycine is $+\frac{1}{2}$.

(c) Half of the amino groups are ionized.

(d) The pH is equal to the pK_a of the carboxyl group.

(e) The pH is equal to the pK_a of the protonated amino group.

(f) Glycine has its maximum buffering capacity.

(g) The *average* net charge of glycine is zero.

(h) The carboxyl group has been completely titrated (first equivalence point).

(i) Glycine is completely titrated (second equivalence point).

(j) The predominant species is $^{+}H_3N—CH_2—COO^{-}$.

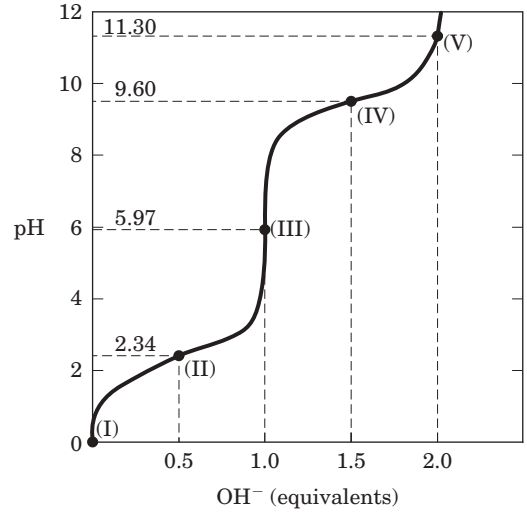
(k) The *average* net charge of glycine is -1.

(l) Glycine is present predominantly as a 50:50 mixture of $^{+}H_3N—CH_2—COOH$ and $^{+}H_3N—CH_2—COO^{-}$.

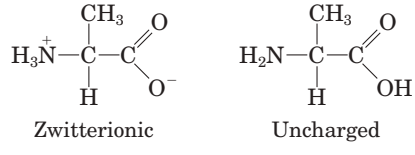
(m) This is the isoelectric point.

(n) This is the end of the titration.

(o) These are the *worst* pH regions for buffering power.



3. How Much Alanine Is Present as the Completely Uncharged Species? At a pH equal to the isoelectric point of alanine, the *net* charge on alanine is zero. Two structures can be drawn that have a net charge of zero, but the predominant form of alanine at its pI is zwitterionic.



(a) Why is alanine predominantly zwitterionic rather than completely uncharged at its pI?

(b) What fraction of alanine is in the completely uncharged form at its pI? Justify your assumptions.

4. Ionization State of Histidine Each ionizable group of an amino acid can exist in one of two states, charged or neutral. The electric charge on the functional group is determined by the relationship between its pK_a and the pH of the solution. This relationship is described by the Henderson-Hasselbalch equation.

(a) Histidine has three ionizable functional groups. Write the equilibrium equations for its three ionizations and assign the proper pK_a for each ionization. Draw the structure of histidine in each ionization state. What is the net charge on the histidine molecule in each ionization state?

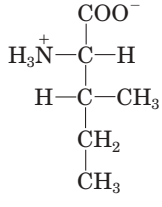
(b) Draw the structures of the predominant ionization state of histidine at pH 1, 4, 8, and 12. Note that the ionization state can be approximated by treating each ionizable group independently.

(c) What is the net charge of histidine at pH 1, 4, 8, and 12? For each pH, will histidine migrate toward the anode (+) or cathode (–) when placed in an electric field?

5. Separation of Amino Acids by Ion-Exchange Chromatography Mixtures of amino acids can be analyzed by first separating the mixture into its components through ion-exchange chromatography. Amino acids placed on a cation-exchange resin (see Fig. 3–17a) containing sulfonate ($—SO_3^-$) groups flow down the column at different rates because of two factors that influence their movement: (1) ionic attraction between the sulfonate residues on the column and positively charged functional groups on the amino acids, and (2) hydrophobic interactions between amino acid side chains and the strongly hydrophobic backbone of the polystyrene resin. For each pair of amino acids listed, determine which will be eluted first from the cation-exchange column by a pH 7.0 buffer.

(a) Asp and Lys
(b) Arg and Met
(c) Glu and Val
(d) Gly and Leu
(e) Ser and Ala

6. Naming the Stereoisomers of Isoleucine The structure of the amino acid isoleucine is



(a) How many chiral centers does it have?
(b) How many optical isomers?
(c) Draw perspective formulas for all the optical isomers of isoleucine.

7. Comparing the pK_a Values of Alanine and Polyalanine The titration curve of alanine shows the ionization of two functional groups with pK_a values of 2.34 and 9.69, corresponding to the ionization of the carboxyl and the protonated amino groups, respectively. The titration of di-, tri-, and larger oligopeptides of alanine also shows the ionization of only two functional groups, although the experimental pK_a values are different. The trend in pK_a values is summarized in the table.

Amino acid or peptide	pK_1	pK_2
Ala	2.34	9.69
Ala–Ala	3.12	8.30
Ala–Ala–Ala	3.39	8.03
Ala–$(Ala)_n$–Ala, $n \geq 4$	3.42	7.94

(a) Draw the structure of Ala–Ala–Ala. Identify the functional groups associated with pK_1 and pK_2.
(b) Why does the value of pK_1 *increase* with each additional Ala residue in the oligopeptide?
(c) Why does the value of pK_2 *decrease* with each additional Ala residue in the oligopeptide?

8. The Size of Proteins What is the approximate molecular weight of a protein with 682 amino acid residues in a single polypeptide chain?

9. The Number of Tryptophan Residues in Bovine Serum Albumin A quantitative amino acid analysis reveals that bovine serum albumin (BSA) contains 0.58% tryptophan (M_r 204) by weight.

(a) Calculate the *minimum* molecular weight of BSA (i.e., assume there is only one Trp residue per protein molecule).
(b) Size-exclusion chromatography of BSA gives a molecular weight estimate of 70,000. How many Trp residues are present in a molecule of serum albumin?

10. Subunit Composition of a Protein A protein has a molecular mass of 400 kDa when measured by size-exclusion chromatography. When subjected to gel electrophoresis in the presence of sodium dodecyl sulfate (SDS), the protein gives three bands with molecular masses of 180, 160, and 60 kDa. When electrophoresis is carried out in the presence of SDS and dithiothreitol, three bands are again formed, this time with molecular masses of 160, 90, and 60 kDa. Determine the subunit composition of the protein.

11. Net Electric Charge of Peptides A peptide has the sequence

Glu–His–Trp–Ser–Gly–Leu–Arg–Pro–Gly

(a) What is the net charge of the molecule at pH 3, 8, and 11? (Use pK_a values for side chains and terminal amino and carboxyl groups as given in Table 3–1.)
(b) Estimate the pI for this peptide.

12. Isoelectric Point of Pepsin Pepsin is the name given to a mix of several digestive enzymes secreted (as larger precursor proteins) by glands that line the stomach. These glands also secrete hydrochloric acid, which dissolves the particulate matter in food, allowing pepsin to enzymatically cleave individual protein molecules. The resulting mixture of food, HCl, and digestive enzymes is known as chyme and has a pH near 1.5. What pI would you predict for the pepsin proteins? What functional groups must be present to confer this pI on pepsin? Which amino acids in the proteins would contribute such groups?

13. Isoelectric Point of Histones Histones are proteins found in eukaryotic cell nuclei, tightly bound to DNA, which has many phosphate groups. The pI of histones is very high, about 10.8. What amino acid residues must be present in relatively large numbers in histones? In what way do these residues contribute to the strong binding of histones to DNA?

14. Solubility of Polypeptides One method for separating polypeptides makes use of their different solubilities. The solubility of large polypeptides in water depends on the relative polarity of their R groups, particularly on the number of ionized groups: the more ionized groups there are, the more soluble the polypeptide. Which of each pair of polypeptides that follow is more soluble at the indicated pH?

(a) $(Gly)_{20}$ or $(Glu)_{20}$ at pH 7.0
(b) $(Lys–Ala)_3$ or $(Phe–Met)_3$ at pH 7.0
(c) $(Ala–Ser–Gly)_5$ or $(Asn–Ser–His)_5$ at pH 6.0
(d) $(Ala–Asp–Gly)_5$ or $(Asn–Ser–His)_5$ at pH 3.0

15. Purification of an Enzyme A biochemist discovers and purifies a new enzyme, generating the purification table below.

Procedure	Total protein (mg)	Activity (units)
1. Crude extract	20,000	4,000,000
2. Precipitation (salt)	5,000	3,000,000
3. Precipitation (pH)	4,000	1,000,000
4. Ion-exchange chromatography	200	800,000
5. Affinity chromatography	50	750,000
6. Size-exclusion chromatography	45	675,000

(a) From the information given in the table, calculate the specific activity of the enzyme after each purification procedure.

(b) Which of the purification procedures used for this enzyme is most effective (i.e., gives the greatest relative increase in purity)?

(c) Which of the purification procedures is least effective?

(d) Is there any indication based on the results shown in the table that the enzyme after step 6 is now pure? What else could be done to estimate the purity of the enzyme preparation?

16. Dialysis A purified protein is in a Hepes (*N*-(2-hydroxyethyl)piperazine-*N*′-(2-ethanesulfonic acid)) buffer at pH 7 with 500 mM NaCl. A sample (1 mL) of the protein solution is placed in a tube made of dialysis membrane and dialyzed against 1 L of the same Hepes buffer with 0 mM NaCl. Small molecules and ions (such as Na^+, Cl^-, and Hepes) can diffuse across the dialysis membrane, but the protein cannot.

(a) Once the dialysis has come to equilibrium, what is the concentration of NaCl in the protein sample? Assume no volume changes occur in the sample during the dialysis.

(b) If the original 1 mL sample were dialyzed twice, successively, against 100 mL of the same Hepes buffer with 0 mM NaCl, what would be the final NaCl concentration in the sample?

17. Peptide Purification At pH 7.0, in what order would the following three peptides be eluted from a column filled with a cation-exchange polymer? Their amino acid compositions are:

Peptide A: Ala 10%, Glu 5%, Ser 5%, Leu 10%, Arg 10%, His 5%, Ile 10%, Phe 5%, Tyr 5%, Lys 10%, Gly 10%, Pro 5%, and Trp 10%.

Peptide B: Ala 5%, Val 5%, Gly 10%, Asp 5%, Leu 5%, Arg 5%, Ile 5%, Phe 5%, Tyr 5%, Lys 5%, Trp 5%, Ser 5%, Thr 5%, Glu 5%, Asn 5%, Pro 10%, Met 5%, and Cys 5%.

Peptide C: Ala 10%, Glu 10%, Gly 5%, Leu 5%, Asp 10%, Arg 5%, Met 5%, Cys 5%, Tyr 5%, Phe 5%, His 5%, Val 5%, Pro 5%, Thr 5%, Ser 5%, Asn 5%, and Gln 5%.

18. Sequence Determination of the Brain Peptide Leucine Enkephalin A group of peptides that influence nerve transmission in certain parts of the brain has been isolated from normal brain tissue. These peptides are known as opioids, because they bind to specific receptors that also bind opiate drugs, such as morphine and naloxone. Opioids thus mimic some of the properties of opiates. Some researchers consider these peptides to be the brain's own painkillers. Using the information below, determine the amino acid sequence of the opioid leucine enkephalin. Explain how your structure is consistent with each piece of information.

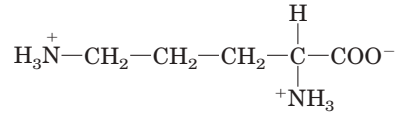
(a) Complete hydrolysis by 6 M HCl at 110 °C followed by amino acid analysis indicated the presence of Gly, Leu, Phe, and Tyr, in a 2:1:1:1 molar ratio.

(b) Treatment of the peptide with 1-fluoro-2,4-dinitrobenzene followed by complete hydrolysis and chromatography indicated the presence of the 2,4-dinitrophenyl derivative of tyrosine. No free tyrosine could be found.

(c) Complete digestion of the peptide with chymotrypsin followed by chromatography yielded free tyrosine and leucine, plus a tripeptide containing Phe and Gly in a 1:2 ratio.

19. Structure of a Peptide Antibiotic from *Bacillus brevis* Extracts from the bacterium *Bacillus brevis* contain a peptide with antibiotic properties. This peptide forms complexes with metal ions and seems to disrupt ion transport across the cell membranes of other bacterial species, killing them. The structure of the peptide has been determined from the following observations.

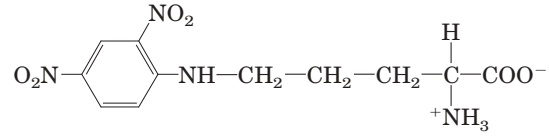
(a) Complete acid hydrolysis of the peptide followed by amino acid analysis yielded equimolar amounts of Leu, Orn, Phe, Pro, and Val. Orn is ornithine, an amino acid not present in proteins but present in some peptides. It has the structure



(b) The molecular weight of the peptide was estimated as about 1,200.

(c) The peptide failed to undergo hydrolysis when treated with the enzyme carboxypeptidase. This enzyme catalyzes the hydrolysis of the carboxyl-terminal residue of a polypeptide unless the residue is Pro or, for some reason, does not contain a free carboxyl group.

(d) Treatment of the intact peptide with 1-fluoro-2,4-dinitrobenzene, followed by complete hydrolysis and chromatography, yielded only free amino acids and the following derivative:



(Hint: The 2,4-dinitrophenyl derivative involves the amino group of a side chain rather than the α-amino group.)

(e) Partial hydrolysis of the peptide followed by chromatographic separation and sequence analysis yielded the following di- and tripeptides (the amino-terminal amino acid is always at the left):

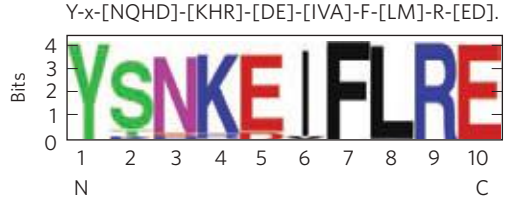
Leu–Phe Phe–Pro Orn–Leu Val–Orn

Val–Orn–Leu Phe–Pro–Val Pro–Val–Orn

Given the above information, deduce the amino acid sequence of the peptide antibiotic. Show your reasoning. When you have arrived at a structure, demonstrate that it is consistent with *each* experimental observation.

20. Efficiency in Peptide Sequencing A peptide with the primary structure Lys–Arg–Pro–Leu–Ile–Asp–Gly–Ala is sequenced by the Edman procedure. If each Edman cycle is 96% efficient, what percentage of the amino acids liberated in the fourth cycle will be leucine? Do the calculation a second time, but assume a 99% efficiency for each cycle.

21. Sequence Comparisons Proteins called molecular chaperones (described in Chapter 4) assist in the process of protein folding. One class of chaperone found in organisms from bacteria to mammals is heat shock protein 90 (Hsp90). All Hsp90 chaperones contain a 10 amino acid "signature sequence," which allows for ready identification of these proteins in sequence databases. Two representations of this signature sequence are shown below.



(a) In this sequence, which amino acid residues are invariant (conserved across all species)?

(b) At which position(s) are amino acids limited to those with positively charged side chains? For each position, which amino acid is more commonly found?

(c) At which positions are substitutions restricted to amino acids with negatively charged side chains? For each position, which amino acid predominates?

(d) There is one position that can be any amino acid, although one amino acid appears much more often than any other. What position is this, and which amino acid appears most often?

22. Chromatographic Methods Three polypeptides, the sequences of which are represented below using the one-letter code for their amino acids, are present in a mixture:

1. ATKNRASCLVPKHGALMFWRHKQLVSDPILQKR-QHILVCRNAAG
2. GPYFGDEPLDVHDEPEEG
3. PHLLSAWKGMEGVGKSQSFAALIVILA

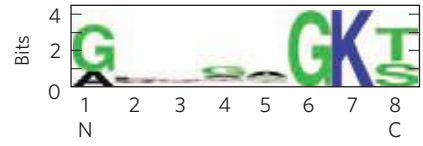
Of the three, which one would migrate most slowly during chromatography through:

(a) an ion-exchange resin; beads coated with positively charged groups?

(b) an ion-exchange resin; beads coated with negatively charged groups?

(c) a size-exclusion (gel-filtration) column designed to separate small peptides such as these?

(d) Which peptide contains the ATP-binding motif shown in the following sequence logo?

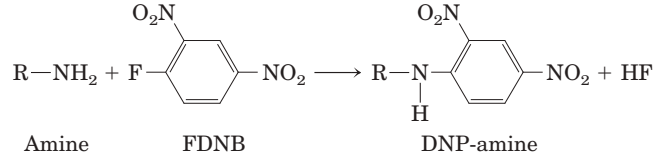


Data Analysis Problem

23. Determining the Amino Acid Sequence of Insulin Figure 3–24 shows the amino acid sequence of bovine insulin. This structure was determined by Frederick Sanger and his coworkers. Most of this work is described in a series of articles published in the *Biochemical Journal* from 1945 to 1955.

When Sanger and colleagues began their work in 1945, it was known that insulin was a small protein consisting of two or four polypeptide chains linked by disulfide bonds. Sanger and his coworkers had developed a few simple methods for studying protein sequences.

Treatment with FDNB. FDNB (1-fluoro-2,4-dinitrobenzene) reacted with free amino (but not amido or guanidino) groups in proteins to produce dinitrophenyl (DNP) derivatives of amino acids:



Acid Hydrolysis. Boiling a protein with 10% HCl for several hours hydrolyzed all of its peptide and amide bonds. Short treatments produced short polypeptides; the longer the treatment, the more complete the breakdown of the protein into its amino acids.

Oxidation of Cysteines. Treatment of a protein with performic acid cleaved all the disulfide bonds and converted all Cys residues to cysteic acid residues (see Fig. 3–28).

Paper Chromatography. This more primitive version of thin-layer chromatography (see Fig. 10–25) separated compounds based on their chemical properties, allowing identification of single amino acids and, in some cases, dipeptides. Thin-layer chromatography also separates larger peptides.

As reported in his first paper (1945), Sanger reacted insulin with FDNB and hydrolyzed the resulting protein. He found many free amino acids, but only three DNP–amino acids: α-DNP-glycine (DNP group attached to the α-amino group); α-DNP-phenylalanine; and ε-DNP-lysine (DNP attached to the ε-amino group). Sanger interpreted these results as showing that insulin had two protein chains: one with Gly at its amino terminus and one with Phe at its amino terminus. One of the two chains also contained a Lys residue, not at the amino terminus. He named the chain beginning with a Gly residue "A" and the chain beginning with Phe "B."

(a) Explain how Sanger's results support his conclusions.

(b) Are the results consistent with the known structure of bovine insulin (see Fig. 3–24)?

In a later paper (1949), Sanger described how he used these techniques to determine the first few amino acids (amino-terminal end) of each insulin chain. To analyze the B chain, for example, he carried out the following steps:

1. Oxidized insulin to separate the A and B chains.
2. Prepared a sample of pure B chain with paper chromatography.
3. Reacted the B chain with FDNB.
4. Gently acid-hydrolyzed the protein so that some small peptides would be produced.
5. Separated the DNP-peptides from the peptides that did not contain DNP groups.
6. Isolated four of the DNP-peptides, which were named B1 through B4.
7. Strongly hydrolyzed each DNP-peptide to give free amino acids.
8. Identified the amino acids in each peptide with paper chromatography.

The results were as follows:

B1: α-DNP-phenylalanine only
B2: α-DNP-phenylalanine; valine
B3: aspartic acid; α-DNP-phenylalanine; valine
B4: aspartic acid; glutamic acid; α-DNP-phenylalanine; valine

(c) Based on these data, what are the first four (amino-terminal) amino acids of the B chain? Explain your reasoning.

(d) Does this result match the known sequence of bovine insulin (see Fig. 3–24)? Explain any discrepancies.

Sanger and colleagues used these and related methods to determine the entire sequence of the A and B chains. Their sequence for the A chain was as follows (amino terminus on left):

$$\overset{1}{\text{Gly}}\text{–Ile–Val–Glx–}\overset{5}{\text{Glx}}\text{–Cys–Cys–Ala–Ser–}\overset{10}{\text{Val}}\text{–}$$
$$\text{Cys–Ser–Leu–Tyr–}\overset{15}{\text{Glx}}\text{–Leu–Glx–Asx–Tyr–}\overset{20}{\text{Cys}}\text{–Asx}$$

Because acid hydrolysis had converted all Asn to Asp and all Gln to Glu, these residues had to be designated Asx and Glx, respectively (exact identity in the peptide unknown). Sanger solved this problem by using protease enzymes that cleave peptide bonds, but not the amide bonds in Asn and Gln residues, to prepare short peptides. He then determined the number of amide groups present in each peptide by measuring the NH_4^+ released when the peptide was acid-hydrolyzed. Some of the results for the A chain are shown below. The peptides may not have been completely pure, so the numbers were approximate—but good enough for Sanger's purposes.

Peptide name	Peptide sequence	Number of amide groups in peptide
Ac1	Cys–Asx	0.7
Ap15	Tyr–Glx–Leu	0.98
Ap14	Tyr–Glx–Leu–Glx	1.06
Ap3	Asx–Tyr–Cys–Asx	2.10
Ap1	Glx–Asx–Tyr–Cys–Asx	1.94
Ap5pa1	Gly–Ile–Val–Glx	0.15
Ap5	Gly–Ile–Val–Glx–Glx–Cys–Cys–Ala–Ser–Val–Cys–Ser–Leu	1.16

(e) Based on these data, determine the amino acid sequence of the A chain. Explain how you reached your answer. Compare it with Figure 3–24.

References

Sanger, F. (1945) The free amino groups of insulin. *Biochem. J.* **39**, 507–515.

Sanger, F. (1949) The terminal peptides of insulin. *Biochem. J.* **45**, 563–574.

4

The Three-Dimensional Structure of Proteins

Proteins are big molecules. The covalent backbone of a typical protein contains hundreds of individual bonds. Because free rotation is possible around many of these bonds, the protein can in principle assume a virtually uncountable number of conformations. However, each protein has a specific chemical or structural function, suggesting that each has a unique three-dimensional structure **(Fig. 4–1)**. How stable is this structure, what factors guide its formation, and what holds it together? By the late 1920s, several proteins had been crystallized, including hemoglobin (M_r 64,500) and the enzyme urease (M_r 483,000). Given that, generally, the ordered array of molecules in a crystal can form only if the molecular units are identical, the finding that many proteins could be crystallized was evidence that even very large proteins are discrete chemical entities with unique structures. This conclusion revolutionized thinking about proteins and their functions, but the insight it provided was incomplete. Protein structure is always malleable in sometimes surprising ways. Changes in structure can be as important to a protein's function as the structure itself.

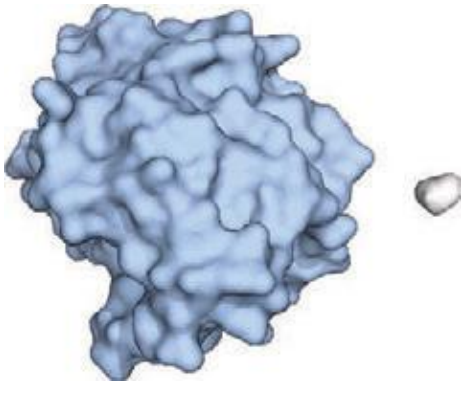


FIGURE 4–1 Structure of the enzyme chymotrypsin, a globular protein. A molecule of glycine (gray) is shown for size comparison. The known three-dimensional structures of proteins are archived in the Protein Data Bank, or PDB (see Box 4–4). The image shown here was made using data from the entry with PDB ID 6GCH.

In this chapter, we examine the structure of proteins. We emphasize six themes. First, the three-dimensional structure or structures taken up by a protein are determined by its amino acid sequence. Second, the function of a typical protein depends on its structure. Third, most isolated proteins exist in one or a small number of stable structural forms. Fourth, the most important forces stabilizing the specific structures maintained by a given protein are noncovalent interactions. Fifth, amid the huge number of unique protein structures, we can recognize some common structural patterns that help to organize our understanding of protein architecture. Sixth, protein structures are not static. All proteins undergo changes in conformation ranging from subtle to quite dramatic. Parts of many proteins have no discernible structure. For some proteins, a lack of definable structure is critical to their function.

4.1 Overview of Protein Structure

The spatial arrangement of atoms in a protein or any part of a protein is called its **conformation**. The possible conformations of a protein or protein segment include any structural state it can achieve without breaking covalent bonds. A change in conformation could occur, for example, by rotation about single bonds. Of the many conformations that are theoretically possible in a protein containing hundreds of single bonds, one or (more commonly) a few generally predominate under

biological conditions. The need for multiple stable conformations reflects the changes that must take place in most proteins as they bind to other molecules or catalyze reactions. The conformations existing under a given set of conditions are usually the ones that are thermodynamically the most stable—that is, having the lowest Gibbs free energy (G). Proteins in any of their functional, folded conformations are called **native** proteins.

For the vast majority of proteins, a particular structure or small set of structures is critical to function. However, in many cases, parts of proteins lack discernible structure. These protein segments are intrinsically disordered. In a few cases, entire proteins are intrinsically disordered, yet fully functional.

What principles determine the most stable conformations of a typical protein? An understanding of protein conformation can be built stepwise from the discussion of primary structure in Chapter 3 through a consideration of secondary, tertiary, and quaternary structures. To this traditional approach we must add the newer emphasis on common and classifiable folding patterns, variously called supersecondary structures, folds, or motifs, which provide an important organizational context to this complex endeavor. We begin by introducing some guiding principles.

A Protein's Conformation Is Stabilized Largely by Weak Interactions

In the context of protein structure, the term **stability** can be defined as the tendency to maintain a native conformation. Native proteins are only marginally stable; the ΔG separating the folded and unfolded states in typical proteins under physiological conditions is in the range of only 20 to 65 kJ/mol. A given polypeptide chain can theoretically assume countless conformations, and as a result the unfolded state of a protein is characterized by a high degree of conformational entropy. This entropy, and the hydrogen-bonding interactions of many groups in the polypeptide chain with the solvent (water), tend to maintain the unfolded state. The chemical interactions that counteract these effects and stabilize the native conformation include disulfide (covalent) bonds and the weak (noncovalent) interactions described in Chapter 2: hydrogen bonds and hydrophobic and ionic interactions.

Many proteins do not have disulfide bonds. The environment within most cells is highly reducing due to high concentrations of reductants such as glutathione, and most sulfhydryls will thus remain in the reduced state. Outside the cell, the environment is often more oxidizing, and disulfide formation is more likely to occur. In eukaryotes, disulfide bonds are found primarily in secreted, extracellular proteins (for example, the hormone insulin). Disulfide bonds are also uncommon in bacterial proteins. However, thermophilic bacteria, as well as the archaea, typically have many proteins with disulfide bonds, which stabilize proteins; this is presumably an adaptation to life at high temperatures.

For all proteins of all organisms, weak interactions are especially important in the folding of polypeptide chains into their secondary and tertiary structures. The association of multiple polypeptides to form quaternary structures also relies on these weak interactions.

About 200 to 460 kJ/mol are required to break a single covalent bond, whereas weak interactions can be disrupted by a mere 0.4 to 30 kJ/mol. Individual covalent bonds, such as disulfide bonds linking separate parts of a single polypeptide chain, are clearly much stronger than individual weak interactions. Yet, because they are so numerous, it is weak interactions that predominate as a stabilizing force in protein structure. In general, the protein conformation with the lowest free energy (that is, the most stable conformation) is the one with the maximum number of weak interactions.

The stability of a protein is not simply the sum of the free energies of formation of the many weak interactions within it. For every hydrogen bond formed in a protein during folding, a hydrogen bond (of similar strength) between the same group and water was broken. The net stability contributed by a given hydrogen bond, or the *difference* in free energies of the folded and unfolded states, may be close to zero. Ionic interactions may be either stabilizing or destabilizing. We must therefore look elsewhere to understand why a particular native conformation is favored.

On carefully examining the contribution of weak interactions to protein stability, we find that **hydrophobic interactions** generally predominate. Pure water contains a network of hydrogen-bonded H_2O molecules. No other molecule has the hydrogen-bonding potential of water, and the presence of other molecules in an aqueous solution disrupts the hydrogen bonding of water. When water surrounds a hydrophobic molecule, the optimal arrangement of hydrogen bonds results in a highly structured shell, or **solvation layer**, of water around the molecule (see Fig. 2–7). The increased order of the water molecules in the solvation layer correlates with an unfavorable decrease in the entropy of the water. However, when nonpolar groups cluster together, the extent of the solvation layer decreases, because each group no longer presents its entire surface to the solution. The result is a favorable increase in entropy. As described in Chapter 2, this increase in entropy is the major thermodynamic driving force for the association of hydrophobic groups in aqueous solution. Hydrophobic amino acid side chains therefore tend to cluster in a protein's interior, away from water (think of an oil droplet in water). The amino acid sequences of most proteins thus feature a significant content of hydrophobic amino acid side chains (especially Leu, Ile, Val, Phe, and Trp). These are positioned so that they are clustered when the protein is folded, forming a hydrophobic protein core.

Under physiological conditions, the formation of hydrogen bonds in a protein is driven largely by this same entropic effect. Polar groups can generally form hydrogen bonds with water and hence are soluble in

water. However, the number of hydrogen bonds per unit mass is generally greater for pure water than for any other liquid or solution, and there are limits to the solubility of even the most polar molecules as their presence causes a net decrease in hydrogen bonding per unit mass. Therefore, a solvation layer also forms to some extent around polar molecules. Even though the energy of formation of an intramolecular hydrogen bond between two polar groups in a macromolecule is largely canceled by the elimination of such interactions between these polar groups and water, the release of structured water as intramolecular interactions form provides an entropic driving force for folding. Most of the net change in free energy as weak interactions form within a protein is therefore derived from the increased entropy in the surrounding aqueous solution resulting from the burial of hydrophobic surfaces. This more than counterbalances the large loss of conformational entropy as a polypeptide is constrained into its folded conformation.

Hydrophobic interactions are clearly important in stabilizing conformation; the interior of a structured protein is generally a densely packed core of hydrophobic amino acid side chains. It is also important that any polar or charged groups in the protein interior have suitable partners for hydrogen bonding or ionic interactions. One hydrogen bond seems to contribute little to the stability of a native structure, but the presence of hydrogen-bonding groups without partners in the hydrophobic core of a protein can be so *destabilizing* that conformations containing these groups are often thermodynamically untenable. The favorable free-energy change resulting from the combination of several such groups with partners in the surrounding solution can be greater than the free-energy difference between the folded and unfolded states. In addition, hydrogen bonds between groups in a protein form cooperatively (formation of one makes the next one more likely) in repeating secondary structures that optimize hydrogen bonding, as described below. In this way, hydrogen bonds often have an important role in guiding the protein-folding process.

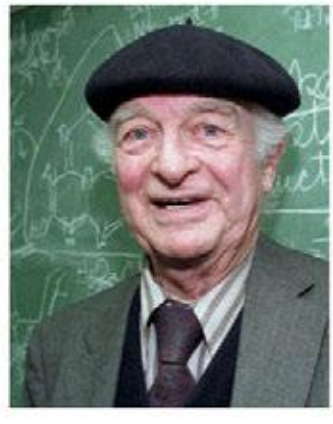
The interaction of oppositely charged groups that form an ion pair, or salt bridge, can have either a stabilizing or destabilizing effect on protein structure. As in the case of hydrogen bonds, charged amino acid side chains interact with water and salts when the protein is unfolded, and the loss of those interactions must be considered when evaluating the effect of a salt bridge on the overall stability of a folded protein. However, the strength of a salt bridge increases as it moves to an environment of lower dielectric constant, ε (p. 50): from the polar aqueous solvent (ε near 80) to the nonpolar protein interior (ε near 4). Salt bridges, especially those that are partly or entirely buried, can thus provide significant stabilization to a protein structure. This trend explains the increased occurrence of buried salt bridges in the proteins of thermophilic organisms. Ionic interactions also limit structural flexibility and confer a uniqueness to a particular protein structure that nonspecific hydrophobic interactions cannot provide.

In the tightly packed atomic environment of a protein, one more type of weak interaction can have a significant effect—van der Waals interactions (p. 54). Van der Waals interactions are dipole-dipole interactions involving the permanent electric dipoles in groups such as carbonyls, transient dipoles derived from fluctuations of the electron cloud surrounding any atom, and dipoles induced by interaction of an atom with another that has a permanent or transient dipole. As atoms approach each other, these dipole-dipole interactions provide an attractive intermolecular force that operates only over a limited intermolecular distance (0.3 to 0.6 nm). Van der Waals interactions are weak and individually contribute little to overall protein stability. However, in a well-packed protein, or in an interaction between a protein and another protein or other molecule at a complementary surface, the number of such interactions can be substantial.

Most of the structural patterns outlined in this chapter reflect two simple rules: (1) hydrophobic residues are largely buried in the protein interior, away from water, and (2) the number of hydrogen bonds and ionic interactions within the protein is maximized, thus reducing the number of hydrogen-bonding and ionic groups that are not paired with a suitable partner. Proteins within membranes (which we examine in Chapter 11) and proteins that are intrinsically disordered or have intrinsically disordered segments follow different rules. This reflects their particular function or environment, but weak interactions are still critical structural elements. For example, soluble but intrinsically disordered protein segments are enriched in amino acid side chains that are charged (especially Arg, Lys, Glu) or small (Gly, Ala), providing little or no opportunity for the formation of a stable hydrophobic core.

The Peptide Bond Is Rigid and Planar

Protein Architecture—Primary Structure Covalent bonds, too, place important constraints on the conformation of a polypeptide. In the late 1930s, Linus Pauling and Robert Corey embarked on a series of studies that laid the foundation for our current understanding of protein structure. They began with a careful analysis of the peptide bond.



Linus Pauling, 1901–1994

Robert Corey, 1897–1971

The α carbons of adjacent amino acid residues are separated by three covalent bonds, arranged as C_α—C—N—C_α. X-ray diffraction studies of crystals of amino acids and of simple dipeptides and tripeptides showed that the peptide C—N bond is somewhat shorter than the C—N bond in a simple amine and that the atoms associated with the peptide bond are coplanar. This indicated a resonance or partial sharing of two pairs of electrons between the carbonyl oxygen and the amide nitrogen **(Fig. 4–2a)**. The oxygen has a partial negative charge and the hydrogen bonded to the nitrogen has a net partial positive charge, setting up a small electric dipole. The six atoms of the **peptide group** lie in a single plane, with the oxygen atom of the carbonyl group trans to the hydrogen atom of the amide nitrogen. From these findings Pauling and Corey concluded that the peptide C—N bonds, because of their partial double-bond character, cannot rotate freely. Rotation is permitted about the N—C_α and the C_α—C bonds. The backbone of a polypeptide chain can thus be pictured as a series of rigid planes, with consecutive planes sharing a common point of rotation at C_α (Fig. 4–2b). The rigid peptide bonds limit the range of conformations possible for a polypeptide chain.

Peptide conformation is defined by three dihedral angles (also known as torsion angles) called ϕ (phi), ψ (psi), and ω (omega), reflecting rotation about each of the three repeating bonds in the peptide backbone. A dihedral angle is the angle at the intersection of two planes. In the case of peptides, the planes are defined by bond vectors in the peptide backbone. Two successive bond vectors describe a plane. Three successive bond vectors describe two planes (the central bond vector is common to both; Fig. 4–2c), and the angle between these two planes is what we measure to describe protein conformation.

KEY CONVENTION: The important dihedral angles in a peptide are defined by the three bond vectors connecting four consecutive main-chain (peptide backbone) atoms (Fig. 4–2c): ϕ involves the C—N—C_α—C bonds (with the rotation occurring about the N—C_α bond), and ψ involves the N—C_α—C—N bonds. Both ϕ and ψ are defined as $\pm 180°$ when the polypeptide is fully extended and all peptide groups are in the same plane (Fig. 4–2d). As one looks down the central bond vector in the direction of the vector arrow (as depicted in Fig. 4–2c for ψ), the dihedral angles increase as the distal

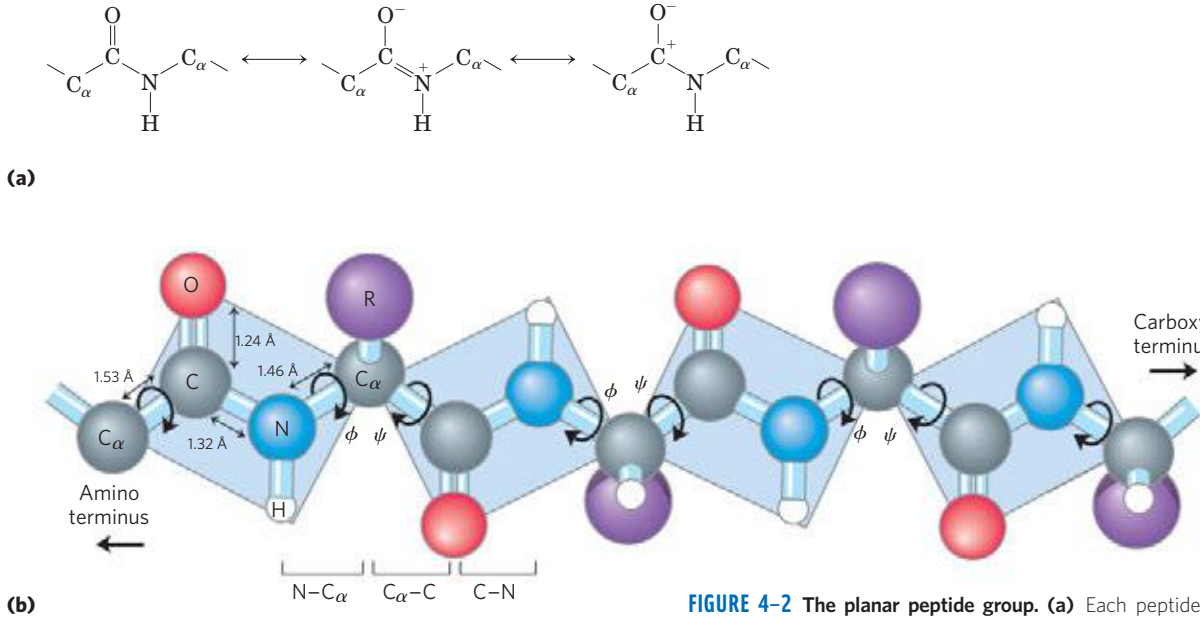


FIGURE 4–2 The planar peptide group. (a) Each peptide bond has some double-bond character due to resonance and cannot rotate. Although the N atom in a peptide bond is often represented with a partial positive charge, careful consideration of bond orbitals and quantum mechanics indicates that the N has a net charge that is neutral or slightly negative. **(b)** Three bonds separate sequential α carbons in a polypeptide chain. The N—C_α and C_α—C bonds can rotate, described by dihedral angles designated ϕ and ψ, respectively. The peptide C—N bond is not free to rotate. Other single bonds in the backbone may also be rotationally hindered, depending on the size and charge of the R groups. **(c)** The atoms and planes defining ψ. **(d)** By convention, ϕ and ψ are 180° (or −180°) when the first and fourth atoms are farthest apart and the peptide is fully extended. As the viewer looks out along the bond undergoing rotation (from either direction), the ϕ and ψ angles increase as the fourth atom rotates clockwise relative to the first. In a protein, some of the conformations shown here (e.g., 0°) are prohibited by steric overlap of atoms. In (b) through (d), the balls representing atoms are smaller than the van der Waals radii for this scale.

(fourth) atom is rotated clockwise (Fig. 4–2d). From the ±180° position, the dihedral angle increases from –180° to 0°, at which point the first and fourth atoms are eclipsed. The rotation can be continued from 0° to +180° (same position as –180°) to bring the structure back to the starting point. The third dihedral angle, ω, is not often considered. It involves the C_α—C—N—C_α bonds. The central bond in this case is the peptide bond, where rotation is constrained. The peptide bond is normally (99.6% of the time) in the trans configuration, constraining ω to a value of ±180°. For a rare cis peptide bond, $\omega = 0°$. ■

In principle, ϕ and ψ can have any value between –180° and +180°, but many values are prohibited by steric interference between atoms in the polypeptide backbone and amino acid side chains. The conformation in which both ϕ and ψ are 0° (Fig. 4–2d) is prohibited for this reason; this conformation is merely a reference point for describing the dihedral angles. Allowed values for ϕ and ψ become evident when ψ is plotted versus ϕ in a **Ramachandran plot** **(Fig. 4–3)**, introduced by G. N. Ramachandran. We will see that Ramachandran plots are very useful tools that are often used to test the quality of three-dimensional protein structures that are deposited in international databases.

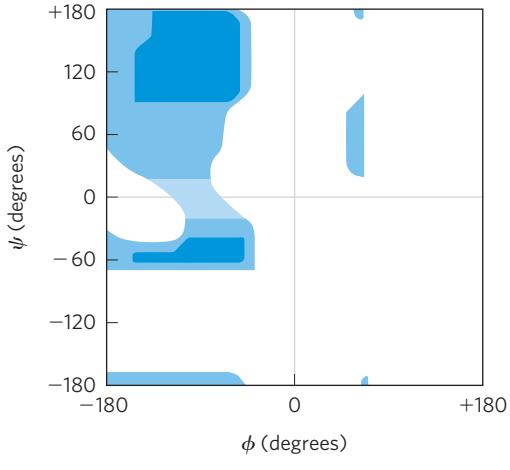


FIGURE 4–3 Ramachandran plot for L-Ala residues. Peptide conformations are defined by the values of ϕ and ψ. Conformations deemed possible are those that involve little or no steric interference, based on calculations using known van der Waals radii and dihedral angles. The areas shaded dark blue represent conformations that involve no steric overlap if the van der Waals radii of each atom are modeled as a hard sphere and thus are fully allowed; medium blue indicates conformations permitted if atoms are allowed to approach each other by an additional 0.1 nm, a slight clash; the lightest blue indicates conformations that are permissible if a very modest flexibility (a few degrees) is allowed in the ω dihedral angle that describes the peptide bond itself (generally constrained to 180°). The white regions are conformations that are not allowed. The asymmetry of the plot results from the L stereochemistry of the amino acid residues. The plots for other L residues with unbranched side chains are nearly identical. Allowed ranges for branched residues such as Val, Ile, and Thr are somewhat smaller than for Ala. The Gly residue, which is less sterically hindered, has a much broader range of allowed conformations. The range for Pro residues is greatly restricted because ϕ is limited by the cyclic side chain to the range of −35° to −85°.

SUMMARY 4.1 Overview of Protein Structure

- A typical protein usually has one or more stable three-dimensional structures, or conformations, that reflect its function. Some proteins have segments that are intrinsically disordered.
- Protein structure is stabilized largely by multiple weak interactions. Hydrophobic interactions, derived from the increase in entropy of the surrounding water when nonpolar molecules or groups are clustered together, are the major contributors to stabilizing the globular form of most soluble proteins. Van der Waals interactions also contribute. Hydrogen bonds and ionic interactions are optimized in the thermodynamically most stable structures.
- Nonpeptide covalent bonds, particularly disulfide bonds, play a role in the stabilization of structure in some proteins.
- The nature of the covalent bonds in the polypeptide backbone places constraints on structure. The peptide bond has a partial double-bond character that keeps the entire six-atom peptide group in a rigid planar configuration. The N—C_α and C_α—C bonds can rotate to define the dihedral angles ϕ and ψ, respectively.
- The Ramachandran plot is a visual description of the combinations of ϕ and ψ dihedral angles that are permitted in a peptide backbone or that are not permitted due to steric constraints.

4.2 Protein Secondary Structure

The term **secondary structure** refers to any chosen segment of a polypeptide chain and describes the local spatial arrangement of its main-chain atoms, without regard to the positioning of its side chains or its relationship to other segments. A *regular* secondary structure occurs when each dihedral angle, ϕ and ψ, remains the same or nearly the same throughout the segment. There are a few types of secondary structure that are particularly stable and occur widely in proteins. The most prominent are the α helix and β conformations; another common type is the β turn. Where a regular pattern is not found, the secondary structure is sometimes referred to as undefined or as a random coil. This last designation, however, does not properly describe the structure of these segments. The path of most of the polypeptide backbone in a typical protein is not random; rather, it is unchanging and highly specific to the structure and function of that particular protein. Our discussion here focuses on the regular structures that are most common.

The α Helix Is a Common Protein Secondary Structure

Protein Architecture–α Helix Pauling and Corey were aware of the importance of hydrogen bonds in orienting polar chemical groups such as the C═O and N—H groups of the peptide bond. They also had the experimental results of William Astbury, who in the 1930s had conducted pioneering x-ray studies of proteins. Astbury demonstrated that the protein that makes up hair and porcupine quills (the fibrous protein α-keratin) has a regular structure that repeats every 5.15 to 5.2 Å. (The angstrom, Å, named after the physicist Anders J. Ångström, is equal to 0.1 nm. Although not an SI unit, it is used universally by structural biologists to describe atomic distances—it is approximately the length of a typical C—H bond.) With this information and their data on the peptide bond, and with the help of precisely constructed models, Pauling and Corey set out to determine the likely conformations of protein molecules.

The first breakthrough came in 1948. Pauling was a visiting lecturer at Oxford University, became ill, and retired to his apartment for several days of rest. Bored with the reading available, Pauling grabbed some paper and pencils to work out a plausible stable structure that could be taken up by a polypeptide chain. The model he developed, and later confirmed in work with Corey and coworker Herman Branson, was the simplest arrangement the polypeptide chain can assume that maximizes the use of internal hydrogen bonding. It is a helical structure, and Pauling and Corey called it the **α helix** (Fig. 4–4). In this structure, the polypeptide backbone is tightly wound around an imaginary axis drawn longitudinally through the middle of the helix, and the R groups of the amino acid residues protrude outward from the helical backbone. The repeating unit is a single turn of the helix, which extends about 5.4 Å along the long axis, slightly greater than the periodicity Astbury observed on x-ray analysis of hair keratin. The backbone atoms of the amino acid residues in the prototypical α helix have a characteristic set of dihedral angles that define the α-helix conformation (Table 4–1), and each helical turn includes 3.6 amino acid residues. The α-helical segments in proteins often deviate slightly from these dihedral angles, and even vary somewhat within a single contiguous segment to produce subtle bends or kinks in the helical axis. Pauling and Corey considered both right- and left-handed variants of the α helix. The subsequent elucidation of the three-dimensional structure of myoglobin and other proteins showed that the right-handed α helix is the common form (Box 4–1). Extended left-handed α helices are theoretically less stable and have not been observed in proteins. The α helix proved to be the predominant structure in

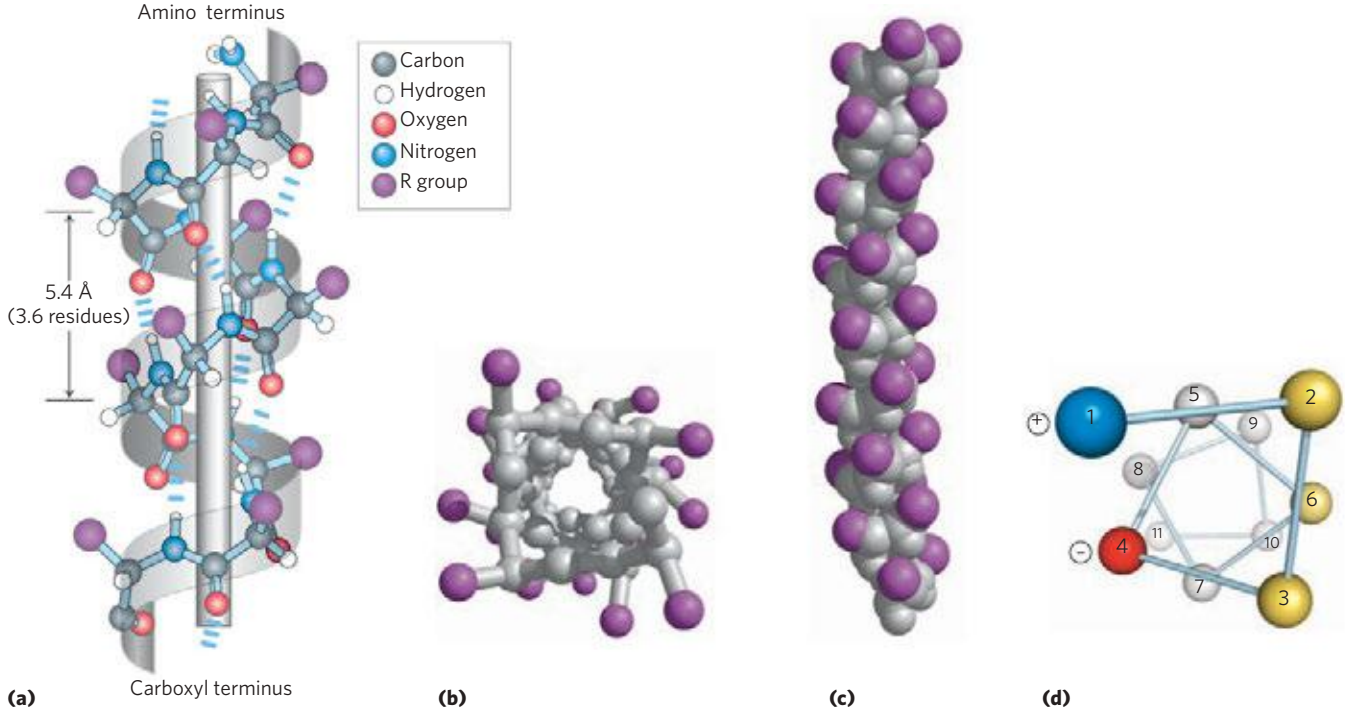


FIGURE 4–4 Models of the α helix, showing different aspects of its structure. (a) Ball-and-stick model showing the intrachain hydrogen bonds. The repeat unit is a single turn of the helix, 3.6 residues. **(b)** The α helix viewed from one end, looking down the longitudinal axis (derived from PDB ID 4TNC). Note the positions of the R groups, represented by purple spheres. This ball-and-stick model, which emphasizes the helical arrangement, gives the false impression that the helix is hollow, because the balls do not represent the van der Waals radii of the individual atoms. **(c)** As this space-filling model shows, the atoms in the center of the α helix are in very close contact. **(d)** Helical wheel projection of an α helix. This representation can be colored to identify surfaces with particular properties. The yellow residues, for example, could be hydrophobic and conform to an interface between the helix shown here and another part of the same or another polypeptide. The red (negative) and blue (positive) residues illustrate the potential for interaction of oppositely charged side chains separated by two residues in the helix.

TABLE 4–1 Idealized ϕ and ψ Angles for Common Secondary Structures in Proteins

Structure	ϕ	ψ
α Helix	−57°	−47°
β Conformation		
Antiparallel	−139°	+135°
Parallel	−119°	+113°
Collagen triple helix	−51°	+153°
β Turn type I		
i + 1*	−60°	−30°
i + 2*	−90°	0°
β Turn type II		
i + 1	−60°	+120°
i + 2	+80°	0°

Note: In real proteins, the dihedral angles often vary somewhat from these idealized values.

*The i+1 and i+2 angles are those for the second and third amino acid residues in the β turn, respectively.

α-keratins. More generally, about one-fourth of all amino acid residues in proteins are found in α helices, the exact fraction varying greatly from one protein to another.

Why does the α helix form more readily than many other possible conformations? The answer lies in part in its optimal use of internal hydrogen bonds. The structure is stabilized by a hydrogen bond between the hydrogen atom attached to the electronegative nitrogen atom of a peptide linkage and the electronegative carbonyl oxygen atom of the fourth amino acid on the amino-terminal side of that peptide bond (Fig. 4–4a). Within the α helix, every peptide bond (except those close to each end of the helix) participates in such hydrogen bonding. Each successive turn of the α helix is held to adjacent turns by three to four hydrogen bonds, conferring significant stability on the overall structure. At the ends of an α-helical segment, there are always three or four amide carbonyl or amino groups that cannot participate in this helical pattern of hydrogen bonding. These may be exposed to the surrounding solvent, where they hydrogen-bond with water, or other parts of the protein may cap the helix to provide the needed hydrogen-bonding partners.

Further experiments have shown that an α helix can form in polypeptides consisting of either L- or D-amino acids. However, all residues must be of one stereoisomeric series; a D-amino acid will disrupt a regular structure consisting of L-amino acids, and vice versa. The most stable form of an α helix consisting of D-amino acids is left-handed.

WORKED EXAMPLE 4–1 Secondary Structure and Protein Dimensions

What is the length of a polypeptide with 80 amino acid residues in a single contiguous α helix?

Solution: An idealized α helix has 3.6 residues per turn and the rise along the helical axis is 5.4 Å. Thus, the rise along the axis for each amino acid residue is 1.5 Å. The length of the polypeptide is therefore 80 residues × 1.5 Å/residue = 120 Å.

Amino Acid Sequence Affects Stability of the α Helix

Not all polypeptides can form a stable α helix. Each amino acid residue in a polypeptide has an intrinsic propensity to form an α helix (Table 4–2), reflecting the properties of the R group and how they affect the capacity of the adjoining main-chain atoms to take up the characteristic ϕ and ψ angles. Alanine shows the greatest tendency to form α helices in most experimental model systems.

The position of an amino acid residue relative to its neighbors is also important. Interactions between amino

BOX 4–1 METHODS Knowing the Right Hand from the Left

There is a simple method for determining whether a helical structure is right-handed or left-handed. Make fists of your two hands with thumbs outstretched and pointing away from you. Looking at your right hand, think of a helix spiraling up your right thumb in the direction in which the other four fingers are curled as shown (clockwise). The resulting helix is right-handed. Your left hand will demonstrate a left-handed helix, which rotates in the counterclockwise direction as it spirals up your thumb.

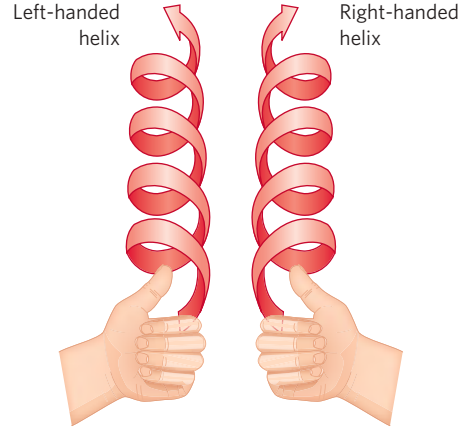


TABLE 4–2 Propensity of Amino Acid Residues to Take Up an α-Helical Conformation

Amino acid	$\Delta\Delta G°$ (kJ/mol)*	Amino acid	$\Delta\Delta G°$ (kJ/mol)*
Ala	0	Leu	0.79
Arg	0.3	Lys	0.63
Asn	3	Met	0.88
Asp	2.5	Phe	2.0
Cys	3	Pro	>4
Gln	1.3	Ser	2.2
Glu	1.4	Thr	2.4
Gly	4.6	Tyr	2.0
His	2.6	Trp	2.0
Ile	1.4	Val	2.1

Sources: Data (except proline) from Bryson, J.W., Betz, S.F., Lu, H.S., Suich, D.J., Zhou, H.X., O'Neil, K.T., & DeGrado, W.F. (1995) Protein design: a hierarchic approach. *Science* 270, 935. Proline data from Myers, J.K., Pace, C.N., & Scholtz, J.M. (1997) Helix propensities are identical in proteins and peptides. *Biochemistry* 36, 10,926.

*$\Delta\Delta G°$ is the difference in free-energy change, relative to that for alanine, required for the amino acid residue to take up the α-helical conformation. Larger numbers reflect greater difficulty taking up the α-helical structure. Data are a composite derived from multiple experiments and experimental systems.

acid side chains can stabilize or destabilize the α-helical structure. For example, if a polypeptide chain has a long block of Glu residues, this segment of the chain will not form an α helix at pH 7.0. The negatively charged carboxyl groups of adjacent Glu residues repel each other so strongly that they prevent formation of the α helix. For the same reason, if there are many adjacent Lys and/or Arg residues, with positively charged R groups at pH 7.0, they also repel each other and prevent formation of the α helix. The bulk and shape of Asn, Ser, Thr, and Cys residues can also destabilize an α helix if they are close together in the chain.

The twist of an α helix ensures that critical interactions occur between an amino acid side chain and the side chain three (and sometimes four) residues away on either side of it. This is clear when the α helix is depicted as a helical wheel (Fig. 4–4d). Positively charged amino acids are often found three residues away from negatively charged amino acids, permitting the formation of an ion pair. Two aromatic amino acid residues are often similarly spaced, resulting in a hydrophobic interaction.

A constraint on the formation of the α helix is the presence of Pro or Gly residues, which have the least proclivity to form α helices. In proline, the nitrogen atom is part of a rigid ring (see Fig. 4–8), and rotation about the N—C_α bond is not possible. Thus, a Pro residue introduces a destabilizing kink in an α helix. In addition, the nitrogen atom of a Pro residue in a peptide linkage has no substituent hydrogen to participate in hydrogen bonds with other residues. For these reasons, proline is only rarely found in an α helix. Glycine occurs infrequently in α helices for a different reason: it has more conformational flexibility than the other amino acid residues. Polymers of glycine tend to take up coiled structures quite different from an α helix.

A final factor affecting the stability of an α helix is the identity of the amino acid residues near the ends of the α-helical segment of the polypeptide. A small electric dipole exists in each peptide bond (Fig. 4–2a). These dipoles are aligned through the hydrogen bonds of the helix, resulting in a net dipole along the helical axis that increases with helix length **(Fig. 4–5)**. The partial positive and negative charges of the helix dipole reside on the peptide amino and carbonyl groups near the amino-terminal and carboxyl-terminal ends, respectively. For this reason, negatively charged amino acids are often found near the amino terminus of the helical segment, where they have a stabilizing interaction with the positive charge of the helix dipole; a positively charged amino acid at the amino-terminal end is destabilizing. The opposite is true at the carboxyl-terminal end of the helical segment.

In summary, five types of constraints affect the stability of an α helix: (1) the intrinsic propensity of an amino acid residue to form an α helix; (2) the interactions between R groups, particularly those spaced three (or four) residues apart; (3) the bulkiness of adjacent R groups; (4) the occurrence of Pro and Gly residues; and (5) interactions between amino acid residues at the ends of the helical segment and the electric dipole inherent to the α helix. The tendency of a given segment of a polypeptide chain to form an α helix therefore depends on the identity and sequence of amino acid residues within the segment.

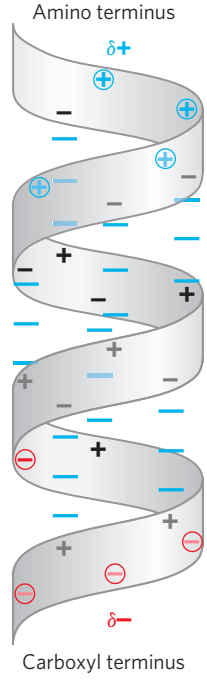


FIGURE 4–5 Helix dipole. The electric dipole of a peptide bond (see Fig. 4–2a) is transmitted along an α-helical segment through the intrachain hydrogen bonds, resulting in an overall helix dipole. In this illustration, the amino and carbonyl constituents of each peptide bond are indicated by + and − symbols, respectively. Non-hydrogen-bonded amino and carbonyl constituents of the peptide bonds near each end of the α-helical region are circled and shown in color.

The β Conformation Organizes Polypeptide Chains into Sheets

Protein Architecture–β Sheet In 1951, Pauling and Corey predicted a second type of repetitive structure, the **β conformation**. This is a more extended conformation of polypeptide chains, and its structure is again defined by backbone atoms arranged according to a characteristic set of dihedral angles (Table 4–1). In the β conformation, the backbone of the polypeptide chain is extended into a zigzag rather than helical structure **(Fig. 4–6)**. The arrangement of several segments side by side, all of which are in the β conformation, is called a **β sheet**. The zigzag structure of the individual polypeptide segments gives rise to a pleated appearance of the overall sheet. Hydrogen bonds form between adjacent segments of polypeptide chain within the sheet. The individual segments that form a β sheet are usually nearby on the polypeptide chain but can also be quite distant from each other in the linear sequence of the polypeptide; they may even be in different polypeptide chains. The R groups of adjacent amino acids protrude from the zigzag structure in opposite directions, creating the alternating pattern seen in the side view in Figure 4–6.

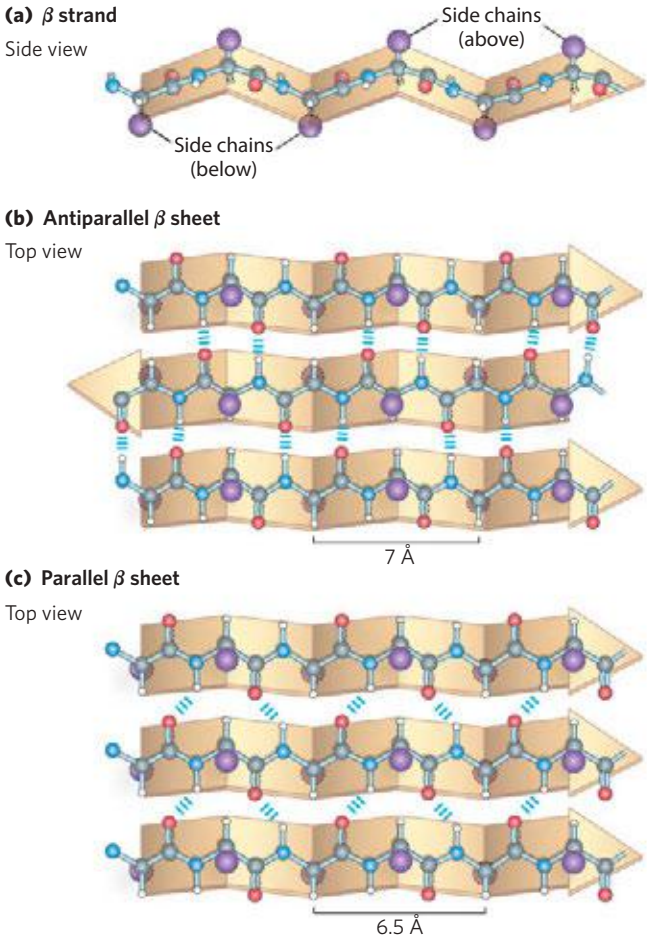


FIGURE 4–6 The β conformation of polypeptide chains. These **(a)** side and **(b, c)** top views reveal the R groups extending out from the β sheet and emphasize the pleated shape formed by the planes of the peptide bonds. (An alternative name for this structure is β-pleated sheet.) Hydrogen-bond cross-links between adjacent chains are also shown. The amino-terminal to carboxyl-terminal orientations of adjacent chains (arrows) can be the same or opposite, forming (b) an antiparallel β sheet or (c) a parallel β sheet.

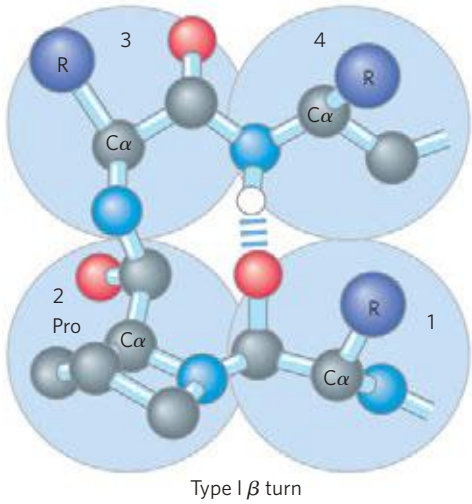
The adjacent polypeptide chains in a β sheet can be either parallel or antiparallel (having the same or opposite amino-to-carboxyl orientations, respectively). The structures are somewhat similar, although the repeat period is shorter for the parallel conformation (6.5 Å, vs. 7 Å for antiparallel) and the hydrogen-bonding patterns are different. The interstrand hydrogen bonds are essentially in-line (see Fig. 2–5) in the antiparallel β sheet, whereas they are distorted or not in-line for the parallel variant. The idealized structures exhibit the bond angles given in Table 4–1; these values vary somewhat in real proteins, resulting in structural variation, as seen above for α helices.

β Turns Are Common in Proteins

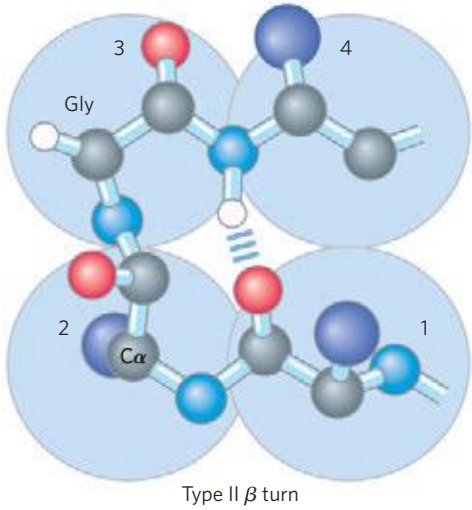
Protein Architecture–β Turn In globular proteins, which have a compact folded structure, some amino acid residues are in turns or loops where the polypeptide chain reverses direction **(Fig. 4–7)**. These are the connecting elements that link successive runs of α helix or β conformation. Particularly common are **β turns** that connect the ends of two adjacent segments of an antiparallel β sheet. The structure is a 180° turn involving four amino acid residues, with the carbonyl oxygen of the first residue forming a hydrogen bond with the amino-group hydrogen of the fourth. The peptide groups of the central two residues do not participate in any inter-residue hydrogen bonding. Several types of β turns have been described, each defined by the φ and ψ angles of the bonds that link the four amino acid residues that make up the particular turn (Table 4–1). Gly and Pro residues often occur in β turns, the former because it is small and flexible, the latter because peptide bonds involving the imino nitrogen of proline readily assume the cis configuration **(Fig. 4–8)**, a form that is particularly amenable to a tight turn. The two types of β turns shown in Figure 4–7 are the most common. Beta turns are often found near the surface of a protein, where the peptide groups of the central two amino acid residues in the turn can hydrogen-bond with water. Considerably less common is the γ turn, a three-residue turn with a hydrogen bond between the first and third residues.

Common Secondary Structures Have Characteristic Dihedral Angles

The α helix and the β conformation are the major repetitive secondary structures in a wide variety of proteins, although other repetitive structures exist in some specialized proteins (an example is collagen; see Fig. 4–13). Every type of secondary structure can be completely described by the dihedral angles φ and ψ associated with each residue. As shown by a Ramachandran plot, the dihedral angles that define the α helix and β conformation fall within a relatively restricted range of

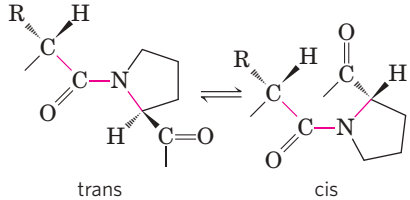


Type I β turn



Type II β turn

FIGURE 4–7 Structures of β turns. Type I and type II β turns are most common, distinguished by the φ and ψ angles taken up by the peptide backbone in the turn (see Table 4–1). Type I turns occur more than twice as frequently as type II. Type II β turns usually have Gly as the third residue. Note the hydrogen bond between the peptide groups of the first and fourth residues of the bends. (Individual amino acid residues are framed by large blue circles. Not all H atoms are shown in these depictions.)



Proline isomers

FIGURE 4–8 Trans and cis isomers of a peptide bond involving the imino nitrogen of proline. Of the peptide bonds between amino acid residues other than Pro, more than 99.95% are in the trans configuration. For peptide bonds involving the imino nitrogen of proline, however, about 6% are in the cis configuration; many of these occur at β turns.

sterically allowed structures **(Fig. 4–9a)**. Most values of φ and ψ taken from known protein structures fall into the expected regions, with high concentrations near the α helix and β conformation values as predicted (Fig. 4–9b). The only amino acid residue often found in a conformation outside these regions is glycine. Because its side chain is small, a Gly residue can take part in many conformations that are sterically forbidden for other amino acids.

Common Secondary Structures Can Be Assessed by Circular Dichroism

Any form of structural asymmetry in a molecule gives rise to differences in absorption of left-handed versus right-handed circularly polarized light. Measurement of this difference is called **circular dichroism (CD)**

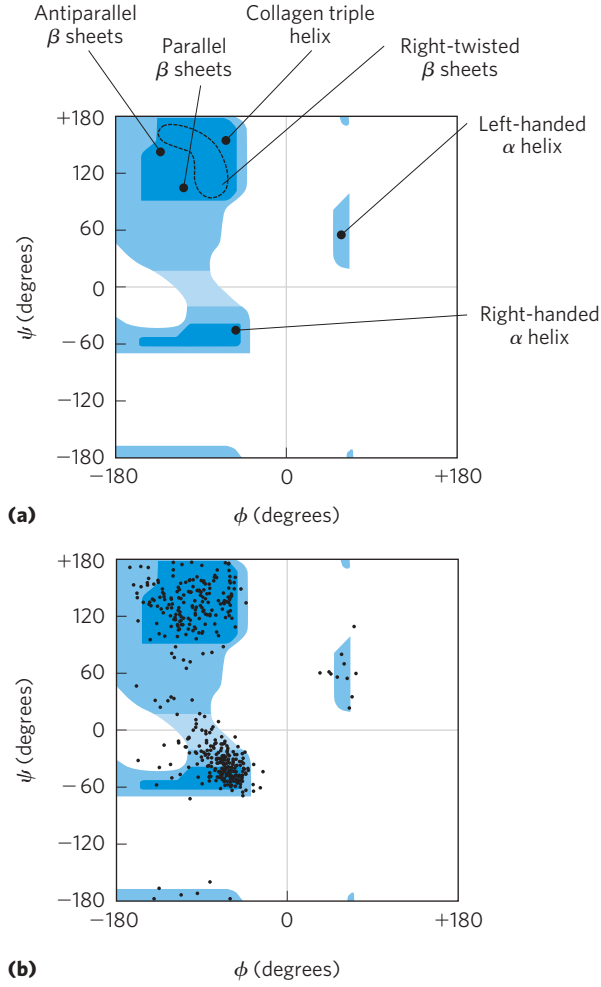


FIGURE 4–9 Ramachandran plots showing a variety of structures. **(a)** The values of φ and ψ for various allowed secondary structures are overlaid on the plot from Figure 4–3. Although left-handed α helices extending over several amino acid residues are theoretically possible, they have not been observed in proteins. **(b)** The values of φ and ψ for all the amino acid residues except Gly in the enzyme pyruvate kinase (isolated from rabbit) are overlaid on the plot of theoretically allowed conformations (Fig. 4–3). The small, flexible Gly residues were excluded because they frequently fall outside the expected (blue) ranges.

spectroscopy. An ordered structure, such as a folded protein, gives rise to an absorption spectrum that can have peaks or regions with both positive and negative values. For proteins, spectra are obtained in the far UV region (190 to 250 nm). The light-absorbing entity, or chromophore, in this region is the peptide bond; a signal is obtained when the peptide bond is in a folded environment. The difference in molar extinction coefficients (see Box 3–1) for left- and right-handed, circularly polarized light ($\Delta\varepsilon$) is plotted as a function of wavelength. The α helix and β conformations have characteristic CD spectra **(Fig. 4–10)**. Using CD spectra, biochemists can determine whether proteins are properly folded, estimate the fraction of the protein that is folded in either of the common secondary structures, and monitor transitions between the folded and unfolded states.

SUMMARY 4.2 Protein Secondary Structure

- Secondary structure is the local spatial arrangement of the main-chain atoms in a selected segment of a polypeptide chain.
- The most common regular secondary structures are the α helix, the β conformation, and β turns.
- The secondary structure of a polypeptide segment can be completely defined if the ϕ and ψ angles are known for all amino acid residues in that segment.
- Circular dichroism spectroscopy is a method for assessing common secondary structure and monitoring folding in proteins.

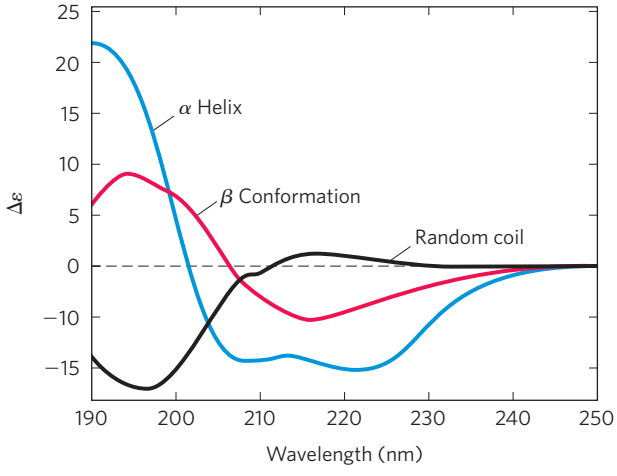


FIGURE 4–10 Circular dichroism spectroscopy. These spectra show polylysine entirely as α helix, as β conformation, or as a denatured, random coil. The *y* axis unit is a simplified version of the units most commonly used in CD experiments. Since the curves are different for α helix, β conformation, and random coil, the CD spectrum for a given protein can provide a rough estimate for the fraction of the protein made up of the two most common secondary structures. The CD spectrum of the native protein can serve as a benchmark for the folded state, useful for monitoring denaturation or conformational changes brought about by changes in solution conditions.

4.3 Protein Tertiary and Quaternary Structures

Protein Architecture—Introduction to Tertiary Structure The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein's **tertiary structure**. Whereas the term "secondary structure" refers to the spatial arrangement of amino acid residues that are adjacent in a segment of a polypeptide, tertiary structure includes *longer-range* aspects of amino acid sequence. Amino acids that are far apart in the polypeptide sequence and are in different types of secondary structure may interact within the completely folded structure of a protein. The location of bends (including β turns) in the polypeptide chain and the direction and angle of these bends are determined by the number and location of specific bend-producing residues, such as Pro, Thr, Ser, and Gly. Interacting segments of polypeptide chains are held in their characteristic tertiary positions by several kinds of weak interactions (and sometimes by covalent bonds such as disulfide cross-links) between the segments.

Some proteins contain two or more separate polypeptide chains, or subunits, which may be identical or different. The arrangement of these protein subunits in three-dimensional complexes constitutes **quaternary structure**.

In considering these higher levels of structure, it is useful to designate two major groups into which many proteins can be classified: **fibrous proteins**, with polypeptide chains arranged in long strands or sheets, and **globular proteins**, with polypeptide chains folded into a spherical or globular shape. The two groups are structurally distinct. Fibrous proteins usually consist largely of a single type of secondary structure, and their tertiary structure is relatively simple. Globular proteins often contain several types of secondary structure. The two groups also differ functionally: the structures that provide support, shape, and external protection to vertebrates are made of fibrous proteins, whereas most enzymes and regulatory proteins are globular proteins.

Fibrous Proteins Are Adapted for a Structural Function

Protein Architecture—Tertiary Structure of Fibrous Proteins α-Keratin, collagen, and silk fibroin nicely illustrate the relationship between protein structure and biological function (Table 4–3). Fibrous proteins share properties that give strength and/or flexibility to the structures in which they occur. In each case, the fundamental structural unit is a simple repeating element of secondary structure. All fibrous proteins are insoluble in water, a property conferred by a high concentration of hydrophobic amino acid residues both in the interior of the protein and on its surface. These hydrophobic surfaces are largely buried as many similar polypeptide chains are packed together to form elaborate supramolecular complexes. The underlying structural simplicity of fibrous

proteins makes them particularly useful for illustrating some of the fundamental principles of protein structure discussed above.

α-Keratin The α-keratins have evolved for strength. Found only in mammals, these proteins constitute almost the entire dry weight of hair, wool, nails, claws, quills, horns, hooves, and much of the outer layer of skin. The α-keratins are part of a broader family of proteins called intermediate filament (IF) proteins. Other IF proteins are found in the cytoskeletons of animal cells. All IF proteins have a structural function and share the structural features exemplified by the α-keratins.

The α-keratin helix is a right-handed α helix, the same helix found in many other proteins. Francis Crick and Linus Pauling in the early 1950s independently suggested that the α helices of keratin were arranged as a coiled coil. Two strands of α-keratin, oriented in parallel (with their amino termini at the same end), are wrapped about each other to form a supertwisted coiled coil. The supertwisting amplifies the strength of the overall structure, just as strands are twisted to make a strong rope **(Fig. 4–11)**. The twisting of the axis of an α helix to form a coiled coil explains the discrepancy between the 5.4 Å per turn predicted for an α helix by Pauling and Corey and the 5.15 to 5.2 Å repeating structure observed in the x-ray diffraction of hair (p. 120). The helical path of the supertwists is left-handed, opposite in sense to the α helix. The surfaces where the two α helices touch are made up of hydrophobic amino acid residues, their R groups meshed together in a regular interlocking pattern. This permits a close packing of the polypeptide chains within the left-handed supertwist. Not surprisingly, α-keratin is rich in the hydrophobic residues Ala, Val, Leu, Ile, Met, and Phe.

An individual polypeptide in the α-keratin coiled coil has a relatively simple tertiary structure, dominated by an α-helical secondary structure with its helical axis twisted in a left-handed superhelix. The intertwining of the two α-helical polypeptides is an example of quaternary structure. Coiled coils of this type are common structural elements in filamentous proteins and in the muscle protein myosin (see Fig. 5–27). The quaternary structure of α-keratin can be quite complex. Many coiled coils can be assembled into large supramolecular complexes, such as the arrangement of α-keratin to form the intermediate filament of hair (Fig. 4–11b).

The strength of fibrous proteins is enhanced by covalent cross-links between polypeptide chains in the multihelical "ropes" and between adjacent chains in a supramolecular assembly. In α-keratins, the cross-links stabilizing quaternary structure are disulfide bonds (Box 4–2). In the hardest and toughest α-keratins, such

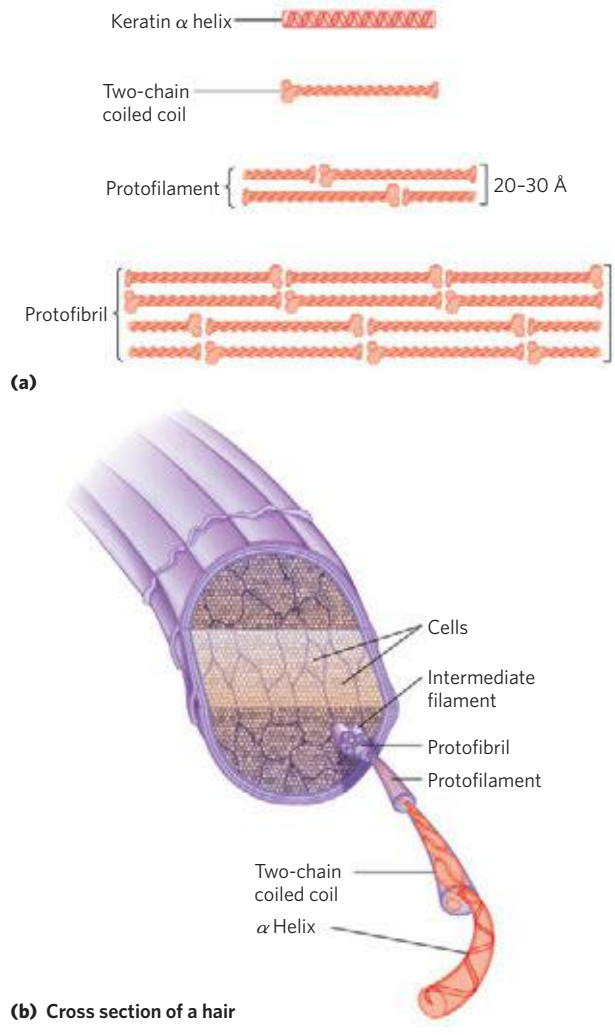


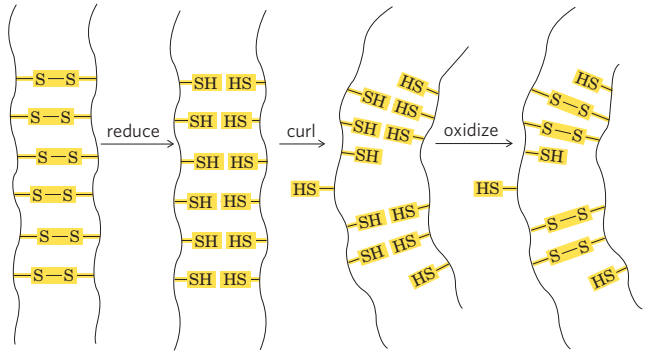
FIGURE 4–11 Structure of hair. (a) Hair α-keratin is an elongated α helix with somewhat thicker elements near the amino and carboxyl termini. Pairs of these helices are interwound in a left-handed sense to form two-chain coiled coils. These then combine in higher-order structures called protofilaments and protofibrils. About four protofibrils—32 strands of α-keratin in all—combine to form an intermediate filament. The individual two-chain coiled coils in the various substructures also seem to be interwound, but the handedness of the interwinding and other structural details are unknown. **(b)** A hair is an array of many α-keratin filaments, made up of the substructures shown in (a).

TABLE 4–3 Secondary Structures and Properties of Some Fibrous Proteins

Structure	Characteristics	Examples of occurrence
α Helix, cross-linked by disulfide bonds	Tough, insoluble protective structures of varying hardness and flexibility	α-Keratin of hair, feathers, nails
β Conformation	Soft, flexible filaments	Silk fibroin
Collagen triple helix	High tensile strength, without stretch	Collagen of tendons, bone matrix

BOX 4–2 Permanent Waving Is Biochemical Engineering

When hair is exposed to moist heat, it can be stretched. At the molecular level, the α helices in the α-keratin of hair are stretched out until they arrive at the fully extended β conformation. On cooling they spontaneously revert to the α-helical conformation. The characteristic "stretchability" of α-keratins, and their numerous disulfide cross-linkages, are the basis of permanent waving. The hair to be waved or curled is first bent around a form of appropriate shape. A solution of a reducing agent, usually a compound containing a thiol or sulfhydryl group (—SH), is then applied with heat. The reducing agent cleaves the cross-linkages by reducing each disulfide bond to form two Cys residues. The moist heat breaks hydrogen bonds and causes the α-helical structure of the polypeptide chains to uncoil. After a time the reducing solution is removed, and an oxidizing agent is added to establish *new* disulfide bonds between pairs of Cys residues of adjacent polypeptide chains, but not the same pairs as before the treatment. After the hair is washed and cooled, the polypeptide chains revert to their α-helical conformation. The hair fibers now curl in the desired fashion because the new disulfide cross-linkages exert some torsion or twist on the bundles of α-helical coils in the hair fibers. The same process can be used to straighten hair that is naturally curly. A permanent wave (or hair straightening) is not truly permanent, because the hair grows; in the new hair replacing the old, the α-keratin has the natural pattern of disulfide bonds.



as those of rhinoceros horn, up to 18% of the residues are cysteines involved in disulfide bonds.

Collagen Like the α-keratins, collagen has evolved to provide strength. It is found in connective tissue such as tendons, cartilage, the organic matrix of bone, and the cornea of the eye. The collagen helix is a unique secondary structure, quite distinct from the α helix. It is left-handed and has three amino acid residues per turn (**Fig. 4–12** and Table 4–1). Collagen is also a coiled coil, but one with distinct tertiary and quaternary structures: three separate polypeptides, called α chains (not to be confused with α helices), are supertwisted about each other (Fig. 4–12c). The superhelical twisting is right-handed in collagen, opposite in sense to the left-handed helix of the α chains.

There are many types of vertebrate collagen. Typically they contain about 35% Gly, 11% Ala, and 21% Pro and 4-Hyp (4-hydroxyproline, an uncommon amino acid; see Fig. 3–8a). The food product gelatin is derived from collagen. It has little nutritional value as a protein, because collagen is extremely low in many amino acids that are essential in the human diet. The unusual amino acid content of collagen is related to structural constraints unique to the collagen helix. The amino acid sequence in collagen is generally a repeating tripeptide unit, Gly–X–Y, where X is often Pro, and Y is often 4-Hyp. Only Gly residues can be accommodated at the very tight junctions between the individual α chains (Fig. 4–12d). The Pro and 4-Hyp residues permit the sharp twisting of the collagen helix. The amino acid sequence and the supertwisted quaternary structure of collagen allow a very close packing of its three polypeptides. 4-Hydroxyproline has a special role in the structure of collagen—and in human history (Box 4–3).

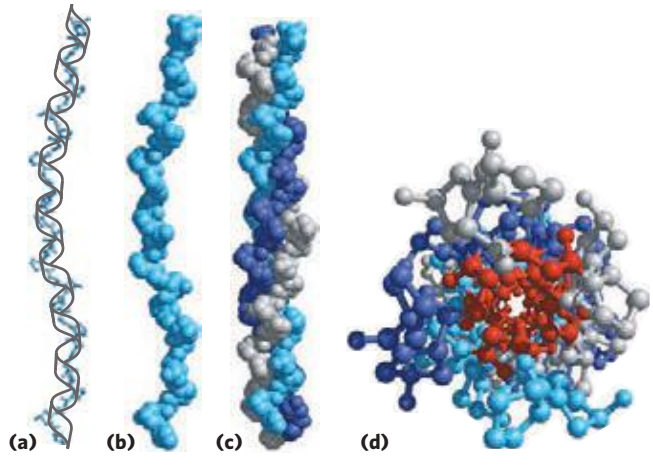


FIGURE 4–12 Structure of collagen. (Derived from PDB ID 1CGD) **(a)** The α chain of collagen has a repeating secondary structure unique to this protein. The repeating tripeptide sequence Gly–X–Pro or Gly–X–4-Hyp adopts a left-handed helical structure with three residues per turn. The repeating sequence used to generate this model is Gly–Pro–4-Hyp. **(b)** Space-filling model of the same α chain. **(c)** Three of these helices (shown here in gray, blue, and purple) wrap around one another with a right-handed twist. **(d)** The three-stranded collagen superhelix shown from one end, in a ball-and-stick representation. Gly residues are shown in red. Glycine, because of its small size, is required at the tight junction where the three chains are in contact. The balls in this illustration do not represent the van der Waals radii of the individual atoms. The center of the three-stranded superhelix is not hollow, as it appears here, but very tightly packed.

BOX 4–3 MEDICINE Why Sailors, Explorers, and College Students Should Eat Their Fresh Fruits and Vegetables

. . . from this misfortune, together with the unhealthiness of the country, where there never falls a drop of rain, we were stricken with the "camp-sickness," which was such that the flesh of our limbs all shrivelled up, and the skin of our legs became all blotched with black, mouldy patches, like an old jack-boot, and proud flesh came upon the gums of those of us who had the sickness, and none escaped from this sickness save through the jaws of death. The signal was this: when the nose began to bleed, then death was at hand . . .

—The Memoirs of the Lord of Joinville, *ca. 1300*

This excerpt describes the plight of Louis IX's army toward the end of the Seventh Crusade (1248–1254), when the scurvy-weakened Crusader army was destroyed by the Egyptians. What was the nature of the malady afflicting these thirteenth-century soldiers?

Scurvy is caused by lack of vitamin C, or ascorbic acid (ascorbate). Vitamin C is required for, among other things, the hydroxylation of proline and lysine in collagen; scurvy is a deficiency disease characterized by general degeneration of connective tissue. Manifestations of advanced scurvy include numerous small hemorrhages caused by fragile blood vessels, tooth loss, poor wound healing and the reopening of old wounds, bone pain and degeneration, and eventually heart failure. Milder cases of vitamin C deficiency are accompanied by fatigue, irritability, and an increased severity of respiratory tract infections. Most animals make large amounts of vitamin C, converting glucose to ascorbate in four enzymatic steps. But in the course of evolution, humans and some other animals—gorillas, guinea pigs, and fruit bats—have lost the last enzyme in this pathway and must obtain ascorbate in their diet. Vitamin C is available in a wide range of fruits and vegetables. Until 1800, however, it was often absent in the dried foods and other food supplies stored for winter or for extended travel.

Scurvy was recorded by the Egyptians in 1500 BCE, and it is described in the fifth century BCE writings of Hippocrates. Yet it did not come to wide public notice until the European voyages of discovery from 1500 to 1800. The first circumnavigation of the globe (1519–1522), led by Ferdinand Magellan, was accomplished only with the loss of more than 80% of his crew to scurvy. During Jacques Cartier's second voyage to explore the St. Lawrence River (1535–1536), his band was threatened with complete disaster until the native Americans taught the men to make a cedar tea that cured and prevented scurvy (it contained vitamin C). Winter outbreaks of scurvy in Europe were gradually eliminated in the nineteenth century as the cultivation of the potato, introduced from South America, became widespread.

In 1747, James Lind, a Scottish surgeon in the Royal Navy, carried out the first controlled clinical study in recorded history. During an extended voyage on the 50-gun warship HMS *Salisbury*, Lind selected 12 sailors suffering from scurvy and separated them into groups of two. All 12 received the same diet, except that each group was given a different remedy for scurvy from among those recommended at the time. The sailors given lemons and oranges recovered and returned to duty. The sailors given boiled apple juice improved slightly. The remainder continued to deteriorate. Lind's *Treatise on the Scurvy* was published in 1753, but inaction persisted in the Royal Navy for another 40 years. In 1795 the British admiralty finally mandated a ration of concentrated lime or lemon juice for all British sailors (hence the name "limeys"). Scurvy continued to be a problem in some other parts of the world until 1932, when Hungarian scientist Albert Szent-Györgyi, and W. A. Waugh and C. G. King at the University of Pittsburgh, isolated and synthesized ascorbic acid.

James Lind, 1716–1794; naval surgeon, 1739–1748

L-Ascorbic acid (vitamin C) is a white, odorless, crystalline powder. It is freely soluble in water and relatively insoluble in organic solvents. In a dry state, away from light, it is stable for a considerable length of time. The appropriate daily intake of this vitamin is still in dispute. The recommended value in the United States is 90 mg (Australia and the United Kingdom recommend 60 mg; Russia recommends 125 mg). Along with citrus fruits and almost all other fresh fruits, good sources of vitamin C include peppers, tomatoes, potatoes, and broccoli. The vitamin C of fruits and vegetables is destroyed by overcooking or prolonged storage.

So why is ascorbate so necessary to good health? Of particular interest to us here is its role in the formation of collagen. As noted in the text, collagen is constructed of the repeating tripeptide unit Gly–X–Y, where X and Y are generally Pro or 4-Hyp—the proline derivative (4*R*)-L-hydroxyproline, which plays an essential role in the folding of collagen and in maintaining its structure. The proline ring is normally found as a mixture of two puckered conformations, called C_γ-endo and C_γ-exo (Fig. 1). The collagen helix structure requires the Pro residue in the Y positions to

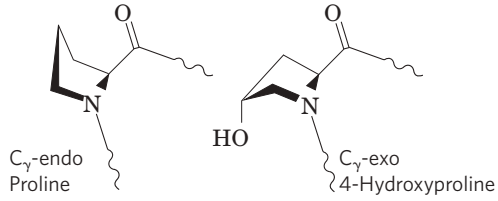


FIGURE 1 The C_γ-endo conformation of proline and the C_γ-exo conformation of 4-hydroxyproline.

be in the C_γ-exo conformation, and it is this conformation that is enforced by the hydroxyl substitution at C-4 in 4-Hyp. The collagen structure also requires that the Pro residue in the X positions have the C_γ-endo conformation, and introduction of 4-Hyp here can destabilize the helix. In the absence of vitamin C, cells cannot hydroxylate the Pro at the Y positions. This leads to collagen instability and the connective tissue problems seen in scurvy.

The hydroxylation of specific Pro residues in procollagen, the precursor of collagen, requires the action of the enzyme prolyl 4-hydroxylase. This enzyme (M_r 240,000) is an $\alpha_2\beta_2$ tetramer in all vertebrates. The proline-hydroxylating activity is found in the α subunits. Each α subunit contains one atom of nonheme iron (Fe^{2+}), and the enzyme is one of a class of hydroxylases that require α-ketoglutarate in their reactions.

In the normal prolyl 4-hydroxylase reaction (Fig. 2a), one molecule of α-ketoglutarate and one of O_2 bind to the enzyme. The α-ketoglutarate is oxidatively decarboxylated to form CO_2 and succinate. The remaining oxygen atom is then used to hydroxylate an appropriate Pro residue in procollagen. No ascorbate is needed in this reaction. However, prolyl 4-hydroxylase also catalyzes an oxidative decarboxylation of α-ketoglutarate that is not coupled to proline hydroxylation (Fig. 2b). During this reaction the heme Fe^{2+} becomes oxidized, inactivating the enzyme and preventing the proline hydroxylation. The ascorbate consumed in the reaction is needed to restore enzyme activity—by reducing the heme iron.

Scurvy remains a problem today, not only in remote regions where nutritious food is scarce but, surprisingly, on U.S. college campuses. The only vegetables consumed by some students are those in tossed salads, and days go by without these young adults consuming fruit. A 1998 study of 230 students at Arizona State University revealed that 10% had serious vitamin C deficiencies, and 2 students had vitamin C levels so low that they probably had scurvy. Only half the students in the study consumed the recommended daily allowance of vitamin C.

Eat your fresh fruits and vegetables.

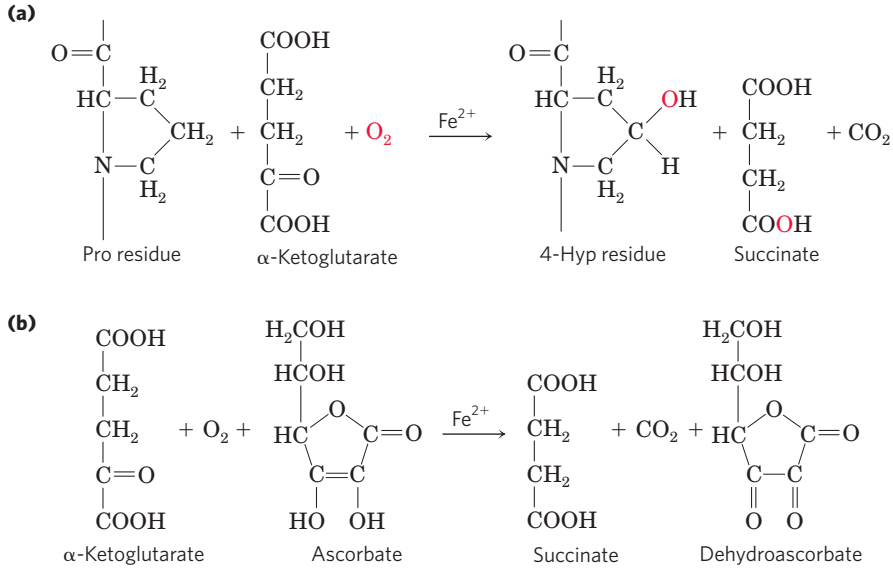
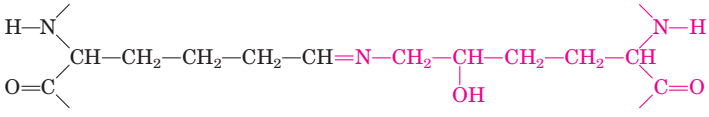


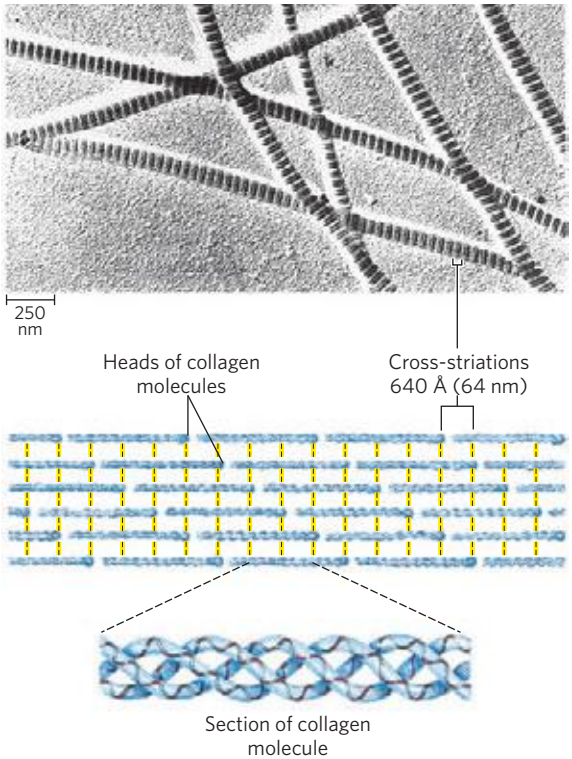
FIGURE 2 Reactions catalyzed by prolyl 4-hydroxylase. **(a)** The normal reaction, coupled to proline hydroxylation, which does not require ascorbate. The fate of the two oxygen atoms from O_2 is shown in red. **(b)** The uncoupled reaction, in which α-ketoglutarate is oxidatively decarboxylated without hydroxylation of proline. Ascorbate is consumed stoichiometrically in this process as it is converted to dehydroascorbate, preventing Fe^{2+} oxidation.

The tight wrapping of the α chains in the collagen triple helix provides tensile strength greater than that of a steel wire of equal cross section. Collagen fibrils **(Fig. 4–13)** are supramolecular assemblies consisting of triple-helical collagen molecules (sometimes referred to as tropocollagen molecules) associated in a variety of ways to provide different degrees of tensile strength. The α chains of collagen molecules and the collagen molecules of fibrils are cross-linked by unusual types of covalent bonds involving Lys, HyLys (5-hydroxylysine; see Fig. 3–8a), or His residues that are present at a few of the X and Y positions. These links create uncommon amino acid residues such as dehydrohydroxylysinonorleucine. The increasingly rigid and brittle character of aging connective tissue results from accumulated covalent cross-links in collagen fibrils.



Dehydrohydroxylysinonorleucine

A typical mammal has more than 30 structural variants of collagen, particular to certain tissues and each somewhat different in sequence and function. Some human genetic defects in collagen structure illustrate the close relationship between amino acid sequence and three-dimensional structure in this protein.



Osteogenesis imperfecta is characterized by abnormal bone formation in babies; at least eight variants of this condition, with different degrees of severity, occur in the human population. Ehlers-Danlos syndrome is characterized by loose joints, and at least six variants occur in humans. The composer Niccolò Paganini (1782–1840) was famed for his seemingly impossible dexterity in playing the violin. He suffered from a variant of Ehlers-Danlos syndrome that rendered him effectively double-jointed. In both disorders, some variants can be lethal, whereas others cause lifelong problems.

All of the variants of both conditions result from the substitution of an amino acid residue with a larger R group (such as Cys or Ser) for a single Gly residue in an α chain in one or another collagen protein (a different Gly residue in each disorder). These single-residue substitutions have a catastrophic effect on collagen function because they disrupt the Gly–X–Y repeat that gives collagen its unique helical structure. Given its role in the collagen triple helix (Fig. 4–12d), Gly cannot be replaced by another amino acid residue without substantial deleterious effects on collagen structure. ■

Silk Fibroin Fibroin, the protein of silk, is produced by insects and spiders. Its polypeptide chains are predominantly in the β conformation. Fibroin is rich in Ala and Gly residues, permitting a close packing of β sheets and an interlocking arrangement of R groups **(Fig. 4–14)**. The overall structure is stabilized by extensive hydrogen bonding between all peptide linkages in the polypeptides of each β sheet and by the optimization of van der Waals interactions between sheets. Silk does not stretch, because the β conformation is already highly extended (Fig. 4–6). However, the structure is flexible, because the sheets are held together by numerous weak interactions rather than by covalent bonds such as the disulfide bonds in α-keratins.

Structural Diversity Reflects Functional Diversity in Globular Proteins

In a globular protein, different segments of the polypeptide chain (or multiple polypeptide chains) fold back on each other, generating a more compact shape than is seen in the fibrous proteins **(Fig. 4–15)**. The folding also provides the structural diversity necessary for proteins to carry out a wide array of biological functions.

FIGURE 4–13 Structure of collagen fibrils. Collagen (M_r 300,000) is a rod-shaped molecule, about 3,000 Å long and only 15 Å thick. Its three helically intertwined α chains may have different sequences; each chain has about 1,000 amino acid residues. Collagen fibrils are made up of collagen molecules aligned in a staggered fashion and cross-linked for strength. The specific alignment and degree of cross-linking vary with the tissue and produce characteristic cross-striations in an electron micrograph. In the example shown here, alignment of the head groups of every fourth molecule produces striations 640 Å (64 nm) apart.

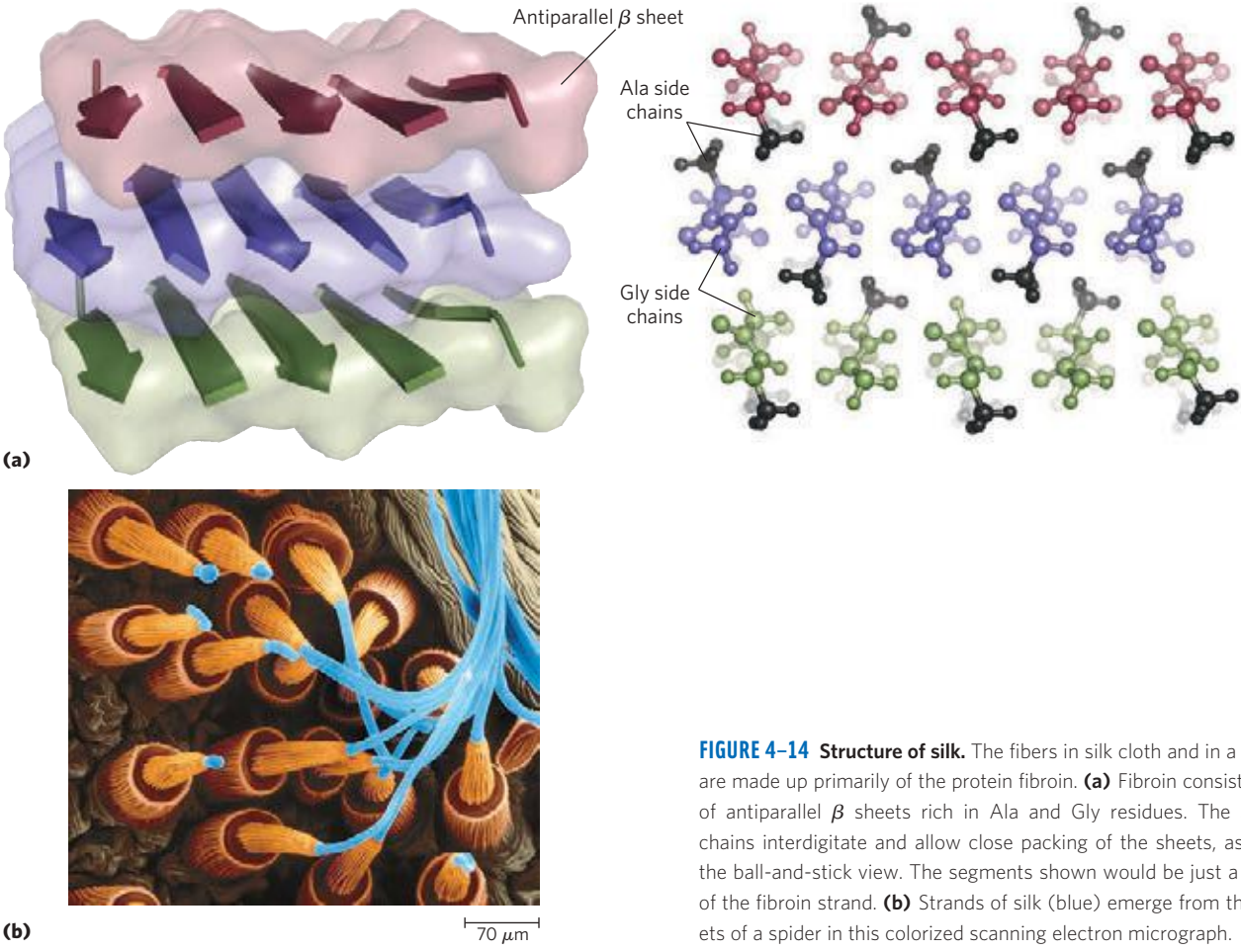


FIGURE 4–14 Structure of silk. The fibers in silk cloth and in a spiderweb are made up primarily of the protein fibroin. **(a)** Fibroin consists of layers of antiparallel β sheets rich in Ala and Gly residues. The small side chains interdigitate and allow close packing of the sheets, as shown in the ball-and-stick view. The segments shown would be just a small part of the fibroin strand. **(b)** Strands of silk (blue) emerge from the spinnerets of a spider in this colorized scanning electron micrograph.

Globular proteins include enzymes, transport proteins, motor proteins, regulatory proteins, immunoglobulins, and proteins with many other functions.

Our discussion of globular proteins begins with the principles gleaned from the first protein structures to be elucidated. This is followed by a detailed description of protein substructure and comparative categorization. Such discussions are possible only because of the vast amount of information available over the Internet from publicly accessible databases, particularly the Protein Data Bank (Box 4–4).

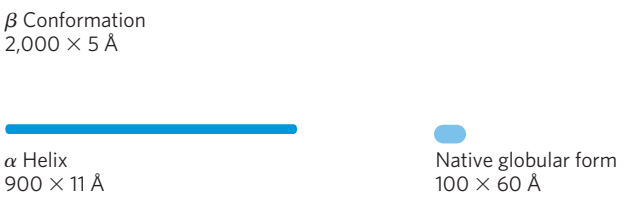


FIGURE 4–15 Globular protein structures are compact and varied. Human serum albumin (M_r 64,500) has 585 residues in a single chain. Given here are the approximate dimensions its single polypeptide chain would have if it occurred entirely in extended β conformation or as an α helix. Also shown is the size of the protein in its native globular form, as determined by x-ray crystallography; the polypeptide chain must be very compactly folded to fit into these dimensions.

Myoglobin Provided Early Clues about the Complexity of Globular Protein Structure

Protein Architecture—Tertiary Structure of Small Globular Proteins, II. Myoglobin The first breakthrough in understanding the three-dimensional structure of a globular protein came from x-ray diffraction studies of myoglobin carried out by John Kendrew and his colleagues in the 1950s. Myoglobin is a relatively small (M_r 16,700), oxygen-binding protein of muscle cells. It functions both to store oxygen and to facilitate oxygen diffusion in rapidly contracting muscle tissue. Myoglobin contains a single polypeptide chain of 153 amino acid residues of known sequence and a single iron protoporphyrin, or heme, group. The same heme group that is found in myoglobin is found in hemoglobin, the oxygen-binding protein of erythrocytes, and is responsible for the deep red-brown color of both myoglobin and hemoglobin. Myoglobin is particularly abundant in the muscles of diving mammals such as the whale, seal, and porpoise—so abundant that the muscles of these animals are brown. Storage and distribution of oxygen by muscle myoglobin permits diving mammals to remain submerged for long periods. The activities of

BOX 4–4 The Protein Data Bank

The number of known three-dimensional protein structures is now in the tens of thousands and more than doubles every couple of years. This wealth of information is revolutionizing our understanding of protein structure, the relation of structure to function, and the evolutionary paths by which proteins arrived at their present state, which can be seen in the family resemblances that come to light as protein databases are sifted and sorted. One of the most important resources available to biochemists is the **Protein Data Bank** (**PDB**; www.pdb.org).

The PDB is an archive of experimentally determined three-dimensional structures of biological macromolecules, containing virtually all of the macromolecular structures (proteins, RNAs, DNAs, etc.) elucidated to date. Each structure is assigned an identifying label (a four-character identifier called the PDB ID). Such labels are provided in the figure legends for every PDB-derived structure illustrated in this text so that students and instructors can explore the same structures on their own. The data files in the PDB describe the spatial coordinates of each atom whose position has been determined (many of the cataloged structures are not complete). Additional data files provide information on how the structure was determined and its accuracy. The atomic coordinates can be converted into an image of the macromolecule by using structure visualization software. Students are encouraged to access the PDB and explore structures, using visualization software linked to the database. Macromolecular structure files can also be downloaded and explored on the desktop, using free software such as Jmol.

myoglobin and other globin molecules are investigated in greater detail in Chapter 5.

Figure 4–16 shows several structural representations of myoglobin, illustrating how the polypeptide chain is folded in three dimensions—its tertiary structure. The red group surrounded by protein is heme. The backbone of the myoglobin molecule consists of eight relatively straight segments of α helix interrupted by bends, some of which are β turns. The longest α helix has 23 amino acid residues and the shortest only 7; all helices are right-handed. More than 70% of the residues in myoglobin are in these α-helical regions. X-ray analysis has revealed the precise position of each of the R groups, which fill up nearly all the space within the folded chain that is not occupied by backbone atoms.

Many important conclusions were drawn from the structure of myoglobin. The positioning of amino acid side chains reflects a structure that derives much of its stability from hydrophobic interactions. Most of the hydrophobic R groups are in the interior of the molecule, hidden from exposure to water. All but two of the polar R groups are located on the outer surface of the molecule, and all are hydrated. The myoglobin molecule is so compact that its interior has room for only four molecules of water. This dense hydrophobic core is typical of globular proteins. The fraction of space occupied by atoms in an organic liquid is 0.4 to 0.6. In a globular protein the fraction is about 0.75, comparable to that in a crystal (in a typical crystal the fraction is 0.70 to 0.78, near the theoretical maximum). In this packed environment, weak interactions strengthen and reinforce each other. For example, the nonpolar side chains in the core are so close together that short-range van der Waals interactions make a significant contribution to stabilizing hydrophobic interactions.

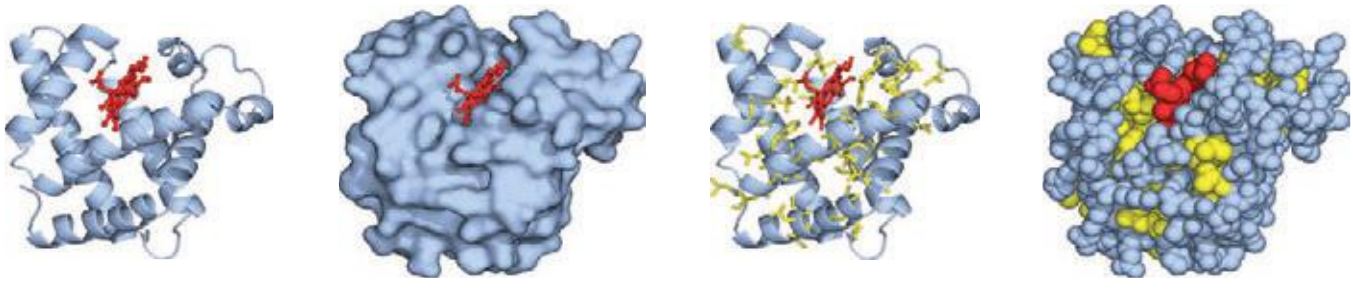


FIGURE 4–16 Tertiary structure of sperm whale myoglobin. (PDB ID 1MBO) Orientation of the protein is similar in (a) through (d); the heme group is shown in red. In addition to illustrating the myoglobin structure, this figure provides examples of several different ways to display protein structure. **(a)** The polypeptide backbone in a ribbon representation of a type introduced by Jane Richardson, which highlights regions of secondary structure. The α-helical regions are evident. **(b)** Surface contour image; this is useful for visualizing pockets in the protein where other molecules might bind. **(c)** Ribbon representation including side chains (yellow) for the hydrophobic residues Leu, Ile, Val, and Phe. **(d)** Space-filling model with all amino acid side chains. Each atom is represented by a sphere encompassing its van der Waals radius. The hydrophobic residues are again shown in yellow; most are buried in the interior of the protein and thus not visible.

Deduction of the structure of myoglobin confirmed some expectations and introduced some new elements of secondary structure. As predicted by Pauling and Corey, all the peptide bonds are in the planar trans configuration. The α helices in myoglobin provided the first direct experimental evidence for the existence of this type of secondary structure. Three of the four Pro residues are found at bends. The fourth Pro residue occurs within an α helix, where it creates a kink necessary for tight helix packing.

The flat heme group rests in a crevice, or pocket, in the myoglobin molecule. The iron atom in the center of the heme group has two bonding (coordination) positions perpendicular to the plane of the heme **(Fig. 4–17)**. One of these is bound to the R group of the His residue at position 93; the other is the site at which an O_2 molecule binds. Within this pocket, the accessibility of the heme group to solvent is highly restricted. This is important for function, because free heme groups in an oxygenated solution are rapidly oxidized from the ferrous (Fe^{2+}) form, which is active in the reversible binding of O_2, to the ferric (Fe^{3+}) form, which does not bind O_2.

As many different myoglobin structures were resolved, investigators were able to observe the structural changes that accompany the binding of oxygen or other molecules and thus, for the first time, to understand the correlation between protein structure and function. Hundreds of proteins have now been subjected to similar analysis. Today, nuclear magnetic resonance (NMR) spectroscopy and other techniques supplement x-ray diffraction data, providing more information on a protein's structure (Box 4–5). In addition, the sequencing of the genomic DNA of many organisms (Chapter 9) has identified thousands of genes that encode proteins of known sequence but, as yet, unknown function; this work continues apace.

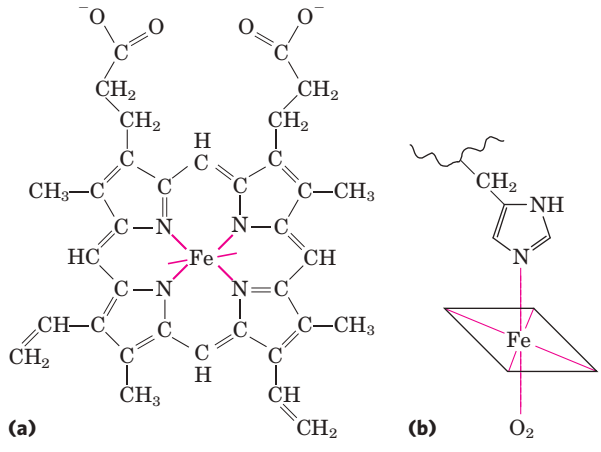


FIGURE 4–17 The heme group. This group is present in myoglobin, hemoglobin, cytochromes, and many other proteins (the heme proteins). **(a)** Heme consists of a complex organic ring structure, protoporphyrin, which binds an iron atom in its ferrous (Fe^{2+}) state. The iron atom has six coordination bonds, four in the plane of, and bonded to, the flat porphyrin molecule and two perpendicular to it. **(b)** In myoglobin and hemoglobin, one of the perpendicular coordination bonds is bound to a nitrogen atom of a His residue. The other is "open" and serves as the binding site for an O_2 molecule.

Globular Proteins Have a Variety of Tertiary Structures

From what we now know about the tertiary structures of hundreds of globular proteins, it is clear that myoglobin illustrates just one of many ways in which a polypeptide chain can fold. Table 4–4 shows the proportions of α helix and β conformations (expressed as percentage of residues in each type) in several small, single-chain, globular proteins. Each of these proteins has a distinct structure, adapted for its particular biological function, but together they share several important properties with myoglobin. Each is folded compactly, and in each case the hydrophobic amino acid side chains are oriented toward the interior (away from water) and the hydrophilic side chains are on the surface. The structures are also stabilized by a multitude of hydrogen bonds and some ionic interactions.

For the beginning student, the very complex tertiary structures of globular proteins—some much larger than myoglobin—are best approached by focusing on common structural patterns, recurring in different and often unrelated proteins. The three-dimensional structure of a typical globular protein can be considered an assemblage of polypeptide segments in the α-helical and β conformations, linked by connecting segments. The structure can then be defined by how these segments stack on one another and how the segments that connect them are arranged.

To understand a complete three-dimensional structure, we need to analyze its folding patterns. We begin by defining two important terms that describe protein structural patterns or elements in a polypeptide chain and then turn to the folding rules.

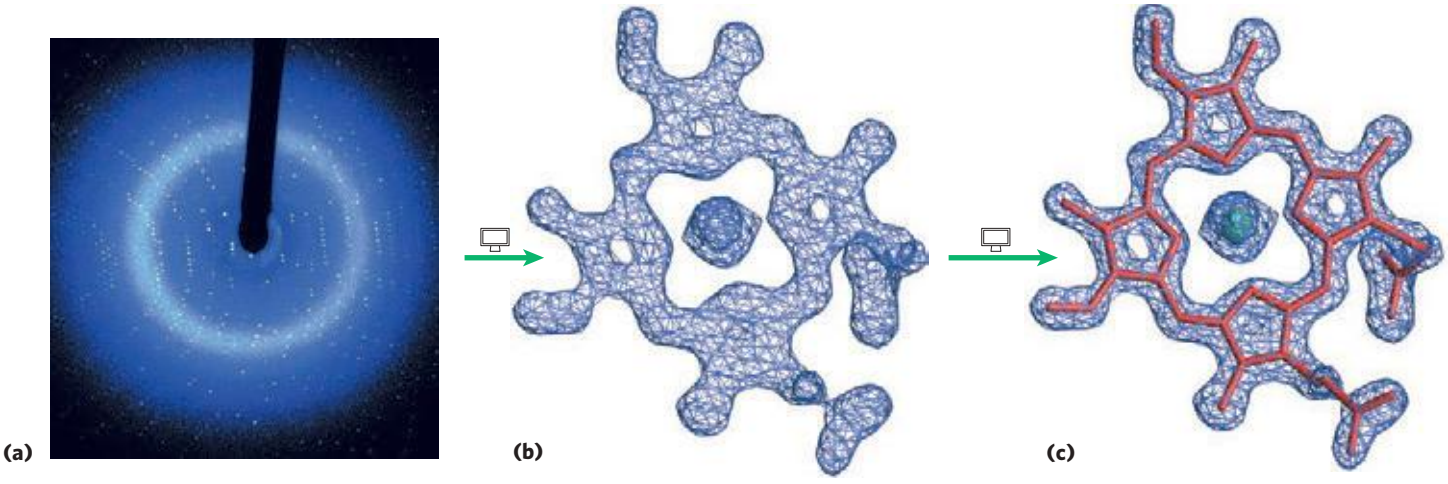
TABLE 4–4 Approximate Proportion of α Helix and β Conformation in Some Single-Chain Proteins

Protein (total residues)	Residues (%)*	
	α Helix	β Conformation
Chymotrypsin (247)	14	45
Ribonuclease (124)	26	35
Carboxypeptidase (307)	38	17
Cytochrome *c* (104)	39	0
Lysozyme (129)	40	12
Myoglobin (153)	78	0

Source: Data from Cantor, C.R. & Schimmel, P.R. (1980) *Biophysical Chemistry*, Part I: *The Conformation of Biological Macromolecules*, p. 100, W. H. Freeman and Company, New York.

*Portions of the polypeptide chains not accounted for by α helix or β conformation consist of bends and irregularly coiled or extended stretches. Segments of α helix and β conformation sometimes deviate slightly from their normal dimensions and geometry.

BOX 4–5 METHODS Methods for Determining the Three-Dimensional Structure of a Protein



X-Ray Diffraction

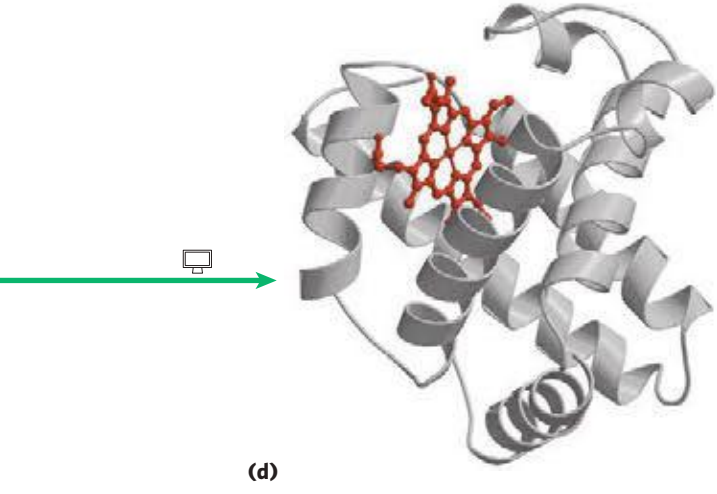
The spacing of atoms in a crystal lattice can be determined by measuring the locations and intensities of spots produced on photographic film by a beam of x rays of given wavelength, after the beam has been diffracted by the electrons of the atoms. For example, x-ray analysis of sodium chloride crystals shows that Na^+ and Cl^- ions are arranged in a simple cubic lattice. The spacing of the different kinds of atoms in complex organic molecules, even very large ones such as proteins, can also be analyzed by x-ray diffraction methods. However, the technique for analyzing crystals of complex molecules is far more laborious than for simple salt crystals. When the repeating pattern of the crystal is a molecule as large as, say, a protein, the numerous atoms in the molecule yield thousands of diffraction spots that must be analyzed by computer.

Consider how images are generated in a light microscope. Light from a point source is focused on an object. The object scatters the light waves, and these scattered waves are recombined by a series of lenses to generate an enlarged image of the object. The smallest object whose structure can be determined by such a system—that is, the resolving power of the microscope—is determined by the wavelength of the light, in this case visible light, with wavelengths in the range of 400 to 700 nm. Objects smaller than half the wavelength of the incident light cannot be resolved. To resolve objects as small as proteins we must use x rays, with wavelengths in the range of 0.7 to 1.5 Å (0.07 to 0.15 nm). However, there are no lenses that can recombine x rays to form an image; instead, the pattern of diffracted x rays is collected directly and an image is reconstructed by mathematical techniques.

The amount of information obtained from x-ray crystallography depends on the degree of structural order in the sample. Some important structural parameters were obtained from early studies of the diffraction patterns of the fibrous proteins arranged in regular arrays in hair and wool. However, the orderly bundles formed by fibrous proteins are not crystals—the molecules are aligned side by side, but not all are oriented in the same direction. More detailed three-dimensional structural information about proteins requires a highly ordered protein crystal. The structures of many proteins are not yet known, simply because they have proved difficult to crystallize. Practitioners have compared making protein crystals to holding together a stack of bowling balls with cellophane tape.

Operationally, there are several steps in x-ray structural analysis (Fig. 1). A crystal is placed in an x-ray beam between the x-ray source and a detector, and a regular array of spots called reflections is generated. The spots are created by the diffracted x-ray beam, and each atom in a molecule makes a contribution to each spot. An electron-density map of the protein is reconstructed from the overall diffraction pattern of spots by a mathematical technique called a Fourier transform. In effect, the computer acts as a "computational lens." A model for the structure is then built that is consistent with the electron-density map.

John Kendrew found that the x-ray diffraction pattern of crystalline myoglobin (isolated from muscles of the sperm whale) is very complex, with nearly 25,000 reflections. Computer analysis of these reflections took place in stages. The resolution improved at each stage, until in 1959 the positions of virtually all the non-hydrogen atoms in the protein had been determined. The amino acid sequence of the protein, obtained by chemical analysis, was consistent with the molecular model. The structures of thousands of proteins, many of them much more complex than



(d)

FIGURE 1 Steps in determining the structure of sperm whale myoglobin by x-ray crystallography. **(a)** X-ray diffraction patterns are generated from a crystal of the protein. **(b)** Data extracted from the diffraction patterns are used to calculate a three-dimensional electron-density map. The electron density of only part of the structure, the heme, is shown here. **(c)** Regions of greatest electron density reveal the location of atomic nuclei, and this information is used to piece together the final structure. Here, the heme structure is modeled into its electron-density map. **(d)** The completed structure of sperm whale myoglobin, including the heme (PDB ID 2MBW).

myoglobin, have since been determined to a similar level of resolution.

The physical environment in a crystal, of course, is not identical to that in solution or in a living cell. A crystal imposes a space and time average on the structure deduced from its analysis, and x-ray diffraction studies provide little information about molecular motion within the protein. The conformation of proteins in a crystal could in principle also be affected by nonphysiological factors such as incidental protein-protein contacts within the crystal. However, when structures derived from the analysis of crystals are compared with structural information obtained by other means (such as NMR, as described below), the crystal-derived structure almost always represents a functional conformation of the protein. X-ray crystallography can be applied successfully to proteins too large to be structurally analyzed by NMR.

Nuclear Magnetic Resonance

An advantage of nuclear magnetic resonance (NMR) studies is that they are carried out on macromolecules in solution, whereas x-ray crystallography is limited to molecules that can be crystallized. NMR can also illuminate the dynamic side of protein structure, including conformational changes, protein folding, and interactions with other molecules.

NMR is a manifestation of nuclear spin angular momentum, a quantum mechanical property of atomic nuclei. Only certain atoms, including ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P, have the kind of nuclear spin that gives rise to an NMR signal. Nuclear spin generates a magnetic dipole. When a strong, static magnetic field is applied to a solution containing a single type of macromolecule, the magnetic dipoles are aligned in the field in one of two orientations, parallel (low energy) or antiparallel (high energy). A short ($\sim$10 μs) pulse of electromagnetic energy of suitable frequency (the resonant frequency, which is in the radio frequency range) is applied at right angles to the nuclei aligned in the magnetic field. Some energy is absorbed as nuclei switch to the high-energy state, and the absorption spectrum that results contains information about the identity of the nuclei and their immediate chemical environment. The data from many such experiments on a sample are averaged, increasing the signal-to-noise ratio, and an NMR spectrum such as that in Figure 2 is generated.

^{1}H is particularly important in NMR experiments because of its high sensitivity and natural abundance. For macromolecules, ^{1}H NMR spectra can become quite complicated. Even a small protein has hundreds of ^{1}H atoms, typically resulting in a one-dimensional NMR spectrum too complex for analysis. Structural analysis of proteins became possible with the advent of two-dimensional NMR techniques (Fig. 3). These methods allow measurement of distance-dependent coupling of nuclear spins in nearby atoms through space (the nuclear Overhauser effect (NOE), in a method dubbed NOESY) or the coupling of nuclear spins in atoms connected by covalent bonds (total correlation spectroscopy, or TOCSY).

(*Continued on next page*)

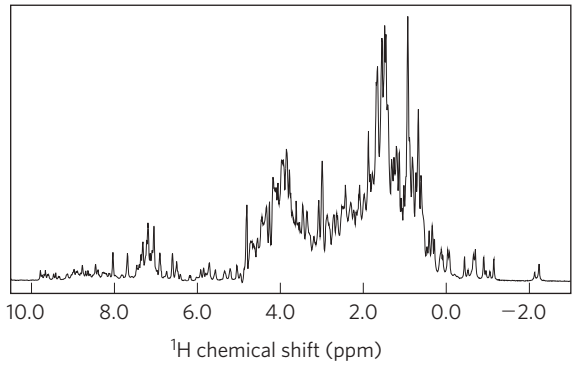


FIGURE 2 One-dimensional NMR spectrum of a globin from a marine blood worm. This protein and sperm whale myoglobin are very close structural analogs, belonging to the same protein structural family and sharing an oxygen-transport function.

BOX 4–5 METHODS Methods for Determining the Three-Dimensional Structure of a Protein (*Continued*)

Translating a two-dimensional NMR spectrum into a complete three-dimensional structure can be a laborious process. The NOE signals provide some information about the distances between individual atoms, but for these distance constraints to be useful, the atoms giving rise to each signal must be identified. Complementary TOCSY experiments can help identify which NOE signals reflect atoms that are linked by covalent bonds. Certain patterns of NOE signals have been associated with secondary structures such as α helices. Genetic engineering (Chapter 9) can be used to prepare proteins that contain the rare isotopes ^{13}C or ^{15}N. The new NMR signals produced by these atoms, and the coupling with ^{1}H signals resulting from these substitutions, help in the assignment of individual ^{1}H NOE signals. The process is also aided by a knowledge of the amino acid sequence of the polypeptide.

To generate a three-dimensional structure, researchers feed the distance constraints into a computer along with known geometric constraints such as chirality, van der Waals radii, and bond lengths and angles. The computer generates a family of closely related structures that represent the range of conformations consistent with the NOE distance constraints (Fig. 3c). The uncertainty in structures generated by NMR is in part a reflection of the molecular vibrations (known as breathing) within a protein structure in solution, discussed in more detail in Chapter 5. Normal experimental uncertainty can also play a role.

Protein structures determined by both x-ray crystallography and NMR generally agree well. In some cases, the precise locations of particular amino acid side chains on the protein exterior are different, often because of effects related to the packing of adjacent protein molecules in a crystal. The two techniques together are at the heart of the rapid increase in the availability of structural information about the macromolecules of living cells.

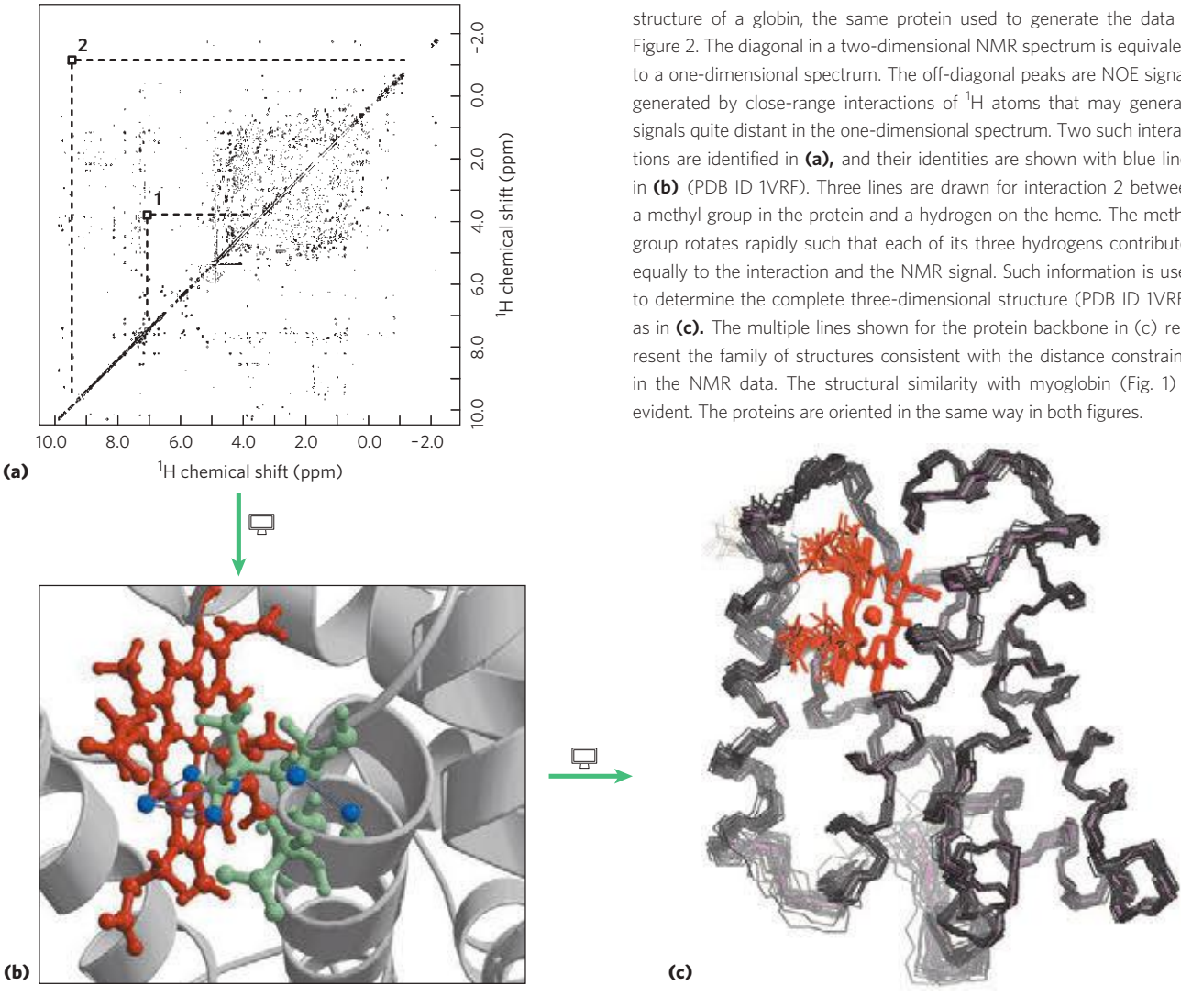


FIGURE 3 Use of two-dimensional NMR to generate a three-dimensional structure of a globin, the same protein used to generate the data in Figure 2. The diagonal in a two-dimensional NMR spectrum is equivalent to a one-dimensional spectrum. The off-diagonal peaks are NOE signals generated by close-range interactions of ^{1}H atoms that may generate signals quite distant in the one-dimensional spectrum. Two such interactions are identified in **(a),** and their identities are shown with blue lines in **(b)** (PDB ID 1VRF). Three lines are drawn for interaction 2 between a methyl group in the protein and a hydrogen on the heme. The methyl group rotates rapidly such that each of its three hydrogens contributes equally to the interaction and the NMR signal. Such information is used to determine the complete three-dimensional structure (PDB ID 1VRE), as in **(c).** The multiple lines shown for the protein backbone in (c) represent the family of structures consistent with the distance constraints in the NMR data. The structural similarity with myoglobin (Fig. 1) is evident. The proteins are oriented in the same way in both figures.

The first term is **motif**, also called a **fold** or (more rarely) **supersecondary structure**. *A motif or fold is a recognizable folding pattern involving two or more elements of secondary structure and the connection(s) between them.* A motif can be very simple, such as two elements of secondary structure folded against each other, and represent only a small part of a protein. An example is a **β-α-β loop (Fig. 4–18a)**. A motif can also be a very elaborate structure involving scores of protein segments folded together, such as the β barrel (Fig. 4–18b). In some cases, a single large motif may comprise the entire protein. The terms "motif" and "fold" are often used interchangeably, although "fold" is applied more commonly to somewhat more complex folding patterns. The terms encompass any advantageous folding pattern and are useful for describing such patterns. The segment defined as a motif or fold may or may not be independently stable. We have already encountered a well-studied motif, the coiled coil of α-keratin, which is also found in some other proteins. The distinctive arrangement of eight α helices in myoglobin is replicated in all globins and is called the globin fold. Note that a motif is not a hierarchical structural element falling between secondary and tertiary structure. It is simply a folding pattern. The synonymous term "supersecondary structure" is thus somewhat misleading because it suggests hierarchy.

The second term for describing structural patterns is **domain**. A domain, as defined by Jane Richardson in 1981, is a part of a polypeptide chain that is independently stable or could undergo movements as a single entity with respect to the entire protein. Polypeptides with more than a few hundred amino acid residues often fold into two or more domains, sometimes with different functions. In many cases, a domain from a large protein will retain its native three-dimensional structure even when separated (for example, by proteolytic cleavage) from the remainder of the polypeptide chain. In a protein with multiple domains, each domain may appear as a distinct globular lobe **(Fig. 4–19)**; more commonly, extensive contacts between domains make individual domains hard to discern. Different domains often have distinct functions, such as the binding of small molecules or interaction with other proteins. Small proteins usually have only one domain (the domain *is* the protein).

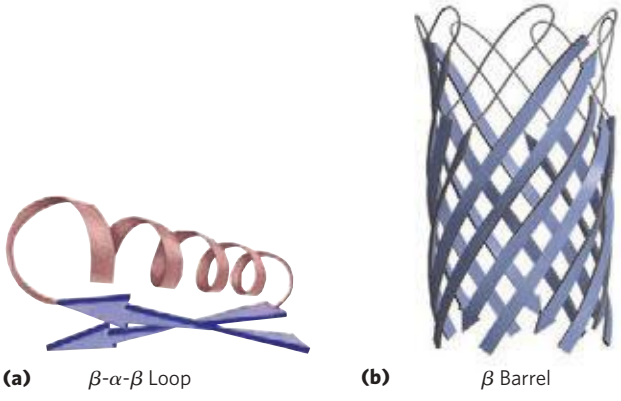


FIGURE 4–18 Motifs. (a) A simple motif, the β-α-β loop. **(b)** A more elaborate motif, the β barrel. This β barrel is a single domain of α-hemolysin (a toxin that kills a cell by creating a hole in its membrane) from the bacterium *Staphylococcus aureus* (derived from PDB ID 7AHL).

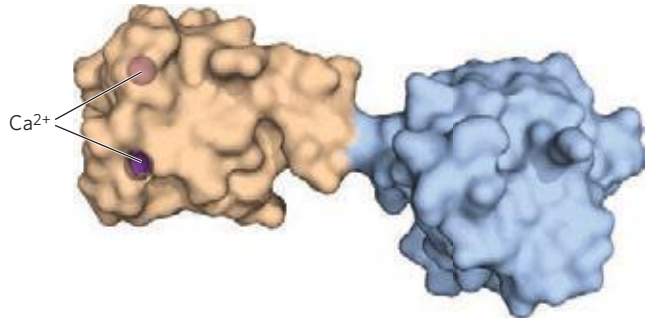


FIGURE 4–19 Structural domains in the polypeptide troponin C. (PDB ID 4TNC) This calcium-binding protein, associated with muscle, has two separate calcium-binding domains, shown here in brown and blue.

Folding of polypeptides is subject to an array of physical and chemical constraints, and several rules have emerged from studies of common protein folding patterns.

1. Hydrophobic interactions make a large contribution to the stability of protein structures. Burial of hydrophobic amino acid R groups so as to exclude water requires at least two layers of secondary structure. Simple motifs, such as the β-α-β loop (Fig. 4–18a), create two such layers.
2. Where they occur together in a protein, α helices and β sheets generally are found in different structural layers. This is because the backbone of a polypeptide segment in the β conformation (Fig. 4–6) cannot readily hydrogen-bond to an α helix that is adjacent to it.
3. Segments adjacent to each other in the amino acid sequence are usually stacked adjacent to each other in the folded structure. Distant segments of a polypeptide may come together in the tertiary structure, but this is not the norm.
4. The β conformation is most stable when the individual segments are twisted slightly in a right-handed sense. This influences both the arrangement of β sheets derived from the twisted segments and the path of the polypeptide connections between them. Two parallel β strands, for example, must be connected by a crossover strand **(Fig. 4–20a)**. In principle, this crossover could have a right- or left-handed conformation, but in proteins it is almost always right-handed. Right-handed connections tend to be shorter than left-handed connections and tend to bend through smaller angles, making them easier to

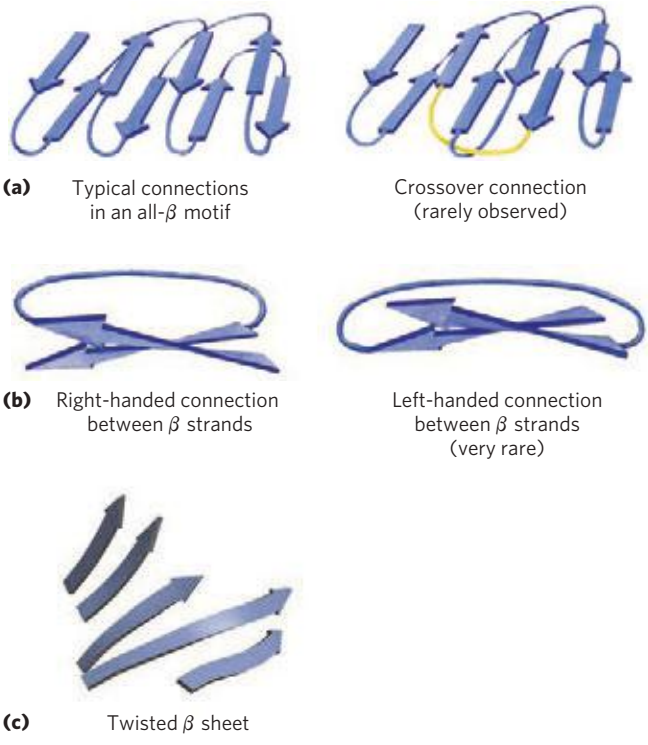


FIGURE 4–20 Stable folding patterns in proteins. (a) Connections between β strands in layered β sheets. The strands here are viewed from one end, with no twisting. Thick lines represent connections at the ends nearest the viewer; thin lines are connections at the far ends of the β strands. The connections at a given end (e.g., near the viewer) rarely cross one another. An example of such a rare crossover is illustrated by the yellow strand in the structure on the right. **(b)** Because of the right-handed twist in β strands, connections between strands are generally right-handed. Left-handed connections must traverse sharper angles and are harder to form. **(c)** This twisted β sheet is from a domain of photolyase (a protein that repairs certain types of DNA damage) from *E. coli* (derived from PDB ID 1DNP). Connecting loops have been removed so as to focus on the folding of the β sheet.

form. The twisting of β sheets also leads to a characteristic twisting of the structure formed by many such segments together, as seen in the β barrel (Fig. 4–18b) and twisted β sheet (Fig. 4–20c), which form the core of many larger structures.

Following these rules, complex motifs can be built up from simple ones. For example, a series of β-α-β loops arranged so that the β strands form a barrel creates a particularly stable and common motif, the **α/β barrel (Fig. 4–21)**. In this structure, each parallel β segment is attached to its neighbor by an α-helical segment. All connections are right-handed. The α/β barrel is found in many enzymes, often with a binding site (for a cofactor or substrate) in the form of a pocket near one end of the barrel. Note that domains with similar folding patterns are said to have the same motif even though their constituent α helices and β sheets may differ in length.

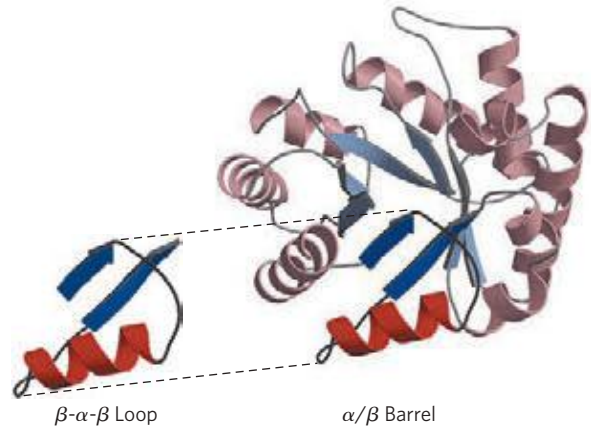


FIGURE 4–21 Constructing large motifs from smaller ones. The α/β barrel is a commonly occurring motif constructed from repetitions of the β-α-β loop motif. This α/β barrel is a domain of pyruvate kinase (a glycolytic enzyme) from rabbit (derived from PDB ID 1PKN).

Protein Motifs Are the Basis for Protein Structural Classification

Protein Architecture—Tertiary Structure of Large Globular Proteins, IV. Structural Classification of Proteins As we have seen, understanding the complexities of tertiary structure is made easier by considering substructures. Taking this idea further, researchers have organized the complete contents of protein databases according to hierarchical levels of structure. All of these databases rely on data and information deposited in the Protein Data Bank. The Structural Classification of Proteins (SCOP) database is a good example of this important trend in biochemistry. At the highest level of classification, the SCOP database (http://scop.mrc-lmb.cam.ac.uk/scop) borrows a scheme already in common use, with four classes of protein structure: **all α**, **all β**, **α/β** (with α and β segments interspersed or alternating), and **$\alpha + \beta$** (with α and β regions somewhat segregated). Each class includes tens to hundreds of different folding arrangements (motifs), built up from increasingly identifiable substructures. Some of the substructure arrangements are very common; others have been found in just one protein. **Figure 4–22** shows a variety of motifs arrayed among the four classes of protein structure; this is just a minute sample of the hundreds of known motifs. The number of folding patterns is not infinite, however. As the rate at which new protein structures are elucidated has increased, the fraction of those structures containing a new motif has steadily declined. Fewer than 1,000 different folds or motifs may exist in all. Figure 4–22 also shows how proteins can be organized based on the presence of the various motifs. The top two levels of organization, **class** and **fold**, are purely structural. Below the fold level (see color key in Fig. 4–22), categorization is based on evolutionary relationships.

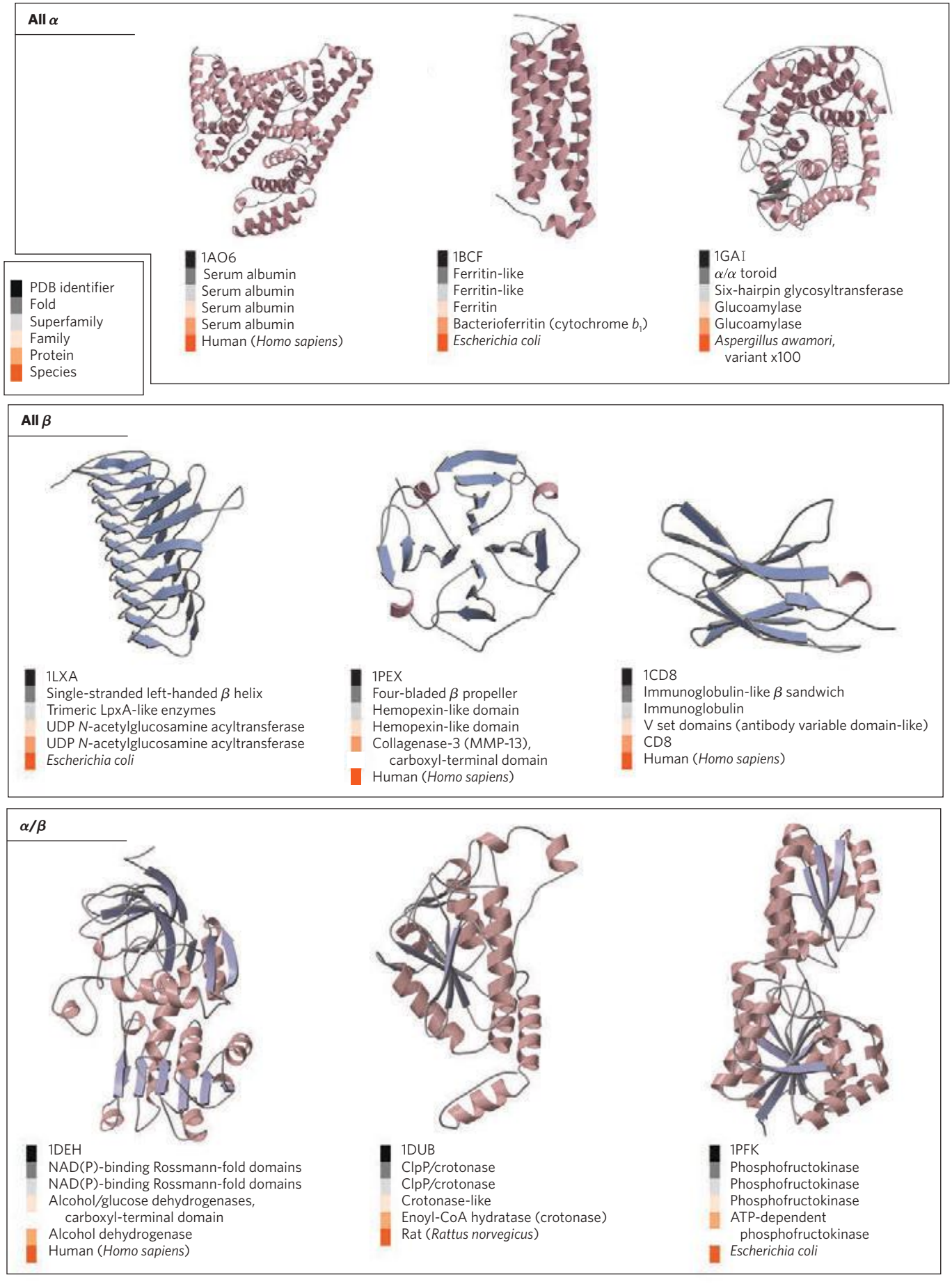


FIGURE 4–22 Organization of proteins based on motifs. Shown here are a few of the hundreds of known stable motifs divided into four classes: all α, all β, α/β, and $\alpha + \beta$. Structural classification data from the SCOP (Structural Classification of Proteins) database (http://scop.mrc-lmb.cam.ac.uk/scop) are also provided (see the color key). The PDB identifier (listed first for each structure) is the unique accession code given to each structure archived in the Protein Data Bank (www.pdb.org). The α/β, barrel (see Fig. 4–21) is another particularly common α/β motif. ***(Continued on next page)***

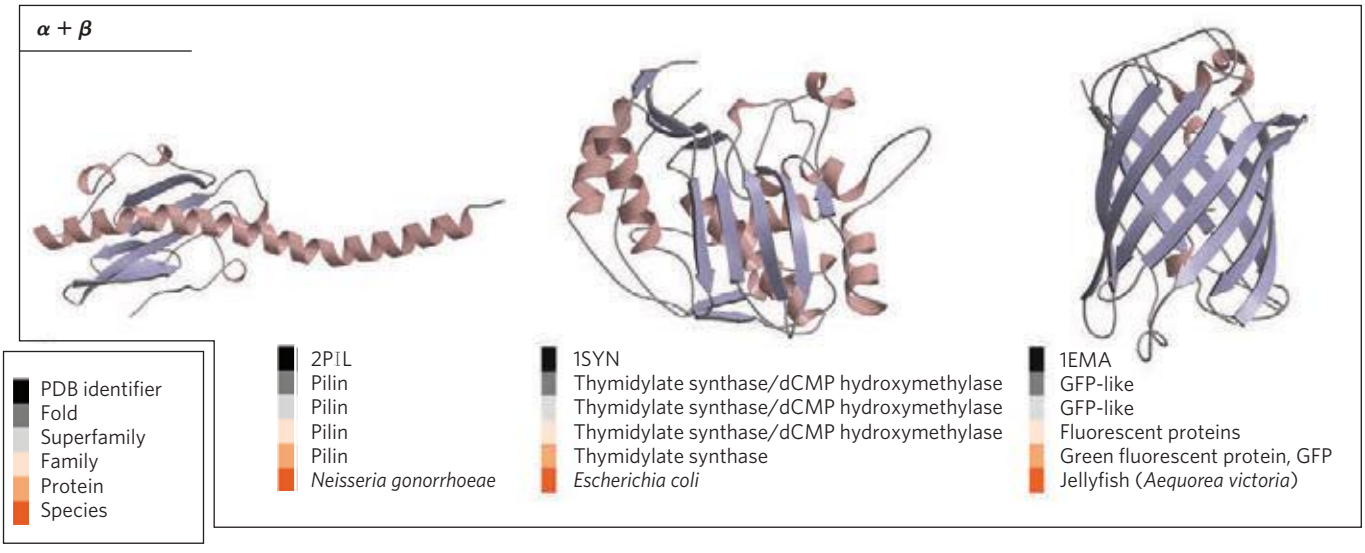


FIGURE 4–22 ***(Continued)***

Many examples of recurring domain or motif structures are available, and these reveal that protein tertiary structure is more reliably conserved than amino acid sequence. The comparison of protein structures can thus provide much information about evolution. Proteins with significant similarity in primary structure and/or with similar tertiary structure and function are said to be in the same **protein family**. A strong evolutionary relationship is usually evident within a protein family. For example, the globin family has many different proteins with both structural and sequence similarity to myoglobin (as seen in the proteins used as examples in Box 4–5 and in Chapter 5). Two or more families with little similarity in amino acid sequence sometimes make use of the same major structural motif and have functional similarities; these families are grouped as **superfamilies**. An evolutionary relationship among families in a superfamily is considered probable, even though time and functional distinctions—that is, different adaptive pressures—may have erased many of the telltale sequence relationships. A protein family may be widespread in all three domains of cellular life, the Bacteria, Archaea, and Eukarya, suggesting an ancient origin. Many proteins involved in intermediary metabolism and the metabolism of nucleic acids and proteins fall into this category. Other families may be present in only a small group of organisms, indicating that the structure arose more recently. Tracing the natural history of structural motifs, using structural classifications in databases such as SCOP, provides a powerful complement to sequence analyses in tracing evolutionary relationships. The SCOP database is curated manually, with the objective of placing proteins in the correct evolutionary framework based on conserved structural features.

Structural motifs become especially important in defining protein families and superfamilies. Improved classification and comparison systems for proteins lead inevitably to the elucidation of new functional relationships. Given the central role of proteins in living systems, these structural comparisons can help illuminate every aspect of biochemistry, from the evolution of individual proteins to the evolutionary history of complete metabolic pathways.

Protein Quaternary Structures Range from Simple Dimers to Large Complexes

Protein Architecture—Quaternary Structure Many proteins have multiple polypeptide subunits (from two to hundreds). The association of polypeptide chains can serve a variety of functions. Many multisubunit proteins have regulatory roles; the binding of small molecules may affect the interaction between subunits, causing large changes in the protein's activity in response to small changes in the concentration of substrate or regulatory molecules (Chapter 6). In other cases, separate subunits take on separate but related functions, such as catalysis and regulation. Some associations, such as the fibrous proteins considered earlier in this chapter and the coat proteins of viruses, serve primarily structural roles. Some very large protein assemblies are the site of complex, multistep reactions. For example, each ribosome, the site of protein synthesis, incorporates dozens of protein subunits along with a number of RNA molecules.

A multisubunit protein is also referred to as a **multimer**. A multimer with just a few subunits is often called an **oligomer**. If a multimer has nonidentical subunits, the overall structure of the protein can be asymmetric and quite complicated. However, most multimers have identical subunits or repeating groups of nonidentical subunits, usually in symmetric arrangements. As noted in Chapter 3, the repeating structural unit in such a multimeric protein, whether a single subunit or a group of subunits, is called a

protomer. Greek letters are sometimes used to distinguish the individual subunits that make up a protomer.

The first oligomeric protein to have its three-dimensional structure determined was hemoglobin (M_r 64,500), which contains four polypeptide chains and four heme prosthetic groups, in which the iron atoms are in the ferrous (Fe^{2+}) state (Fig. 4–17). The protein portion, the globin, consists of two α chains (141 residues each) and two β chains (146 residues each). Note that in this case, α and β do not refer to secondary structures. In a practice that can be confusing to the beginning student, the Greek letters α and β (and γ and δ, and others) are often used to distinguish two different kinds of subunits within a multisubunit protein, regardless of what kinds of secondary structure may predominate in the subunits. Because hemoglobin is four times as large as myoglobin, much more time and effort were required to solve its three-dimensional structure by x-ray analysis, finally achieved by Max Perutz, John Kendrew, and their colleagues in 1959. The subunits of hemoglobin are arranged in symmetric pairs **(Fig. 4–23)**, each pair having one α and one β subunit. Hemoglobin can therefore be described either as a tetramer or as a dimer of $\alpha\beta$ protomers. The role these distinct subunits play in hemoglobin function is discussed extensively in Chapter 5.

Max Perutz, 1914–2002 (left), and John Kendrew, 1917–1997

Some Proteins or Protein Segments Are Intrinsically Disordered

In spite of decades of progress in the understanding of protein structure, many proteins cannot be crystallized, making it difficult to determine their three-dimensional structure by methods now considered classical (see Box 4–5). Even where crystallization succeeds, parts of the protein are often sufficiently disordered within the crystal that the determined structure does not include those parts. Sometimes, this is due to subtle features of the structure that render crystallization difficult. However, the reason can be more straightforward: some proteins or protein segments lack an ordered structure in solution.

The concept that some proteins function in the absence of a definable structure is a product of the reassessment of data involving many different proteins. As many as a third of all human proteins may be unstructured or have significant unstructured segments. All organisms have some proteins that fall into this category. **Intrinsically disordered proteins** have properties that are distinct from classical structured proteins. They lack a hydrophobic core, and instead are characterized by high densities of charged amino acid residues such as Lys, Arg, and Glu. Pro residues are also prominent, as they tend to disrupt ordered structures.

Structural disorder and high charge density can facilitate the function of some proteins as spacers, insulators, or linkers in larger structures. Other disordered proteins are scavengers, binding up ions and small molecules in solution and serving as reservoirs or garbage dumps. However, many intrinsically disordered proteins are at the heart of important protein interaction networks. The lack of an ordered structure can facilitate a kind of functional promiscuity, allowing one protein to interact with multiple partners. Some intrinsically disordered proteins act to inhibit the action of other proteins by an unusual mechanism: wrapping around their protein targets. One disordered protein may have several or even dozens of protein partners. The structural disorder allows the inhibitor protein to wrap around the multiple

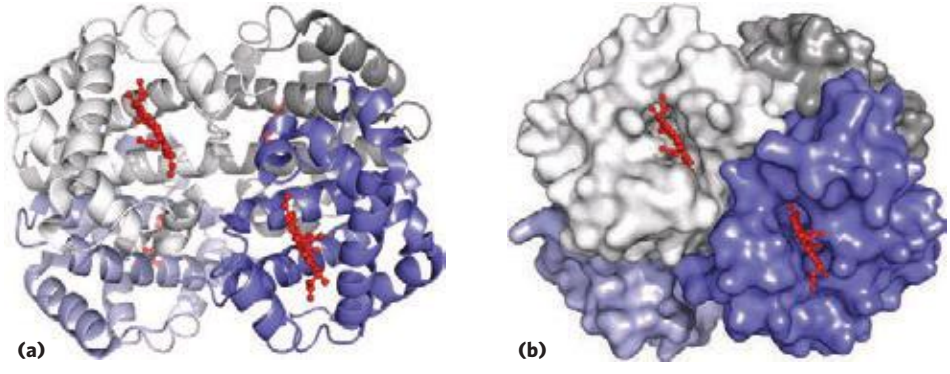


FIGURE 4–23 Quaternary structure of deoxyhemoglobin. (PDB ID 2HHB) X-ray diffraction analysis of deoxyhemoglobin (hemoglobin without oxygen molecules bound to the heme groups) shows how the four polypeptide subunits are packed together. **(a)** A ribbon representation reveals the secondary structural elements of the structure and the positioning of all the heme cofactors. **(b)** A surface contour model shows the pockets in which the heme cofactors are bound and helps to visualize subunit packing. The α subunits are shown in shades of gray; the β subunits in shades of blue. Note that the heme groups (red) are relatively far apart.

targets in different ways. The intrinsically disordered protein p27 plays a key role in controlling mammalian cell division. This protein lacks definable structure in solution. It wraps around and thus inhibits the action of several enzymes called protein kinases (see Chapter 6) that facilitate cell division. The flexible structure of p27 allows it to accommodate itself to its different target proteins. Human tumor cells, which are simply cells that have lost the capacity to control cell division normally, generally have reduced levels of p27; the lower the levels of p27, the poorer the prognosis for the cancer patient. Similarly, intrinsically disordered proteins are often present as hubs or scaffolds at the center of protein networks that constitute signaling pathways. These proteins, or parts of them, may interact with many different binding partners. They often take on a structure when they interact with other proteins, but the structure they assume may vary with different binding partners. The mammalian protein p53 is also critical in the control of cell division. It features both structured and unstructured segments, and the different segments interact with dozens of other proteins. An unstructured region of p53 at the carboxyl terminus interacts with at least four different binding partners and assumes a different structure in each of the complexes **(Fig. 4–24)**.

SUMMARY 4.3 Protein Tertiary and Quaternary Structures

- Tertiary structure is the complete three-dimensional structure of a polypeptide chain. Many proteins fall into one of two general classes of proteins based on tertiary structure: fibrous and globular.
- Fibrous proteins, which serve mainly structural roles, have simple repeating elements of secondary structure.
- Globular proteins have more complicated tertiary structures, often containing several types of secondary structure in the same polypeptide chain. The first globular protein structure to be determined, by x-ray diffraction methods, was that of myoglobin.
- The complex structures of globular proteins can be analyzed by examining folding patterns called motifs (also called folds or supersecondary structures). The thousands of known protein structures are generally assembled from a repertoire of only a few hundred motifs. Domains are regions of a polypeptide chain that can fold stably and independently.

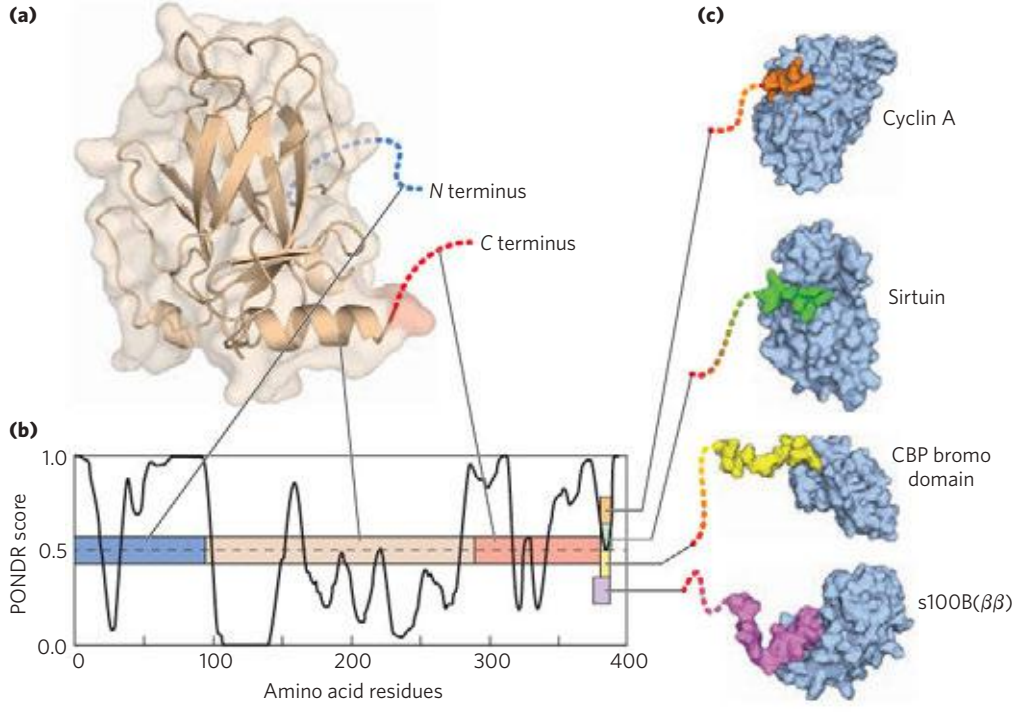


FIGURE 4–24 Binding of the intrinsically disordered carboxyl terminus of p53 protein to its binding partners. (a) The p53 protein is made up of several different segments (PDB ID 1XQH). Only the central domain is well ordered. **(b)** The linear sequence of the p53 protein is depicted as a colored bar. The overlaid graph presents a plot of the PONDR (Predictor of Natural Disordered Regions) score versus the protein sequence. PONDR is one of the best available algorithms for predicting the likelihood that a given amino acid residue is in a region of intrinsic disorder, based on the surrounding amino acid sequence and amino acid composition. A score of 1.0 indicates a probability of 100% that a protein will be disordered. In the actual protein structure, the tan central domain is ordered. The amino-terminal (blue) and carboxyl-terminal (red) regions are disordered. The very end of the carboxyl-terminal region has multiple binding partners and folds when it binds to each of them; however, the three-dimensional structure that is assumed when binding occurs is different for each of the interactions shown, and thus the color of this carboxyl-terminal segment (11 to 20 residues) is shown in a different color in each complex (cyclin A, PDB ID 1H26; sirtuin, PDB ID 1MA3; CBP bromo domain, PDB ID 1JSP; s100B(*ββ*), PDB ID 1DT7).

- Quaternary structure results from interactions between the subunits of multisubunit (multimeric) proteins or large protein assemblies. Some multimeric proteins have a repeated unit consisting of a single subunit or a group of subunits, each unit called a protomer.
- Some proteins or protein segments are intrinsically disordered, lacking definable structure. These proteins have distinctive amino acid compositions that allow a more flexible structure. Some of these disordered proteins function as structural components or scavengers; others can interact with many different protein partners, serving as versatile inhibitors or as central components of protein interaction networks.

4.4 Protein Denaturation and Folding

Proteins lead a surprisingly precarious existence. As we have seen, a native protein conformation is only marginally stable. In addition, most proteins must maintain conformational flexibility to function. The continual maintenance of the active set of cellular proteins required under a given set of conditions is called **proteostasis**. Cellular proteostasis requires the coordinated function of pathways for protein synthesis and folding, the refolding of proteins that are partially unfolded, and the sequestration and degradation of proteins that have been irreversibly unfolded. In all cells, these networks involve hundreds of enzymes and specialized proteins.

As seen in **Figure 4–25**, the life of a protein encompasses much more than its synthesis and later degradation. The marginal stability of most proteins can produce a tenuous balance between folded and unfolded states. As proteins are synthesized on ribosomes (Chapter 27), they must fold into their native conformations. Sometimes this occurs spontaneously, but more often it occurs with the assistance of specialized enzymes and complexes called chaperones. Many of these same folding helpers function to refold proteins that become transiently unfolded. Proteins that are not properly folded often have exposed hydrophobic surfaces that render them "sticky," leading to the formation of inactive aggregates. These aggregates may lack their normal function but are not inert; their accumulation in cells lies at the heart of diseases ranging from diabetes to Parkinson and Alzheimer diseases. Not surprisingly, all cells have elaborate pathways for recycling and/or degrading proteins that are irreversibly misfolded.

The transitions between the folded and unfolded states, and the network of pathways that control these transitions, now become our focus.

Loss of Protein Structure Results in Loss of Function

Protein structures have evolved to function in particular cellular environments. Conditions different from those in the cell can result in protein structural changes, large and small. A loss of three-dimensional structure sufficient to cause loss of function is called **denaturation**. The denatured state does not necessarily equate with complete unfolding of the protein and randomization of conformation. Under most conditions, denatured proteins exist in a set of partially folded states.

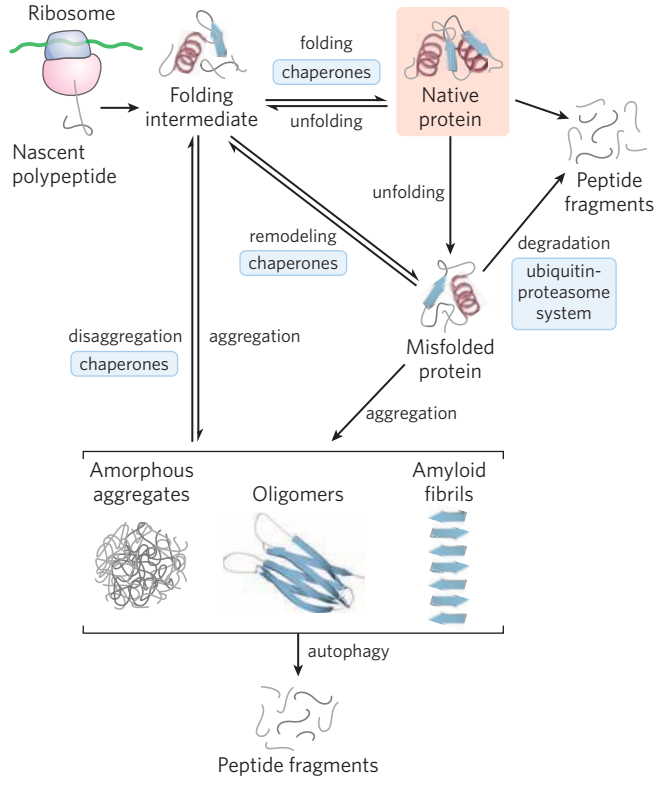


FIGURE 4–25 Pathways that contribute to proteostasis. Three kinds of processes contribute to proteostasis, in some cases with multiple contributing pathways. First, proteins are synthesized on a ribosome. Second, multiple pathways contribute to protein folding, many of which involve the activity of complexes called chaperones. Chaperones (including chaperonins) also contribute to the refolding of proteins that are partially and transiently unfolded. Finally, proteins that are irreversibly unfolded are subject to sequestration and degradation by several additional pathways. Partially unfolded proteins and protein-folding intermediates that escape the quality-control activities of the chaperones and degradative pathways may aggregate, forming both disordered aggregates and ordered amyloid-like aggregates that contribute to disease and aging processes.

Most proteins can be denatured by heat, which has complex effects on many weak interactions in a protein (primarily on the hydrogen bonds). If the temperature is increased slowly, a protein's conformation generally remains intact until an abrupt loss of structure (and function) occurs over a narrow temperature range **(Fig. 4–26)**. The abruptness of the change suggests that unfolding is a cooperative process: loss of structure in one part of the protein destabilizes other parts. The effects of heat on proteins are not readily predictable. The very heat-stable proteins of thermophilic bacteria and archaea have evolved to function at the temperature of hot springs (~100 °C). Yet the structures of these proteins often differ only slightly from those of

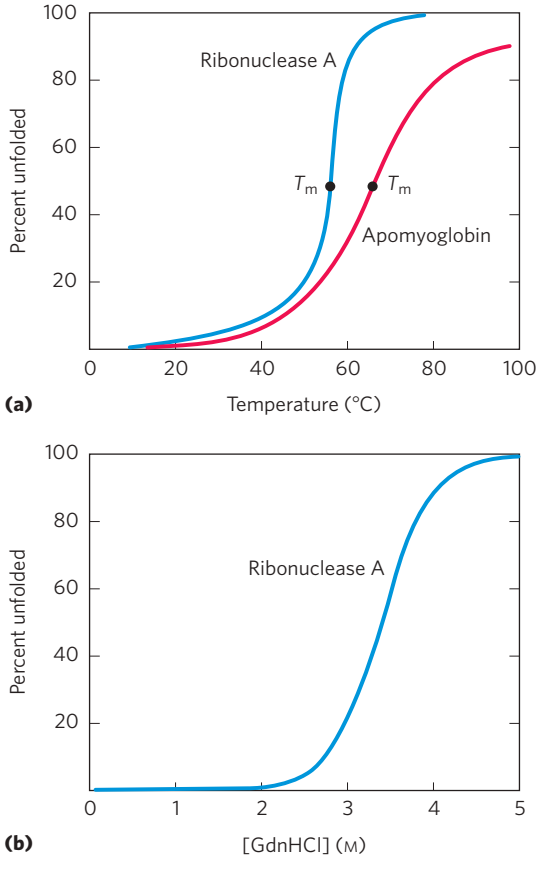


FIGURE 4–26 Protein denaturation. Results are shown for proteins denatured by two different environmental changes. In each case, the transition from the folded to the unfolded state is abrupt, suggesting cooperativity in the unfolding process. **(a)** Thermal denaturation of horse apomyoglobin (myoglobin without the heme prosthetic group) and ribonuclease A (with its disulfide bonds intact; see Fig. 4–27). The midpoint of the temperature range over which denaturation occurs is called the melting temperature, or T_m. Denaturation of apomyoglobin was monitored by circular dichroism (see Fig. 4–10), which measures the amount of helical structure in the protein. Denaturation of ribonuclease A was tracked by monitoring changes in the intrinsic fluorescence of the protein, which is affected by changes in the environment of Trp residues. **(b)** Denaturation of disulfide-intact ribonuclease A by guanidine hydrochloride (GdnHCl), monitored by circular dichroism.

homologous proteins derived from bacteria such as *Escherichia coli*. How these small differences promote structural stability at high temperatures is imperfectly understood.

Proteins can also be denatured by extremes of pH, by certain miscible organic solvents such as alcohol or acetone, by certain solutes such as urea and guanidine hydrochloride, or by detergents. Each of these denaturing agents represents a relatively mild treatment in the sense that no covalent bonds in the polypeptide chain are broken. Organic solvents, urea, and detergents act primarily by disrupting the hydrophobic interactions that make up the stable core of globular proteins; urea also disrupts hydrogen bonds; extremes of pH alter the net charge on the protein, causing electrostatic repulsion and the disruption of some hydrogen bonding. The denatured structures resulting from these various treatments are not necessarily the same.

Denaturation often leads to protein precipitation, a consequence of protein aggregate formation as exposed hydrophobic surfaces associate. The aggregates are often highly disordered. The protein precipitate that is seen after boiling an egg white is one example. More-ordered aggregates are also observed in some proteins, as we shall see.

Amino Acid Sequence Determines Tertiary Structure

The tertiary structure of a globular protein is determined by its amino acid sequence. The most important proof of this came from experiments showing that denaturation of some proteins is reversible. Certain globular proteins denatured by heat, extremes of pH, or denaturing reagents will regain their native structure and their biological activity if returned to conditions in which the native conformation is stable. This process is called **renaturation**.

A classic example is the denaturation and renaturation of ribonuclease A, demonstrated by Christian Anfinsen in the 1950s. Purified ribonuclease A denatures completely in a concentrated urea solution in the presence of a reducing agent. The reducing agent cleaves the four disulfide bonds to yield eight Cys residues, and the urea disrupts the stabilizing hydrophobic interactions, thus freeing the entire polypeptide from its folded conformation. Denaturation of ribonuclease is accompanied by a complete loss of catalytic activity. When the urea and the reducing agent are removed, the randomly coiled, denatured ribonuclease spontaneously refolds into its correct tertiary structure, with full restoration of its catalytic activity **(Fig. 4–27)**. The refolding of ribonuclease is so accurate that the four intrachain disulfide bonds are re-formed in the same positions in the renatured molecule as in the native ribonuclease. Later, similar results were obtained using chemically synthesized, catalytically active ribonuclease A. This eliminated the possibility that some minor contaminant in Anfinsen's purified ribonuclease preparation might have contributed to the renaturation of the enzyme, thus dispelling any remaining doubt that this enzyme folds spontaneously.

The Anfinsen experiment provided the first evidence that the amino acid sequence of a polypeptide chain contains all the information required to fold the chain into its native, three-dimensional structure. Subsequent work has shown that only a minority of proteins, many of them small and inherently stable, will fold spontaneously into their native form. Even though all proteins have the potential to fold into their native structure, many require some assistance.

Polypeptides Fold Rapidly by a Stepwise Process

In living cells, proteins are assembled from amino acids at a very high rate. For example, *E. coli* cells can make a complete, biologically active protein molecule containing 100 amino acid residues in about 5 seconds at 37 °C.

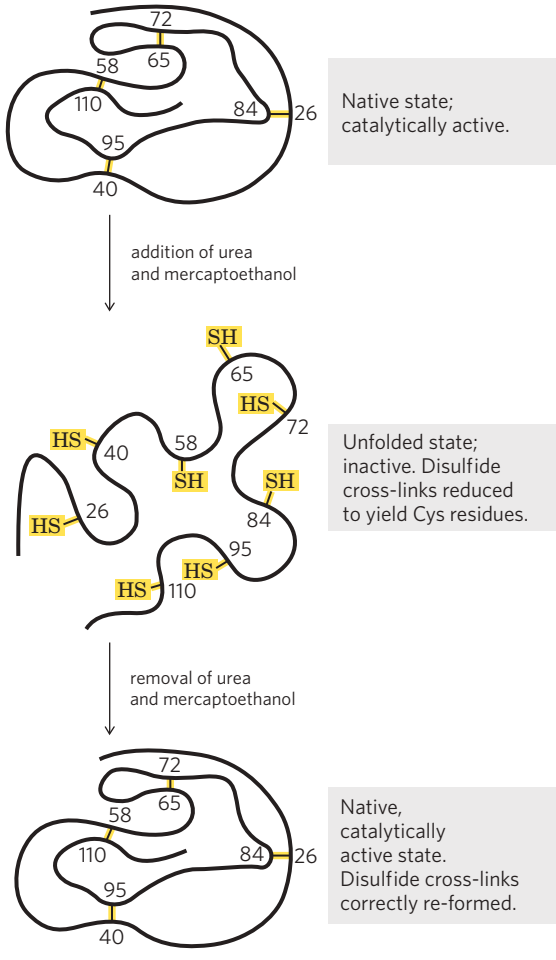


FIGURE 4–27 Renaturation of unfolded, denatured ribonuclease. Urea denatures the ribonuclease, and mercaptoethanol ($HOCH_2CH_2SH$) reduces and thus cleaves the disulfide bonds to yield eight Cys residues. Renaturation involves reestablishing the correct disulfide cross-links.

However, the synthesis of peptide bonds on the ribosome is not enough; the protein must fold.

How does the polypeptide chain arrive at its native conformation? Let's assume conservatively that each of the amino acid residues could take up 10 different conformations on average, giving 10^{100} different conformations for the polypeptide. Let's also assume that the protein folds spontaneously by a random process in which it tries out all possible conformations around every single bond in its backbone until it finds its native, biologically active form. If each conformation were sampled in the shortest possible time ($\sim 10^{-13}$ second, or the time required for a single molecular vibration), it would take about 10^{77} years to sample all possible conformations. Clearly, protein folding is not a completely random, trial-and-error process. There must be shortcuts. This problem was first pointed out by Cyrus Levinthal in 1968 and is sometimes called Levinthal's paradox.

The folding pathway of a large polypeptide chain is unquestionably complicated. However, rapid progress has been made in this field, sufficient to produce robust algorithms that can often predict the structure of smaller proteins on the basis of their amino acid sequences. The major folding pathways are hierarchical. Local secondary structures form first. Certain amino acid sequences fold readily into α helices or β sheets, guided by constraints such as those reviewed in our discussion of secondary structure. Ionic interactions, involving charged groups that are often near one another in the linear sequence of the polypeptide chain, can play an important role in guiding these early folding steps. Assembly of local structures is followed by longer-range interactions between, say, two elements of secondary structure that come together to form stable folded structures. Hydrophobic interactions play a significant role throughout the process, as the aggregation of nonpolar amino acid side chains provides an entropic stabilization to intermediates and, eventually, to the final folded structure. The process continues until complete domains form and the entire polypeptide is folded **(Fig. 4–28)**. Notably, proteins dominated by close-range interactions (between pairs of residues generally located near each other in the polypeptide sequence) tend to fold faster than proteins with more complex folding patterns and many long-range interactions between different segments. As larger proteins with multiple domains are synthesized, domains near the amino terminus (which are synthesized

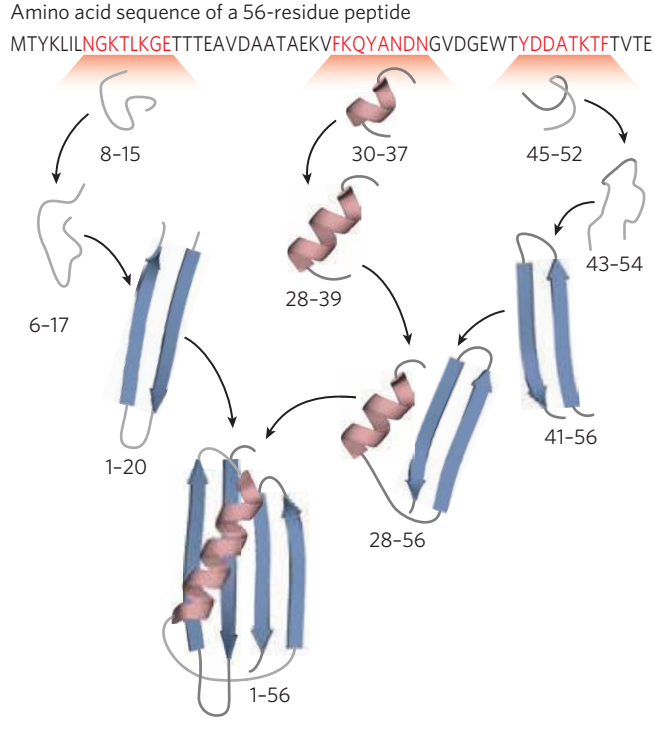


FIGURE 4–28 A protein-folding pathway as defined for a small protein. A hierarchical pathway is shown, based on computer modeling. Small regions of secondary structure are assembled first and then gradually incorporated into larger structures. The program used for this model has been highly successful in predicting the three-dimensional structure of small proteins from their amino acid sequence. The numbers indicate the amino acid residues in this 56 residue peptide that have acquired their final structure in each of the steps shown.

first) may fold before the entire polypeptide has been assembled.

Thermodynamically, the folding process can be viewed as a kind of free-energy funnel **(Fig. 4–29)**. The unfolded states are characterized by a high degree of conformational entropy and relatively high free energy. As folding proceeds, the narrowing of the funnel reflects the decrease in the conformational space that must be searched as the protein approaches its native state. Small depressions along the sides of the free-energy funnel represent semistable intermediates that can briefly slow the folding process. At the bottom of the funnel, an ensemble of folding intermediates has been reduced to a single native conformation (or one of a small set of native conformations). The funnels can have a variety of shapes depending on the complexity of the folding pathway, the existence of semistable intermediates, and the potential for particular intermediates to assemble into aggregates of misfolded proteins (Fig. 4–29).

Thermodynamic stability is not evenly distributed over the structure of a protein—the molecule has regions of relatively high stability and others of low or negligible stability. For example, a protein may have two stable domains joined by a segment that is entirely disordered. Regions of low stability may allow a protein to alter its conformation between two or more states. As we shall see in the next two chapters, variations in the stability of regions within a protein are often essential to protein function. Intrinsically disordered proteins or protein segments do not fold at all.

As our understanding of protein folding and protein structure improves, increasingly sophisticated computer programs for predicting the structure of proteins from their amino acid sequence are being developed. Prediction of protein structure is a specialty field of bioinformatics, and progress in this area is monitored with a biennial test called the CASP (Critical Assessment of Structural Prediction) competition. Entrants from around the world vie to predict the structure of an assigned protein (whose structure has been determined but not yet published). The most successful teams are invited to present their results at a CASP conference. The success of these efforts is improving rapidly.

Some Proteins Undergo Assisted Folding

Not all proteins fold spontaneously as they are synthesized in the cell. Folding for many proteins requires **chaperones**, proteins that interact with partially folded or improperly folded polypeptides, facilitating correct folding pathways or providing microenvironments in which folding can occur. Several types of molecular chaperones are found in organisms ranging from bacteria to humans. Two major families of chaperones, both well studied, are the **Hsp70** family and the **chaperonins**.

The Hsp70 family of proteins generally have a molecular weight near 70,000 and are more abundant in

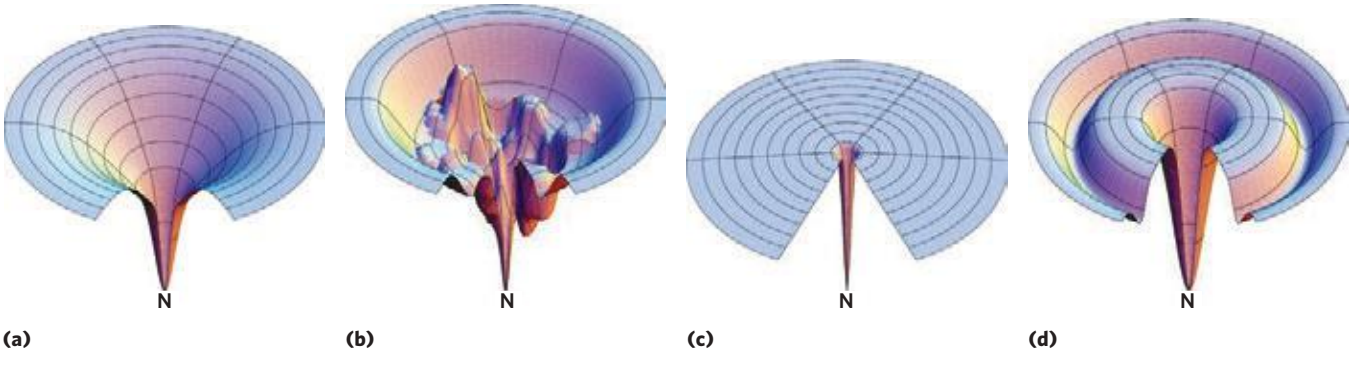


FIGURE 4–29 The thermodynamics of protein folding depicted as free-energy funnels. As proteins fold, the conformational space that can be explored by the structure is constrained. This is modeled as a three-dimensional thermodynamic funnel, with ΔG represented as depth and with the native structure (N) at the bottom (lowest free-energy point) of the funnel. The funnel for a given protein can have a variety of shapes, depending on the number and types of folding intermediates in the folding pathways. Any folding intermediate with significant stability and a finite lifetime would be represented as a local free-energy minimum—a depression on the surface of the funnel. **(a)** A simple but relatively wide and smooth funnel represents a protein that has multiple folding pathways (that is, the order in which different parts of the protein fold would be somewhat random), but it assumes its three-dimensional structure with no folding intermediates that have significant stability. **(b)** This funnel represents a more typical protein that has multiple possible folding intermediates with significant stability on the multiple pathways leading to the native structure. **(c)** A protein with one stable native structure, essentially no other folded intermediates with significant stability, and only one or a very few productive folding pathways is shown as a funnel with one narrow depression leading to the native form. **(d)** A protein with folding intermediates of substantial stability on virtually every pathway leading to the native state (that is, a protein in which a particular motif or domain always folds quickly, but other parts of the protein fold more slowly and in a random order) is depicted by a funnel with a major depression surrounding the depression leading to the native form.

cells stressed by elevated temperatures (hence, *h*eat *s*hock *p*roteins of M_r 70,000, or Hsp70). Hsp70 proteins bind to regions of unfolded polypeptides that are rich in hydrophobic residues. These chaperones thus "protect" both proteins subject to denaturation by heat and new peptide molecules being synthesized (and not yet folded). Hsp70 proteins also block the folding of certain proteins that must remain unfolded until they have been translocated across a membrane (as described in Chapter 27). Some chaperones also facilitate the quaternary assembly of oligomeric proteins. The Hsp70 proteins bind to and release polypeptides in a cycle that uses energy from ATP hydrolysis and involves several other proteins (including a class called Hsp40). **Figure 4–30** illustrates chaperone-assisted folding as elucidated for the eukaryotic Hsp70 and Hsp40 chaperones. The binding of an unfolded polypeptide by an Hsp70 chaperone may break up a protein aggregate or prevent the formation of a new one. When the bound polypeptide is released, it has a chance to resume folding to its native structure. If folding does not occur rapidly enough, the polypeptide may be bound again and the process repeated. Alternatively, the Hsp70-bound polypeptide may be delivered to a chaperonin.

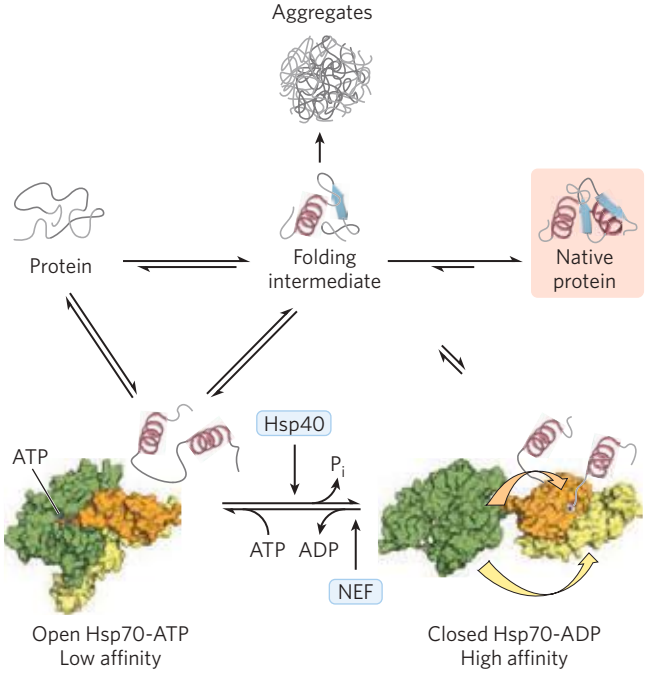


FIGURE 4–30 Chaperones in protein folding. The pathway by which Hsp70-class chaperones bind and release polypeptides is illustrated for the eukaryotic chaperones Hsp70 and Hsp40. The chaperones do not actively promote the folding of the substrate protein, but instead prevent aggregation of unfolded peptides. The unfolded or partly folded proteins bind first to the open, ATP-bound form of Hsp70 (PDB ID 2QXL). Hsp40 then interacts with this complex and triggers ATP hydrolysis that produces the closed form of the complex (derived from PDB IDs 2KHO and 1DKZ), where the domains colored orange and yellow come together like the two parts of a jaw, trapping parts of the unfolded protein inside. Dissociation of ADP and recycling of the Hsp70 requires interaction with another protein, nucleotide-exchange factor (NEF). For a population of polypeptide molecules, some fraction of the molecules released after the transient binding of partially folded proteins by Hsp70 will take up the native conformation. The remainder are rebound by Hsp70 or diverted to the chaperonin system (Hsp60; see Fig. 4–31). In bacteria, the Hsp70 and Hsp40 chaperones are called DnaK and DnaJ, respectively. DnaK and DnaJ were first identified as proteins required for in vitro replication of certain viral DNA molecules (hence the "Dna" designation).

Chaperonins are elaborate protein complexes required for the folding of some cellular proteins that do not fold spontaneously. In *E. coli*, an estimated 10% to 15% of cellular proteins require the resident chaperonin system, called GroEL/GroES, for folding under normal conditions (up to 30% require this assistance when the cells are heat stressed). The analogous chaperonin system in eukaryotes is called Hsp60. The chaperonins first became known when they were found to be necessary for the growth of certain bacterial viruses (hence the designation "Gro"). This family of proteins is structured as a series of multisubunit rings, forming two chambers oriented back to back. An unfolded protein is first bound to an exposed hydrophobic surface near the apical end of one GroEL chamber. The protein is then trapped within the chamber when it is capped transiently by the GroES "lid" **(Fig. 4–31)**. GroEL undergoes substantial conformational changes, coupled to slow ATP hydrolysis, which also regulates the binding and release of GroES. Inside the chamber, a protein has about 10 seconds to fold—the time required for the bound ATP to hydrolyze. Constraining a protein within the chamber prevents inappropriate protein aggregation and also restricts the conformational space that a polypeptide chain can explore as it folds. The protein is released when the GroES cap dissociates but can rebound rapidly for another round if folding has not been completed. The two chambers in a GroEL complex alternate in binding and releasing unfolded polypeptide substrates. In eukaryotes, the Hsp60 system utilizes a similar process to fold proteins. However, in place of the GroES lid, protrusions from the apical domains of the subunits flex and close over the chamber. The ATP hydrolytic cycle is also slower in the Hsp60 complexes, giving the proteins constrained inside more time to fold.

Finally, the folding pathways of some proteins require two enzymes that catalyze isomerization reactions. **Protein disulfide isomerase (PDI)** is a widely distributed enzyme that catalyzes the interchange, or shuffling, of disulfide bonds until the bonds of the native conformation are formed. Among its functions, PDI catalyzes the elimination of folding intermediates with inappropriate disulfide cross-links. **Peptide prolyl cis-trans isomerase (PPI)** catalyzes the interconversion of the cis and trans isomers of Pro residue peptide bonds (Fig. 4–8), which can be a slow step in the folding of proteins that contain some Pro peptide bonds in the cis conformation.

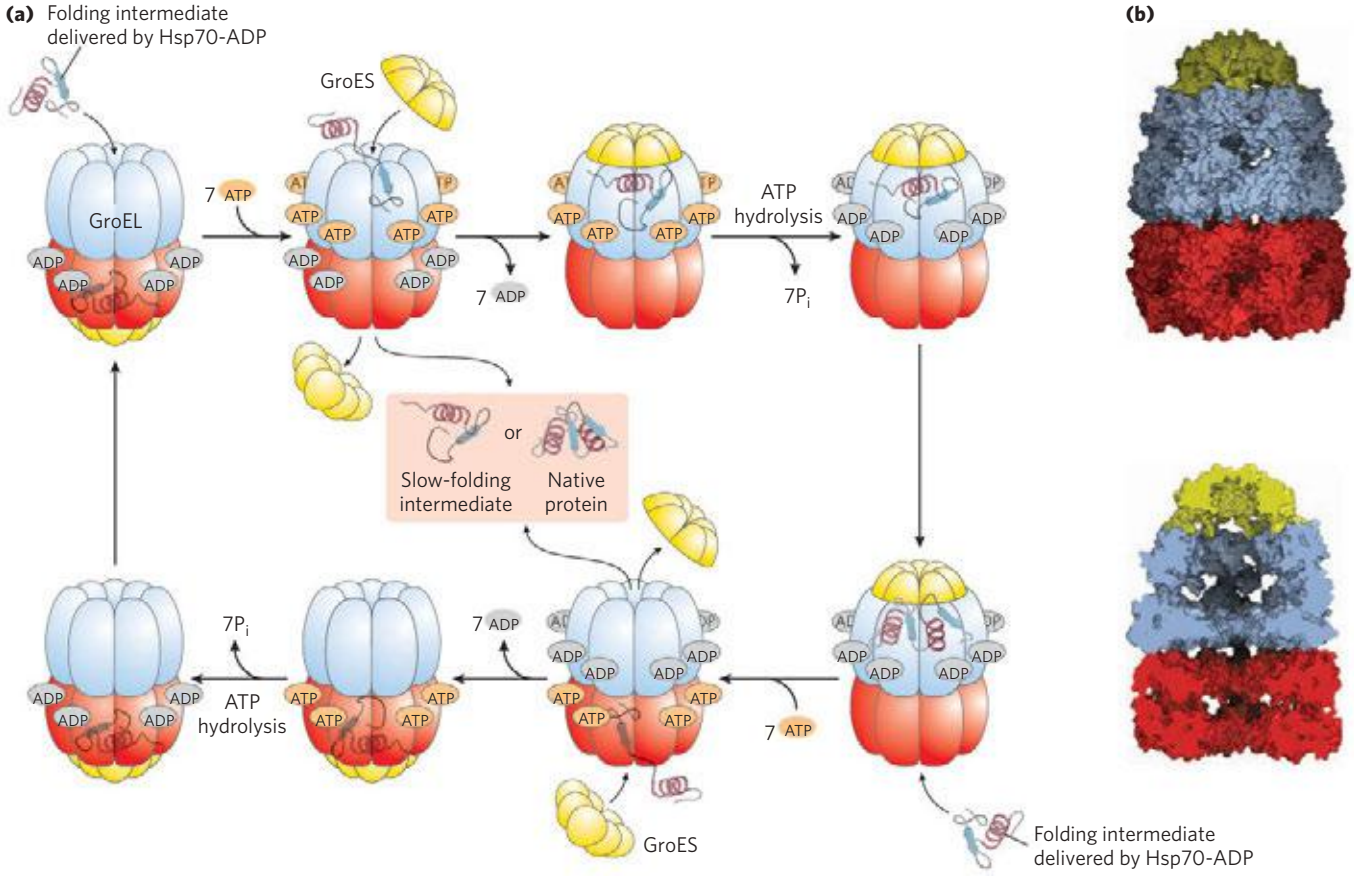


FIGURE 4–31 Chaperonins in protein folding. (a) A proposed pathway for the action of the *E. coli* chaperonins GroEL (a member of the Hsp60 protein family) and GroES. Each GroEL complex consists of two large chambers formed by two heptameric rings (each subunit M_r 57,000). GroES, also a heptamer (subunit M_r 10,000), blocks one of the GroEL chambers after an unfolded protein is bound inside. The chamber with the unfolded protein is referred to as cis; the opposite one is trans. Folding occurs within the cis chamber, during the time it takes to hydrolyze the 7 ATP bound to the subunits in the heptameric ring. The GroES and the ADP molecules then dissociate, and the protein is released. The two chambers of the GroEL/Hsp60 systems alternate in the binding and facilitated folding of client proteins. **(b)** Surface and cutaway images of the GroEL/GroES complex (PDB ID 1AON). The cutaway (below) illustrates the large interior space within which other proteins are bound.

Defects in Protein Folding Provide the Molecular Basis for a Wide Range of Human Genetic Disorders

Despite the many processes that assist in protein folding, misfolding does occur. In fact, protein misfolding is a substantial problem in all cells, and a quarter or more of all polypeptides synthesized may be destroyed because they do not fold correctly. In some cases, the misfolding causes or contributes to the development of serious disease.

Many conditions, including type 2 diabetes, Alzheimer disease, Huntington disease, and Parkinson disease, are associated with a misfolding mechanism: a soluble protein that is normally secreted from the cell is secreted in a misfolded state and converted into an insoluble extracellular **amyloid** fiber. The diseases are collectively referred to as **amyloidoses**. The fibers are highly ordered and unbranched, with a diameter of 7 to 10 nm and a high degree of β-sheet structure. The β segments are oriented perpendicular to the axis of the fiber. In some amyloid fibers the overall structure features two layers of β sheet, such as that shown for amyloid-β peptide in **Figure 4–32**.

Many proteins can take on the amyloid fibril structure as an alternative to their normal folded conformations, and most of these proteins have a concentration of aromatic amino acid residues in a core region of β sheet or α helix. The proteins are secreted in an incompletely folded conformation. The core (or some part of it) folds into a β sheet before the rest of the protein folds correctly, and the β sheets from two or more incompletely folded protein molecules associate to begin forming an amyloid fibril. The fibril grows in the extracellular space. Other parts of the protein then fold differently, remaining on the outside of the β-sheet core in the growing fibril. The effect of aromatic residues in stabilizing the structure is shown in Figure 4–32c. Because most of the protein molecules fold normally, the onset of symptoms in the amyloidoses is often very slow. If a person inherits a mutation such as substitution with an aromatic residue at a position that favors formation of amyloid fibrils, disease symptoms may begin at an earlier age.

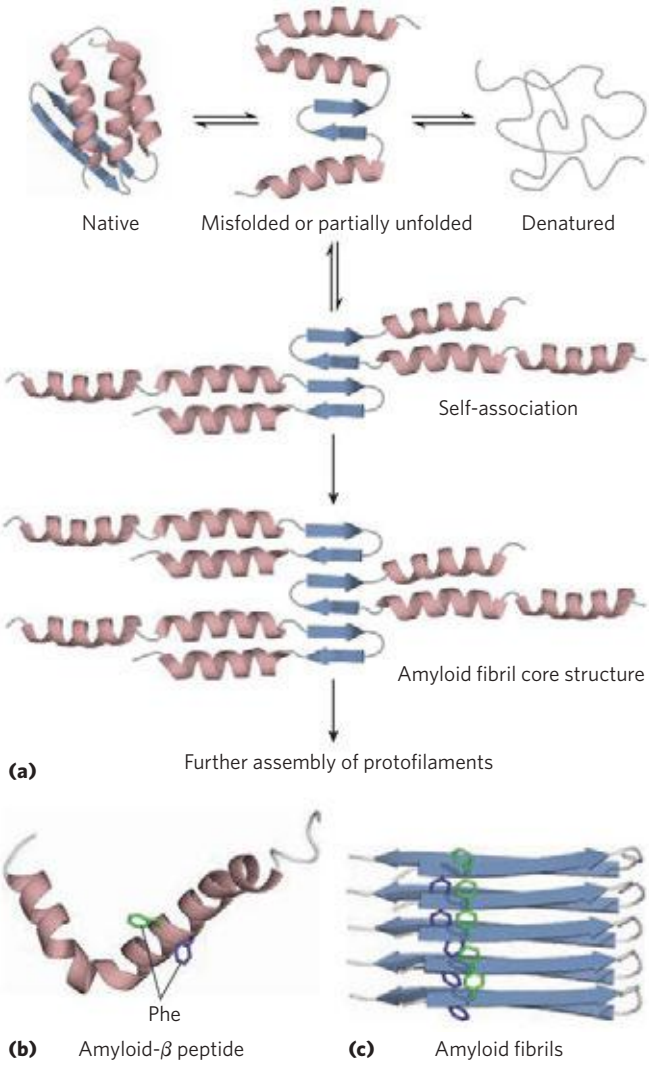


FIGURE 4–32 Formation of disease-causing amyloid fibrils. (a) Protein molecules whose normal structure includes regions of β sheet undergo partial folding. In a small number of the molecules, before folding is complete, the β-sheet regions of one polypeptide associate with the same region in another polypeptide, forming the nucleus of an amyloid. Additional protein molecules slowly associate with the amyloid and extend it to form a fibril. **(b)** The amyloid-β peptide begins as two α-helical segments of a larger protein. Proteolytic cleavage of this larger protein leaves the relatively unstable amyloid-β peptide, which loses its α-helical structure. It can then assemble slowly into amyloid fibrils **(c)**, which contribute to the characteristic plaques on the exterior of nervous tissue in people with Alzheimer disease. The aromatic side chains shown here play a significant role in stabilizing the amyloid structure. Amyloid is rich in β sheet, with the β strands arranged perpendicular to the axis of the amyloid fibril. Amyloid-β peptide takes the form of two layers of extended parallel β sheet. Some amyloid-forming peptides may fold to form left-handed β-helices (see Fig. 4–22).

In eukaryotes, proteins destined for secretion undergo their initial folding in the endoplasmic reticulum (ER; see pathway in Chapter 27). When stress conditions arise, or when protein synthesis threatens to overwhelm the protein-folding capacity of the ER, unfolded proteins can accumulate. These conditions trigger the unfolded protein response (UPR). A set of transcriptional regulators that constitute the UPR bring the various systems into alignment by increasing the concentration of chaperones in the ER or decreasing the rate of overall protein synthesis, or both. Amyloid aggregates that form before the UPR can come into play may be removed. Some are degraded by **autophagy**. In this process, they are first encapsulated in a membrane, then the contents of the resulting vesicle are degraded after the vesicle docks with a cytosolic lysosome. Alternatively, misfolded proteins can be degraded by a system of proteases called the ubiquitin-proteasome system (described in Chapter 27). Defects in any of these systems decrease the capacity to deal with misfolded proteins and increase the propensity for development of amyloid-related diseases.

Some amyloidoses are systemic, involving many tissues. Primary systemic amyloidosis is caused by deposition of fibrils consisting of misfolded immunoglobulin light chains (see Chapter 5), or fragments of light chains derived from proteolytic degradation. The mean age of onset is about 65 years. Patients have symptoms including fatigue, hoarseness, swelling, and weight loss, and many die within the first year after diagnosis. The kidneys or heart are often most affected. Some amyloidoses are associated with other types of disease. People with certain chronic infectious or inflammatory diseases such as rheumatoid arthritis, tuberculosis, cystic fibrosis, and some cancers can experience a sharp increase in secretion of an amyloid-prone polypeptide called serum amyloid A (SAA) protein. This protein, or fragments of it, deposits in the connective tissue of the spleen, kidney, and liver, and around the heart. People with this condition, known as secondary systemic amyloidosis, have a wide range of symptoms, depending on the organs initially affected. The disease is generally fatal within a few years. More than 80 amyloidoses are associated with mutations in transthyretin (a protein that binds to and transports thyroid hormones, distributing them throughout the body and brain). A variety of mutations in this protein lead to amyloid deposition concentrated around different tissues, thus producing different symptoms. Amyloidoses are also associated with inherited mutations in the proteins lysozyme, fibrinogen A α chain, and apolipoproteins A-I and A-II; all of these proteins are described in later chapters.

Some amyloid diseases are associated with particular organs. The amyloid-prone protein is generally secreted only by the affected tissue, and its locally high concentration leads to amyloid deposition around that tissue (although some of the protein may be distributed systemically). One common site of amyloid deposition is near the pancreatic islet β cells, responsible for insulin secretion and regulation of glucose metabolism (see Fig. 23–26). Secretion by β cells of a small (37 amino acid) peptide called islet amyloid polypeptide (IAPP), or amylin, can lead to amyloid deposits around the islets, gradually destroying the cells. A healthy human adult has 1 to 1.5 million

BOX 4–6 MEDICINE Death by Misfolding: The Prion Diseases

A misfolded brain protein seems to be the causative agent of several rare degenerative brain diseases in mammals. Perhaps the best known of these is bovine spongiform encephalopathy (BSE; also known as mad cow disease). Related diseases include kuru and Creutzfeldt-Jakob disease in humans, scrapie in sheep, and chronic wasting disease in deer and elk. These diseases are also referred to as spongiform encephalopathies, because the diseased brain frequently becomes riddled with holes (Fig. 1). Progressive deterioration of the brain leads to a spectrum of neurological symptoms, including weight loss, erratic behavior, problems with posture, balance, and coordination, and loss of cognitive function. The diseases are fatal.

In the 1960s, investigators found that preparations of the disease-causing agents seemed to lack nucleic acids. At this time, Tikvah Alper suggested that the agent was a protein. Initially, the idea seemed heretical. All disease-causing agents known up to that time—viruses, bacteria, fungi, and so on—contained nucleic acids, and their virulence was related to genetic reproduction and propagation. However, four decades of investigations, pursued most notably by Stanley Prusiner, have provided evidence that spongiform encephalopathies are different.

The infectious agent has been traced to a single protein (M_r 28,000), which Prusiner dubbed **prion** protein (PrP). The name was derived from *pro*teinaceous *in*fectious, but Prusiner thought that "prion" sounded better than "proin." Prion protein is a normal constituent of brain tissue in all mammals. Its role is not known in detail, but it may have a molecular signaling function. Strains of mice lacking the gene for PrP (and thus the protein itself) suffer no obvious ill effects. Illness occurs only when the normal cellular PrP, or PrP^{C}, occurs in an altered conformation called PrP^{Sc} (Sc denotes scrapie). The structure

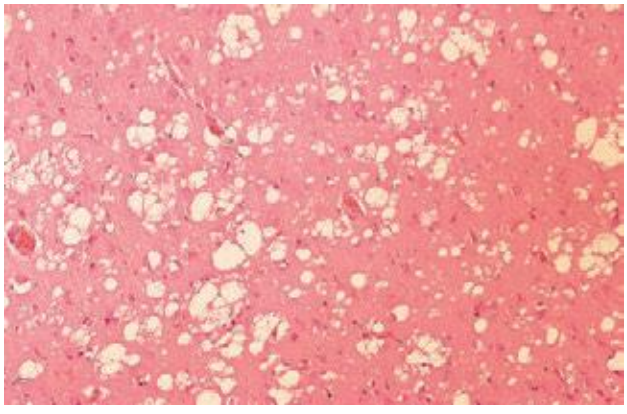


FIGURE 1 Stained section of cerebral cortex from autopsy of a patient with Creutzfeldt-Jakob disease shows spongiform (vacuolar) degeneration, the most characteristic neurohistological feature. The yellowish vacuoles are intracellular and occur mostly in pre- and postsynaptic processes of neurons. The vacuoles in this section vary in diameter from 20 to 100 μm.

pancreatic β cells. With progressive loss of these cells, glucose homeostasis is affected and eventually, when 50% or more of the cells are lost, the condition matures into type 2 (non–insulin-dependent) diabetes mellitus.

The amyloid deposition diseases that trigger neurodegeneration, particularly in older adults, are a special class of localized amyloidoses. Alzheimer disease is associated with extracellular amyloid deposition by neurons, involving the amyloid-β peptide (Fig. 4–32b), derived from a larger transmembrane protein (amyloid-β precursor protein) found in most human tissues. When it is part of the larger protein, the peptide is composed of two α-helical segments spanning the membrane. When the external and internal domains are cleaved off by dedicated proteases, the relatively unstable amyloid-β peptide leaves the membrane and loses its α-helical structure. It can then take the form of two layers of extended parallel β sheet, which can slowly assemble into amyloid fibrils (Fig. 4–32c). Deposits of these amyloid fibers seem to be the primary cause of Alzheimer disease, but a second type of amyloidlike aggregation, involving a protein called tau, also occurs intracellularly (in neurons) in people with Alzheimer disease. Inherited mutations in the tau protein do not result in Alzheimer disease, but they cause a frontotemporal dementia and parkinsonism (a condition with symptoms resembling Parkinson disease) that can be equally devastating.

Several other neurodegenerative conditions involve intracellular aggregation of misfolded proteins. In Parkinson disease, the misfolded form of the protein α-synuclein aggregates into spherical filamentous masses called Lewy bodies. Huntington disease involves the protein huntingtin, which has a long polyglutamine repeat. In some individuals, the polyglutamine repeat is longer than normal and a more subtle type of intracellular aggregation occurs. Notably, when the mutant human proteins involved in Parkinson and Huntington diseases are expressed in *Drosophila melanogaster*, the flies display neurodegeneration expressed as eye deterioration, tremors, and early death. All of these symptoms are highly suppressed if expression of the Hsp70 chaperone is also increased.

of PrP^C features two α helices. The structure of PrP^{Sc} is very different, with much of the structure converted to amyloid-like β sheets (Fig. 2). The interaction of PrP^{Sc} with PrP^C converts the latter to PrP^{Sc}, initiating a domino effect in which more and more of the brain protein converts to the disease-causing form. The mechanism by which the presence of PrP^{Sc} leads to spongiform encephalopathy is not understood.

In inherited forms of prion diseases, a mutation in the gene encoding PrP produces a change in one amino acid residue that is believed to make the conversion of PrP^C to PrP^{Sc} more likely. A complete understanding of prion diseases awaits new information on how prion protein affects brain function. Structural information about PrP is beginning to provide insights into the molecular process that allows the prion proteins to interact so as to alter their conformation (Fig. 2).

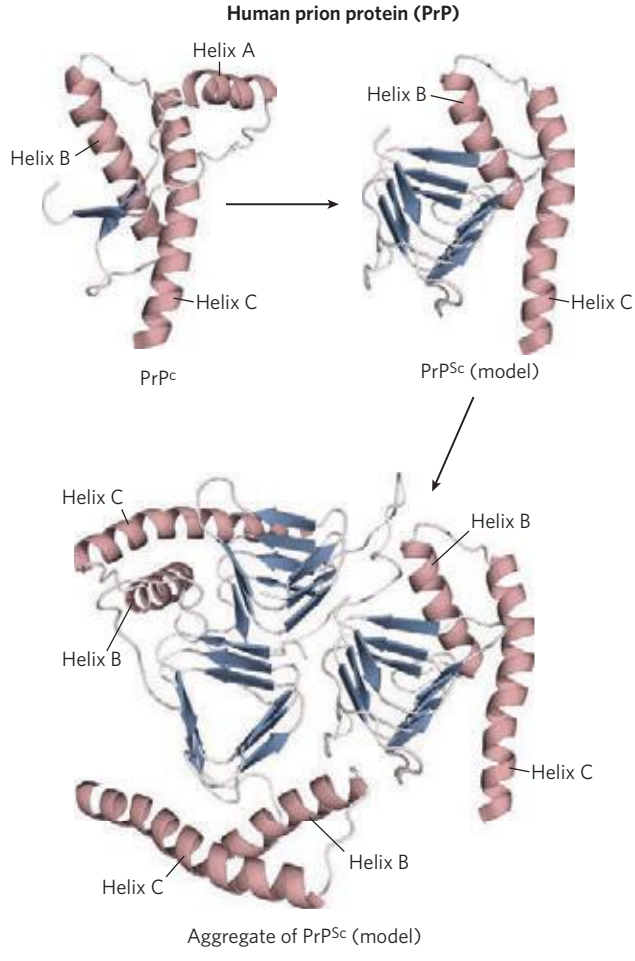


FIGURE 2 Structure of the globular domain of human PrP (PDB ID 1QLX) and models of the misfolded, disease-causing conformation PrP^{Sc}, and an aggregate of PrP^{Sc}. The α helices are labeled to help illustrate the conformation change. Helix A is incorporated into the β-sheet structure of the misfolded conformation.

Protein misfolding need not lead to amyloid formation to cause serious disease. For example, cystic fibrosis is caused by defects in a membrane-bound protein called *c*ystic *f*ibrosis *t*ransmembrane conductance *r*egulator (CFTR), which acts as a channel for chloride ions. The most common cystic fibrosis–causing mutation is the deletion of a Phe residue at position 508 in CFTR, which causes improper protein folding. Most of this protein is then degraded and its normal function is lost (see Box 11–2). Many of the disease-related mutations in collagen (p. 130) also cause defective folding. A particularly remarkable type of protein misfolding is seen in the prion diseases (Box 4–6). ■

SUMMARY 4.4 Protein Denaturation and Folding

- The maintenance of the steady-state collection of active cellular proteins required under a particular set of conditions—called proteostasis—involves an elaborate set of pathways and processes that fold, refold, and degrade polypeptide chains.
- The three-dimensional structure and the function of most proteins can be destroyed by denaturation, demonstrating a relationship between structure and function. Some denatured proteins can renature spontaneously to form biologically active protein, showing that tertiary structure is determined by amino acid sequence.
- Protein folding in cells is generally hierarchical. Initially, regions of secondary structure may form, followed by folding into motifs and domains. Large ensembles of folding intermediates are rapidly brought to a single native conformation.
- For many proteins, folding is facilitated by Hsp70 chaperones and by chaperonins. Disulfide bond formation and the cis-trans isomerization of Pro peptide bonds are catalyzed by specific enzymes.
- Protein misfolding is the molecular basis of a wide range of human diseases, including the amyloidoses.

Key Terms

Terms in bold are defined in the glossary.

conformation 115
native conformation 116
hydrophobic interactions 116
solvation layer 116
peptide group 118
Ramachandran plot 119
secondary structure 119
α helix 120
β conformation 123
β sheet 123
β turn 123
circular dichroism (CD) spectroscopy 124
tertiary structure 125
quaternary structure 125
fibrous proteins 125
globular proteins 125
α-keratin 126
collagen 127
silk fibroin 130
Protein Data Bank (PDB) 132
motif 137
fold 137
domain 137
protein family 140
multimer 140
oligomer 140
protomer 141
intrinsically disordered proteins 141
proteostasis 143
denaturation 143
renaturation 144
chaperone 146
Hsp70 146
chaperonin 146
protein disulfide isomerase (PDI) 147
peptide prolyl cis-trans isomerase (PPI) 147
amyloid 148
amyloidoses 148
autophagy 149
prion 150

Further Reading

General

Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science* **181**, 223–230.
The author reviews his classic work on ribonuclease.

Creighton, T.E. (1993) *Proteins: Structures and Molecular Properties,* 2nd edn, W. H. Freeman and Company, New York.
A comprehensive and authoritative source.

Kendrew, J.C. (1961) The three-dimensional structure of a protein molecule. *Sci. Am.* **205** (December), 96–111.
Describes how the structure of myoglobin was determined and what was learned from it.

Richardson, J.S. (1981) The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167–339.
An outstanding summary of protein structural patterns and principles; the author originated the very useful "ribbon" representations of protein structure.

Secondary, Tertiary, and Quaternary Structures

Beeby, M., O'Connor, B.D., Ryttersgaard, C., Boutz, D.R., Perry, L.J., & Yeates, T.O. (2005) The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol.* **3**, e309.

Brown, J.H. (2006) Breaking symmetry in protein dimers: designs and function. *Protein Sci.* **15**, 1–13.

Dunker, A.K. & Kriwacki, R.W. (2011) The orderly chaos of proteins. *Sci. Am.* **304** (April), 68–73.
A nice summary of the work on proteins that lack intrinsic structure.

Herráez, A. (2006) Biomolecules in the computer. *Biochem. Mol. Biol. Educ.* **34**, 255–261.

McPherson, A. (1989) Macromolecular crystals. *Sci. Am.* **260** (March), 62–69.
A description of how macromolecules such as proteins are crystallized.

Milner-White, E.J. (1997) The partial charge of the nitrogen atom in peptide bonds. *Protein Sci.* **6**, 2477–2482.

Ponting, C.P. & Russell, R.R. (2002) The natural history of protein domains. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 45–71.
An explanation of how structural databases can be used to explore evolution.

Protein Denaturation and Folding

Chiti, F. & Dobson, C.M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366.

Dill, K.A., Ozkan, S.B., Shell, M.S., & Weikl, T.R. (2008) The protein folding problem. *Annu. Rev. Biophys.* **37**, 289–316.

Gazit, E. (2005) Mechanisms of amyloid fibril self-assembly and inhibition. *FEBS J.* **272**, 5971–5978.

Hartl, F.U., Bracher, A., & Hayer-Hartl, M. (2011) Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332.

Hoppener, J.W.M. & Lips, C.J.M. (2006) Role of islet amyloid in type 2 diabetes mellitus. *Int. J. Biochem. Cell Biol.* **38**, 726–736.

Kapinga, H.H. & Craig, E.A. (2010) The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat. Rev. Mol. Cell. Biol.* **11**, 579–592.

Norrby, E. (2011) Prions and protein-folding diseases. *J. Intern. Med.* **270**, 1–14.

Prusiner, S.B. (1995) The prion diseases. *Sci. Am.* **272** (January), 48–57.
A good summary of the evidence leading to the prion hypothesis.

Selkoe, D.J. (2003) Folding proteins in fatal ways. *Nature* **426**, 900–904.
A good summary of amyloidoses.

Tang, Y., Chang, H., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M., Hartl, F., & Hayer-Hartl, M. (2006) Structural features of the GroEL-GroES nanocage required for rapid folding of encapsulated protein. *Cell* **125**, 903–914.

Tyedmers, J., Mogk, A., & Bukau, B. (2010) Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.* **11**, 777–788.

Problems

1. Properties of the Peptide Bond In x-ray studies of crystalline peptides, Linus Pauling and Robert Corey found that the C—N bond in the peptide link is intermediate in length (1.32 Å) between a typical C—N single bond (1.49 Å) and a C═N double bond (1.27 Å). They also found that the peptide bond is planar (all four atoms attached to the C—N group are located in the same plane) and that the two α-carbon atoms attached to the C—N are always trans to each other (on opposite sides of the peptide bond).

(a) What does the length of the C—N bond in the peptide linkage indicate about its strength and its bond order (i.e., whether it is single, double, or triple)?

(b) What do the observations of Pauling and Corey tell us about the ease of rotation about the C—N peptide bond?

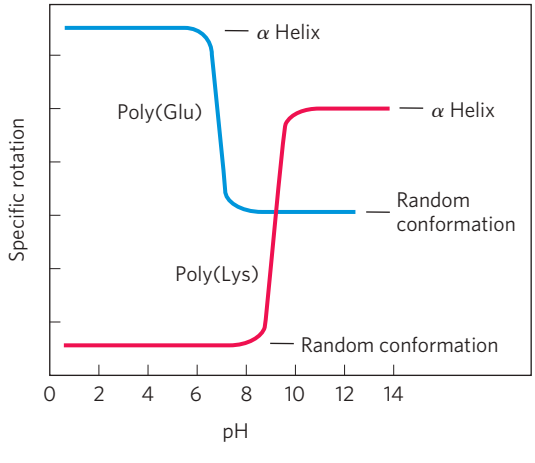
2. Structural and Functional Relationships in Fibrous Proteins William Astbury discovered that the x-ray diffraction pattern of wool shows a repeating structural unit spaced about 5.2 Å along the length of the wool fiber. When he steamed and stretched the wool, the x-ray pattern showed a new repeating structural unit at a spacing of 7.0 Å. Steaming and stretching the wool and then letting it shrink gave an x-ray pattern consistent with the original spacing of about 5.2 Å. Although these observations provided important clues to the molecular structure of wool, Astbury was unable to interpret them at the time.

(a) Given our current understanding of the structure of wool, interpret Astbury's observations.

(b) When wool sweaters or socks are washed in hot water or heated in a dryer, they shrink. Silk, on the other hand, does not shrink under the same conditions. Explain.

3. Rate of Synthesis of Hair α-Keratin Hair grows at a rate of 15 to 20 cm/yr. All this growth is concentrated at the base of the hair fiber, where α-keratin filaments are synthesized inside living epidermal cells and assembled into ropelike structures (see Fig. 4–11). The fundamental structural element of α-keratin is the α helix, which has 3.6 amino acid residues per turn and a rise of 5.4 Å per turn (see Fig. 4–4a). Assuming that the biosynthesis of α-helical keratin chains is the rate-limiting factor in the growth of hair, calculate the rate at which peptide bonds of α-keratin chains must be synthesized (peptide bonds per second) to account for the observed yearly growth of hair.

4. Effect of pH on the Conformation of α-Helical Secondary Structures The unfolding of the α helix of a polypeptide to a randomly coiled conformation is accompanied by a large decrease in a property called specific rotation, a measure of a solution's capacity to rotate circularly polarized light. Polyglutamate, a polypeptide made up of only L-Glu residues, has the α-helical conformation at pH 3. When the pH is raised to 7, there is a large decrease in the specific rotation of the solution. Similarly, polylysine (L-Lys residues) is an α helix at pH 10, but when the pH is lowered to 7 the specific rotation also decreases, as shown by the following graph.



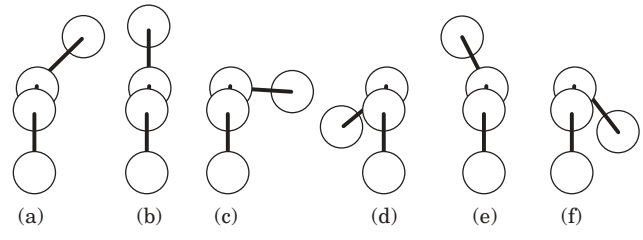
What is the explanation for the effect of the pH changes on the conformations of poly(Glu) and poly(Lys)? Why does the transition occur over such a narrow range of pH?

5. Disulfide Bonds Determine the Properties of Many Proteins Some natural proteins are rich in disulfide bonds, and their mechanical properties (tensile strength, viscosity, hardness, etc.) are correlated with the degree of disulfide bonding.

(a) Glutenin, a wheat protein rich in disulfide bonds, is responsible for the cohesive and elastic character of dough made from wheat flour. Similarly, the hard, tough nature of tortoise shell is due to the extensive disulfide bonding in its α-keratin. What is the molecular basis for the correlation between disulfide-bond content and mechanical properties of the protein?

(b) Most globular proteins are denatured and lose their activity when briefly heated to 65 °C. However, globular proteins that contain multiple disulfide bonds often must be heated longer at higher temperatures to denature them. One such protein is bovine pancreatic trypsin inhibitor (BPTI), which has 58 amino acid residues in a single chain and contains three disulfide bonds. On cooling a solution of denatured BPTI, the activity of the protein is restored. What is the molecular basis for this property?

6. Dihedral Angles A series of torsion angles, ϕ and ψ, that might be taken up by the peptide backbone is shown below. Which of these closely correspond to ϕ and ψ for an idealized collagen triple helix? Refer to Figure 4–9 as a guide.



7. Amino Acid Sequence and Protein Structure Our growing understanding of how proteins fold allows researchers to make predictions about protein structure based on primary amino acid sequence data. Consider the following amino acid sequence.

1	2	3	4	5	6	7	8	9	10
Ile–	Ala–	His–	Thr–	Tyr–	Gly–	Pro–	Phe–	Glu–	Ala–
11	12	13	14	15	16	17	18	19	20
Ala–	Met–	Cys–	Lys–	Trp–	Glu–	Ala–	Gln–	Pro–	Asp–
21	22	23	24	25	26	27	28		
Gly–	Met–	Glu–	Cys–	Ala–	Phe–	His–	Arg		

(a) Where might bends or β turns occur?

(b) Where might intrachain disulfide cross-linkages be formed?

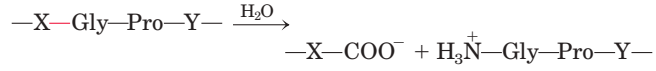
(c) Assuming that this sequence is part of a larger globular protein, indicate the probable location (the external surface or interior of the protein) of the following amino acid residues: Asp, Ile, Thr, Ala, Gln, Lys. Explain your reasoning. (Hint: See the hydropathy index in Table 3–1.)

8. Bacteriorhodopsin in Purple Membrane Proteins Under the proper environmental conditions, the salt-loving archaeon *Halobacterium halobium* synthesizes a membrane protein (M_r 26,000) known as bacteriorhodopsin, which is purple because it contains retinal (see Fig. 10–21). Molecules of this protein aggregate into "purple patches" in the cell membrane. Bacteriorhodopsin acts as a light-activated proton pump that provides energy for cell functions. X-ray analysis of this protein reveals that it consists of seven parallel α-helical segments, each of which traverses the bacterial cell membrane (thickness 45 Å). Calculate the minimum number of amino acid residues necessary for one segment of α helix to traverse the membrane completely. Estimate the fraction of the bacteriorhodopsin protein that is involved in membrane-spanning helices. (Use an average amino acid residue weight of 110.)

9. Protein Structure Terminology Is myoglobin a motif, a domain, or a complete three-dimensional structure?

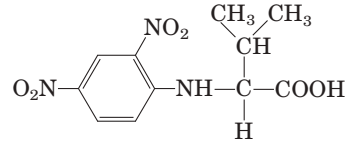
10. Interpreting Ramachandran Plots Examine the two proteins labeled (a) and (b) below. Which of the two Ramachandran plots, labeled (c) and (d), is more likely to be derived from which protein? Why?

11. Pathogenic Action of Bacteria That Cause Gas Gangrene The highly pathogenic anaerobic bacterium *Clostridium perfringens* is responsible for gas gangrene, a condition in which animal tissue structure is destroyed. This bacterium secretes an enzyme that efficiently catalyzes the hydrolysis of the peptide bond indicated in red:

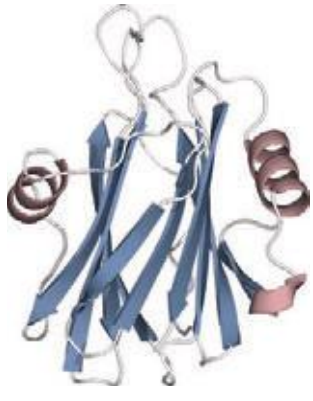


where X and Y are any of the 20 common amino acids. How does the secretion of this enzyme contribute to the invasiveness of this bacterium in human tissues? Why does this enzyme not affect the bacterium itself?

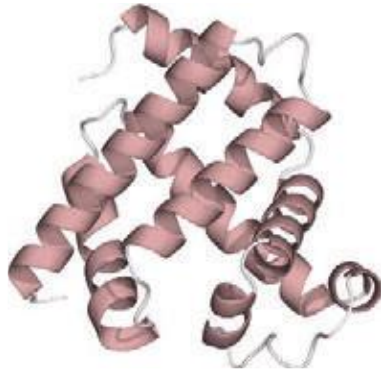
12. Number of Polypeptide Chains in a Multisubunit Protein A sample (660 mg) of an oligomeric protein of M_r 132,000 was treated with an excess of 1-fluoro-2,4-dinitrobenzene (Sanger's reagent) under slightly alkaline conditions until the chemical reaction was complete. The peptide bonds of the protein were then completely hydrolyzed by heating it with concentrated HCl. The hydrolysate was found to contain 5.5 mg of the following compound:



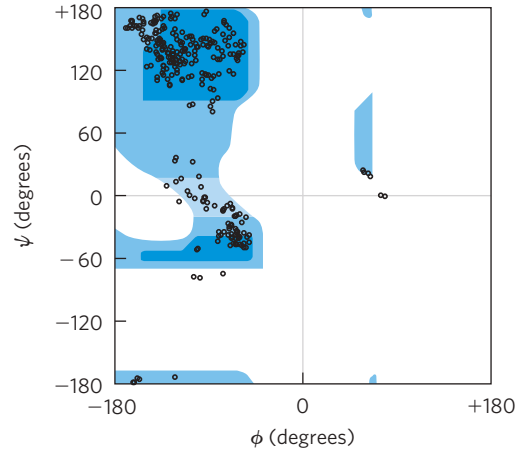
(a)



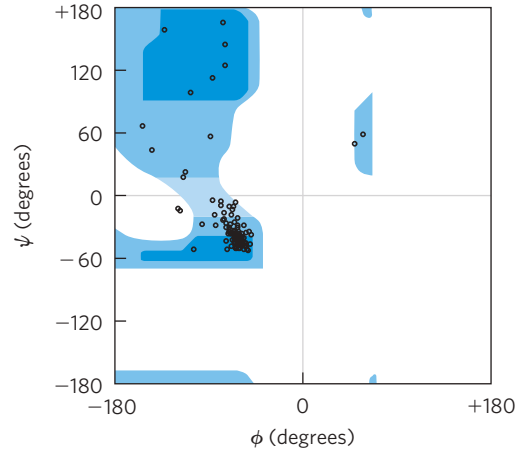
(b)



(c)



(d)



2,4-Dinitrophenyl derivatives of the α-amino groups of other amino acids could not be found.

(a) Explain how this information can be used to determine the number of polypeptide chains in an oligomeric protein.

(b) Calculate the number of polypeptide chains in this protein.

(c) What other protein analysis technique could you employ to determine whether the polypeptide chains in this protein are similar or different?

13. Predicting Secondary Structure Which of the following peptides is more likely to take up an α-helical structure, and why?

(a) LKAENDEAARAMSEA

(b) CRAGGFPWDQPGTSN

14. Amyloid Fibers in Disease Several small aromatic molecules, such as phenol red (used as a nontoxic drug model), have been shown to inhibit the formation of amyloid in laboratory model systems. A goal of the research on these small aromatic compounds is to find a drug that would efficiently inhibit the formation of amyloid in the brain in people with incipient Alzheimer disease.

(a) Suggest why molecules with aromatic substituents would disrupt the formation of amyloid.

(b) Some researchers have suggested that a drug used to treat Alzheimer disease may also be effective in treating type 2 (non–insulin-dependent) diabetes mellitus. Why might a single drug be effective in treating these two different conditions?

Using the Web

15. Protein Modeling on the Internet A group of patients with Crohn disease (an inflammatory bowel disease) underwent biopsies of their intestinal mucosa in an attempt to identify the causative agent. Researchers identified a protein that was present at higher levels in patients with Crohn disease than in patients with an unrelated inflammatory bowel disease or in unaffected controls. The protein was isolated, and the following *partial* amino acid sequence was obtained (reads left to right):

EAELCPDRCI	HSFQNLGIQC	VKKRDLEQAI
SQRIQTNNNP	FQVPIEEQRG	DYDLNAVRLC
FQVTVRDPSG	RPLRLPPVLP	HPIFDNRAPN
TAELKICRVN	RNSGSCLGGD	EIFLLCDKVQ
KEDIEVYFTG	PGWEARGSFS	QADVHRQVAI
VFRTPPYADP	SLQAPVRVSM	QLRRPSDREL
SEPMEFQYLP	DTDDRHRIEE	KRKRTYETFK
SIMKKSPFSG	PTDPRPPPRR	IAVPSRSSAS
VPKPAPQPYP		

(a) You can identify this protein using a protein database on the Internet. Some good places to start include Protein Information Resource (PIR; http://pir.georgetown.edu), Structural Classification of Proteins (SCOP; http://scop.mrc-lmb.cam.ac.uk/scop), and Prosite (http://prosite.expasy.org).

At your selected database site, follow links to the sequence comparison engine. Enter about 30 residues from the protein sequence in the appropriate search field and submit it for analysis. What does this analysis tell you about the identity of the protein?

(b) Try using different portions of the amino acid sequence. Do you always get the same result?

(c) A variety of websites provide information about the three-dimensional structure of proteins. Find information about the protein's secondary, tertiary, and quaternary structure using database sites such as the Protein Data Bank (PDB; www.pdb.org) or SCOP.

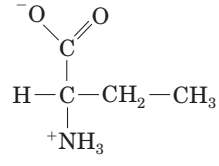
(d) In the course of your Web searches, what did you learn about the cellular function of the protein?

Data Analysis Problem

16. Mirror-Image Proteins As noted in Chapter 3, "The amino acid residues in protein molecules are exclusively L stereoisomers." It is not clear whether this selectivity is necessary for proper protein function or is an accident of evolution. To explore this question, Milton and colleagues (1992) published a study of an enzyme made entirely of D stereoisomers. The enzyme they chose was HIV protease, a proteolytic enzyme made by HIV that converts inactive viral preproteins to their active forms.

Previously, Wlodawer and coworkers (1989) had reported the complete chemical synthesis of HIV protease from L-amino acids (the L-enzyme), using the process shown in Figure 3–32. Normal HIV protease contains two Cys residues at positions 67 and 95. Because chemical synthesis of proteins containing Cys is technically difficult, Wlodawer and colleagues substituted the synthetic amino acid L-α-amino-n-butyric acid (Aba) for the two Cys residues in the protein. In the authors' words, this was done to "reduce synthetic difficulties associated with Cys deprotection and ease product handling."

(a) The structure of Aba is shown below. Why was this a suitable substitution for a Cys residue? Under what circumstances would it not be suitable?



L-α-Amino-n-butyric acid

Wlodawer and coworkers denatured the newly synthesized protein by dissolving it in 6 M guanidine HCl and then allowed it to fold slowly by dialyzing away the guanidine against a neutral buffer (10% glycerol, 25 mM $NaPO_4$, pH 7).

(b) There are many reasons to predict that a protein synthesized, denatured, and folded in this manner would not be active. Give three such reasons.

(c) Interestingly, the resulting L-protease was active. What does this finding tell you about the role of disulfide bonds in the native HIV protease molecule?

In their new study, Milton and coworkers synthesized HIV protease from D-amino acids, using the same protocol as the earlier study (Wlodawer et al.). Formally, there are three possibilities for the folding of the D-protease: it would give (1) the

same shape as the L-protease, (2) the mirror image of the L-protease, or (3) something else, possibly inactive.

(d) For each possibility, decide whether or not it is a likely outcome and defend your position.

In fact, the D-protease was active: it cleaved a particular synthetic substrate and was inhibited by specific inhibitors. To examine the structure of the D- and L-enzymes, Milton and coworkers tested both forms for activity with D and L forms of a chiral peptide substrate and for inhibition by D and L forms of a chiral peptide-analog inhibitor. Both forms were also tested for inhibition by the achiral inhibitor Evans blue. The findings are given in the table.

			Inhibition		
	Substrate hydrolysis		Peptide inhibitor		Evans blue
HIV Protease	D-substrate	L-substrate	D-inhibitor	L-inhibitor	(achiral)
L-protease	−	+	−	+	+
D-protease	+	−	+	−	+

(e) Which of the three models proposed above is supported by these data? Explain your reasoning.

(f) Why does Evans blue inhibit both forms of the protease?

(g) Would you expect chymotrypsin to digest the D-protease? Explain your reasoning.

(h) Would you expect total synthesis from D-amino acids followed by renaturation to yield active enzyme for any enzyme? Explain your reasoning.

References

Milton, R.C., Milton, S.C., & Kent, S.B. (1992) Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show demonstration of reciprocal chiral substrate specificity. *Science* **256**, 1445–1448.

Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., & Kent, S.B. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **245**, 616–621.

5

Protein Function

Knowing the three-dimensional structure of a protein is an important part of understanding how the protein functions. However, the structure shown in two dimensions on a page is deceptively static. Proteins are dynamic molecules whose functions almost invariably depend on interactions with other molecules, and these interactions are affected in physiologically important ways by sometimes subtle, sometimes striking changes in protein conformation. In this chapter, we explore how proteins interact with other molecules and how their interactions are related to dynamic protein structure. The importance of molecular interactions to a protein's function can hardly be overemphasized. In Chapter 4, we saw that the function of fibrous proteins as structural elements of cells and tissues depends on stable, long-term quaternary interactions between identical polypeptide chains. As we shall see in this chapter, the functions of many other proteins involve interactions with a variety of different molecules. Most of these interactions are fleeting, though they may be the basis of complex physiological processes such as oxygen transport, immune function, and muscle contraction—the topics we examine in this chapter. The proteins that carry out these processes illustrate the following key principles of protein function, some of which will be familiar from the previous chapter:

> The functions of many proteins involve the reversible binding of other molecules. A molecule bound reversibly by a protein is called a **ligand**. A ligand may be any kind of molecule, including another protein. The transient nature of protein-ligand interactions is critical to life, allowing an organism to respond rapidly and reversibly to changing environmental and metabolic circumstances.

A ligand binds at a site on the protein called the **binding site**, which is complementary to the ligand in size, shape, charge, and hydrophobic or hydrophilic character. Furthermore, the interaction is specific: the protein can discriminate among the thousands of different molecules in its environment and selectively bind only one or a few. A given protein may have separate binding sites for several different ligands. These specific molecular interactions are crucial in maintaining the high degree of order in a living system. (This discussion excludes the binding of water, which may interact weakly and nonspecifically with many parts of a protein. In Chapter 6, we consider water as a specific ligand for many enzymes.)

Proteins are flexible. Changes in conformation may be subtle, reflecting molecular vibrations and small movements of amino acid residues throughout the protein. A protein flexing in this way is sometimes said to "breathe." Changes in conformation may also be quite dramatic, with major segments of the protein structure moving as much as several nanometers. Specific conformational changes are frequently essential to a protein's function.

The binding of a protein and ligand is often coupled to a conformational change in the protein that makes the binding site more complementary to the ligand, permitting tighter binding. The structural adaptation that occurs between protein and ligand is called **induced fit**.

In a multisubunit protein, a conformational change in one subunit often affects the conformation of other subunits.

Interactions between ligands and proteins may be regulated, usually through specific interactions with one or more additional ligands. These other ligands may cause conformational changes in the protein that affect the binding of the first ligand.

Enzymes represent a special case of protein function. Enzymes bind and chemically transform other molecules—they catalyze reactions. The molecules acted upon by enzymes are called reaction **substrates** rather than ligands, and the ligand-binding site is called the **catalytic site** or **active site**. In this chapter we emphasize the noncatalytic functions of proteins. In Chapter 6 we consider catalysis by enzymes, a central topic in biochemistry. You will see that the themes of this chapter—binding, specificity, and conformational change—are continued in the next chapter, with the added element of proteins participating in chemical transformations.

5.1 Reversible Binding of a Protein to a Ligand: Oxygen-Binding Proteins

Myoglobin and hemoglobin may be the most-studied and best-understood proteins. They were the first proteins for which three-dimensional structures were determined, and these two molecules illustrate almost every aspect of that most central of biochemical processes: the reversible binding of a ligand to a protein. This classic model of protein function tells us a great deal about how proteins work. **Oxygen-Binding Proteins–Myoglobin: Oxygen Storage**

Oxygen Can Bind to a Heme Prosthetic Group

Oxygen is poorly soluble in aqueous solutions (see Table 2–3) and cannot be carried to tissues in sufficient quantity if it is simply dissolved in blood serum. Diffusion of oxygen through tissues is also ineffective over distances greater than a few millimeters. The evolution of larger, multicellular animals depended on the evolution of proteins that could transport and store oxygen. However, none of the amino acid side chains in proteins are suited for the reversible binding of oxygen molecules. This role is filled by certain transition metals, among them iron and copper, that have a strong tendency to bind oxygen. Multicellular organisms exploit the properties of metals, most commonly iron, for oxygen transport. However, free iron promotes the formation of highly reactive oxygen species such as hydroxyl radicals that can damage DNA and other macromolecules. Iron used in cells is therefore bound in forms that sequester it and/or make it less reactive. In multicellular organisms—especially those in which iron, in its oxygen-carrying capacity, must be transported over large distances—iron is often incorporated into a protein-bound prosthetic group called **heme** (or haem). (Recall from Chapter 3 that a prosthetic group is a compound permanently associated with a protein that contributes to the protein's function.)

Heme consists of a complex organic ring structure, **protoporphyrin**, to which is bound a single iron atom in its ferrous (Fe^{2+}) state **(Fig. 5–1)**. The iron atom has six coordination bonds, four to nitrogen atoms that are part of the flat **porphyrin ring** system and two perpendicular to the porphyrin. The coordinated nitrogen atoms (which have an electron-donating character) help prevent conversion of the heme iron to the ferric (Fe^{3+}) state. Iron in the Fe^{2+} state binds oxygen reversibly; in the Fe^{3+} state it does not bind oxygen. Heme is found in many oxygen-transporting proteins, as well as in some proteins, such as the cytochromes, that participate in oxidation-reduction (electron-transfer) reactions (Chapter 19).

Free heme molecules (heme not bound to protein) leave Fe^{2+} with two "open" coordination bonds. Simultaneous reaction of one O_2 molecule with two free heme molecules (or two free Fe^{2+}) can result in irreversible conversion of Fe^{2+} to Fe^{3+}. In heme-containing proteins, this reaction is prevented by sequestering each

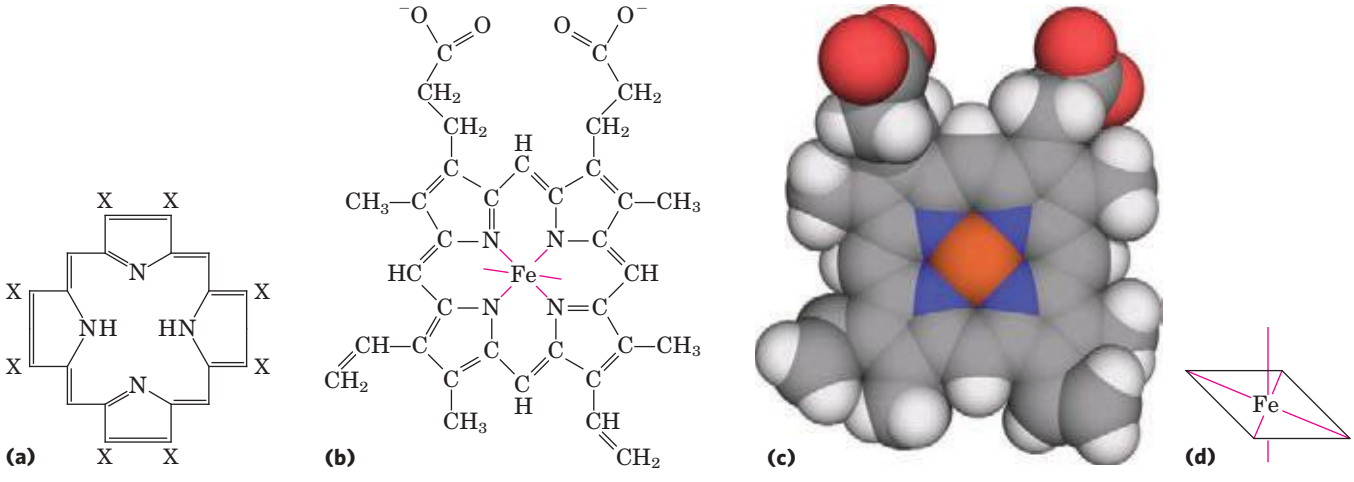


FIGURE 5–1 Heme. The heme group is present in myoglobin, hemoglobin, and many other proteins, designated heme proteins. Heme consists of a complex organic ring structure, protoporphyrin IX, with a bound iron atom in its ferrous (Fe^{2+}) state. **(a)** Porphyrins, of which protoporphyrin IX is only one example, consist of four pyrrole rings linked by methene bridges, with substitutions at one or more of the positions denoted X. **(b, c)** Two representations of heme (derived from PDB ID 1CCR). The iron atom of heme has six coordination bonds: four in the plane of, and bonded to, the flat porphyrin ring system, and **(d)** two perpendicular to it.

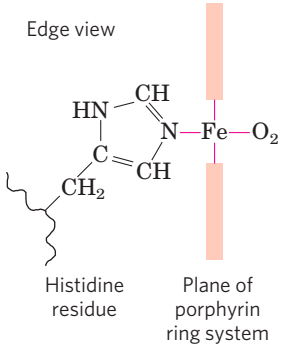


FIGURE 5–2 The heme group viewed from the side. This view shows the two coordination bonds to Fe^{2+} that are perpendicular to the porphyrin ring system. One is occupied by a His residue, sometimes called the proximal His; the other is the binding site for oxygen. The remaining four coordination bonds are in the plane of, and bonded to, the flat porphyrin ring system.

heme deep within the protein structure. Thus, access to the two open coordination bonds is restricted. One of these two coordination bonds is occupied by a side-chain nitrogen of a His residue. The other is the binding site for molecular oxygen (O_2) **(Fig. 5–2)**. When oxygen binds, the electronic properties of heme iron change; this accounts for the change in color from the dark purple of oxygen-depleted venous blood to the bright red of oxygen-rich arterial blood. Some small molecules, such as carbon monoxide (CO) and nitric oxide (NO), coordinate to heme iron with greater affinity than does O_2. When a molecule of CO is bound to heme, O_2 is excluded, which is why CO is highly toxic to aerobic organisms (a topic explored later, in Box 5–1). By surrounding and sequestering heme, oxygen-binding proteins regulate the access of CO and other small molecules to the heme iron.

Globins Are a Family of Oxygen-Binding Proteins

The **globins** are a widespread family of proteins, all having similar primary and tertiary structures. Globins are commonly found in eukaryotes of all classes and even in some bacteria. Most function in oxygen transport or storage, although some play a role in the sensing of oxygen, nitric oxide, or carbon monoxide. The simple nematode worm *Caenorhabditis elegans* has genes encoding 33 different globins. In humans and other mammals, there are at least four kinds of globins. The monomeric myoglobin facilitates oxygen diffusion in muscle tissue. Myoglobin is particularly abundant in the muscles of diving marine mammals such as seals and whales, where it also has an oxygen-storage function for prolonged excursions undersea. The tetrameric hemoglobin is responsible for oxygen transport in the blood stream. The monomeric neuroglobin is expressed largely in neurons and helps to protect the brain from hypoxia (low oxygen) or ischemia (restricted blood supply). Cytoglobin, another monomeric globin, is found at high levels in a range of tissues, but its function is unknown.

Myoglobin Has a Single Binding Site for Oxygen

Myoglobin (M_r 16,700; abbreviated Mb) is a single polypeptide of 153 amino acid residues with one molecule of heme. As is typical for a globin polypeptide, myoglobin is made up of eight α-helical segments connected by bends **(Fig. 5–3)**. About 78% of the amino acid residues in the protein are found in these α helices.

Any detailed discussion of protein function inevitably involves protein structure. In the case of myoglobin, we first introduce some structural conventions peculiar to globins. As seen in Figure 5–3, the helical segments are named A through H. An individual amino acid residue is designated either by its position in the amino acid sequence or by its location in the sequence of a particular α-helical segment. For example, the His residue coordinated to the heme in myoglobin, His^{93} (the 93rd residue from the amino-terminal end of the myoglobin polypeptide sequence), is also called His F8 (the 8th residue in α helix F). The bends in the structure are designated AB, CD, EF, FG, and so forth, reflecting the α-helical segments they connect.

Protein-Ligand Interactions Can Be Described Quantitatively

The function of myoglobin depends on the protein's ability not only to bind oxygen but also to release it when and where it is needed. Function in biochemistry

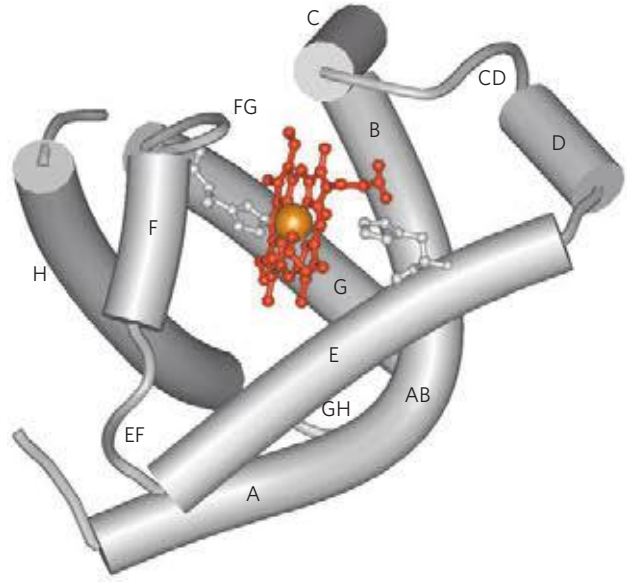


FIGURE 5–3 Myoglobin. (PDB ID 1MBO) The eight α-helical segments (shown here as cylinders) are labeled A through H. Nonhelical residues in the bends that connect them are labeled AB, CD, EF, and so forth, indicating the segments they interconnect. A few bends, including BC and DE, are abrupt and do not contain any residues; these are not normally labeled. (The short segment visible between D and E is an artifact of the computer representation.) The heme is bound in a pocket made up largely of the E and F helices, although amino acid residues from other segments of the protein also participate.

often revolves around a reversible protein-ligand interaction of this type. A quantitative description of this interaction is therefore a central part of many biochemical investigations.

In general, the reversible binding of a protein (P) to a ligand (L) can be described by a simple **equilibrium expression**:

$$P + L \rightleftharpoons PL \qquad (5–1)$$

The reaction is characterized by an equilibrium constant, K_a, such that

$$K_a = \frac{[PL]}{[P][L]} = \frac{k_a}{k_d} \qquad (5–2)$$

where k_a and k_d are rate constants (more on these below). The term $\boldsymbol{K_a}$ is an **association constant** (not to be confused with the K_a that denotes an acid dissociation constant; p. 62) that describes the equilibrium between the complex and the unbound components of the complex. The association constant provides a measure of the affinity of the ligand L for the protein. K_a has units of M^{-1}; a higher value of K_a corresponds to a higher affinity of the ligand for the protein.

The equilibrium term K_a is also equivalent to the ratio of the rates of the forward (association) and reverse (dissociation) reactions that form the PL complex. The association rate is described by a rate constant k_a, and dissociation by the rate constant k_d. As discussed further in the next chapter, rate constants are proportionality constants, describing the fraction of a pool of reactant that reacts in a given amount of time. When the reaction involves one molecule, such as the dissociation reaction PL $\longrightarrow$ P + L, the reaction is *first order* and the rate constant (k_d) has units of reciprocal time (s^{-1}). When the reaction involves two molecules, such as the association reaction P + L $\longrightarrow$ PL, it is called *second order*, and its rate constant (k_a) has units of $M^{-1}\ s^{-1}$.

KEY CONVENTION: Equilibrium constants are denoted with a capital K and rate constants with a lower case k. ■

A rearrangement of the first part of Equation 5–2 shows that the ratio of bound to free protein is directly proportional to the concentration of free ligand:

$$K_a[L] = \frac{[PL]}{[P]} \qquad (5–3)$$

When the concentration of the ligand is much greater than the concentration of ligand-binding sites, the binding of the ligand by the protein does not appreciably change the concentration of free (unbound) ligand—that is, [L] remains constant. This condition is broadly applicable to most ligands that bind to proteins in cells and simplifies our description of the binding equilibrium.

We can now consider the binding equilibrium from the standpoint of the fraction, θ (theta), of ligand-binding sites on the protein that are occupied by ligand:

$$\theta = \frac{\text{binding sites occupied}}{\text{total binding sites}} = \frac{[PL]}{[PL] + [P]} \qquad (5–4)$$

Substituting $K_a[L][P]$ for [PL] (see Eqn 5–3) and rearranging terms gives

$$\theta = \frac{K_a[L][P]}{K_a[L][P] + [P]} = \frac{K_a[L]}{K_a[L] + 1} = \frac{[L]}{[L] + \dfrac{1}{K_a}} \qquad (5–5)$$

The value of K_a can be determined from a plot of θ versus the concentration of free ligand, [L] **(Fig. 5–4a)**. Any equation of the form $x = y/(y + z)$ describes a hyperbola, and θ is thus found to be a hyperbolic function of [L]. The fraction of ligand-binding sites occupied approaches saturation asymptotically as [L] increases. The [L] at which half of the available ligand-binding sites are occupied (that is, $\theta = 0.5$) corresponds to $1/K_a$.

It is more common (and intuitively simpler), however, to consider the **dissociation constant**, $\boldsymbol{K_d}$, which is the reciprocal of K_a ($K_d = 1/K_a$) and is given in units of molar concentration (M). K_d is the equilibrium constant

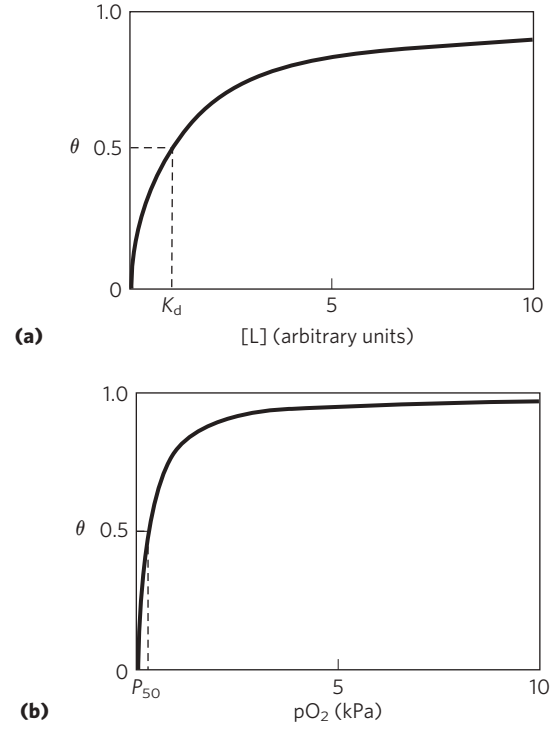
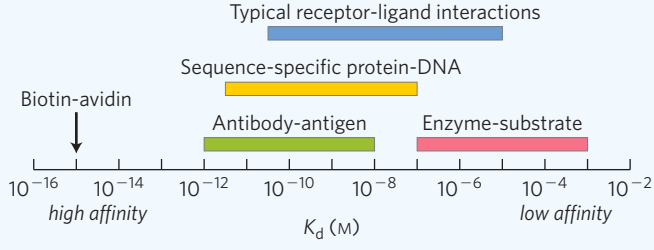


FIGURE 5–4 Graphical representations of ligand binding. The fraction of ligand-binding sites occupied, θ, is plotted against the concentration of free ligand. Both curves are rectangular hyperbolas. **(a)** A hypothetical binding curve for a ligand L. The [L] at which half of the available ligand-binding sites are occupied is equivalent to $1/K_a$, or K_d. The curve has a horizontal asymptote at $\theta = 1$ and a vertical asymptote (not shown) at $[L] = -1/K_a$. **(b)** A curve describing the binding of oxygen to myoglobin. The partial pressure of O_2 in the air above the solution is expressed in kilopascals (kPa). Oxygen binds tightly to myoglobin, with a P_{50} of only 0.26 kPa.

TABLE 5–1 Some Protein Dissociation Constants

Protein	Ligand	K_d (M)*
Avidin (egg white)	Biotin	1×10^{-15}
Insulin receptor (human)	Insulin	1×10^{-10}
Anti-HIV immunoglobulin (human)†	gp41 (HIV-1 surface protein)	4×10^{-10}
Nickel-binding protein (*E. coli*)	Ni^{2+}	1×10^{-7}
Calmodulin (rat)‡	Ca^{2+}	3×10^{-6}
		2×10^{-5}



Color bars indicate the range of dissociation constants typical of various classes of interactions in biological systems. A few interactions, such as that between the protein avidin and the enzyme cofactor biotin, fall outside the normal ranges. The avidin-biotin interaction is so tight it may be considered irreversible. Sequence-specific protein-DNA interactions reflect proteins that bind to a particular sequence of nucleotides in DNA, as opposed to general binding to any DNA site.

*A reported dissociation constant is valid only for the particular solution conditions under which it was measured. K_d values for a protein-ligand interaction can be altered, sometimes by several orders of magnitude, by changes in the solution's salt concentration, pH, or other variables.

†This immunoglobulin was isolated as part of an effort to develop a vaccine against HIV. Immunoglobulins (described later in the chapter) are highly variable, and the K_d reported here should not be considered characteristic of all immunoglobulins.

‡Calmodulin has four binding sites for calcium. The values shown reflect the highest- and lowest-affinity binding sites observed in one set of measurements.

for the release of ligand. The relevant expressions change to

$$K_d = \frac{[P][L]}{[PL]} = \frac{k_d}{k_a} \qquad (5–6)$$

$$[PL] = \frac{[P][L]}{K_d} \qquad (5–7)$$

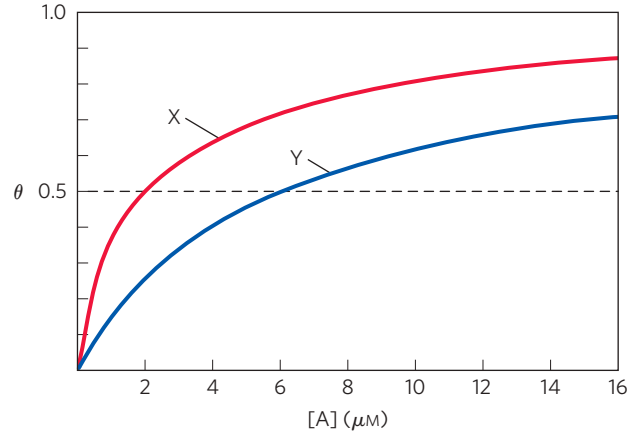
$$\theta = \frac{[L]}{[L] + K_d} \qquad (5–8)$$

When [L] equals K_d, half of the ligand-binding sites are occupied. As [L] falls below K_d, progressively less of the protein has ligand bound to it. In order for 90% of the available ligand-binding sites to be occupied, [L] must be nine times greater than K_d.

In practice, K_d is used much more often than K_a to express the affinity of a protein for a ligand. Note that a lower value of K_d corresponds to a higher affinity of ligand for the protein. The mathematics can be reduced to simple statements: K_d is equivalent to the molar concentration of ligand at which half of the available ligand-binding sites are occupied. At this point, the protein is said to have reached half-saturation with respect to ligand binding. The more tightly a protein binds a ligand, the lower the concentration of ligand required for half the binding sites to be occupied, and thus the lower the value of K_d. Some representative dissociation constants are given in Table 5–1; the scale shows typical ranges for dissociation constants found in biological systems.

WORKED EXAMPLE 5–1 Receptor-Ligand Dissociation Constants

Two proteins, X and Y, bind to the same ligand, A, with the binding curves shown below.



What is the dissociation constant, K_d, for each protein? Which protein (X or Y) has a greater affinity for ligand A?

Solution: We can determine the dissociation constants by inspecting the graph. Since θ represents the fraction of binding sites occupied by ligand, the concentration of ligand at which half the binding sites are occupied—that is, the point where the binding curve crosses the line where $\theta = 0.5$—is the dissociation constant. For X, $K_d = 2\ \mu M$; for Y, $K_d = 6\ \mu M$. Because X is half-saturated at a lower [A], it has a higher affinity for the ligand.

The binding of oxygen to myoglobin follows the patterns discussed above. However, because oxygen is a gas, we must make some minor adjustments to the equations so that laboratory experiments can be carried out more conveniently. We first substitute the concentration of dissolved oxygen for [L] in Equation 5–8 to give

$$\theta = \frac{[O_2]}{[O_2] + K_d} \qquad (5–9)$$

As for any ligand, K_d equals the $[O_2]$ at which half of the available ligand-binding sites are occupied, or $[O_2]_{0.5}$. Equation 5–9 thus becomes

$$\theta = \frac{[O_2]}{[O_2] + [O_2]_{0.5}} \qquad (5–10)$$

In experiments using oxygen as a ligand, it is the partial pressure of oxygen (pO_2) in the gas phase above the solution that is varied, because this is easier to measure than the concentration of oxygen dissolved in the solution. The concentration of a volatile substance in solution is always proportional to the local partial pressure of the gas. So, if we define the partial pressure of oxygen at $[O_2]_{0.5}$ as P_{50}, substitution in Equation 5–10 gives

$$\theta = \frac{pO_2}{pO_2 + P_{50}} \qquad (5–11)$$

A binding curve for myoglobin that relates θ to pO_2 is shown in Figure 5–4b.

Protein Structure Affects How Ligands Bind

The binding of a ligand to a protein is rarely as simple as the above equations would suggest. The interaction is greatly affected by protein structure and is often accompanied by conformational changes. For example, the specificity with which heme binds its various ligands is altered when the heme is a component of myoglobin. Carbon monoxide binds to free heme molecules more than 20,000 times better than does O_2 (that is, the K_d or P_{50} for CO binding to free heme is more than 20,000 times lower than that for O_2), but it binds only about 200 times better than O_2 when the heme is bound in myoglobin. The difference may be partly explained by steric hindrance. When O_2 binds to free heme, the axis of the oxygen molecule is positioned at an angle to the Fe—O bond **(Fig. 5–5a)**. In contrast, when CO binds to free heme, the Fe, C, and O atoms lie in a straight line

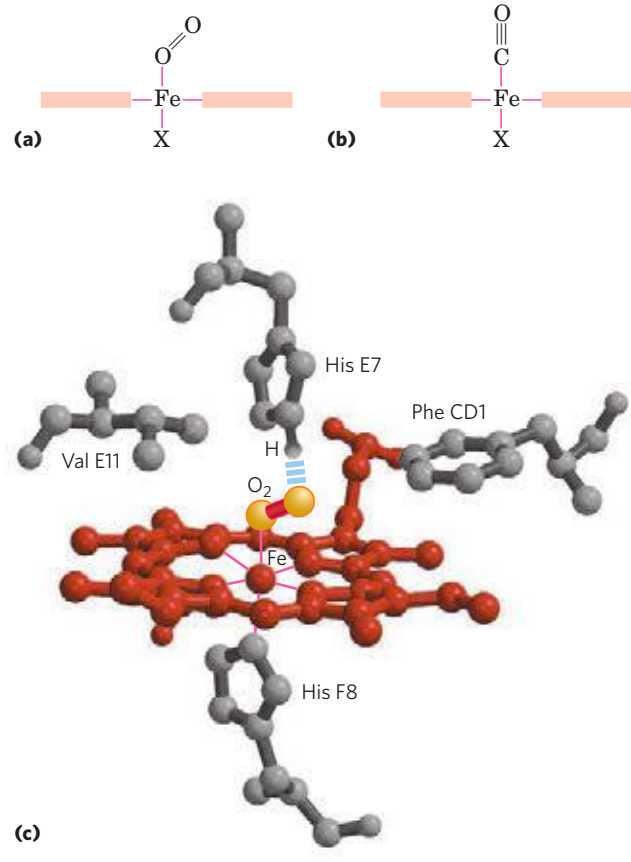


FIGURE 5–5 Steric effects caused by ligand binding to the heme of myoglobin. (a) Oxygen binds to heme with the O_2 axis at an angle, a binding conformation readily accommodated by myoglobin. **(b)** Carbon monoxide binds to free heme with the CO axis perpendicular to the plane of the porphyrin ring. When binding to the heme in myoglobin, CO is forced to adopt a slight angle because the perpendicular arrangement is sterically blocked by His E7, the distal His. This effect weakens the binding of CO to myoglobin. **(c)** Another view of the heme of myoglobin (derived from PDB ID 1MBO), showing the arrangement of key amino acid residues around the heme. The bound O_2 is hydrogen-bonded to the distal His, His E7 (His^{64}), further facilitating the binding of O_2.

(Fig. 5–5b). In both cases, the binding reflects the geometry of hybrid orbitals in each ligand. In myoglobin, His^{64} (His E7), on the O_2-binding side of the heme, is too far away to coordinate with the heme iron, but it does interact with a ligand bound to heme. This residue, called the *distal His* (as distinct from the *proximal His*, His F8), forms a hydrogen bond with O_2 (Fig. 5–5c) but may help preclude the linear binding of CO, providing one explanation for the selectively diminished binding of CO to heme in myoglobin (and hemoglobin). A reduction in CO binding is physiologically important, because CO is a low-level byproduct of cellular metabolism. Other factors, not yet well-defined, also may modulate the interaction of heme with CO in these proteins.

The binding of O_2 to the heme in myoglobin also depends on molecular motions, or "breathing," in the protein structure. The heme molecule is deeply buried in the folded polypeptide, with no direct path for oxygen to move from the surrounding solution to the ligand-binding site. If the protein were rigid, O_2 could

not enter or leave the heme pocket at a measurable rate. However, rapid molecular flexing of the amino acid side chains produces transient cavities in the protein structure, and O_2 makes its way in and out by moving through these cavities. Computer simulations of rapid structural fluctuations in myoglobin suggest that there are many such pathways. One major route is provided by rotation of the side chain of the distal His (His^{64}), which occurs on a nanosecond (10^{-9} s) time scale. Even subtle conformational changes can be critical for protein activity.

In neuroglobin, cytoglobin, and some globins found in plants and invertebrates, the distal His (His E7) is directly coordinated with the heme iron. In these globins, the oxygen or other ligand must displace the distal His in the process of binding.

Hemoglobin Transports Oxygen in Blood

Oxygen-Binding Proteins—Hemoglobin: Oxygen Transport Nearly all the oxygen carried by whole blood in animals is bound and transported by hemoglobin in erythrocytes (red blood cells). Normal human erythrocytes are small (6 to 9 μm in diameter), biconcave disks. They are formed from precursor stem cells called **hemocytoblasts**. In the maturation process, the stem cell produces daughter cells that form large amounts of hemoglobin and then lose their intracellular organelles—nucleus, mitochondria, and endoplasmic reticulum. Erythrocytes are thus incomplete, vestigial cells, unable to reproduce and, in humans, destined to survive for only about 120 days. Their main function is to carry hemoglobin, which is dissolved in the cytosol at a very high concentration (~34% by weight).

In arterial blood passing from the lungs through the heart to the peripheral tissues, hemoglobin is about 96% saturated with oxygen. In the venous blood returning to the heart, hemoglobin is only about 64% saturated. Thus, each 100 mL of blood passing through a tissue releases about one-third of the oxygen it carries, or 6.5 mL of O_2 gas at atmospheric pressure and body temperature.

Myoglobin, with its hyperbolic binding curve for oxygen (Fig. 5–4b), is relatively insensitive to small changes in the concentration of dissolved oxygen and so functions well as an oxygen-storage protein. Hemoglobin, with its multiple subunits and O_2-binding sites, is better suited to oxygen transport. As we shall see, interactions between the subunits of a multimeric protein can permit a highly sensitive response to small changes in ligand concentration. Interactions among the subunits in hemoglobin cause conformational changes that alter the affinity of the protein for oxygen. The modulation of oxygen binding allows the O_2-transport protein to respond to changes in oxygen demand by tissues.

Hemoglobin Subunits Are Structurally Similar to Myoglobin

Hemoglobin (M_r 64,500; abbreviated Hb) is roughly spherical, with a diameter of nearly 5.5 nm. It is a tetrameric protein containing four heme prosthetic groups,

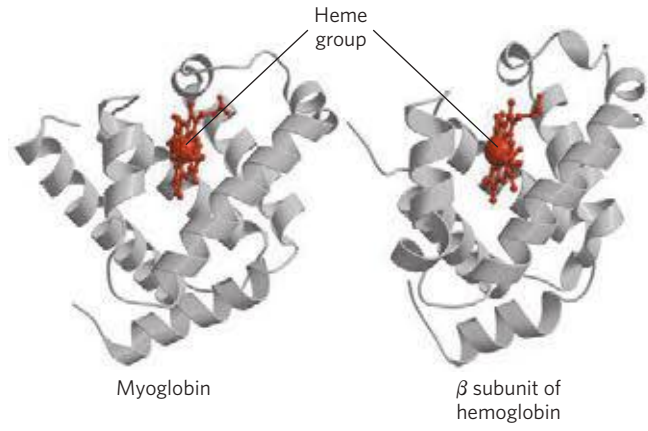


FIGURE 5–6 Comparison of the structures of myoglobin (PDB ID 1MBO) and the β subunit of hemoglobin (derived from PDB ID 1HGA).

one associated with each polypeptide chain. Adult hemoglobin contains two types of globin, two α chains (141 residues each) and two β chains (146 residues each). Although fewer than half of the amino acid residues are identical in the polypeptide sequences of the α and β subunits, the three-dimensional structures of the two types of subunits are very similar. Furthermore, their structures are very similar to that of myoglobin **(Fig. 5–6)**, even though the amino acid sequences of the three polypeptides are identical at only 27 positions **(Fig. 5–7)**. All three polypeptides are members of the globin family of proteins. The helix-naming convention described for myoglobin is also applied to the hemoglobin polypeptides, except that the α subunit lacks the short D helix. The heme-binding pocket is made up largely of the E and F helices in each of the subunits.

The quaternary structure of hemoglobin features strong interactions between unlike subunits. The $\alpha_1\beta_1$ interface (and its $\alpha_2\beta_2$ counterpart) involves more than 30 residues, and its interaction is sufficiently strong that although mild treatment of hemoglobin with urea tends to disassemble the tetramer into $\alpha\beta$ dimers, these dimers remain intact. The $\alpha_1\beta_2$ (and $\alpha_2\beta_1$) interface involves 19 residues **(Fig. 5–8)**. Hydrophobic interactions predominate at all the interfaces, but there are also many hydrogen bonds and a few ion pairs (or salt bridges), whose importance is discussed below.

Hemoglobin Undergoes a Structural Change on Binding Oxygen

X-ray analysis has revealed two major conformations of hemoglobin: the **R state** and the **T state**. Although oxygen binds to hemoglobin in either state, it has a significantly higher affinity for hemoglobin in the R state. Oxygen binding stabilizes the R state. When oxygen is absent experimentally, the T state is more stable and is thus the predominant conformation of **deoxyhemoglobin**. T and R originally denoted "tense" and "relaxed," respectively, because the T state is stabilized by a greater number of ion pairs, many of which lie at the $\alpha_1\beta_2$ (and $\alpha_2\beta_1$) interface **(Fig. 5–9)**. The binding of O_2

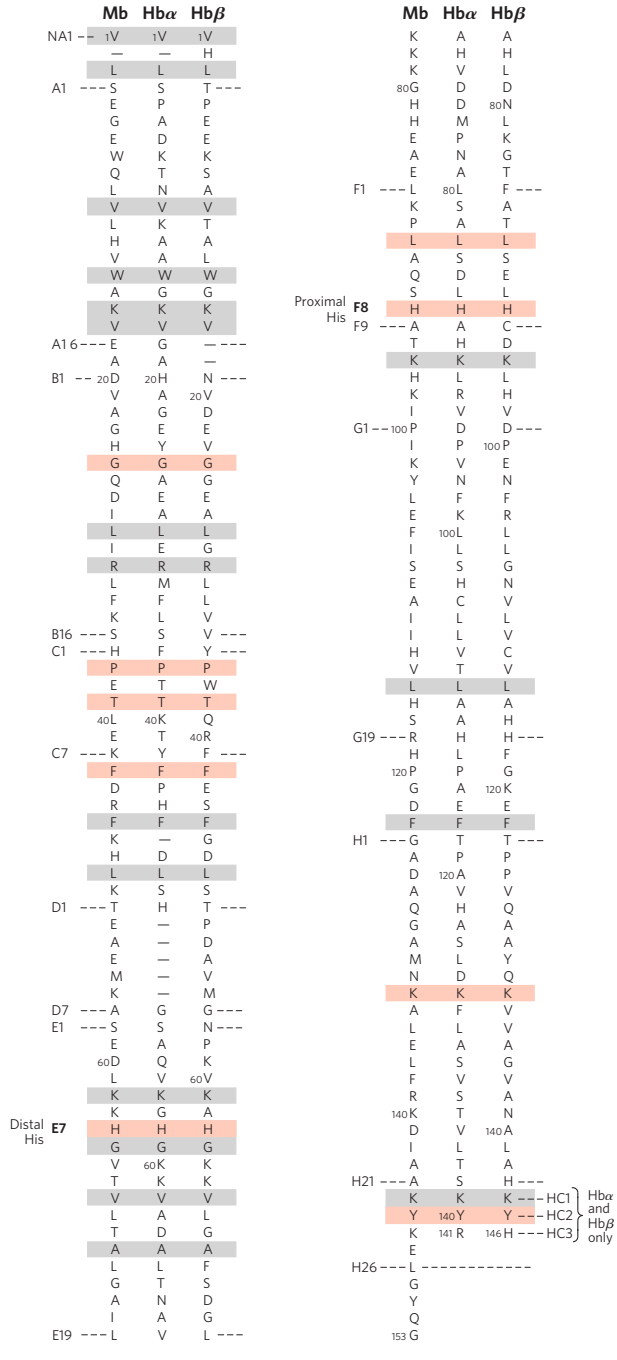


FIGURE 5–7 The amino acid sequences of whale myoglobin and the α and β chains of human hemoglobin. Dashed lines mark helix boundaries. To align the sequences optimally, short gaps must be introduced into both Hb sequences where a few amino acids are present in the other, compared sequences. With the exception of the missing D helix in Hbα, this alignment permits the use of the helix lettering convention that emphasizes the common positioning of amino acid residues that are identical in all three structures (shaded). Residues shaded in light red are conserved in all known globins. Note that the common helix-letter-and-number designation for amino acids does not necessarily correspond to a common position in the linear sequence of amino acids in the polypeptides. For example, the distal His residue is His E7 in all three structures, but corresponds to His^{64}, His^{58}, and His^{63} in the linear sequences of Mb, Hbα, and Hbβ, respectively. Nonhelical residues at the amino and carboxyl termini, beyond the first (A) and last (H) α-helical segments, are labeled NA and HC, respectively.

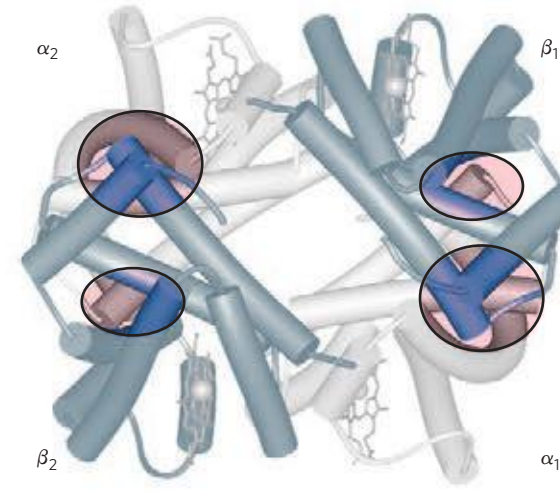


FIGURE 5–8 Dominant interactions between hemoglobin subunits. (PDB ID 1HGA) In this representation, α subunits are light and β subunits are dark. The strongest subunit interactions (highlighted) occur between unlike subunits. When oxygen binds, the $\alpha_1\beta_1$ contact changes little, but there is a large change at the $\alpha_1\beta_2$ contact, with several ion pairs broken.

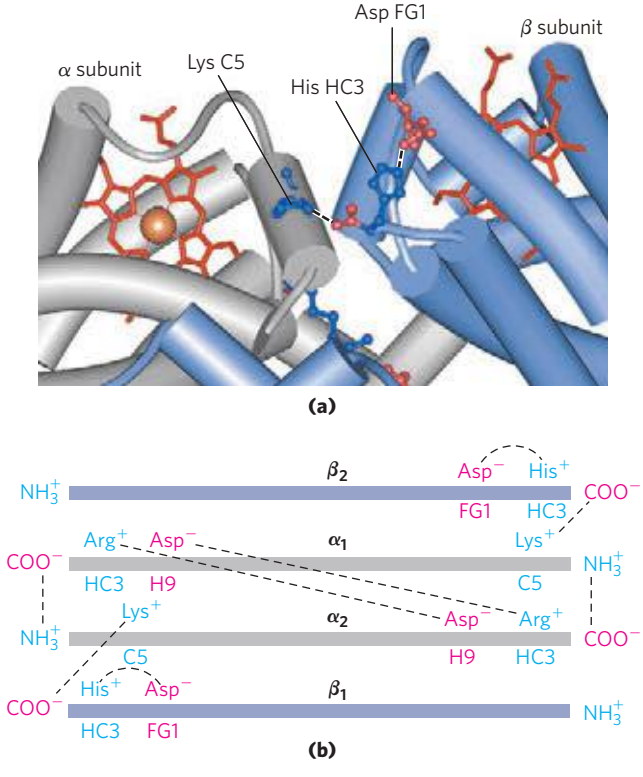


FIGURE 5–9 Some ion pairs that stabilize the T state of deoxyhemoglobin. (a) Close-up view of a portion of a deoxyhemoglobin molecule in the T state (PDB ID 1HGA). Interactions between the ion pairs His HC3 and Asp FG1 of the β subunit (blue) and between Lys C5 of the α subunit (gray) and His HC3 (its α-carboxyl group) of the β subunit are shown with dashed lines. (Recall that HC3 is the carboxyl-terminal residue of the β subunit.) **(b)** Interactions between these ion pairs, and between others not shown in (a), are schematized in this representation of the extended polypeptide chains of hemoglobin.

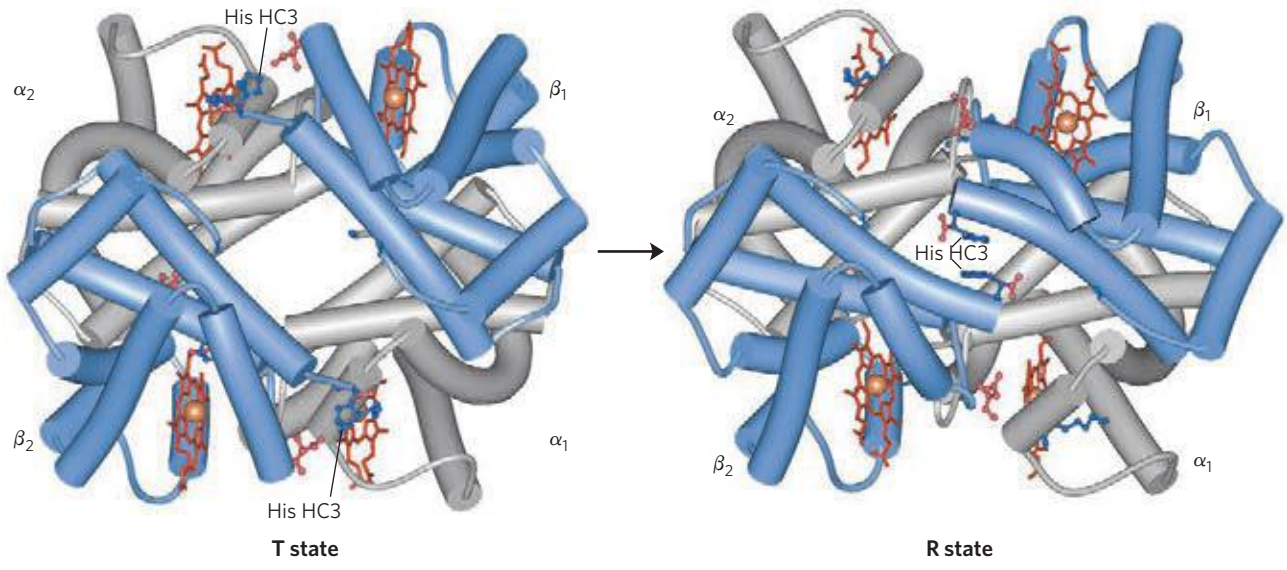


FIGURE 5–10 The T → R transition. (PDB ID 1HGA and 1BBB) In these depictions of deoxyhemoglobin, as in Figure 5–9, the β subunits are blue and the α subunits are gray. Positively charged side chains and chain termini involved in ion pairs are shown in blue, their negatively charged partners in red. The Lys C5 of each α subunit and Asp FG1 of each β subunit are visible but not labeled (compare Fig. 5–9a). Note that the molecule is oriented slightly differently than in Figure 5–9. The transition from the T state to the R state shifts the subunit pairs substantially, affecting certain ion pairs. Most noticeably, the His HC3 residues at the carboxyl termini of the β subunits, which are involved in ion pairs in the T state, rotate in the R state toward the center of the molecule, where they are no longer in ion pairs. Another dramatic result of the T → R transition is a narrowing of the pocket between the β subunits.

to a hemoglobin subunit in the T state triggers a change in conformation to the R state. When the entire protein undergoes this transition, the structures of the individual subunits change little, but the $\alpha\beta$ subunit pairs slide past each other and rotate, narrowing the pocket between the β subunits **(Fig. 5–10)**. In this process, some of the ion pairs that stabilize the T state are broken and some new ones are formed.

Max Perutz proposed that the T → R transition is triggered by changes in the positions of key amino acid side chains surrounding the heme. In the T state, the porphyrin is slightly puckered, causing the heme iron to protrude somewhat on the proximal His (His F8) side. The binding of O_2 causes the heme to assume a more planar conformation, shifting the position of the proximal His and the attached F helix **(Fig. 5–11)**. These changes lead to adjustments in the ion pairs at the $\alpha_1\beta_2$ interface.

Hemoglobin Binds Oxygen Cooperatively

Hemoglobin must bind oxygen efficiently in the lungs, where the pO_2 is about 13.3 kPa, and release oxygen in the tissues, where the pO_2 is about 4 kPa. Myoglobin, or any protein that binds oxygen with a hyperbolic binding curve, would be ill-suited to this function, for the reason illustrated in **Figure 5–12**. A protein that bound O_2 with high affinity would bind it efficiently in the lungs but would not release much of it in the tissues. If the protein bound oxygen with a sufficiently low affinity to release it in the tissues, it would not pick up much oxygen in the lungs.

Hemoglobin solves the problem by undergoing a transition from a low-affinity state (the T state) to a high-affinity state (the R state) as more O_2 molecules are bound. As a result, hemoglobin has a hybrid S-shaped, or sigmoid, binding curve for oxygen (Fig. 5–12). A single-subunit protein with a single ligand-binding site cannot produce a sigmoid binding curve—even if binding elicits a conformational change—because each molecule of ligand binds independently and cannot affect ligand binding to another molecule. In contrast, O_2 binding to individual subunits of hemoglobin can alter the affinity for O_2 in adjacent subunits. The first molecule of O_2 that interacts with deoxyhemoglobin binds weakly, because it binds to a subunit in the T state. Its binding, however,

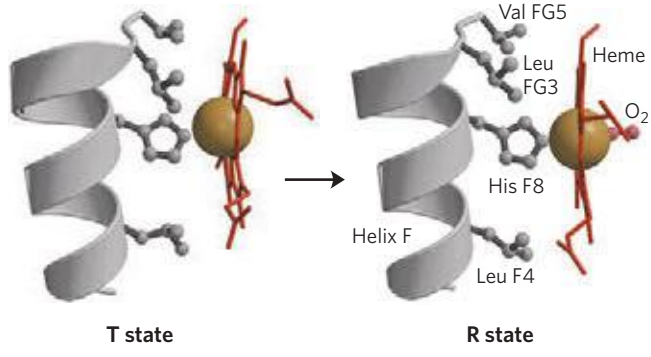


FIGURE 5–11 Changes in conformation near heme on O_2 binding to deoxyhemoglobin. (Derived from PDB ID 1HGA and 1BBB) The shift in the position of helix F when heme binds O_2 is thought to be one of the adjustments that triggers the T → R transition.

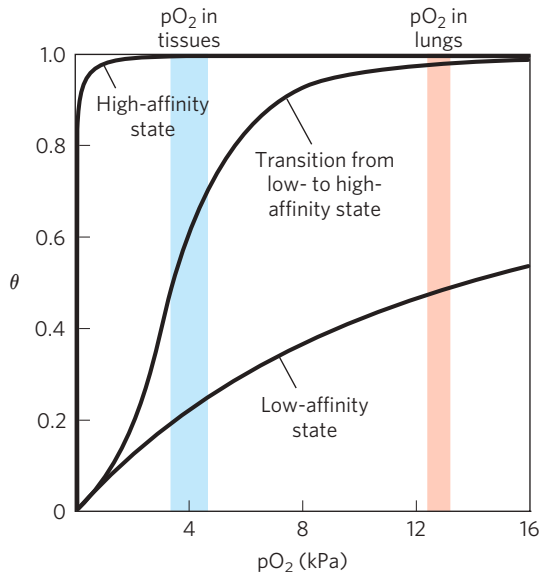


FIGURE 5–12 A sigmoid (cooperative) binding curve. A sigmoid binding curve can be viewed as a hybrid curve reflecting a transition from a low-affinity to a high-affinity state. Because of its cooperative binding, as manifested by a sigmoid binding curve, hemoglobin is more sensitive to the small differences in O_2 concentration between the tissues and the lungs, allowing it to bind oxygen in the lungs (where pO_2 is high) and release it in the tissues (where pO_2 is low).

leads to conformational changes that are communicated to adjacent subunits, making it easier for additional molecules of O_2 to bind. In effect, the T → R transition occurs more readily in the second subunit once O_2 is bound to the first subunit. The last (fourth) O_2 molecule binds to a heme in a subunit that is already in the R state, and hence it binds with much higher affinity than the first molecule.

An **allosteric protein** is one in which the binding of a ligand to one site affects the binding properties of another site on the same protein. The term "allosteric" derives from the Greek *allos*, "other," and *stereos*, "solid" or "shape." Allosteric proteins are those having "other shapes," or conformations, induced by the binding of ligands referred to as modulators. The conformational changes induced by the modulator(s) interconvert more-active and less-active forms of the protein. The modulators for allosteric proteins may be either inhibitors or activators. When the normal ligand and modulator are identical, the interaction is termed **homotropic**. When the modulator is a molecule other than the normal ligand, the interaction is **heterotropic**. Some proteins have two or more modulators and therefore can have both homotropic and heterotropic interactions.

Cooperative binding of a ligand to a multimeric protein, such as we observe with the binding of O_2 to hemoglobin, is a form of allosteric binding. The binding of one ligand affects the affinities of any remaining unfilled binding sites, and O_2 can be considered as both a ligand and an activating homotropic modulator. There is only one binding site for O_2 on each subunit, so the allosteric effects giving rise to cooperativity are mediated by conformational changes transmitted from one subunit to another by subunit-subunit interactions. A sigmoid binding curve is diagnostic of cooperative binding. It permits a much more sensitive response to ligand concentration and is important to the function of many multisubunit proteins. The principle of allostery extends readily to regulatory enzymes, as we shall see in Chapter 6.

Cooperative conformational changes depend on variations in the structural stability of different parts of a protein, as described in Chapter 4. The binding sites of an allosteric protein typically consist of stable segments in proximity to relatively unstable segments, with the latter capable of frequent changes in conformation or intrinsic disorder **(Fig. 5–13)**. When a ligand binds, the moving parts of the protein's binding site may be stabilized in a particular conformation, affecting the conformation of adjacent polypeptide

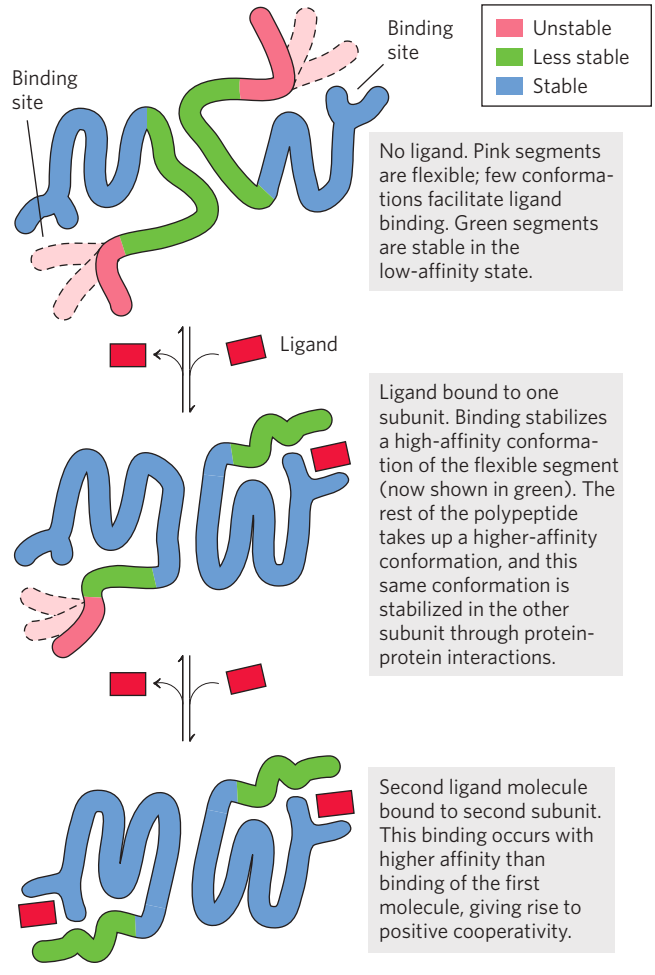


FIGURE 5–13 Structural changes in a multisubunit protein undergoing cooperative binding to ligand. Structural stability is not uniform throughout a protein molecule. Shown here is a hypothetical dimeric protein, with regions of high (blue), medium (green), and low (red) stability. The ligand-binding sites are composed of both high- and low-stability segments, so affinity for ligand is relatively low. The conformational changes that occur as ligand binds convert the protein from a low- to a high-affinity state, a form of induced fit.

subunits. If the entire binding site were highly stable, then few structural changes could occur in this site or be propagated to other parts of the protein when a ligand binds.

As is the case with myoglobin, ligands other than oxygen can bind to hemoglobin. An important example is carbon monoxide, which binds to hemoglobin about 250 times better than does oxygen. Human exposure to CO can have tragic consequences (Box 5–1).

Cooperative Ligand Binding Can Be Described Quantitatively

Cooperative binding of oxygen by hemoglobin was first analyzed by Archibald Hill in 1910. From this work came a general approach to the study of cooperative ligand binding to multisubunit proteins.

For a protein with n binding sites, the equilibrium of Equation 5–1 becomes

$$\mathrm{P} + n\mathrm{L} \rightleftharpoons \mathrm{PL}_n \qquad (5\text{–}12)$$

and the expression for the association constant becomes

$$K_\mathrm{a} = \frac{[\mathrm{PL}_n]}{[\mathrm{P}][\mathrm{L}]^n} \qquad (5\text{–}13)$$

The expression for θ (see Eqn 5–8) is

$$\theta = \frac{[\mathrm{L}]^n}{[\mathrm{L}]^n + K_\mathrm{d}} \qquad (5\text{–}14)$$

Rearranging, then taking the log of both sides, yields

$$\frac{\theta}{1-\theta} = \frac{[\mathrm{L}]^n}{K_\mathrm{d}} \qquad (5\text{–}15)$$

$$\log\left(\frac{\theta}{1-\theta}\right) = n \log [\mathrm{L}] - \log K_\mathrm{d} \qquad (5\text{–}16)$$

where $K_\mathrm{d} = [L]^n_{0.5}$.

Equation 5–16 is the **Hill equation**, and a plot of $\log [\theta/(1-\theta)]$ versus $\log [\mathrm{L}]$ is called a **Hill plot**. Based on the equation, the Hill plot should have a slope of n. However, the experimentally determined slope actually reflects not the number of binding sites but the degree of interaction between them. The slope of a Hill plot is therefore denoted by n_H, the **Hill coefficient**, which is a measure of the degree of cooperativity. If n_H equals 1, ligand binding is not cooperative, a situation that can arise even in a multisubunit protein if the subunits do not communicate. An n_H of greater than 1 indicates positive cooperativity in ligand binding. This is the situation observed in hemoglobin, in which the binding of one molecule of ligand facilitates the binding of others. The theoretical upper limit for n_H is reached when $n_\mathrm{H} = n$. In this case the binding would be completely cooperative: all binding sites on the protein would bind ligand simultaneously, and no protein molecules partially saturated with ligand would be present under any conditions. This limit is never reached in practice, and the measured value of n_H is always less than the actual number of ligand-binding sites in the protein.

An n_H of less than 1 indicates negative cooperativity, in which the binding of one molecule of ligand *impedes* the binding of others. Well-documented cases of negative cooperativity are rare.

To adapt the Hill equation to the binding of oxygen to hemoglobin we must again substitute $\mathrm{pO_2}$ for [L] and P^n_{50} for K_d:

$$\log\left(\frac{\theta}{1-\theta}\right) = n \log \mathrm{pO_2} - n \log P_{50} \qquad (5\text{–}17)$$

Hill plots for myoglobin and hemoglobin are given in **Figure 5–14**.

Two Models Suggest Mechanisms for Cooperative Binding

Biochemists now know a great deal about the T and R states of hemoglobin, but much remains to be learned about how the T → R transition occurs. Two models for the cooperative binding of ligands to proteins with multiple binding sites have greatly influenced thinking about this problem.

The first model was proposed by Jacques Monod, Jeffries Wyman, and Jean-Pierre Changeux in 1965, and is called the **MWC model** or the **concerted**

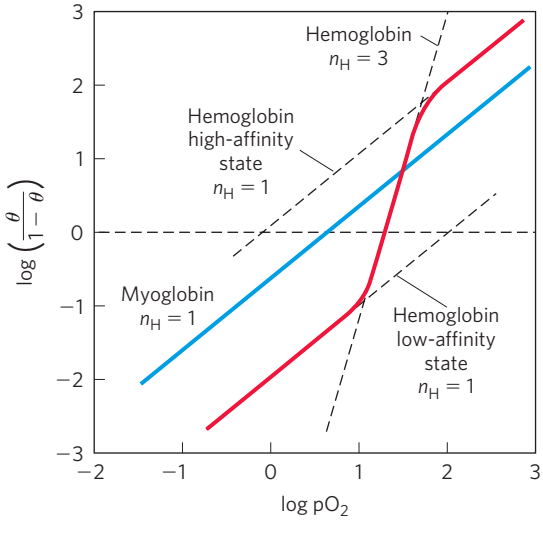


FIGURE 5–14 Hill plots for oxygen binding to myoglobin and hemoglobin. When $n_\mathrm{H} = 1$, there is no evident cooperativity. The maximum degree of cooperativity observed for hemoglobin corresponds approximately to $n_\mathrm{H} = 3$. Note that while this indicates a high level of cooperativity, n_H is less than n, the number of $\mathrm{O_2}$-binding sites in hemoglobin. This is normal for a protein that exhibits allosteric binding behavior.

BOX 5–1 MEDICINE Carbon Monoxide: A Stealthy Killer

Lake Powell, Arizona, August 2000. A family was vacationing in a rented houseboat. They turned on the electrical generator to power an air conditioner and a television. About 15 minutes later, two brothers, aged 8 and 11, jumped off the swim deck at the stern. Situated immediately below the deck was the exhaust port for the generator. Within two minutes, both boys were overcome by the carbon monoxide in the exhaust, which had become concentrated in the space under the deck. Both drowned. These deaths, along with a series of deaths in the 1990s that were linked to houseboats of similar design, eventually led to the recall and redesign of the generator exhaust assembly.

Carbon monoxide (CO), a colorless, odorless gas, is responsible for more than half of yearly deaths due to poisoning worldwide. CO has an approximately 250-fold greater affinity for hemoglobin than does oxygen. Consequently, relatively low levels of CO can have substantial and tragic effects. When CO combines with hemoglobin, the complex is referred to as carboxyhemoglobin, or COHb.

Some CO is produced by natural processes, but locally high levels generally result only from human activities. Engine and furnace exhausts are important sources, as CO is a byproduct of the incomplete combustion of fossil fuels. In the United States alone, nearly 4,000 people succumb to CO poisoning each year, both accidentally and intentionally. Many of the accidental deaths involve undetected CO buildup in enclosed spaces, such as when a household furnace malfunctions or leaks, venting CO into a home. However, CO poisoning can also occur in open spaces, as unsuspecting people at work or play inhale the exhaust from generators, outboard motors, tractor engines, recreational vehicles, or lawn mowers.

Carbon monoxide levels in the atmosphere are rarely dangerous, ranging from less than 0.05 part per million (ppm) in remote and uninhabited areas to 3 to 4 ppm in some cities of the northern hemisphere. In the United States, the government-mandated (Occupational Safety and Health Administration, OSHA) limit for CO at worksites is 50 ppm for people working an eight-hour shift. The tight binding of CO to hemoglobin means that COHb can accumulate over time as people are exposed to a constant low-level source of CO.

In an average, healthy individual, 1% or less of the total hemoglobin is complexed as COHb. Since CO is a product of tobacco smoke, many smokers have COHb levels in the range of 3% to 8% of total hemoglobin, and the levels can rise to 15% for chain-smokers. COHb levels equilibrate at 50% in people who breathe air containing 570 ppm of CO for several hours. Reliable methods have been developed that relate CO content in the atmosphere to COHb levels in the blood (Fig. 1). In tests of houseboats with a generator exhaust like the one responsible for the Lake Powell deaths, CO levels reached 6,000 to 30,000 ppm under the swim deck, and atmospheric O_2 levels under the deck declined from 21% to 12%. Even above the swim deck, CO levels of up to 7,200 ppm were detected, high enough to cause death within a few minutes.

How is a human affected by COHb? At levels of less than 10% of total hemoglobin, symptoms are rarely observed. At 15%, the individual experiences mild headaches. At 20% to 30%, the headache is severe and is generally accompanied by nausea, dizziness, confusion,

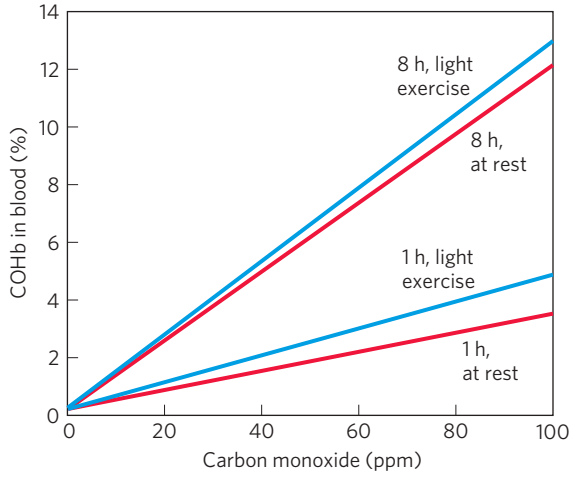


FIGURE 1 Relationship between levels of COHb in blood and concentration of CO in the surrounding air. Four different conditions of exposure are shown, comparing the effects of short versus extended exposure, and exposure at rest versus exposure during light exercise.

model (Fig. 5–15a). The concerted model assumes that the subunits of a cooperatively binding protein are functionally identical, that each subunit can exist in (at least) two conformations, and that all subunits undergo the transition from one conformation to the other simultaneously. In this model, no protein has individual subunits in different conformations. The two conformations are in equilibrium. The ligand can bind to either conformation, but binds each with different affinity. Successive binding of ligand molecules to the low-affinity conformation (which is more stable in the absence of ligand) makes a transition to the high-affinity conformation more likely.

In the second model, the **sequential model** (Fig. 5–15b), proposed in 1966 by Daniel Koshland and colleagues, ligand binding can induce a change of

disorientation, and some visual disturbances; these symptoms are generally reversed if the individual is treated with oxygen. At COHb levels of 30% to 50%, the neurological symptoms become more severe, and at levels near 50%, the individual loses consciousness and can sink into coma. Respiratory failure may follow. With prolonged exposure, some damage becomes permanent. Death normally occurs when COHb levels rise above 60%. Autopsy on the boys who died at Lake Powell revealed COHb levels of 59% and 52%.

Binding of CO to hemoglobin is affected by many factors, including exercise (Fig. 1) and changes in air pressure related to altitude. Because of their higher base levels of COHb, smokers exposed to a source of CO often develop symptoms faster than nonsmokers. Individuals with heart, lung, or blood diseases that reduce the availability of oxygen to tissues may also experience symptoms at lower levels of CO exposure. Fetuses are at particular risk for CO poisoning, because fetal hemoglobin has a somewhat higher affinity for CO than adult hemoglobin. Cases of CO exposure have been recorded in which the fetus died but the mother recovered.

It may seem surprising that the loss of half of one's hemoglobin to COHb can prove fatal—we know that people with any of several anemic conditions manage to function reasonably well with half the usual complement of active hemoglobin. However, the binding of CO to hemoglobin does more than remove protein from the pool available to bind oxygen. It also affects the affinity of the remaining hemoglobin subunits for oxygen. As CO binds to one or two subunits of a hemoglobin tetramer, the affinity for O_2 is increased substantially in the remaining subunits (Fig. 2). Thus, a hemoglobin tetramer with two bound CO molecules can efficiently bind O_2 in the lungs—but it releases very little of it in the tissues. Oxygen deprivation in the tissues rapidly becomes severe. To add to the problem, the effects of CO are not limited to interference with hemoglobin function. CO binds to other heme proteins and a variety of metalloproteins. The effects of these interactions are not yet well understood, but they may be responsible for some of the longer-term effects of acute but nonfatal CO poisoning.

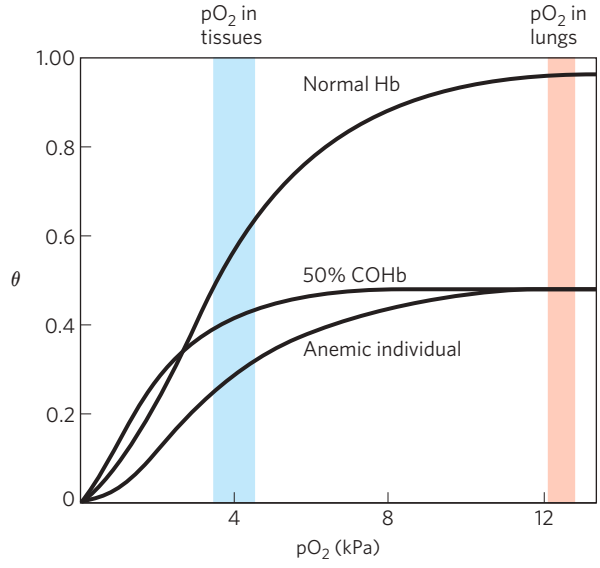


FIGURE 2 Several oxygen-binding curves: for normal hemoglobin, hemoglobin from an anemic individual with only 50% of her hemoglobin functional, and hemoglobin from an individual with 50% of his hemoglobin subunits complexed with CO. The pO_2 in human lungs and tissues is indicated.

When CO poisoning is suspected, rapid evacuation of the person away from the CO source is essential, but this does not always result in rapid recovery. When an individual is moved from the CO-polluted site to a normal, outdoor atmosphere, O_2 begins to replace the CO in hemoglobin—but the COHb levels drop only slowly. The half-time is 2 to 6.5 hours, depending on individual and environmental factors. If 100% oxygen is administered with a mask, the rate of exchange can be increased about fourfold; the half-time for O_2-CO exchange can be reduced to tens of minutes if 100% oxygen at a pressure of 3 atm (303 kPa) is supplied. Thus, rapid treatment by a properly equipped medical team is critical.

Carbon monoxide detectors in all homes are highly recommended. This is a simple and inexpensive measure to avoid possible tragedy. After completing the research for this box, we immediately purchased several new CO detectors for our homes.

conformation in an individual subunit. A conformational change in one subunit makes a similar change in an adjacent subunit, as well as the binding of a second ligand molecule, more likely. There are more potential intermediate states in this model than in the concerted model. The two models are not mutually exclusive; the concerted model may be viewed as the "all-or-none" limiting case of the sequential model. In Chapter 6 we use these models to investigate allosteric enzymes.

Hemoglobin Also Transports H^+ and CO_2

In addition to carrying nearly all the oxygen required by cells from the lungs to the tissues, hemoglobin carries two end products of cellular respiration—H^+ and CO_2—from the tissues to the lungs and the kidneys, where

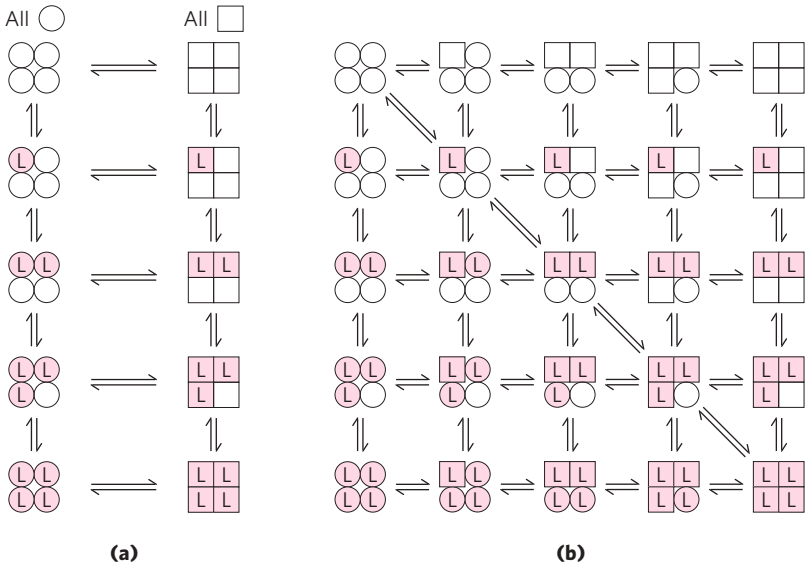


FIGURE 5–15 Two general models for the interconversion of inactive and active forms of a protein during cooperative ligand binding. Although the models may be applied to any protein—including any enzyme (Chapter 6)—that exhibits cooperative binding, we show here four subunits because the model was originally proposed for hemoglobin. **(a)** In the concerted, or all-or-none, model (MWC model), all subunits are postulated to be in the same conformation, either all ○ (low affinity or inactive) or all □ (high affinity or active). Depending on the equilibrium, K_{eq}, between ○ and □ forms, the binding of one or more ligand molecules (L) will pull the equilibrium toward the □ form. Subunits with bound L are shaded. **(b)** In the sequential model, each individual subunit can be in either the ○ or □ form. A very large number of conformations is thus possible.

they are excreted. The CO_2, produced by oxidation of organic fuels in mitochondria, is hydrated to form bicarbonate:

$$CO_2 + H_2O \rightleftharpoons H^+ + HCO_3^-$$

This reaction is catalyzed by **carbonic anhydrase**, an enzyme particularly abundant in erythrocytes. Carbon dioxide is not very soluble in aqueous solution, and bubbles of CO_2 would form in the tissues and blood if it were not converted to bicarbonate. As you can see from the reaction catalyzed by carbonic anhydrase, the hydration of CO_2 results in an increase in the H^+ concentration (a decrease in pH) in the tissues. The binding of oxygen by hemoglobin is profoundly influenced by pH and CO_2 concentration, so the interconversion of CO_2 and bicarbonate is of great importance to the regulation of oxygen binding and release in the blood.

Hemoglobin transports about 40% of the total H^+ and 15% to 20% of the CO_2 formed in the tissues to the lungs and kidneys. (The remainder of the H^+ is absorbed by the plasma's bicarbonate buffer; the remainder of the CO_2 is transported as dissolved HCO_3^- and CO_2.) The binding of H^+ and CO_2 is inversely related to the binding of oxygen. At the relatively low pH and high CO_2 concentration of peripheral tissues, the affinity of hemoglobin for oxygen decreases as H^+ and CO_2 are bound, and O_2 is released to the tissues. Conversely, in the capillaries of the lung, as CO_2 is excreted and the blood pH consequently rises, the affinity of hemoglobin for oxygen increases and the protein binds more O_2 for transport to the peripheral tissues. This effect of pH and CO_2 concentration on the binding and release of oxygen by hemoglobin is called the **Bohr effect**, after Christian Bohr, the Danish physiologist (and father of physicist Niels Bohr) who discovered it in 1904.

The binding equilibrium for hemoglobin and one molecule of oxygen can be designated by the reaction

$$Hb + O_2 \rightleftharpoons HbO_2$$

but this is not a complete statement. To account for the effect of H^+ concentration on this binding equilibrium, we rewrite the reaction as

$$HHb^+ + O_2 \rightleftharpoons HbO_2 + H^+$$

where HHb^+ denotes a protonated form of hemoglobin. This equation tells us that the O_2-saturation curve of hemoglobin is influenced by the H^+ concentration **(Fig. 5–16)**. Both O_2 and H^+ are bound by hemoglobin, but with inverse affinity. When the oxygen concentration is high, as in the lungs, hemoglobin binds O_2 and releases protons. When the oxygen concentration is

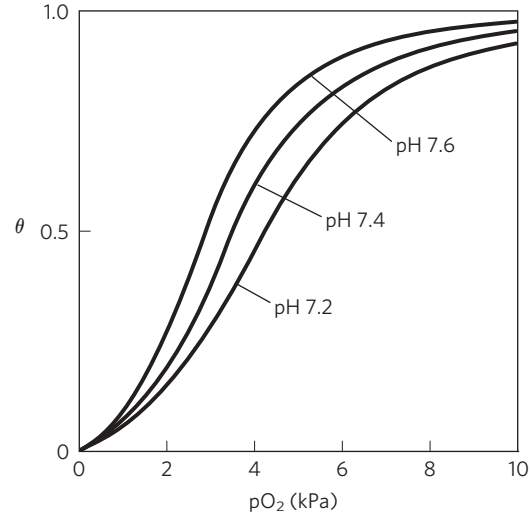
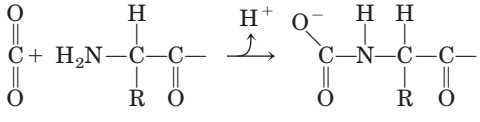


FIGURE 5–16 Effect of pH on oxygen binding to hemoglobin. The pH of blood is 7.6 in the lungs and 7.2 in the tissues. Experimental measurements on hemoglobin binding are often performed at pH 7.4.

low, as in the peripheral tissues, H^+ is bound and O_2 is released.

Oxygen and H^+ are not bound at the same sites in hemoglobin. Oxygen binds to the iron atoms of the hemes, whereas H^+ binds to any of several amino acid residues in the protein. A major contribution to the Bohr effect is made by His^{146} (His HC3) of the β subunits. When protonated, this residue forms one of the ion pairs—to Asp^{94} (Asp FG1)—that helps stabilize deoxyhemoglobin in the T state (Fig. 5–9). The ion pair stabilizes the protonated form of His HC3, giving this residue an abnormally high pK_a in the T state. The pK_a falls to its normal value of 6.0 in the R state because the ion pair cannot form, and this residue is largely unprotonated in oxyhemoglobin at pH 7.6, the blood pH in the lungs. As the concentration of H^+ rises, protonation of His HC3 promotes release of oxygen by favoring a transition to the T state. Protonation of the amino-terminal residues of the α subunits, certain other His residues, and perhaps other groups has a similar effect.

Thus we see that the four polypeptide chains of hemoglobin communicate with each other not only about O_2 binding to their heme groups but also about H^+ binding to specific amino acid residues. And there is still more to the story. Hemoglobin also binds CO_2, again in a manner inversely related to the binding of oxygen. Carbon dioxide binds as a carbamate group to the α-amino group at the amino-terminal end of each globin chain, forming carbaminohemoglobin:

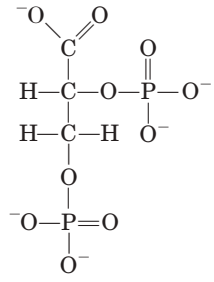


This reaction produces H^+, contributing to the Bohr effect. The bound carbamates also form additional salt bridges (not shown in Fig. 5–9) that help to stabilize the T state and promote the release of oxygen.

When the concentration of carbon dioxide is high, as in peripheral tissues, some CO_2 binds to hemoglobin and the affinity for O_2 decreases, causing its release. Conversely, when hemoglobin reaches the lungs, the high oxygen concentration promotes binding of O_2 and release of CO_2. It is the capacity to communicate ligand-binding information from one polypeptide subunit to the others that makes the hemoglobin molecule so beautifully adapted to integrating the transport of O_2, CO_2, and H^+ by erythrocytes.

Oxygen Binding to Hemoglobin Is Regulated by 2,3-Bisphosphoglycerate

The interaction of **2,3-bisphosphoglycerate (BPG)** with hemoglobin molecules further refines the function of hemoglobin, and provides an example of heterotropic allosteric modulation.


2,3-Bisphosphoglycerate

BPG is present in relatively high concentrations in erythrocytes. When hemoglobin is isolated, it contains substantial amounts of bound BPG, which can be difficult to remove completely. In fact, the O_2-binding curves for hemoglobin that we have examined to this point were obtained in the presence of bound BPG. 2,3-Bisphosphoglycerate is known to greatly reduce the affinity of hemoglobin for oxygen—there is an inverse relationship between the binding of O_2 and the binding of BPG. We can therefore describe another binding process for hemoglobin:

$$HbBPG + O_2 \rightleftharpoons HbO_2 + BPG$$

BPG binds at a site distant from the oxygen-binding site and regulates the O_2-binding affinity of hemoglobin in relation to the pO_2 in the lungs. BPG is important in the physiological adaptation to the lower pO_2 at high altitudes. For a healthy human at sea level, the binding of O_2 to hemoglobin is regulated such that the amount of O_2 delivered to the tissues is nearly 40% of the maximum that could be carried by the blood **(Fig. 5–17).** Imagine that such a person is suddenly transported from sea level to an altitude of 4,500 meters, where the pO_2 is considerably lower. The delivery of O_2 to the tissues is now reduced. However, after just a few hours at the higher altitude, the BPG concentration in the blood has begun to rise, leading to a decrease in the affinity of hemoglobin for oxygen. This adjustment in the BPG level has only a small effect on the binding of O_2 in the lungs but a considerable effect on the release of O_2 in the tissues. As a result, the delivery of oxygen to the tissues is restored to nearly 40% of the O_2 that can be transported by the blood. The situation is reversed when the person returns to sea level. The BPG concentration in erythrocytes also increases in people suffering from **hypoxia**, lowered oxygenation of peripheral tissues due to inadequate functioning of the lungs or circulatory system.

The site of BPG binding to hemoglobin is the cavity between the β subunits in the T state **(Fig. 5–18)**. This cavity is lined with positively charged amino acid residues that interact with the negatively charged groups of BPG. Unlike O_2, only one molecule of BPG is bound to each hemoglobin tetramer. BPG lowers hemoglobin's affinity for oxygen by stabilizing the T state. The transition to the R state narrows the binding pocket for BPG, precluding BPG binding. In the absence

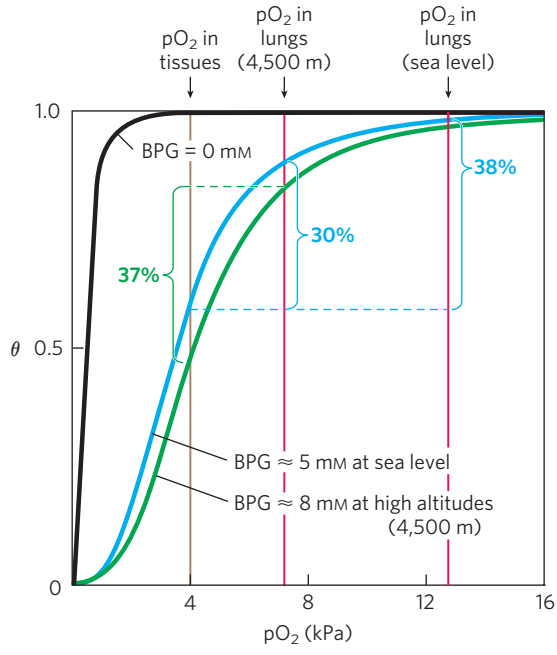


FIGURE 5–17 Effect of BPG on oxygen binding to hemoglobin. The BPG concentration in normal human blood is about 5 mM at sea level and about 8 mM at high altitudes. Note that hemoglobin binds to oxygen quite tightly when BPG is entirely absent, and the binding curve seems to be hyperbolic. In reality, the measured Hill coefficient for O_2-binding cooperativity decreases only slightly (from 3 to about 2.5) when BPG is removed from hemoglobin, but the rising part of the sigmoid curve is confined to a very small region close to the origin. At sea level, hemoglobin is nearly saturated with O_2 in the lungs, but just over 60% saturated in the tissues, so the amount of O_2 released in the tissues is about 38% of the maximum that can be carried in the blood. At high altitudes, O_2 delivery declines by about one-fourth, to 30% of maximum. An increase in BPG concentration, however, decreases the affinity of hemoglobin for O_2, so approximately 37% of what can be carried is again delivered to the tissues.

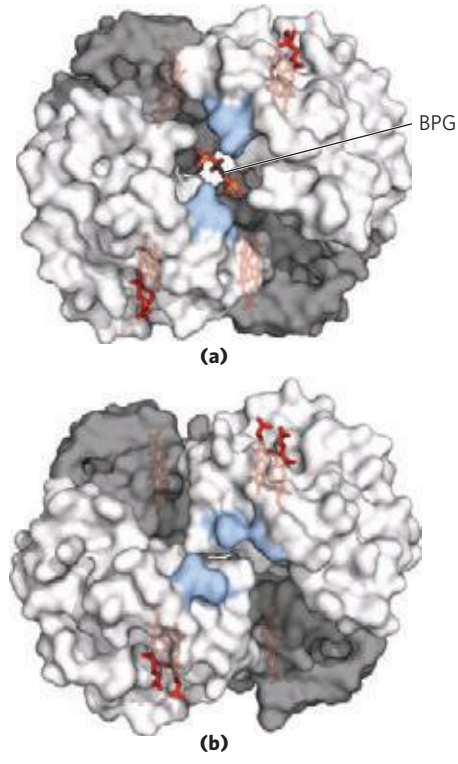


FIGURE 5–18 Binding of BPG to deoxyhemoglobin. (a) BPG binding stabilizes the T state of deoxyhemoglobin (PDB ID 1B86). The negative charges of BPG interact with several positively charged groups (shown in blue in this surface contour image) that surround the pocket between the β subunits on the surface of deoxyhemoglobin in the T state. **(b)** The binding pocket for BPG disappears on oxygenation, following transition to the R state (PDB ID 1BBB). (Compare with Fig. 5–10.)

of BPG, hemoglobin is converted to the R state more easily.

Regulation of oxygen binding to hemoglobin by BPG has an important role in fetal development. Because a fetus must extract oxygen from its mother's blood, fetal hemoglobin must have greater affinity than the maternal hemoglobin for O_2. The fetus synthesizes γ subunits rather than β subunits, forming $\alpha_2\gamma_2$ hemoglobin. This tetramer has a much lower affinity for BPG than normal adult hemoglobin, and a correspondingly higher affinity for O_2. **Oxygen-Binding Proteins—Hemoglobin Is Susceptible to Allosteric Regulation**

Sickle-Cell Anemia Is a Molecular Disease of Hemoglobin

The hereditary human disease sickle-cell anemia demonstrates strikingly the importance of amino acid sequence in determining the secondary, tertiary, and quaternary structures of globular proteins, and thus their biological functions. Almost 500 genetic variants of hemoglobin are known to occur in the human population; all but a few are quite rare. Most variations consist of differences in a single amino acid residue. The effects on hemoglobin structure and function are often minor but can sometimes be extraordinary. Each hemoglobin variation is the product of an altered gene. The variant genes are called alleles. Because humans generally have two copies of each gene, an individual may have two copies of one allele (thus being homozygous for that gene) or one copy of each of two different alleles (thus heterozygous).

Sickle-cell anemia occurs in individuals who inherit the allele for sickle-cell hemoglobin from both parents. The erythrocytes of these individuals are fewer and also abnormal. In addition to an unusually large number of immature cells, the blood contains many long, thin, sickle-shaped erythrocytes **(Fig. 5–19)**. When hemoglobin from sickle cells (called hemoglobin S) is deoxygenated, it becomes insoluble and forms polymers that

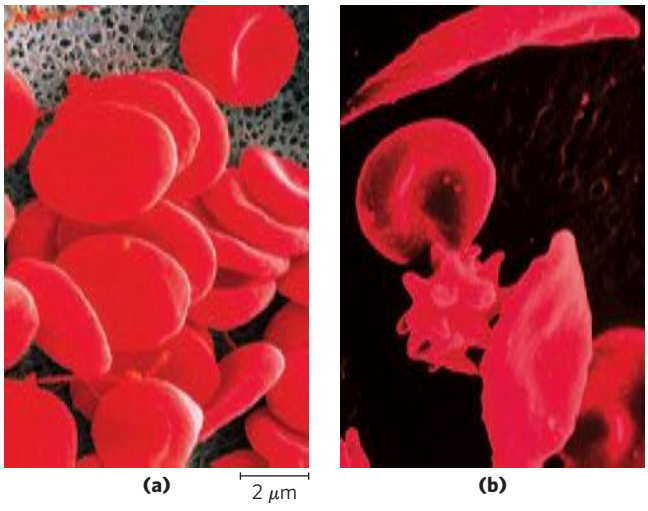


FIGURE 5–19 A comparison of **(a)** uniform, cup-shaped, normal erythrocytes with **(b)** the variably shaped erythrocytes seen in sickle-cell anemia, which range from normal to spiny or sickle-shaped.

aggregate into tubular fibers **(Fig. 5–20)**. Normal hemoglobin (hemoglobin A) remains soluble on deoxygenation. The insoluble fibers of deoxygenated hemoglobin S cause the deformed, sickle shape of the erythrocytes, and the proportion of sickled cells increases greatly as blood is deoxygenated.

The altered properties of hemoglobin S result from a single amino acid substitution, a Val instead of a Glu residue at position 6 in the two β chains. The R group of valine has no electric charge, whereas glutamate has a negative charge at pH 7.4. Hemoglobin S therefore has two fewer negative charges than hemoglobin A (one fewer on each β chain). Replacement of the Glu residue by Val creates a "sticky" hydrophobic contact point at position 6 of the β chain, which is on the outer surface of the molecule. These sticky spots cause deoxyhemoglobin S molecules to associate abnormally with each other, forming the long, fibrous aggregates characteristic of this disorder.

Oxygen-Binding Proteins—Defects in Hb Lead to Serious Genetic Disease

Sickle-cell anemia, as we have noted, occurs in individuals homozygous for the sickle-cell allele of the gene encoding the β subunit of hemoglobin. Individuals who receive the sickle-cell allele from only one parent and are thus heterozygous experience a milder condition called sickle-cell trait; only about 1% of their erythrocytes become sickled on deoxygenation. These individuals may live completely normal lives if they avoid vigorous exercise and other stresses on the circulatory system.

Sickle-cell anemia is life-threatening and painful. People with this disease suffer repeated crises brought on by physical exertion. They become weak, dizzy, and short of breath, and they also experience heart murmurs and an increased pulse rate. The hemoglobin content of their blood is only about half the normal value of 15 to 16 g/100 mL, because sickled cells are very fragile and rupture easily; this results in anemia ("lack of blood"). An even more serious consequence is that capillaries become blocked by the long, abnormally shaped cells, causing severe pain and interfering with normal organ function—a major factor in the early death of many people with the disease.

Without medical treatment, people with sickle-cell anemia usually die in childhood. Curiously, the frequency of the sickle-cell allele in populations is unusually high in

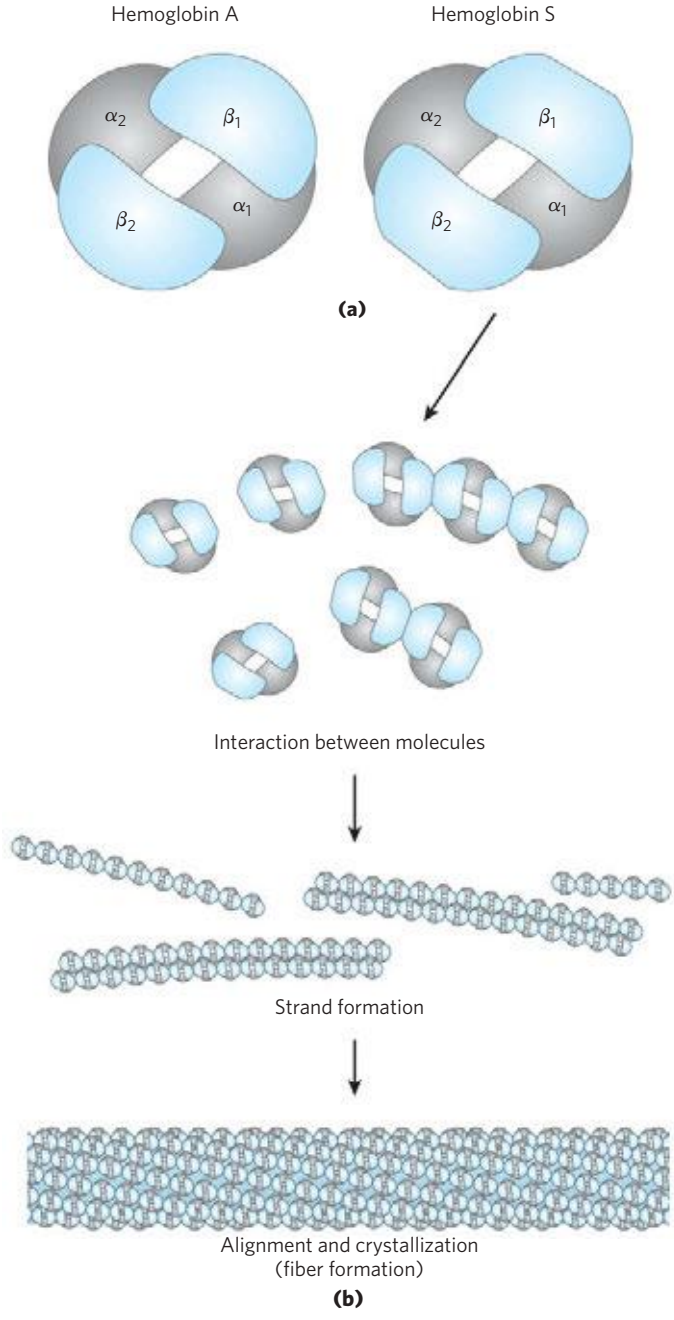


FIGURE 5–20 **Normal and sickle-cell hemoglobin. (a)** Subtle differences between the conformations of hemoglobin A and hemoglobin S result from a single amino acid change in the β chains. **(b)** As a result of this change, deoxyhemoglobin S has a hydrophobic patch on its surface, which causes the molecules to aggregate into strands that align into insoluble fibers.

certain parts of Africa. Investigation into this matter led to the finding that in heterozygous individuals, the allele confers a small but significant resistance to lethal forms of malaria. Natural selection has resulted in an allele population that balances the deleterious effects of the homozygous condition against the resistance to malaria afforded by the heterozygous condition. ■

SUMMARY 5.1 Reversible Binding of a Protein to a Ligand: Oxygen-Binding Proteins

- Protein function often entails interactions with other molecules. A protein binds a molecule, known as a ligand, at its binding site. Proteins may undergo conformational changes when a ligand binds, a process called induced fit. In a multisubunit protein, the binding of a ligand to one subunit may affect ligand binding to other subunits. Ligand binding can be regulated.
- Myoglobin contains a heme prosthetic group, which binds oxygen. Heme consists of a single atom of Fe^{2+} coordinated within a porphyrin. Oxygen binds to myoglobin reversibly; this simple reversible binding can be described by an association constant K_a or a dissociation constant K_d. For a monomeric protein such as myoglobin, the fraction of binding sites occupied by a ligand is a hyperbolic function of ligand concentration.
- Normal adult hemoglobin has four heme-containing subunits, two α and two β, similar in structure to each other and to myoglobin. Hemoglobin exists in two interchangeable structural states, T and R. The T state is most stable when oxygen is not bound. Oxygen binding promotes transition to the R state.
- Oxygen binding to hemoglobin is both allosteric and cooperative. As O_2 binds to one binding site, the hemoglobin undergoes conformational changes that affect the other binding sites—an example of allosteric behavior. Conformational changes between the T and R states, mediated by subunit-subunit interactions, result in cooperative binding; this is described by a sigmoid binding curve and can be analyzed by a Hill plot.
- Two major models have been proposed to explain the cooperative binding of ligands to multisubunit proteins: the concerted model and the sequential model.
- Hemoglobin also binds H^+ and CO_2, resulting in the formation of ion pairs that stabilize the T state and lessen the protein's affinity for O_2 (the Bohr effect). Oxygen binding to hemoglobin is also modulated by 2,3-bisphosphoglycerate, which binds to and stabilizes the T state.
- Sickle-cell anemia is a genetic disease caused by a single amino acid substitution (Glu^6 to Val^6) in each β chain of hemoglobin. The change produces a hydrophobic patch on the surface of the hemoglobin that causes the molecules to aggregate into bundles of fibers. This homozygous condition results in serious medical complications.

5.2 Complementary Interactions between Proteins and Ligands: The Immune System and Immunoglobulins

We have seen how the conformations of oxygen-binding proteins affect and are affected by the binding of small ligands (O_2 or CO) to the heme group. However, most protein-ligand interactions do not involve a prosthetic group. Instead, the binding site for a ligand is more often like the hemoglobin binding site for BPG—a cleft in the protein lined with amino acid residues, arranged to make the binding interaction highly specific. Effective discrimination between ligands is the norm at binding sites, even when the ligands have only minor structural differences.

All vertebrates have an immune system capable of distinguishing molecular "self" from "nonself" and then destroying what is identified as nonself. In this way, the immune system eliminates viruses, bacteria, and other pathogens and molecules that may pose a threat to the organism. On a physiological level, the immune response is an intricate and coordinated set of interactions among many classes of proteins, molecules, and cell types. At the level of individual proteins, the immune response demonstrates how an acutely sensitive and specific biochemical system is built upon the reversible binding of ligands to proteins.

The Immune Response Features a Specialized Array of Cells and Proteins

Immunity is brought about by a variety of **leukocytes** (white blood cells), including **macrophages** and **lymphocytes**, all of which develop from undifferentiated stem cells in the bone marrow. Leukocytes can leave the bloodstream and patrol the tissues, each cell producing one or more proteins capable of recognizing and binding to molecules that might signal an infection.

The immune response consists of two complementary systems, the humoral and cellular immune systems. The **humoral immune system** (Latin *humor*, "fluid") is directed at bacterial infections and extracellular viruses (those found in the body fluids), but can also respond to individual foreign proteins. The **cellular immune system** destroys host cells infected by viruses and also destroys some parasites and foreign tissues.

At the heart of the humoral immune response are soluble proteins called **antibodies** or **immunoglobulins**, often abbreviated **Ig**. Immunoglobulins bind bacteria, viruses, or large molecules identified as foreign and target them for destruction. Making up 20% of blood

protein, the immunoglobulins are produced by **B lymphocytes**, or **B cells**, so named because they complete their development in the *b*one marrow.

The agents at the heart of the cellular immune response are a class of **T lymphocytes**, or **T cells** (so called because the latter stages of their development occur in the *t*hymus), known as **cytotoxic T cells** (**T_C cells**, also called killer T cells). Recognition of infected cells or parasites involves proteins called **T-cell receptors** on the surface of T_C cells. Receptors are proteins, usually found on the outer surface of cells and extending through the plasma membrane; they recognize and bind extracellular ligands, triggering changes inside the cell.

In addition to cytotoxic T cells, there are **helper T cells (T_H cells)**, whose function it is to produce soluble signaling proteins called cytokines, which include the interleukins. T_H cells interact with macrophages. The T_H cells participate only indirectly in the destruction of infected cells and pathogens, stimulating the selective proliferation of those T_C and B cells that can bind to a particular antigen. This process, called **clonal selection**, increases the number of immune system cells that can respond to a particular pathogen. The importance of T_H cells is dramatically illustrated by the epidemic produced by HIV (human immunodeficiency virus), the virus that causes AIDS (acquired immune deficiency syndrome). The primary targets of HIV infection are T_H cells. Elimination of these cells progressively incapacitates the entire immune system. Table 5–2 summarizes the functions of some leukocytes of the immune system.

Each recognition protein of the immune system, either a T-cell receptor or an antibody produced by a B cell, specifically binds some particular chemical structure, distinguishing it from virtually all others. Humans are capable of producing more than 10^8 different antibodies with distinct binding specificities. Given this extraordinary diversity, any chemical structure on the surface of a virus or invading cell will most likely be recognized and bound by one or more antibodies. Antibody diversity is derived from random reassembly of a set of immunoglobulin gene segments through genetic recombination mechanisms that are discussed in Chapter 25 (see Fig. 25–42).

TABLE 5–2 Some Types of Leukocytes Associated with the Immune System

Cell type	Function
Macrophages	Ingest large particles and cells by phagocytosis
B lymphocytes (B cells)	Produce and secrete antibodies
T lymphocytes (T cells)	
Cytotoxic (killer) T cells (T_C)	Interact with infected host cells through receptors on T-cell surface
Helper T cells (T_H)	Interact with macrophages and secrete cytokines (interleukins) that stimulate T_C, T_H, and B cells to proliferate.

A specialized lexicon is used to describe the unique interactions between antibodies or T-cell receptors and the molecules they bind. Any molecule or pathogen capable of eliciting an immune response is called an **antigen**. An antigen may be a virus, a bacterial cell wall, or an individual protein or other macromolecule. A complex antigen may be bound by several different antibodies. An individual antibody or T-cell receptor binds only a particular molecular structure within the antigen, called its **antigenic determinant** or **epitope**.

It would be unproductive for the immune system to respond to small molecules that are common intermediates and products of cellular metabolism. Molecules of M_r <5,000 are generally not antigenic. However, when small molecules are covalently attached to large proteins in the laboratory, they can be used to elicit an immune response. These small molecules are called **haptens**. The antibodies produced in response to protein-linked haptens will then bind to the same small molecules in their free form. Such antibodies are sometimes used in the development of analytical tests described later in this chapter or as a specific ligand in affinity chromatography (see Fig. 3–18c). We now turn to a more detailed description of antibodies and their binding properties.

Antibodies Have Two Identical Antigen-Binding Sites

Immunoglobulin G (IgG) is the major class of antibody molecule and one of the most abundant proteins in the blood serum. IgG has four polypeptide chains: two large ones, called heavy chains, and two light chains, linked by noncovalent and disulfide bonds into a complex of M_r 150,000. The heavy chains of an IgG molecule interact at one end, then branch to interact separately with the light chains, forming a Y-shaped molecule **(Fig. 5–21)**. At the "hinges" separating the base of an IgG molecule from its branches, the immunoglobulin can be cleaved with proteases. Cleavage with the protease papain liberates the basal fragment, called **Fc** because it usually *c*rystallizes readily, and the two branches, called **Fab**, the *a*ntigen-*b*inding fragments. Each branch has a single antigen-binding site.

The fundamental structure of immunoglobulins was first established by Gerald Edelman and Rodney Porter. Each chain is made up of identifiable domains; some are constant in sequence and structure from one IgG to the next, others are variable. The constant domains have a characteristic structure known as the **immunoglobulin fold**, a well-conserved structural motif in the all-β class of proteins (Chapter 4).

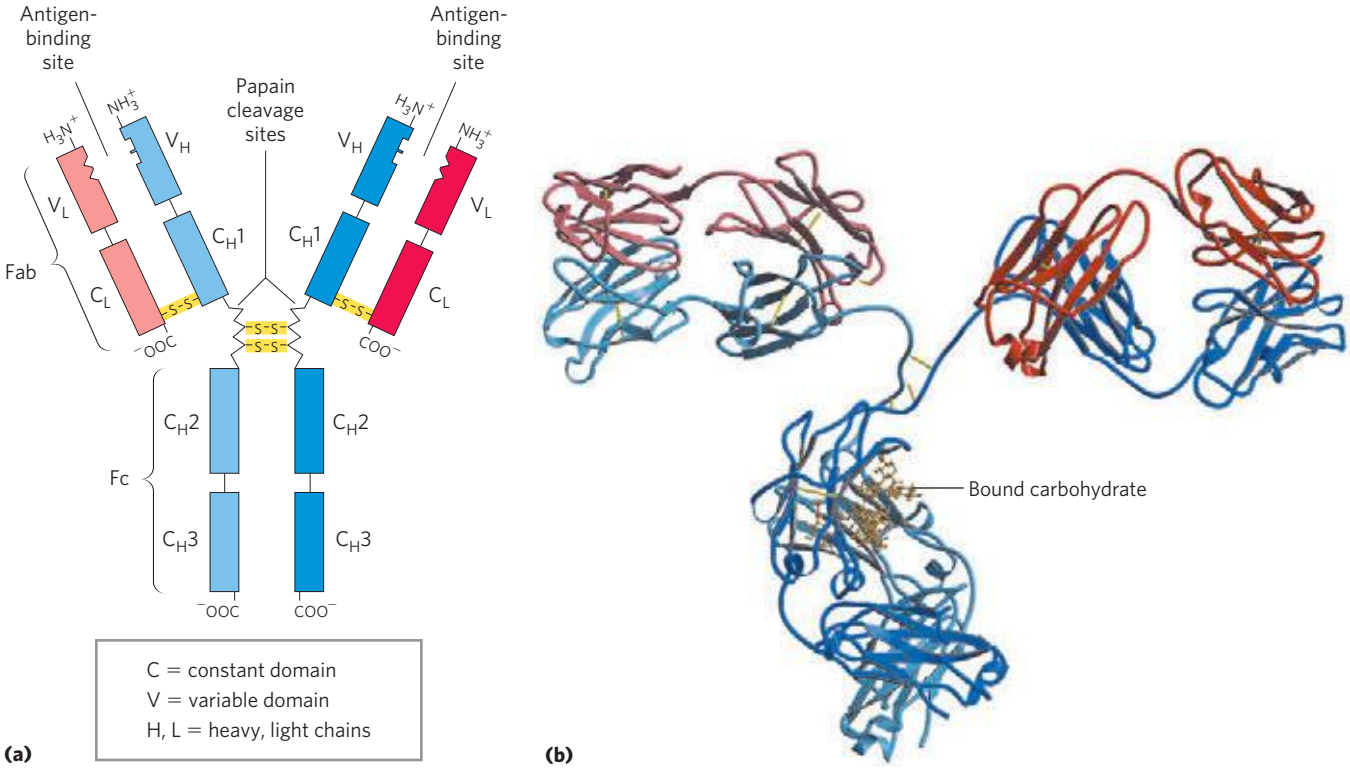


FIGURE 5–21 Immunoglobulin G. (a) Pairs of heavy and light chains combine to form a Y-shaped molecule. Two antigen-binding sites are formed by the combination of variable domains from one light (V_L) and one heavy (V_H) chain. Cleavage with papain separates the Fab and Fc portions of the protein in the hinge region. The Fc portion of the molecule also contains bound carbohydrate (shown in (b)). **(b)** A ribbon model of the first complete IgG molecule to be crystallized and structurally analyzed (PDB ID 1IGT). Although the molecule has two identical heavy chains (two shades of blue) and two identical light chains (two shades of red), it crystallized in the asymmetric conformation shown here. Conformational flexibility may be important to the function of immunoglobulins.

There are three of these constant domains in each heavy chain and one in each light chain. The heavy and light chains also have one variable domain each, in which most of the variability in amino acid sequence is found. The variable domains associate to create the antigen-binding site (Fig. 5–21), allowing formation of an antigen-antibody complex **(Fig. 5–22)**.

In many vertebrates, IgG is but one of five classes of immunoglobulins. Each class has a characteristic type of heavy chain, denoted α, δ, ε, γ, and μ for IgA, IgD, IgE, IgG, and IgM, respectively. Two types of light chain, κ and λ, occur in all classes of immunoglobulins. The overall structures of **IgD** and **IgE** are similar to that of IgG. **IgM** occurs either in a monomeric, membrane-bound form or in a secreted form that is a cross-linked pentamer of this basic structure **(Fig. 5–23)**. **IgA**, found principally in secretions such as saliva, tears, and milk, can be a monomer, dimer, or trimer. IgM is the first antibody to

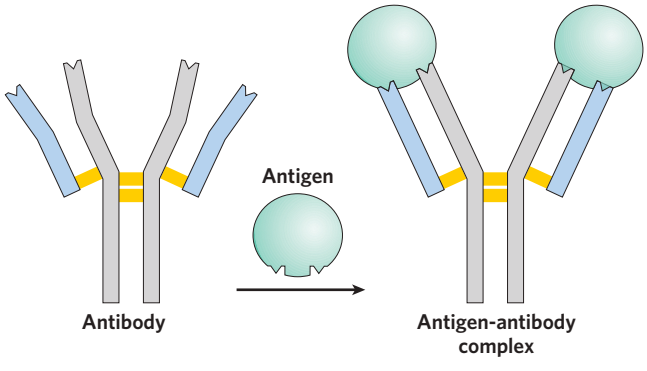


FIGURE 5–22 Binding of IgG to an antigen. To generate an optimal fit for the antigen, the binding sites of IgG often undergo slight conformational changes. Such induced fit is common to many protein-ligand interactions.

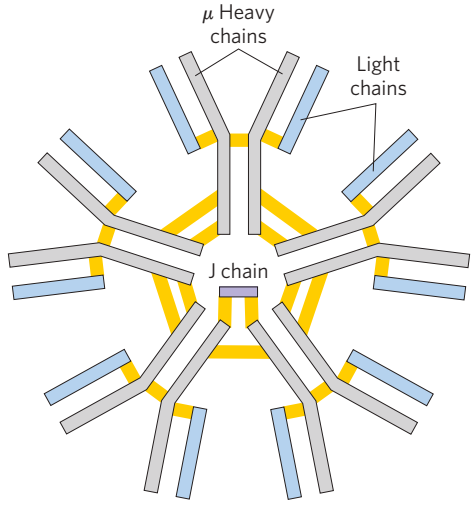


FIGURE 5–23 IgM pentamer of immunoglobulin units. The pentamer is cross-linked with disulfide bonds (yellow). The J chain is a polypeptide of M_r 20,000 found in both IgA and IgM.

be made by B lymphocytes and the major antibody in the early stages of a primary immune response. Some B cells soon begin to produce IgD (with the same antigen-binding site as the IgM produced by the same cell), but the particular function of IgD is less clear.

The IgG described above is the major antibody in secondary immune responses, which are initiated by a class of B cells called memory B cells. As part of the organism's ongoing immunity to antigens already encountered and dealt with, IgG is the most abundant immunoglobulin in the blood. When IgG binds to an invading bacterium or virus, it activates certain leukocytes such as macrophages to engulf and destroy the invader, and also activates some other parts of the immune response. Receptors on the macrophage surface recognize and bind the Fc region of IgG. When these Fc receptors bind an antibody-pathogen complex, the macrophage engulfs the complex by phagocytosis **(Fig. 5–24)**.

IgE plays an important role in the allergic response, interacting with basophils (phagocytic leukocytes) in the blood and with histamine-secreting cells called mast cells, which are widely distributed in tissues. This immunoglobulin binds, through its Fc region, to special Fc receptors on the basophils or mast cells. In this form, IgE serves as a receptor for antigen. If antigen is bound, the cells are induced to secrete histamine and other biologically active amines that cause dilation and increased permeability of blood vessels. These effects on the blood vessels are thought to facilitate the movement of immune system cells and proteins to sites of inflammation. They also produce the symptoms normally associated with allergies. Pollen or other allergens are recognized as foreign, triggering an immune response normally reserved for pathogens. ■

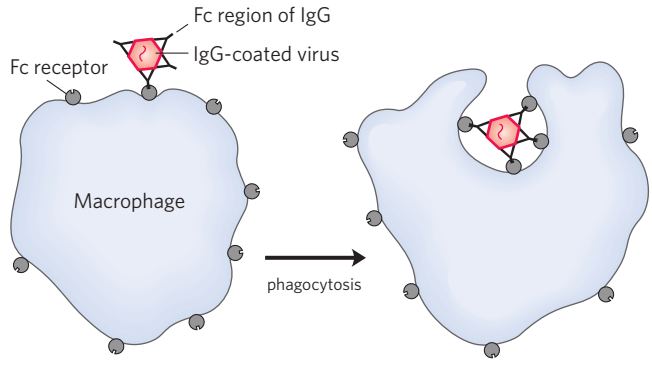
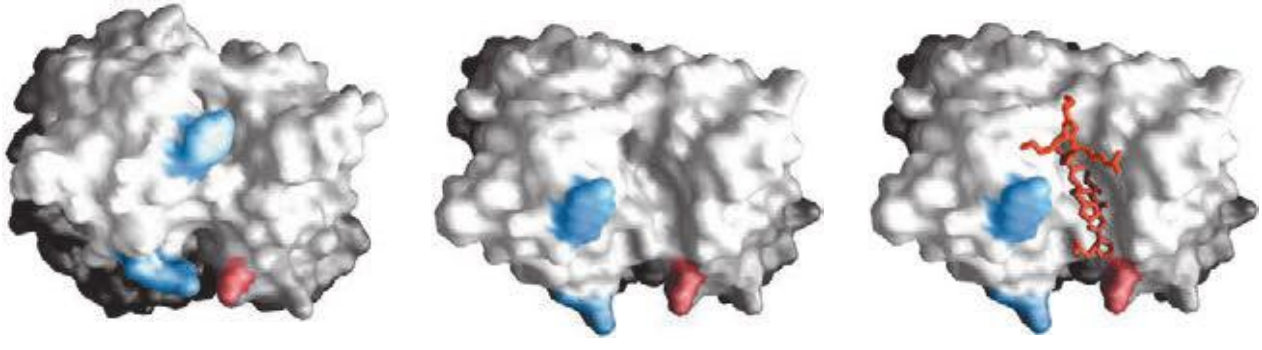


FIGURE 5–24 Phagocytosis of an antibody-bound virus by a macrophage. The Fc regions of antibodies bound to the virus now bind to Fc receptors on the surface of a macrophage, triggering the macrophage to engulf and destroy the virus.

Antibodies Bind Tightly and Specifically to Antigen

The binding specificity of an antibody is determined by the amino acid residues in the variable domains of its heavy and light chains. Many residues in these domains are variable, but not equally so. Some, particularly those lining the antigen-binding site, are hypervariable—especially likely to differ. Specificity is conferred by chemical complementarity between the antigen and its specific binding site, in terms of shape and the location of charged, nonpolar, and hydrogen-bonding groups. For example, a binding site with a negatively charged group may bind an antigen with a positive charge in the complementary position. In many instances, complementarity is achieved interactively as the structures of antigen and binding site influence each other as they come closer together. Conformational changes in the antibody and/or the antigen then allow the complementary groups to interact fully. This is an example of induced fit. The complex of a peptide derived from HIV (a model antigen) and an Fab molecule, shown in **Figure 5–25**, illustrates some of these properties. The changes in structure observed on antigen binding are particularly striking in this example.

A typical antibody-antigen interaction is quite strong, characterized by K_d values as low as 10^{-10} M



(a) Conformation with no antigen bound

(b) Antigen bound (but not shown)

(c) Antigen bound (shown)

FIGURE 5–25 Induced fit in the binding of an antigen to IgG. The molecule here, shown in surface contour, is the Fab fragment of an IgG. The antigen is a small peptide derived from HIV. Two residues in the heavy chain (blue) and one in the light chain (red) are colored to provide visual points of reference. **(a)** View of the Fab fragment in the absence of antigen, looking down on the antigen-binding site (PDB ID 1GGC). **(b)** The same view, but with the Fab fragment in the "bound" conformation (PDB ID 1GGI); the antigen is omitted to provide an unobstructed view of the altered binding site. Note how the binding cavity has enlarged and several groups have shifted position. **(c)** The same view as (b), but with the antigen in the binding site, pictured as a red stick structure.

(recall that a lower K_d corresponds to a stronger binding interaction; see Table 5–1). The K_d reflects the energy derived from the various ionic, hydrogen-bonding, hydrophobic, and van der Waals interactions that stabilize the binding. The binding energy required to produce a K_d of 10^{-10} M is about 65 kJ/mol.

The Antibody-Antigen Interaction Is the Basis for a Variety of Important Analytical Procedures

The extraordinary binding affinity and specificity of antibodies make them valuable analytical reagents. Two types of antibody preparations are in use: polyclonal and monoclonal. **Polyclonal antibodies** are those produced by many different B lymphocytes responding to one antigen, such as a protein injected into an animal. Cells in the population of B lymphocytes produce antibodies that bind specific, different epitopes within the antigen. Thus, polyclonal preparations contain a mixture of antibodies that recognize different parts of the protein. **Monoclonal antibodies**, in contrast, are synthesized by a population of identical B cells (a **clone**) grown in cell culture. These antibodies are homogeneous, all recognizing the same epitope. The techniques for producing monoclonal antibodies were developed by Georges Köhler and Cesar Milstein.

Georges Köhler, 1946–1995

Cesar Milstein, 1927–2002

The specificity of antibodies has practical uses. A selected antibody can be covalently attached to a resin and used in a chromatography column of the type shown in Figure 3–17c . When a mixture of proteins is added to the column, the antibody specifically binds its target protein and retains it on the column while other proteins are washed through. The target protein can then be eluted from the resin by a salt solution or some other agent. This is a powerful protein analytical tool.

In another versatile analytical technique, an antibody is attached to a radioactive label or some other reagent that makes it easy to detect. When the antibody binds the target protein, the label reveals the presence of the protein in a solution or its location in a gel or even a living cell. Several variations of this procedure are illustrated in **Figure 5–26**.

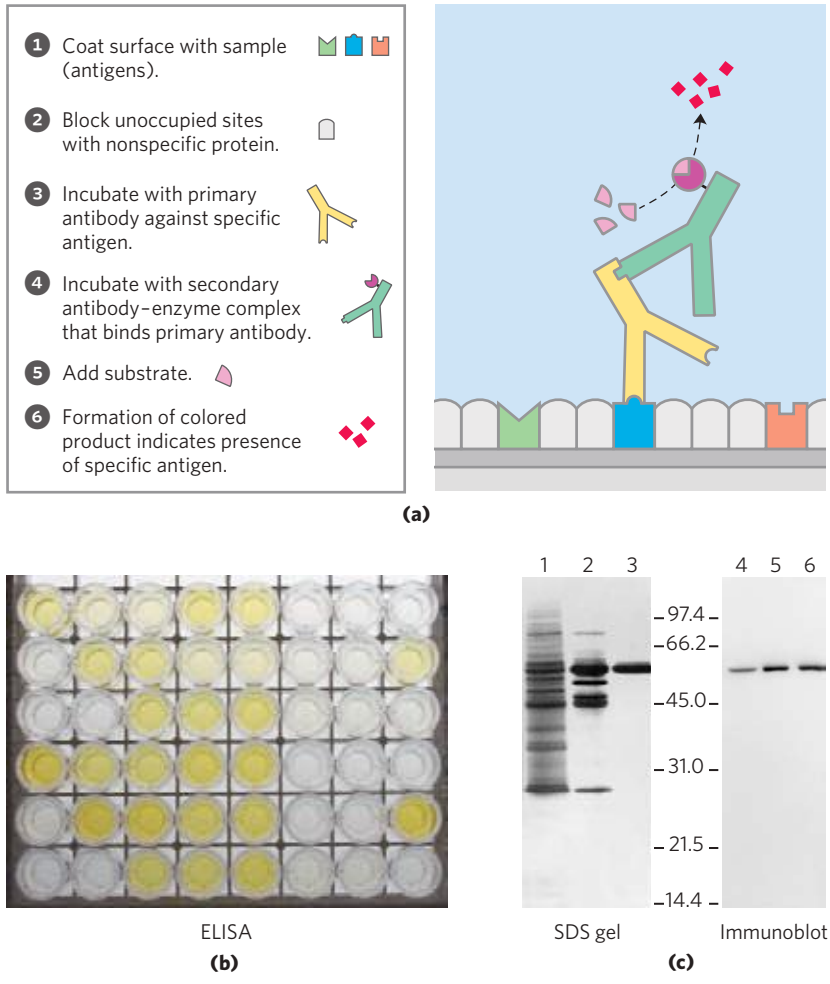


FIGURE 5–26 Antibody techniques. The specific reaction of an antibody with its antigen is the basis of several techniques that identify and quantify a specific protein in a complex sample. **(a)** A schematic representation of the general method. **(b)** An ELISA to test for the presence of herpes simplex virus (HSV) antibodies in blood samples. Wells were coated with an HSV antigen, to which antibodies against HSV will bind. The second antibody is anti-human IgG linked to horseradish peroxidase. Following completion of the steps shown in (a), blood samples with greater amounts of HSV antibody turn brighter yellow. **(c)** An immunoblot. Lanes 1 to 3 are from an SDS gel; samples from successive stages in the purification of a protein kinase were separated and stained with Coomassie blue. Lanes 4 to 6 show the same samples, but these were electrophoretically transferred to a nitrocellulose membrane after separation on an SDS gel. The membrane was then "probed" with antibody against the protein kinase. The numbers between the SDS gel and the immunoblot indicate M_r in thousands.

An **ELISA** (enzyme-linked immunosorbent assay) can be used to rapidly screen for and quantify an antigen in a sample (Fig. 5–26b). Proteins in the sample are adsorbed to an inert surface, usually a 96-well polystyrene plate. The surface is washed with a solution of an inexpensive nonspecific protein (often casein from nonfat dry milk powder) to block proteins introduced in subsequent steps from adsorbing to unoccupied sites. The surface is then treated with a solution containing the primary antibody—an antibody against the protein of interest. Unbound antibody is washed away, and the surface is treated with a solution containing a secondary antibody—antibody against the primary antibody—linked to an enzyme that catalyzes a reaction that forms a colored product. After unbound secondary antibody is washed away, the substrate of the antibody-linked enzyme is added. Product formation (monitored as color intensity) is proportional to the concentration of the protein of interest in the sample.

In an **immunoblot assay**, also called a **Western blot** (Fig. 5–26c), proteins that have been separated by gel electrophoresis are transferred electrophoretically to a nitrocellulose membrane. The membrane is blocked (as described above for ELISA), then treated successively with primary antibody, secondary antibody linked to enzyme, and substrate. A colored precipitate forms only along the band containing the protein of interest. Immunoblotting allows the detection of a minor component in a sample and provides an approximation of its molecular weight. Immunoblotting

We will encounter other aspects of antibodies in later chapters. They are extremely important in medicine and can tell us much about the structure of proteins and the action of genes.

SUMMARY 5.2 Complementary Interactions between Proteins and Ligands: The Immune System and Immunoglobulins

- The immune response is mediated by interactions among an array of specialized leukocytes and their associated proteins. T lymphocytes produce T-cell receptors. B lymphocytes produce immunoglobulins. In a process called clonal selection, helper T cells induce the proliferation of B cells and cytotoxic T cells that produce immunoglobulins or of T-cell receptors that bind to a specific antigen.
- Humans have five classes of immunoglobulins, each with different biological functions. The most abundant class is IgG, a Y-shaped protein with two heavy and two light chains. The domains near the upper ends of the Y are hypervariable within the broad population of IgGs and form two antigen-binding sites.
- A given immunoglobulin generally binds to only a part, called the epitope, of a large antigen. Binding often involves a conformational change in the IgG, an induced fit to the antigen.
- The exquisite binding specificity of immunoglobins is exploited in analytical techniques such as the ELISA and the immunoblot.

5.3 Protein Interactions Modulated by Chemical Energy: Actin, Myosin, and Molecular Motors

Organisms move. Cells move. Organelles and macromolecules within cells move. Most of these movements arise from the activity of a fascinating class of protein-based molecular motors. Fueled by chemical energy, usually derived from ATP, large aggregates of motor proteins undergo cyclic conformational changes that accumulate into a unified, directional force—the tiny force that pulls apart chromosomes in a dividing cell, and the immense force that levers a pouncing, quarter-ton jungle cat into the air.

The interactions among motor proteins, as you might predict, feature complementary arrangements of ionic, hydrogen-bonding, hydrophobic, and van der Waals interactions at protein binding sites. In motor proteins, however, these interactions achieve exceptionally high levels of spatial and temporal organization.

Motor proteins underlie the contraction of muscles, the migration of organelles along microtubules, the rotation of bacterial flagella, and the movement of some proteins along DNA. Proteins called kinesins and dyneins move along microtubules in cells, pulling along organelles or reorganizing chromosomes during cell division. An interaction of dynein with microtubules brings about the motion of eukaryotic flagella and cilia. Flagellar motion in bacteria involves a complex rotational motor at the base of the flagellum (see Fig. 19–41). Helicases, polymerases, and other proteins move along DNA as they carry out their functions in DNA metabolism (Chapter 25). Here, we focus on the well-studied example of the contractile proteins of vertebrate skeletal muscle as a paradigm for how proteins translate chemical energy into motion.

The Major Proteins of Muscle Are Myosin and Actin

The contractile force of muscle is generated by the interaction of two proteins, myosin and actin. These proteins are arranged in filaments that undergo transient interactions and slide past each other to bring about contraction. Together, actin and myosin make up more than 80% of the protein mass of muscle.

Myosin (M_r 520,000) has six subunits: two heavy chains (each of M_r 220,000) and four light chains (each of M_r 20,000). The heavy chains account for much of the overall structure. At their carboxyl termini, they are arranged as extended α helices, wrapped

around each other in a fibrous, left-handed coiled coil similar to that of α-keratin **(Fig. 5–27a)**. At its amino terminus, each heavy chain has a large globular domain containing a site where ATP is hydrolyzed. The light chains are associated with the globular domains. When myosin is treated briefly with the protease trypsin, much of the fibrous tail is cleaved off, dividing the protein into components called light and heavy meromyosin (Fig. 5–27b). The globular domain—called myosin subfragment 1, or S1, or simply the myosin head group—is liberated from heavy meromyosin by cleavage with papain. The S1 fragment is the motor domain that makes muscle contraction possible. S1 fragments can be crystallized, and their overall structure as determined by Ivan Rayment and Hazel Holden is shown in Figure 5–27c.

In muscle cells, molecules of myosin aggregate to form structures called **thick filaments (Fig. 5–28a)**.

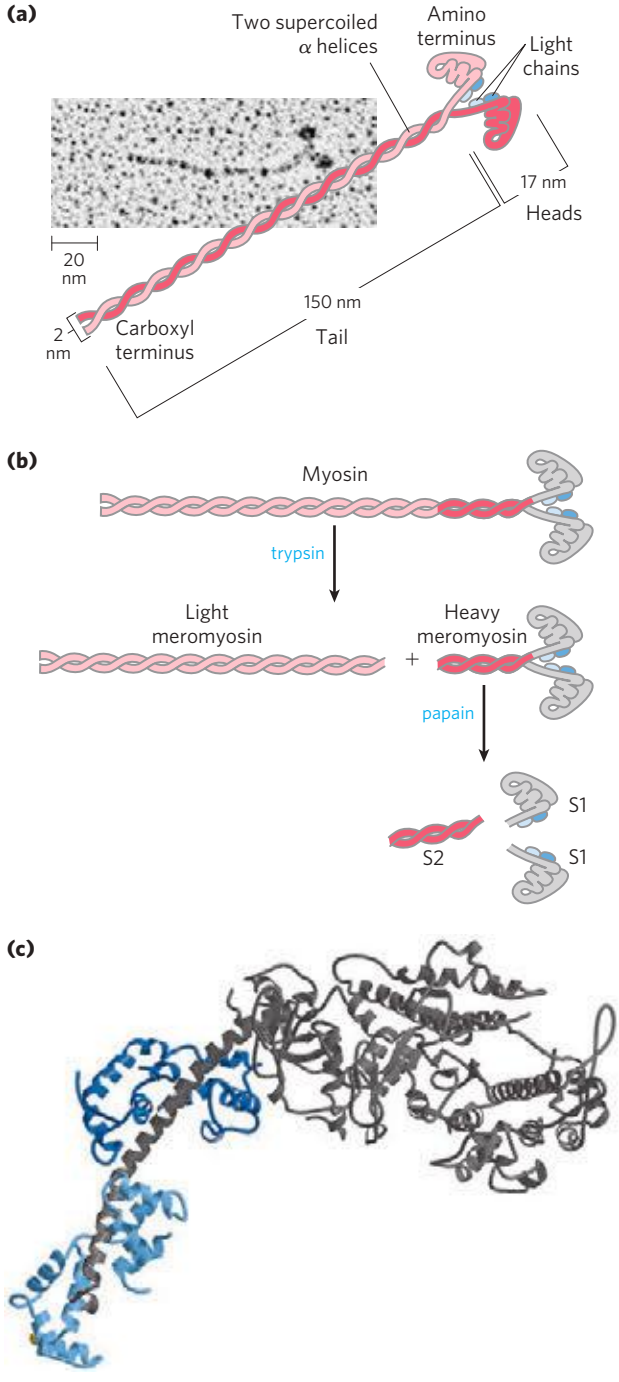


FIGURE 5–27 Myosin. (a) Myosin has two heavy chains (in two shades of pink), the carboxyl termini forming an extended coiled coil (tail) and the amino termini having globular domains (heads). Two light chains (blue) are associated with each myosin head. **(b)** Cleavage with trypsin and papain separates the myosin heads (S1 fragments) from the tails. **(c)** Ribbon representation of the myosin S1 fragment (from coordinates supplied by Ivan Rayment). The heavy chain is in gray, the two light chains in two shades of blue.

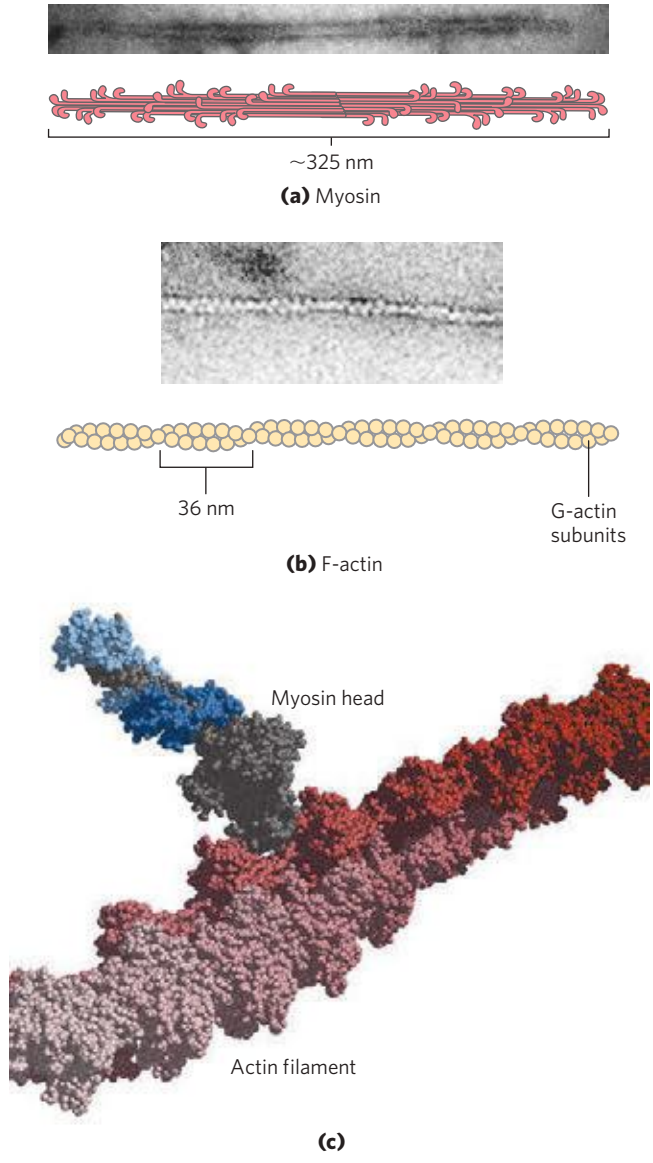


FIGURE 5–28 The major components of muscle. (a) Myosin aggregates to form a bipolar structure called a thick filament. **(b)** F-actin is a filamentous assemblage of G-actin monomers that polymerize two by two, giving the appearance of two filaments spiraling about one another in a right-handed fashion. **(c)** Space-filling model of an actin filament (shades of red) with one myosin head (gray and two shades of blue) bound to an actin monomer within the filament (from coordinates supplied by Ivan Rayment).

These rodlike structures are the core of the contractile unit. Within a thick filament, several hundred myosin molecules are arranged with their fibrous "tails" associated to form a long bipolar structure. The globular domains project from either end of this structure, in regular stacked arrays.

The second major muscle protein, **actin**, is abundant in almost all eukaryotic cells. In muscle, molecules of monomeric actin, called G-actin (*g*lobular actin; M_r 42,000), associate to form a long polymer called F-actin (*f*ilamentous actin). The **thin filament** consists of F-actin (Fig. 5–28b), along with the proteins troponin and tropomyosin (discussed below). The filamentous parts of thin filaments assemble as successive monomeric actin molecules add to one end. On addition, each monomer binds ATP, then hydrolyzes it to ADP, so every actin molecule in the filament is complexed to ADP. This ATP hydrolysis by actin functions only in the assembly of the filaments; it does not contribute directly to the energy expended in muscle contraction. Each actin monomer in the thin filament can bind tightly and specifically to one myosin head group (Fig. 5–28c).

Additional Proteins Organize the Thin and Thick Filaments into Ordered Structures

Skeletal muscle consists of parallel bundles of **muscle fibers**, each fiber a single, very large, multinucleated cell, 20 to 100 μm in diameter, formed from many cells fused together; a single fiber often spans the length of the muscle. Each fiber contains about 1,000 **myofibrils**, 2 μm in diameter, each consisting of a vast number of regularly arrayed thick and thin filaments complexed to other proteins **(Fig. 5–29)**. A system of flat membranous vesicles called the **sarcoplasmic reticulum** surrounds each myofibril. Examined under the electron microscope, muscle fibers reveal alternating regions of high and low electron density, called the **A bands** and **I bands** (Fig. 5–29b, c). The A and I bands arise from the arrangement of thick and thin filaments, which are aligned and partially overlapping. The I band is the region of the bundle that in cross section would contain only thin filaments. The darker A band stretches the length of the thick filament and includes the region where parallel thick and thin filaments overlap. Bisecting the I band is a thin structure called the **Z disk**, perpendicular to the thin filaments and serving as an anchor to which the thin filaments are attached. The A band, too, is bisected by a thin line, the **M line** or M disk, a region of high electron density in the middle of the thick filaments. The entire contractile unit, consisting of bundles of thick filaments interleaved at either end with bundles of thin filaments, is called the **sarcomere**. The arrangement of interleaved bundles allows the thick and thin filaments to slide past each other (by a mechanism discussed below), causing a progressive shortening of each sarcomere **(Fig. 5–30)**.

The thin actin filaments are attached at one end to the Z disk in a regular pattern. The assembly includes the minor muscle proteins ***α*-actinin**, **desmin**, and

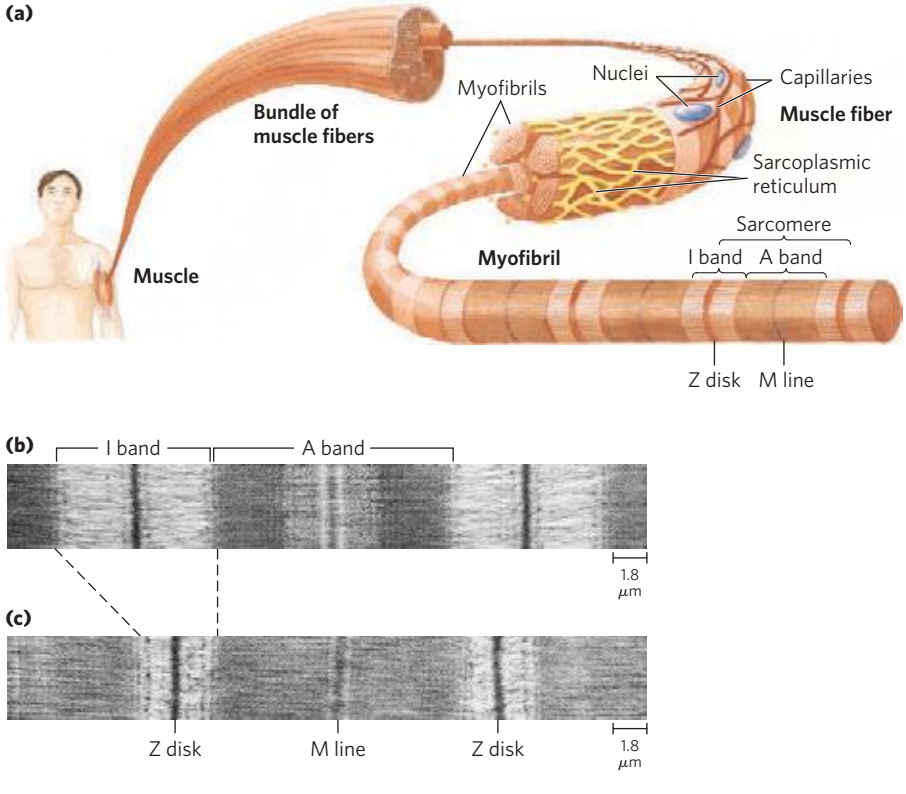


FIGURE 5–29 Skeletal muscle. (a) Muscle fibers consist of single, elongated, multinucleated cells that arise from the fusion of many precursor cells. The fibers are made up of many myofibrils (only six are shown here for simplicity) surrounded by the membranous sarcoplasmic reticulum. The organization of thick and thin filaments in a myofibril gives it a striated appearance. When muscle contracts, the I bands narrow and the Z disks come closer together, as seen in electron micrographs of **(b)** relaxed and **(c)** contracted muscle.

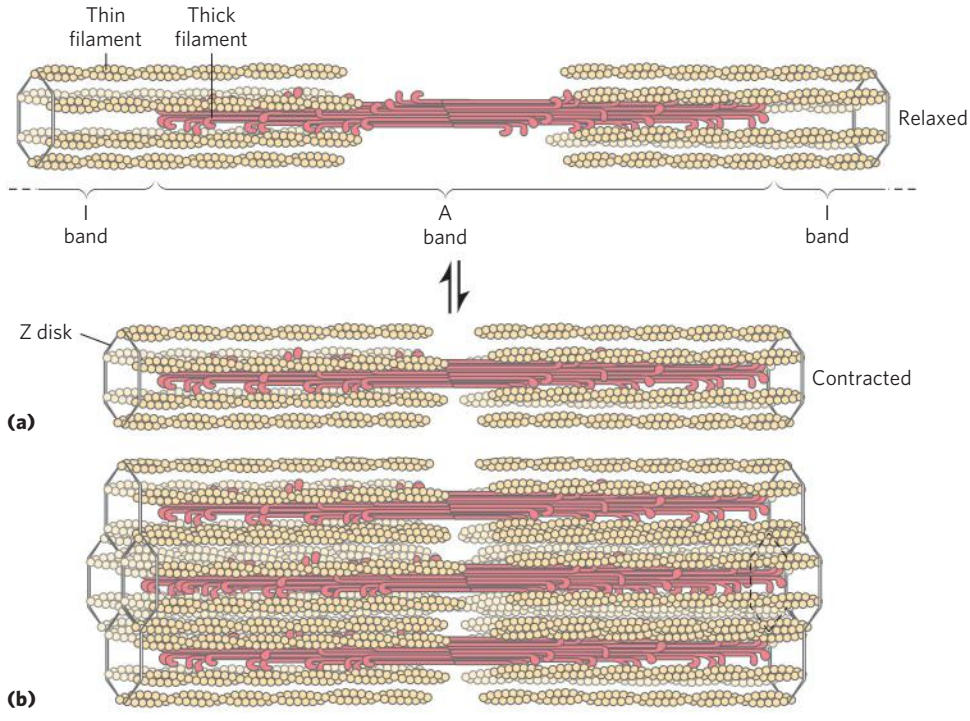


FIGURE 5–30 Muscle contraction. Thick filaments are bipolar structures created by the association of many myosin molecules. **(a)** Muscle contraction occurs by the sliding of the thick and thin filaments past each other so that the Z disks in neighboring I bands draw closer together. **(b)** The thick and thin filaments are interleaved such that each thick filament is surrounded by six thin filaments.

vimentin. Thin filaments also contain a large protein called **nebulin** (~7,000 amino acid residues), thought to be structured as an α helix that is long enough to span the length of the filament. The M line similarly organizes the thick filaments. It contains the proteins **paramyosin**, **C-protein**, and **M-protein**. Another class of proteins called **titins**, the largest single polypeptide chains discovered thus far (the titin of human cardiac muscle has 26,926 amino acid residues), link the thick filaments to the Z disk, providing additional organization to the overall structure. Among their structural functions, the proteins nebulin and titin are believed to act as "molecular rulers," regulating the length of the thin and thick filaments, respectively. Titin extends from the Z disk to the M line, regulating the length of the sarcomere itself and preventing overextension of the muscle. The characteristic sarcomere length varies from one muscle tissue to the next in a vertebrate, largely due to the different titin variants in the tissues.

Myosin Thick Filaments Slide along Actin Thin Filaments

The interaction between actin and myosin, like that between all proteins and ligands, involves weak bonds. When ATP is not bound to myosin, a face on the myosin head group binds tightly to actin **(Fig. 5–31)**. When ATP binds to myosin and is hydrolyzed to ADP and phosphate, a coordinated and cyclic series of conformational changes occurs in which myosin releases the F-actin subunit and binds another subunit farther along the thin filament.

The cycle has four major steps (Fig. 5–31). In step ❶, ATP binds to myosin and a cleft in the myosin molecule opens, disrupting the actin-myosin interaction so that the bound actin is released. ATP is then hydrolyzed in step ❷, causing a conformational change in the protein to a "high-energy" state that moves the myosin head and changes its orientation in relation to the actin thin filament. Myosin then binds weakly to an F-actin subunit closer to the Z disk than the one just released. As the phosphate product of ATP hydrolysis is released from myosin in step ❸, another conformational change occurs in which the myosin cleft closes, strengthening the myosin-actin binding. This is followed quickly by step ❹, a "power stroke" during which the conformation of the myosin head returns to the original resting state, its orientation relative to the bound actin changing so as to pull the tail of the myosin toward the Z disk. ADP is then released to complete the cycle. Each cycle generates about 3 to 4 pN (piconewtons) of force and moves the thick filament 5 to 10 nm relative to the thin filament.

Because there are many myosin heads in a thick filament, at any given moment some (probably 1% to 3%) are bound to thin filaments. This prevents thick filaments from slipping backward when an individual myosin head releases the actin subunit to which it was bound. The thick filament thus actively slides forward past the adjacent thin filaments. This process, coordinated among the many sarcomeres in a muscle fiber, brings about muscle contraction.

The interaction between actin and myosin must be regulated so that contraction occurs only in response to appropriate signals from the nervous system. The

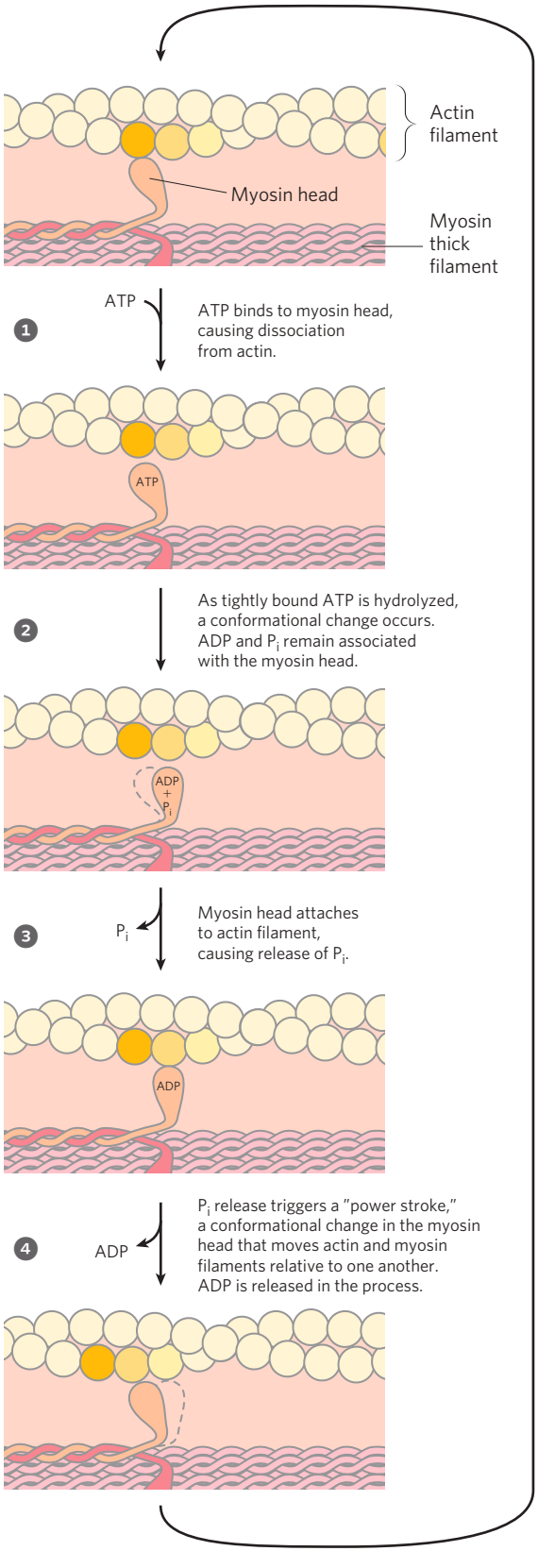


FIGURE 5–31 Molecular mechanism of muscle contraction. Conformational changes in the myosin head that are coupled to stages in the ATP hydrolytic cycle cause myosin to successively dissociate from one actin subunit, then associate with another farther along the actin filament. In this way the myosin heads slide along the thin filaments, drawing the thick filament array into the thin filament array (see Fig. 5-30).

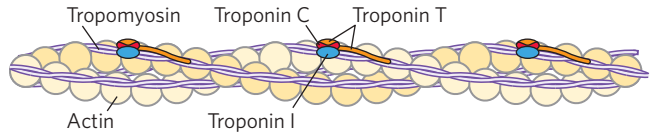


FIGURE 5–32 Regulation of muscle contraction by tropomyosin and troponin. Tropomyosin and troponin are bound to F-actin in the thin filaments. In the relaxed muscle, these two proteins are arranged around the actin filaments so as to block the binding sites for myosin. Tropomyosin is a two-stranded coiled coil of α helices, the same structural motif as in α-keratin (see Fig. 4-11). It forms head-to-tail polymers twisting around the two actin chains. Troponin is attached to the actin-tropomyosin complex at regular intervals of 38.5 nm. Troponin consists of three different subunits: I, C, and T. Troponin I prevents binding of the myosin head to actin; troponin C has a binding site for Ca^{2+}; and troponin T links the entire troponin complex to tropomyosin. When the muscle receives a neural signal to initiate contraction, Ca^{2+} is released from the sarcoplasmic reticulum (see Fig. 5-29a) and binds to troponin C. This causes a conformational change in troponin C, which alters the positions of troponin I and tropomyosin so as to relieve the inhibition by troponin I and allow muscle contraction.

regulation is mediated by a complex of two proteins, **tropomyosin** and **troponin (Fig. 5–32)**. Tropomyosin binds to the thin filament, blocking the attachment sites for the myosin head groups. Troponin is a Ca^{2+}-binding protein. A nerve impulse causes release of Ca^{2+} from the sarcoplasmic reticulum. The released Ca^{2+} binds to troponin (another protein-ligand interaction) and causes a conformational change in the tropomyosin-troponin complexes, exposing the myosin-binding sites on the thin filaments. Contraction follows.

Working skeletal muscle requires two types of molecular functions that are common in proteins—binding and catalysis. The actin-myosin interaction, a protein-ligand interaction like that of immunoglobulins with antigens, is reversible and leaves the participants unchanged. When ATP binds myosin, however, it is hydrolyzed to ADP and P_i. Myosin is not only an actin-binding protein, it is also an ATPase—an enzyme. The function of enzymes in catalyzing chemical transformations is the topic of the next chapter.

SUMMARY 5.3 Protein Interactions Modulated by Chemical Energy: Actin, Myosin, and Molecular Motors

- Protein-ligand interactions achieve a special degree of spatial and temporal organization in motor proteins. Muscle contraction results from choreographed interactions between myosin and actin, coupled to the hydrolysis of ATP by myosin.
- Myosin consists of two heavy and four light chains, forming a fibrous coiled coil (tail) domain and a globular (head) domain. Myosin molecules are organized into thick filaments, which slide past thin filaments composed largely of actin. ATP hydrolysis in myosin is coupled to

a series of conformational changes in the myosin head, leading to dissociation of myosin from one F-actin subunit and its eventual reassociation with another, farther along the thin filament. The myosin thus slides along the actin filaments.

- Muscle contraction is stimulated by the release of Ca^{2+} from the sarcoplasmic reticulum. The Ca^{2+} binds to the protein troponin, leading to a conformational change in a troponin-tropomyosin complex that triggers the cycle of actin-myosin interactions.

Key Terms

Terms in bold are defined in the glossary.

ligand 157
binding site 157
induced fit 157
heme 158
porphyrin 158
globins 159
equilibrium expression 160
association constant, K_a 160
dissociation constant, K_d 160
allosteric protein 166
Hill equation 167
Bohr effect 170
lymphocytes 174
antibody 174
immunoglobulin 174
B lymphocyte or **B cell** 175
T lymphocyte or **T cell** 175
antigen 175
epitope 175
hapten 175
immunoglobulin fold 175
polyclonal antibodies 178
monoclonal antibodies 178
ELISA 179
immunoblotting 179
Western blotting 179
myosin 179
actin 181
sarcomere 181

Further Reading

Oxygen-Binding Proteins

Changeux, J.P. & Edelstein, S.J. (2005) Allosteric mechanisms of signal transduction *Science* **308**, 1424–1428.

Koder, R.L., Anderson, J.L.R., Solomon, L.A., Reddy, K.S., Moser, C.C., & Dutton, P.L. (2009) Design and engineering of an O_2 transport protein. *Nature* **458**, 305–309.

Koshland, D.E., Jr., Nemethy, G., & Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365–385.
The paper that introduced the sequential model.

Laberge, M. & Yonetani, T. (2007) Common dynamics of globin family proteins. *IUBMB Life* **59**, 528–534.

Monod, J., Wyman, J., & Changeux, J.-P. (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118.
The concerted model was first proposed in this landmark paper.

Perutz, M.F., Wilkinson, A.J., Paoli, M., & Dodson, G.G. (1998) The stereochemical mechanism of the cooperative effects in hemoglobin revisited. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 1–34.

Immune System Proteins

Cooper, M.D. & Alder, M.N. (2006) The evolution of adaptive immune systems. *Cell* **124**, 815–822.
An interesting essay tracing the origins of our immune system.

Flajnik, M.F. & Kasahara, M. (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* **11**, 47–59.

Kindt, T.J., Osborne, B.A., & Goldsby, R.A. (2007) *Kuby Immunology*, 6th edn, W. H. Freeman and Company, New York.

Ploegh, H.L. (1998) Viral strategies of immune evasion. *Science* **280**, 248–253.

Yewdell, J.W. & Haeryfar, S.M.M. (2005) Understanding presentation of viral antigens to CD8(+) T cells in vivo: the key to rational vaccine design. *Annu. Rev. Immunol.* **23**, 651–682.

Molecular Motors

Geeves, M.A. & Holmes, K.C. (1999) Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* **68**, 687–728.

Geigel, C. & Schmidt, C.F. (2011) Moving into the cell: single-molecule studies of molecular motors in complex environments. *Nat. Rev. Mol. Cell Biol.* **12**, 163–176.

Huxley, H.E. (1998) Getting to grips with contraction: the interplay of structure and biochemistry. *Trends Biochem. Sci.* **23**, 84–87.
An interesting historical perspective on deciphering the mechanism of muscle contraction.

Molloy, J.E. & Veigel, C. (2003) Myosin motors walk the walk. *Science* **300**, 2045–2046.

Rayment, I. (1996) The structural basis of the myosin ATPase activity. *J. Biol. Chem.* **271**, 15,850–15,853.
Examines the muscle contraction mechanism from a structural perspective.

Rayment, I. & Holden, H.M. (1994) The three-dimensional structure of a molecular motor. *Trends Biochem. Sci.* **19**, 129–134.

Vale, R.D. (2003) The molecular motor toolbox for intracellular transport. *Cell* **112**, 467–480.

Problems

1. Relationship between Affinity and Dissociation Constant Protein A has a binding site for ligand X with a K_d of 10^{-6} M. Protein B has a binding site for ligand X with a K_d of 10^{-9} M. Which protein has a higher affinity for ligand X? Explain your reasoning. Convert the K_d to K_a for both proteins.

2. Negative Cooperativity Which of the following situations would produce a Hill plot with $n_H < 1.0$? Explain your reasoning in each case.

(a) The protein has multiple subunits, each with a single ligand-binding site. Binding of ligand to one site decreases the binding affinity of other sites for the ligand.

(b) The protein is a single polypeptide with two ligand-binding sites, each having a different affinity for the ligand.

(c) The protein is a single polypeptide with a single ligand-binding site. As purified, the protein preparation is heterogeneous, containing some protein molecules that are partially denatured and thus have a lower binding affinity for the ligand.

3. Hemoglobin's Affinity for Oxygen What is the effect of the following changes on the O_2 affinity of hemoglobin? (a) A drop in the pH of blood plasma from 7.4 to 7.2. (b) A decrease

in the partial pressure of CO_2 in the lungs from 6 kPa (holding one's breath) to 2 kPa (normal). (c) An increase in the BPG level from 5 mM (normal altitudes) to 8 mM (high altitudes). (d) An increase in CO from 1.0 part per million (ppm) in a normal indoor atmosphere to 30 ppm in a home that has a malfunctioning or leaking furnace.

4. Reversible Ligand Binding I The protein calcineurin binds to the protein calmodulin with an association rate of $8.9 \times 10^3\ M^{-1}s^{-1}$ and an overall dissociation constant, K_d, of 10 nM. Calculate the dissociation rate, k_d, including appropriate units.

5. Reversible Ligand Binding II Three membrane receptor proteins bind tightly to a hormone. Based on the data in the table below, (a) what is the K_d for hormone binding by protein 2? (Include appropriate units.) (b) Which of these proteins binds *most* tightly to this hormone?

Hormone concentration (nM)	θ		
	Protein 1	Protein 2	Protein 3
0.2	0.048	0.29	0.17
0.5	0.11	0.5	0.33
1	0.2	0.67	0.5
4	0.5	0.89	0.8
10	0.71	0.95	0.91
20	0.83	0.97	0.95
50	0.93	0.99	0.98

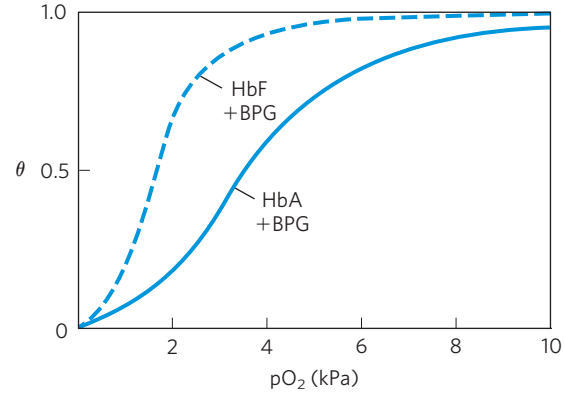
6. Cooperativity in Hemoglobin Under appropriate conditions, hemoglobin dissociates into its four subunits. The isolated α subunit binds oxygen, but the O_2-saturation curve is hyperbolic rather than sigmoid. In addition, the binding of oxygen to the isolated α subunit is not affected by the presence of H^+, CO_2, or BPG. What do these observations indicate about the source of the cooperativity in hemoglobin?

7. Comparison of Fetal and Maternal Hemoglobins Studies of oxygen transport in pregnant mammals show that the O_2-saturation curves of fetal and maternal blood are markedly different when measured under the same conditions. Fetal erythrocytes contain a structural variant of hemoglobin, HbF, consisting of two α and two γ subunits ($\alpha_2\gamma_2$), whereas maternal erythrocytes contain HbA ($\alpha_2\beta_2$).

(a) Which hemoglobin has a higher affinity for oxygen under physiological conditions, HbA or HbF? Explain.

(b) What is the physiological significance of the different O_2 affinities?

(c) When all the BPG is carefully removed from samples of HbA and HbF, the measured O_2-saturation curves (and consequently the O_2 affinities) are displaced to the left. However, HbA now has a greater affinity for oxygen than does HbF. When BPG is reintroduced, the O_2-saturation curves return to normal, as shown in the graph. What is the effect of BPG on the O_2 affinity of hemoglobin? How can the above information be used to explain the different O_2 affinities of fetal and maternal hemoglobin?



8. Hemoglobin Variants There are almost 500 naturally occurring variants of hemoglobin. Most are the result of a single amino acid substitution in a globin polypeptide chain. Some variants produce clinical illness, though not all variants have deleterious effects. A brief sample follows.

HbS (sickle-cell Hb): substitutes a Val for a Glu on the surface
Hb Cowtown: eliminates an ion pair involved in T-state stabilization
Hb Memphis: substitutes one uncharged polar residue for another of similar size on the surface
Hb Bibba: substitutes a Pro for a Leu involved in an α helix
Hb Milwaukee: substitutes a Glu for a Val
Hb Providence: substitutes an Asn for a Lys that normally projects into the central cavity of the tetramer
Hb Philly: substitutes a Phe for a Tyr, disrupting hydrogen bonding at the $\alpha_1\beta_1$ interface

Explain your choices for each of the following:

(a) The Hb variant *least* likely to cause pathological symptoms.

(b) The variant(s) most likely to show pI values different from that of HbA on an isoelectric focusing gel.

(c) The variant(s) most likely to show a decrease in BPG binding and an increase in the overall affinity of the hemoglobin for oxygen.

9. Oxygen Binding and Hemoglobin Structure A team of biochemists uses genetic engineering to modify the interface region between hemoglobin subunits. The resulting hemoglobin variants exist in solution primarily as $\alpha\beta$ dimers (few, if any, $\alpha_2\beta_2$ tetramers form). Are these variants likely to bind oxygen more weakly or more tightly? Explain your answer.

10. Reversible (but Tight) Binding to an Antibody An antibody binds to an antigen with a K_d of 5×10^{-8} M. At what concentration of antigen will θ be (a) 0.2, (b) 0.5, (c) 0.6, (d) 0.8?

11. Using Antibodies to Probe Structure-Function Relationships in Proteins A monoclonal antibody binds to G-actin but not to F-actin. What does this tell you about the epitope recognized by the antibody?

12. The Immune System and Vaccines A host organism needs time, often days, to mount an immune response against a new antigen, but memory cells permit a rapid response to pathogens previously encountered. A vaccine to

protect against a particular viral infection often consists of weakened or killed virus or isolated proteins from a viral protein coat. When injected into a human patient, the vaccine generally does not cause an infection and illness, but it effectively "teaches" the immune system what the viral particles look like, stimulating the production of memory cells. On subsequent infection, these cells can bind to the virus and trigger a rapid immune response. Some pathogens, including HIV, have developed mechanisms to evade the immune system, making it difficult or impossible to develop effective vaccines against them. What strategy could a pathogen use to evade the immune system? Assume that a host's antibodies and/or T-cell receptors are available to bind to any structure that might appear on the surface of a pathogen and that, once bound, the pathogen is destroyed.

13. How We Become a "Stiff" When a vertebrate dies, its muscles stiffen as they are deprived of ATP, a state called rigor mortis. Explain the molecular basis of the rigor state.

14. Sarcomeres from Another Point of View The symmetry of thick and thin filaments in a sarcomere is such that six thin filaments ordinarily surround each thick filament in a hexagonal array. Draw a cross section (transverse cut) of a myofibril at the following points: (a) at the M line; (b) through the I band; (c) through the dense region of the A band; (d) through the less dense region of the A band, adjacent to the M line (see Fig. 5–29b, c).

Using the Web

15. Lysozyme and Antibodies To fully appreciate how proteins function in a cell, it is helpful to have a three-dimensional view of how proteins interact with other cellular components. Fortunately, this is possible using Web-based protein databases and three-dimensional molecular viewing utilities such as Jmol, a free and user-friendly molecular viewer that is compatible with most browsers and operating systems.

In this exercise you will examine the interactions between the enzyme lysozyme (Chapter 4) and the Fab portion of the anti-lysozyme antibody. Use the PDB identifier 1FDL to explore the structure of the IgG1 Fab fragment–lysozyme complex (antibody-antigen complex). To answer the following questions, use the information on the Structure Summary page at the Protein Data Bank (www.rcsb.org), and view the structure using Jmol or a similar viewer.

(a) Which chains in the three-dimensional model correspond to the antibody fragment and which correspond to the antigen, lysozyme?

(b) What type of secondary structure predominates in this Fab fragment?

(c) How many amino acid residues are in the heavy and light chains of the Fab fragment? In lysozyme? Estimate the percentage of the lysozyme that interacts with the antigen-binding site of the antibody fragment.

(d) Identify the specific amino acid residues in lysozyme and in the variable regions of the Fab heavy and light chains that are situated at the antigen-antibody interface. Are the residues contiguous in the primary sequence of the polypeptide chains?

16. Exploring Reversible Interactions of Proteins and Ligands with Living Graphs Use the living graphs for Equations 5–8, 5–11, 5–14, and 5–16 to work through the following exercises.

(a) Reversible binding of a ligand to a simple protein, without cooperativity. For Equation 5–8, set up a plot of θ versus [L] (vertical and horizontal axes, respectively). Examine the plots generated when K_d is set at 5, 10, 20, and 100 μM. Higher affinity of the protein for the ligand means more binding at lower ligand concentrations. Suppose that four different proteins exhibit these four different K_d values for ligand L. Which protein would have the highest affinity for L?

Examine the plot generated when $K_d = 10$ μM. How much does θ increase when [L] increases from 0.2 to 0.4 μM? How much does θ increase when [L] increases from 40 to 80 μM?

You can do the same exercise for Equation 5–11. Convert [L] to pO_2 and K_d to P_{50}. Examine the curves generated when P_{50} is set at 0.5, 1, 2, and 10 kPa. For the curve generated when $P_{50} = 1$ kPa, how much does θ change when the pO_2 increases from 0.02 to 0.04 kPa? From 4 to 8 kPa?

(b) Cooperative binding of a ligand to a multisubunit protein. Using Equation 5–14, generate a binding curve for a protein and ligand with $K_d = 10$ μM and $n = 3$. Note the altered definition of K_d in Equation 5–16. On the same plot, add a curve for a protein with $K_d = 20$ μM and $n = 3$. Now see how both curves change when you change to $n = 4$. Generate Hill plots (Eqn 5–16) for each of these cases. For $K_d = 10$ μM and $n = 3$, what is θ when [L] = 20 μM?

(c) Explore these equations further by varying all the parameters used above.

Data Analysis Problem

17. Protein Function During the 1980s, the structures of actin and myosin were known only at the resolution shown in Figure 5–28a, b. Although researchers knew that the S1 portion of myosin binds to actin and hydrolyzes ATP, there was a substantial debate about where in the myosin molecule the contractile force was generated. At the time, two competing models were proposed for the mechanism of force generation in myosin.

In the "hinge" model, S1 bound to actin, but the pulling force was generated by contraction of the "hinge region" in the myosin tail. The hinge region is in the heavy meromyosin portion of the myosin molecule, near where trypsin cleaves off light meromyosin (see Fig. 5–27b). This is roughly the point labeled "Two supercoiled α helices" in Figure 5–27a. In the "S1" model, the pulling force was generated in the S1 "head" itself and the tail was just for structural support.

Many experiments had been performed but provided no conclusive evidence. In 1987, James Spudich and his colleagues at Stanford University published a study that, although not conclusive, went a long way toward resolving this controversy.

Recombinant DNA techniques were not sufficiently developed to address this issue in vivo, so Spudich and colleagues used an interesting in vitro motility assay. The alga *Nitella* has extremely long cells, often several centimeters in length and about 1 mm in diameter. These cells have actin fibers that run

along their long axes, and the cells can be cut open along their length to expose the actin fibers. Spudich and his group had observed that plastic beads coated with myosin would "walk" along these fibers in the presence of ATP, just as myosin would do in contracting muscle.

For these experiments, they used a more well-defined method for attaching the myosin to the beads. The "beads" were clumps of killed bacterial (*Staphylococcus aureus*) cells. These cells have a protein on their surface that binds to the Fc region of antibody molecules (Fig. 5–21a). The antibodies, in turn, bind to several (unknown) places along the tail of the myosin molecule. When bead-antibody-myosin complexes were prepared with intact myosin molecules, they would move along *Nitella* actin fibers in the presence of ATP.

(a) Sketch a diagram showing what a bead-antibody-myosin complex might look like at the molecular level.

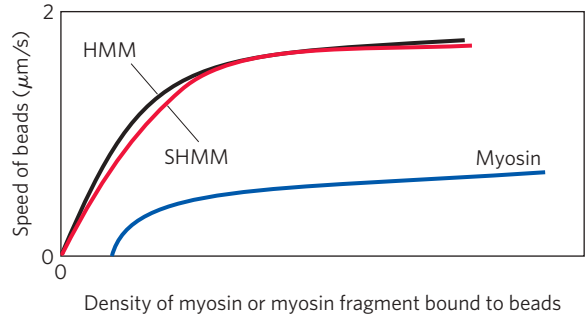
(b) Why was ATP required for the beads to move along the actin fibers?

(c) Spudich and coworkers used antibodies that bound to the myosin tail. Why would this experiment have failed if they had used an antibody that bound to the part of S1 that normally binds to actin? Why would this experiment have failed if they had used an antibody that bound to actin?

To help focus in on the part of myosin responsible for force production, Spudich and his colleagues used trypsin to produce two partial myosin molecules (see Fig. 5–27b): (1) heavy meromyosin (HMM), made by briefly digesting myosin with trypsin; HMM consists of S1 and the part of the tail that includes the hinge; and (2) short heavy meromyosin (SHMM), made from a more extensive digestion of HMM with trypsin; SHMM consists of S1 and a shorter part of the tail that does not include the hinge. Brief digestion of myosin with trypsin produces HMM and light meromyosin, by cleavage of a single specific peptide bond in the myosin molecule.

(d) Why might trypsin attack this peptide bond first rather than other peptide bonds in myosin?

Spudich and colleagues prepared bead-antibody-myosin complexes with varying amounts of myosin, HMM, and SHMM and measured their speeds along *Nitella* actin fibers in the presence of ATP. The graph below sketches their results.



(e) Which model ("S1" or "hinge") is consistent with these results? Explain your reasoning.

(f) Provide a plausible explanation for why the speed of the beads increased with increasing myosin density.

(g) Provide a plausible explanation for why the speed of the beads reached a plateau at high myosin density.

The more extensive trypsin digestion required to produce SHMM had a side effect: another specific cleavage of the myosin polypeptide backbone in addition to the cleavage in the tail. This second cleavage was in the S1 head.

(h) Based on this information, why is it surprising that SHMM was still capable of moving beads along actin fibers?

(i) As it turns out, the tertiary structure of the S1 head remains intact in SHMM. Provide a plausible explanation of how the protein remains intact and functional even though the polypeptide backbone has been cleaved and is no longer continuous.

Reference

Hynes, T.R., Block, S.M., White, B.T., & Spudich, J.A. (1987) Movement of myosin fragments in vitro: domains involved in force production. *Cell* **48**, 953–963.

this page left intentionally blank

6

Enzymes

There are two fundamental conditions for life. First, the organism must be able to self-replicate (a topic considered in Part III); second, it must be able to catalyze chemical reactions efficiently and selectively. The central importance of catalysis may seem surprising, but it is easy to demonstrate. As described in Chapter 1, living systems make use of energy from the environment. Many of us, for example, consume substantial amounts of sucrose—common table sugar—as a kind of fuel, usually in the form of sweetened foods and drinks. The conversion of sucrose to CO_2 and H_2O in the presence of oxygen is a highly exergonic process, releasing free energy that we can use to think, move, taste, and see. However, a bag of sugar can remain on the shelf for years without any obvious conversion to CO_2 and H_2O. Although this chemical process is thermodynamically favorable, it is very slow. Yet when sucrose is consumed by a human (or almost any other organism), it releases its chemical energy in seconds. The difference is catalysis. Without catalysis, chemical reactions such as sucrose oxidation could not occur on a useful time scale, and thus could not sustain life.

In this chapter, then, we turn our attention to the reaction catalysts of biological systems: enzymes, the most remarkable and highly specialized proteins. Enzymes have extraordinary catalytic power, often far greater than that of synthetic or inorganic catalysts. They have a high degree of specificity for their substrates, they accelerate chemical reactions tremendously, and they function in aqueous solutions under very mild conditions of temperature and pH. Few nonbiological catalysts have all these properties.

Enzymes are central to every biochemical process. Acting in organized sequences, they catalyze the hundreds of stepwise reactions that degrade nutrient molecules, conserve and transform chemical energy, and make biological macromolecules from simple precursors.

The study of enzymes has immense practical importance. In some diseases, especially inheritable genetic disorders, there may be a deficiency or even a total absence of one or more enzymes. Other disease conditions may be caused by excessive activity of an enzyme. Measurements of the activities of enzymes in blood plasma, erythrocytes, or tissue samples are important in diagnosing certain illnesses. Many drugs act through interactions with enzymes. Enzymes are also important practical tools in chemical engineering, food technology, and agriculture.

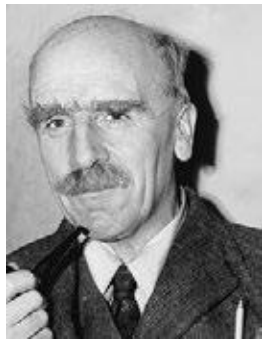
We begin with descriptions of the properties of enzymes and the principles underlying their catalytic power, then introduce enzyme kinetics, a discipline that provides much of the framework for any discussion of enzymes. Specific examples of enzyme mechanisms are then provided, illustrating principles introduced earlier in the chapter. We end with a discussion of how enzyme activity is regulated.

6.1 An Introduction to Enzymes

Much of the history of biochemistry is the history of enzyme research. Biological catalysis was first recognized and described in the late 1700s, in studies on the digestion of meat by secretions of the stomach. Research continued in the 1800s with examinations of the conversion of starch to sugar by saliva and various plant extracts. In the 1850s, Louis Pasteur concluded that fermentation of sugar into alcohol by yeast is catalyzed by "ferments." He postulated that these ferments were inseparable from the structure of living yeast cells; this view, called vitalism, prevailed for decades. Then in 1897 Eduard Buchner discovered that yeast extracts could ferment sugar to alcohol, proving that fermentation was promoted by molecules that continued to function when removed from cells. Buchner's experiment at once marked the


Eduard Buchner, 1860–1917


James Sumner, 1887–1955


J. B. S. Haldane, 1892–1964

end of vitalistic notions and the dawn of the science of biochemistry. Frederick W. Kühne later gave the name **enzymes** (from the Greek *enzymos*, "leavened") to the molecules detected by Buchner.

The isolation and crystallization of urease by James Sumner in 1926 was a breakthrough in early enzyme studies. Sumner found that urease crystals consisted entirely of protein, and he postulated that all enzymes are proteins. In the absence of other examples, this idea remained controversial for some time. Only in the 1930s was Sumner's conclusion widely accepted, after John Northrop and Moses Kunitz crystallized pepsin, trypsin, and other digestive enzymes and found them also to be proteins. During this period, J. B. S. Haldane wrote a treatise titled *Enzymes*. Although the molecular nature of enzymes was not yet fully appreciated, Haldane made the remarkable suggestion that weak bonding interactions between an enzyme and its substrate might be used to catalyze a reaction. This insight lies at the heart of our current understanding of enzymatic catalysis.

Since the latter part of the twentieth century, thousands of enzymes have been purified, their structures elucidated, and their mechanisms explained.

Most Enzymes Are Proteins

With the exception of a small group of catalytic RNA molecules (Chapter 26), all enzymes are proteins. Their catalytic activity depends on the integrity of their native protein conformation. If an enzyme is denatured or dissociated into its subunits, catalytic activity is usually lost. If an enzyme is broken down into its component amino acids, its catalytic activity is always destroyed. Thus the primary, secondary, tertiary, and quaternary structures of protein enzymes are essential to their catalytic activity.

Enzymes, like other proteins, have molecular weights ranging from about 12,000 to more than 1 million. Some enzymes require no chemical groups for activity other than their amino acid residues. Others require an additional chemical component called a **cofactor**—either one or more inorganic ions, such as Fe^{2+}, Mg^{2+}, Mn^{2+}, or Zn^{2+} (Table 6–1), or a complex organic or metalloorganic molecule called a **coenzyme**. Coenzymes act as transient carriers of specific functional groups (Table 6–2). Most are derived from vitamins, organic nutrients required in small amounts in the diet. We consider coenzymes in more detail as we encounter them in the metabolic pathways discussed in Part II. Some enzymes require *both* a coenzyme and one or more metal ions for activity. A coenzyme or metal ion that is very tightly or even covalently bound to the enzyme protein is called a **prosthetic group**. A complete, catalytically active enzyme together with its bound coenzyme and/or metal ions is called a **holoenzyme**. The protein part of such an enzyme is called the **apoenzyme** or **apoprotein**. Finally, some enzyme proteins are modified covalently by phosphorylation, glycosylation, and other processes. Many of these alterations are involved in the regulation of enzyme activity.

Enzymes Are Classified by the Reactions They Catalyze

Many enzymes have been named by adding the suffix "-ase" to the name of their substrate or to a word or phrase describing their activity. Thus urease catalyzes hydrolysis of urea, and DNA polymerase catalyzes the polymerization of nucleotides to form DNA. Other enzymes were named by their discoverers for a broad function, before the specific reaction catalyzed was known. For example, an enzyme known to act in the digestion of foods was named pepsin, from the Greek *pepsis*, "digestion," and lysozyme was named for its ability to lyse (break down) bacterial cell walls. Still others were named for their source: trypsin, named in part

TABLE 6–1 Some Inorganic Ions That Serve as Cofactors for Enzymes

Ions	Enzymes
Cu^{2+}	Cytochrome oxidase
Fe^{2+} or Fe^{3+}	Cytochrome oxidase, catalase, peroxidase
K^{+}	Pyruvate kinase
Mg^{2+}	Hexokinase, glucose 6-phosphatase, pyruvate kinase
Mn^{2+}	Arginase, ribonucleotide reductase
Mo	Dinitrogenase
Ni^{2+}	Urease
Zn^{2+}	Carbonic anhydrase, alcohol dehydrogenase, carboxypeptidases A and B

TABLE 6–2 Some Coenzymes That Serve as Transient Carriers of Specific Atoms or Functional Groups

Coenzyme	Examples of chemical groups transferred	Dietary precursor in mammals
Biocytin	CO_2	Biotin
Coenzyme A	Acyl groups	Pantothenic acid and other compounds
5′-Deoxyadenosylcobalamin (coenzyme B_{12})	H atoms and alkyl groups	Vitamin B_{12}
Flavin adenine dinucleotide	Electrons	Riboflavin (vitamin B_2)
Lipoate	Electrons and acyl groups	Not required in diet
Nicotinamide adenine dinucleotide	Hydride ion ($:H^-$)	Nicotinic acid (niacin)
Pyridoxal phosphate	Amino groups	Pyridoxine (vitamin B_6)
Tetrahydrofolate	One-carbon groups	Folate
Thiamine pyrophosphate	Aldehydes	Thiamine (vitamin B_1)

Note: The structures and modes of action of these coenzymes are described in Part II.

from the Greek *tryein*, "to wear down," was obtained by rubbing pancreatic tissue with glycerin. Sometimes the same enzyme has two or more names, or two different enzymes have the same name. Because of such ambiguities, and the ever-increasing number of newly discovered enzymes, biochemists, by international agreement, have adopted a system for naming and classifying enzymes. This system divides enzymes into six classes, each with subclasses, based on the type of reaction catalyzed (Table 6–3). Each enzyme is assigned a four-part classification number and a systematic name, which identifies the reaction it catalyzes. As an example, the formal systematic name of the enzyme catalyzing the reaction

$$\text{ATP} + \text{D-glucose} \longrightarrow \text{ADP} + \text{D-glucose 6-phosphate}$$

is ATP:glucose phosphotransferase, which indicates that it catalyzes the transfer of a phosphoryl group from ATP to glucose. Its Enzyme Commission number (E.C. number) is 2.7.1.1. The first number (2) denotes the class name (transferase); the second number (7), the subclass (phosphotransferase); the third number (1), a phosphotransferase with a hydroxyl group as acceptor; and the fourth number (1), D-glucose as the phosphoryl group acceptor. For many enzymes, a trivial name is more frequently used—in this case, hexokinase. A complete list and description of the thousands of known enzymes is maintained by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (www.chem.qmul.ac.uk/iubmb/enzyme). This chapter is devoted primarily to principles and properties common to all enzymes.

SUMMARY 6.1 An Introduction to Enzymes

- Life depends on powerful and specific catalysts: enzymes. Almost every biochemical reaction is catalyzed by an enzyme.
- With the exception of a few catalytic RNAs, all known enzymes are proteins. Many require nonprotein coenzymes or cofactors for their catalytic function.
- Enzymes are classified according to the type of reaction they catalyze. All enzymes have formal E.C. numbers and names, and most have trivial names.

TABLE 6–3 International Classification of Enzymes

Class no.	Class name	Type of reaction catalyzed
1	Oxidoreductases	Transfer of electrons (hydride ions or H atoms)
2	Transferases	Group transfer reactions
3	Hydrolases	Hydrolysis reactions (transfer of functional groups to water)
4	Lyases	Cleavage of C—C, C—O, C—N, or other bonds by elimination, leaving double bonds or rings, or addition of groups to double bonds
5	Isomerases	Transfer of groups within molecules to yield isomeric forms
6	Ligases	Formation of C—C, C—S, C—O, and C—N bonds by condensation reactions coupled to cleavage of ATP or similar cofactor

6.2 How Enzymes Work

The enzymatic catalysis of reactions is essential to living systems. Under biologically relevant conditions, uncatalyzed reactions tend to be slow—most biological molecules are quite stable in the neutral-pH, mild-temperature, aqueous environment inside cells. Furthermore, many common chemical processes are unfavorable or unlikely in the cellular environment, such as the transient formation of unstable charged intermediates or the collision of two or more molecules in the precise orientation required for reaction. Reactions required to digest food, send nerve signals, or contract a muscle simply do not occur at a useful rate without catalysis.

An enzyme circumvents these problems by providing a specific environment within which a given reaction can occur more rapidly. The distinguishing feature of an enzyme-catalyzed reaction is that it takes place within the confines of a pocket on the enzyme called the **active site** **(Fig. 6–1)**. The molecule that is bound in the active site and acted upon by the enzyme is called the **substrate**. The surface of the active site is lined with amino acid residues with substituent groups that bind the substrate and catalyze its chemical transformation. Often, the active site encloses a substrate, sequestering it completely from solution. The enzyme-substrate complex, whose existence was first proposed by Charles-Adolphe Wurtz in 1880, is central to the action of enzymes. It is also the starting point for mathematical treatments that define the kinetic behavior of enzyme-catalyzed reactions and for theoretical descriptions of enzyme mechanisms.

Enzymes Affect Reaction Rates, Not Equilibria

A simple enzymatic reaction might be written

$$E + S \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P \qquad (6\text{–}1)$$

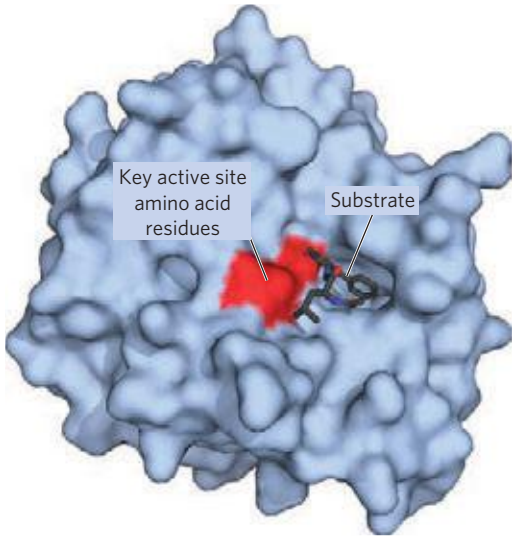


FIGURE 6–1 Binding of a substrate to an enzyme at the active site. The enzyme chymotrypsin with bound substrate (PDB ID 7GCH). Some key active-site amino acid residues appear as a red splotch on the enzyme surface.

where E, S, and P represent the enzyme, substrate, and product; ES and EP are transient complexes of the enzyme with the substrate and with the product.

To understand catalysis, we must first appreciate the important distinction between reaction equilibria and reaction rates. The function of a catalyst is to increase the *rate* of a reaction. Catalysts do not affect reaction *equilibria*. (Recall that a reaction is at equilibrium when there is no net change in the concentrations of reactants or products.) Any reaction, such as $S \rightleftharpoons P$, can be described by a reaction coordinate diagram **(Fig. 6–2)**, a picture of the energy changes during the reaction. As discussed in Chapter 1, energy in biological systems is described in terms of free energy, G. In the coordinate diagram, the free energy of the system is plotted against the progress of the reaction (the reaction coordinate). The starting point for either the forward or the reverse reaction is called the **ground state**, the contribution to the free energy of the system by an average molecule (S or P) under a given set of conditions.

KEY CONVENTION: To describe the free-energy changes for reactions, chemists define a standard set of conditions (temperature 298 K; partial pressure of each gas 1 atm, or 101.3 kPa; concentration of each solute 1 M) and express the free-energy change for a reacting system under these conditions as $\Delta G°$, the **standard free-energy change**. Because biochemical systems commonly involve H^+ concentrations far below 1 M, biochemists define a **biochemical standard free-energy change**, $\Delta G'°$, the standard free-energy change *at pH 7.0*; we employ this definition throughout the book. A more complete definition of $\Delta G'°$ is given in Chapter 13. ■

The equilibrium between S and P reflects the difference in the free energies of their ground states. In the example shown in Figure 6–2, the free energy of the ground state of P is lower than that of S, so $\Delta G'°$ for

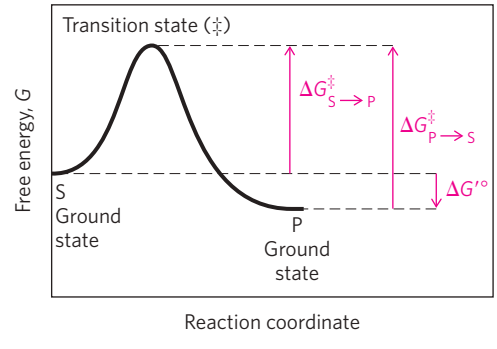


FIGURE 6–2 Reaction coordinate diagram. The free energy of the system is plotted against the progress of the reaction S ⟶ P. A diagram of this kind is a description of the energy changes during the reaction, and the horizontal axis (reaction coordinate) reflects the progressive chemical changes (e.g., bond breakage or formation) as S is converted to P. The activation energies, $\Delta G^{\ddagger}$, for the S ⟶ P and P ⟶ S reactions are indicated. $\Delta G'°$ is the overall standard free-energy change in the direction S ⟶ P.

the reaction is negative (the reaction is exergonic) and at equilibrium there is more P than S (the equilibrium favors P). The position and direction of equilibrium are *not* affected by any catalyst.

A favorable equilibrium does not mean that the S → P conversion will occur at a detectable rate. The *rate* of a reaction is dependent on an entirely different parameter. There is an energy barrier between S and P: the energy required for alignment of reacting groups, formation of transient unstable charges, bond rearrangements, and other transformations required for the reaction to proceed in either direction. This is illustrated by the energy "hill" in Figures 6–2 and 6–3. To undergo reaction, the molecules must overcome this barrier and therefore must be raised to a higher energy level. At the top of the energy hill is a point at which decay to the S or P state is equally probable (it is downhill either way). This is called the **transition state**. The transition state is not a chemical species with any significant stability and should not be confused with a reaction intermediate (such as ES or EP). It is simply a fleeting molecular moment in which events such as bond breakage, bond formation, and charge development have proceeded to the precise point at which decay to either substrate or product is equally likely. The difference between the energy levels of the ground state and the transition state is the **activation energy**, $\Delta G^{\ddagger}$. The rate of a reaction reflects this activation energy: a higher activation energy corresponds to a slower reaction. Reaction rates can be increased by raising the temperature and/or pressure, thereby increasing the number of molecules with sufficient energy to overcome the energy barrier. Alternatively, the activation energy can be lowered by adding a catalyst **(Fig. 6–3)**. *Catalysts enhance reaction rates by lowering activation energies.*

Enzymes are no exception to the rule that catalysts do not affect reaction equilibria. The bidirectional arrows in Equation 6–1 make this point: any enzyme that catalyzes the reaction S → P also catalyzes the reaction P → S. The role of enzymes is to *accelerate* the interconversion of S and P. The enzyme is not used up in the process, and the equilibrium point is unaffected. However, the reaction reaches equilibrium much faster when the appropriate enzyme is present, because the rate of the reaction is increased.

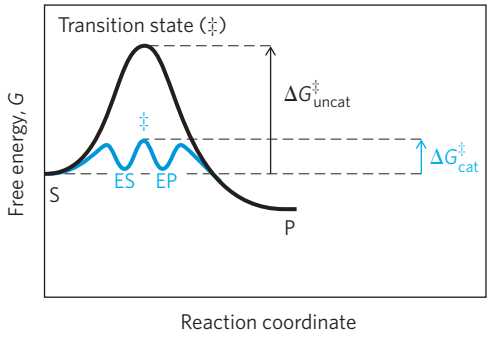


FIGURE 6–3 Reaction coordinate diagram comparing enzyme-catalyzed and uncatalyzed reactions. In the reaction S → P, the ES and EP intermediates occupy minima in the energy progress curve of the enzyme-catalyzed reaction. The terms $\Delta G^{\ddagger}_{uncat}$ and $\Delta G^{\ddagger}_{cat}$ correspond to the activation energy for the uncatalyzed reaction and the overall activation energy for the catalyzed reaction, respectively. The activation energy is lower when the enzyme catalyzes the reaction.

This general principle is illustrated in the conversion of sucrose and oxygen to carbon dioxide and water:

$$C_{12}H_{22}O_{11} + 12O_2 \rightleftharpoons 12CO_2 + 11H_2O$$

This conversion, which takes place through a series of separate reactions, has a very large and negative $\Delta G'^{\circ}$, and at equilibrium the amount of sucrose present is negligible. Yet sucrose is a stable compound, because the activation energy barrier that must be overcome before sucrose reacts with oxygen is quite high. Sucrose can be stored in a container with oxygen almost indefinitely without reacting. In cells, however, sucrose is readily broken down to CO_2 and H_2O in a series of reactions catalyzed by enzymes. These enzymes not only accelerate the reactions, they organize and control them so that much of the energy released is recovered in other chemical forms and made available to the cell for other tasks. The reaction pathway by which sucrose (and other sugars) is broken down is the primary energy-yielding pathway for cells, and the enzymes of this pathway allow the reaction sequence to proceed on a biologically useful time scale.

Any reaction may have several steps, involving the formation and decay of transient chemical species called **reaction intermediates**.* A reaction intermediate is any species on the reaction pathway that has a finite chemical lifetime (longer than a molecular vibration, $\sim 10^{-13}$ second). When the S⇌P reaction is catalyzed by an enzyme, the ES and EP complexes can be considered intermediates, even though S and P are stable chemical species (Eqn 6–1); the ES and EP complexes occupy valleys in the reaction coordinate diagram (Fig. 6–3). Additional, less stable chemical intermediates often exist in the course of an enzyme-catalyzed reaction. The interconversion of two sequential reaction intermediates thus constitutes a reaction step. When several steps occur in a reaction, the overall rate is determined by the step (or steps) with the highest activation energy; this is called the **rate-limiting step**. In a simple case, the rate-limiting step is the highest-energy point in the diagram for interconversion of S and P. In practice, the rate-limiting step can vary with reaction conditions, and for many enzymes several steps may have similar activation energies, which means they are all partially rate-limiting.

*In this chapter, *step* and *intermediate* refer to chemical species in the reaction pathway of a single enzyme-catalyzed reaction. In the context of metabolic pathways involving many enzymes (discussed in Part II), these terms are used somewhat differently. An entire enzymatic reaction is often referred to as a "step" in a pathway, and the product of one enzymatic reaction (which is the substrate for the next enzyme in the pathway) is referred to as a pathway "intermediate."

Activation energies are energy barriers to chemical reactions. These barriers are crucial to life itself. The rate at which a molecule undergoes a particular reaction decreases as the activation barrier for that reaction increases. Without such energy barriers, complex macromolecules would revert spontaneously to much simpler molecular forms, and the complex and highly ordered structures and metabolic processes of cells could not exist. Over the course of evolution, enzymes have developed to lower activation energies *selectively* for reactions that are needed for cell survival.

Reaction Rates and Equilibria Have Precise Thermodynamic Definitions

Reaction *equilibria* are inextricably linked to the standard free-energy change for the reaction, $\Delta G'^\circ$, and reaction *rates* are linked to the activation energy, $\Delta G^\ddagger$. A basic introduction to these thermodynamic relationships is the next step in understanding how enzymes work.

An equilibrium such as $S \rightleftharpoons P$ is described by an **equilibrium constant**, $\boldsymbol{K_{eq}}$, or simply K (p. 25). Under the standard conditions used to compare biochemical processes, an equilibrium constant is denoted K'_{eq} (or K'):

$$K'_{eq} = \frac{[P]}{[S]} \qquad (6\text{–}2)$$

From thermodynamics, the relationship between K'_{eq} and $\Delta G'^\circ$ can be described by the expression

$$\Delta G'^\circ = -RT \ln K'_{eq} \qquad (6\text{–}3)$$

where R is the gas constant, 8.315 J/mol · K, and T is the absolute temperature, 298 K (25 °C). Equation 6–3 is developed and discussed in more detail in Chapter 13. The important point here is that the equilibrium constant is directly related to the overall standard free-energy change for the reaction (Table 6–4). A large negative value for $\Delta G'^\circ$ reflects a favorable reaction equilibrium (one in which there is much more product than substrate at equilibrium)—but as already noted, this does not mean the reaction will proceed at a rapid rate.

TABLE 6–4 Relationship between K'_{eq} and $\Delta G'^\circ$

K'_{eq}	$\Delta G'^\circ$ (kJ/mol)
10^{-6}	34.2
10^{-5}	28.5
10^{-4}	22.8
10^{-3}	17.1
10^{-2}	11.4
10^{-1}	5.7
1	0.0
10^{1}	−5.7
10^{2}	−11.4
10^{3}	−17.1

Note: The relationship is calculated from $\Delta G'^\circ = -RT \ln K'_{eq}$ (Eqn 6–3).

The rate of any reaction is determined by the concentration of the reactant (or reactants) and by a **rate constant**, usually denoted by k. For the unimolecular reaction $S \longrightarrow P$, the rate (or velocity) of the reaction, V—representing the amount of S that reacts per unit time—is expressed by a **rate equation**:

$$V = k[S] \qquad (6\text{–}4)$$

In this reaction, the rate depends only on the concentration of S. This is called a first-order reaction. The factor k is a proportionality constant that reflects the probability of reaction under a given set of conditions (pH, temperature, and so forth). Here, k is a first-order rate constant and has units of reciprocal time, such as s^{-1}. If a first-order reaction has a rate constant k of 0.03 s^{-1}, this may be interpreted (qualitatively) to mean that 3% of the available S will be converted to P in 1 s. A reaction with a rate constant of 2,000 s^{-1} will be over in a small fraction of a second. If a reaction rate depends on the concentration of two different compounds, or if the reaction is between two molecules of the same compound, the reaction is second order and k is a second-order rate constant, with units of $M^{-1}s^{-1}$. The rate equation then becomes

$$V = k[S_1][S_2] \qquad (6\text{–}5)$$

From transition-state theory we can derive an expression that relates the magnitude of a rate constant to the activation energy:

$$k = \frac{\mathbf{k}T}{h} e^{-\Delta G^\ddagger/RT} \qquad (6\text{–}6)$$

where $\mathbf{k}$ is the Boltzmann constant and h is Planck's constant. The important point here is that the relationship between the rate constant k and the activation energy $\Delta G^\ddagger$ is inverse and exponential. In simplified terms, this is the basis for the statement that a lower activation energy means a faster reaction rate.

Now we turn from *what* enzymes do to *how* they do it.

A Few Principles Explain the Catalytic Power and Specificity of Enzymes

Enzymes are extraordinary catalysts. The rate enhancements they bring about are in the range of 5 to 17 orders of magnitude (Table 6–5). Enzymes are also very specific, readily discriminating between substrates with quite similar structures. How can these enormous and highly selective rate enhancements be explained? What is the source of the energy for the dramatic lowering of the activation energies for specific reactions?

The answer to these questions has two distinct but interwoven parts. The first lies in the rearrangement of covalent bonds during an enzyme-catalyzed reaction. Chemical reactions of many types take place between

TABLE 6–5 Some Rate Enhancements Produced by Enzymes

Cyclophilin	10^5
Carbonic anhydrase	10^7
Triose phosphate isomerase	10^9
Carboxypeptidase A	10^{11}
Phosphoglucomutase	10^{12}
Succinyl-CoA transferase	10^{13}
Urease	10^{14}
Orotidine monophosphate decarboxylase	10^{17}

substrates and enzymes' functional groups (specific amino acid side chains, metal ions, and coenzymes). Catalytic functional groups on an enzyme may form a transient covalent bond with a substrate and activate it for reaction, or a group may be transiently transferred from the substrate to the enzyme. In many cases, these reactions occur only in the enzyme active site. Covalent interactions between enzymes and substrates lower the activation energy (and thereby accelerate the reaction) by providing an alternative, lower-energy reaction path. The specific types of rearrangements that occur are described in Section 6.4.

The second part of the explanation lies in the *noncovalent* interactions between enzyme and substrate. Recall from Chapter 4 that weak, noncovalent interactions help stabilize protein structure and protein-protein interactions. These same interactions are critical to the formation of complexes between proteins and small molecules, including enzyme substrates. Much of the energy required to lower activation energies is derived from weak, noncovalent interactions between substrate and enzyme. What really sets enzymes apart from most other catalysts is the formation of a specific ES complex. The interaction between substrate and enzyme in this complex is mediated by the same forces that stabilize protein structure, including hydrogen bonds and hydrophobic and ionic interactions (Chapter 4). Formation of each weak interaction in the ES complex is accompanied by release of a small amount of free energy that stabilizes the interaction. The energy derived from enzyme-substrate interaction is called **binding energy**, ΔG_B. Its significance extends beyond a simple stabilization of the enzyme-substrate interaction. *Binding energy is a major source of free energy used by enzymes to lower the activation energies of reactions.*

Two fundamental and interrelated principles provide a general explanation for how enzymes use noncovalent binding energy:

1. Much of the catalytic power of enzymes is ultimately derived from the free energy released in forming many weak bonds and interactions between an enzyme and its substrate. This binding energy contributes to specificity as well as to catalysis.
2. Weak interactions are optimized in the reaction transition state; enzyme active sites are complementary not to the substrates per se but to the transition states through which substrates pass as they are converted to products during an enzymatic reaction.

These themes are critical to an understanding of enzymes, and they now become our primary focus.

Weak Interactions between Enzyme and Substrate Are Optimized in the Transition State

How does an enzyme use binding energy to lower the activation energy for a reaction? Formation of the ES complex is not the explanation in itself, although some of the earliest considerations of enzyme mechanisms began with this idea. Studies on enzyme specificity carried out by Emil Fischer led him to propose, in 1894, that enzymes were structurally complementary to their substrates, so that they fit together like a lock and key **(Fig. 6–4)**. This elegant idea, that a specific (exclusive) interaction between two biological molecules is mediated

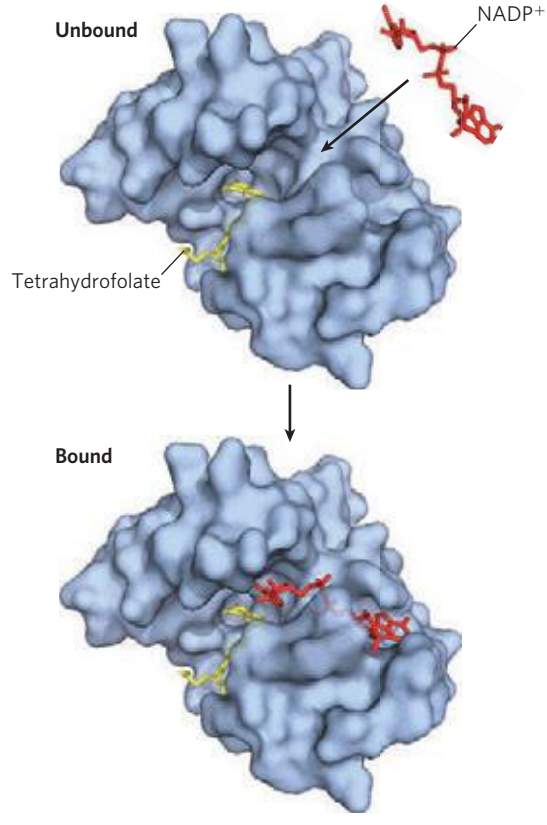


FIGURE 6–4 Complementary shapes of a substrate and its binding site on an enzyme. The enzyme dihydrofolate reductase with its substrate $NADP^+$, unbound and bound; another bound substrate, tetrahydrofolate, is also visible (PDB ID 1RA2). In this model, the $NADP^+$ binds to a pocket that is complementary to it in shape and ionic properties, an illustration of Emil Fischer's "lock and key" hypothesis of enzyme action. In reality, the complementarity between protein and ligand (in this case, substrate) is rarely perfect, as we saw in Chapter 5.

by molecular surfaces with complementary shapes, has greatly influenced the development of biochemistry, and such interactions lie at the heart of many biochemical processes. However, the "lock and key" hypothesis can be misleading when applied to enzymatic catalysis. An enzyme completely complementary to its substrate would be a very poor enzyme, as we can demonstrate.

Consider an imaginary reaction, the breaking of a magnetized metal stick. The uncatalyzed reaction is shown in **Figure 6–5a**. Let's examine two imaginary enzymes—two "stickases"—that could catalyze this reaction, both of which employ magnetic forces as a paradigm for the binding energy used by real enzymes. We first design an enzyme perfectly complementary to the substrate (Fig. 6–5b). The active site of this stickase is a pocket lined with magnets. To react (break), the stick must reach the transition state of the reaction, but the stick fits so tightly in the active site that it cannot bend, because bending would eliminate some of the magnetic interactions between stick and enzyme. Such an enzyme *impedes* the reaction, stabilizing the substrate instead. In a reaction coordinate diagram (Fig. 6–5b), this kind of ES complex would correspond to an energy trough from which the substrate would have difficulty escaping. Such an enzyme would be useless.

The modern notion of enzymatic catalysis, first proposed by Michael Polanyi (1921) and Haldane (1930), was elaborated by Linus Pauling in 1946 and by

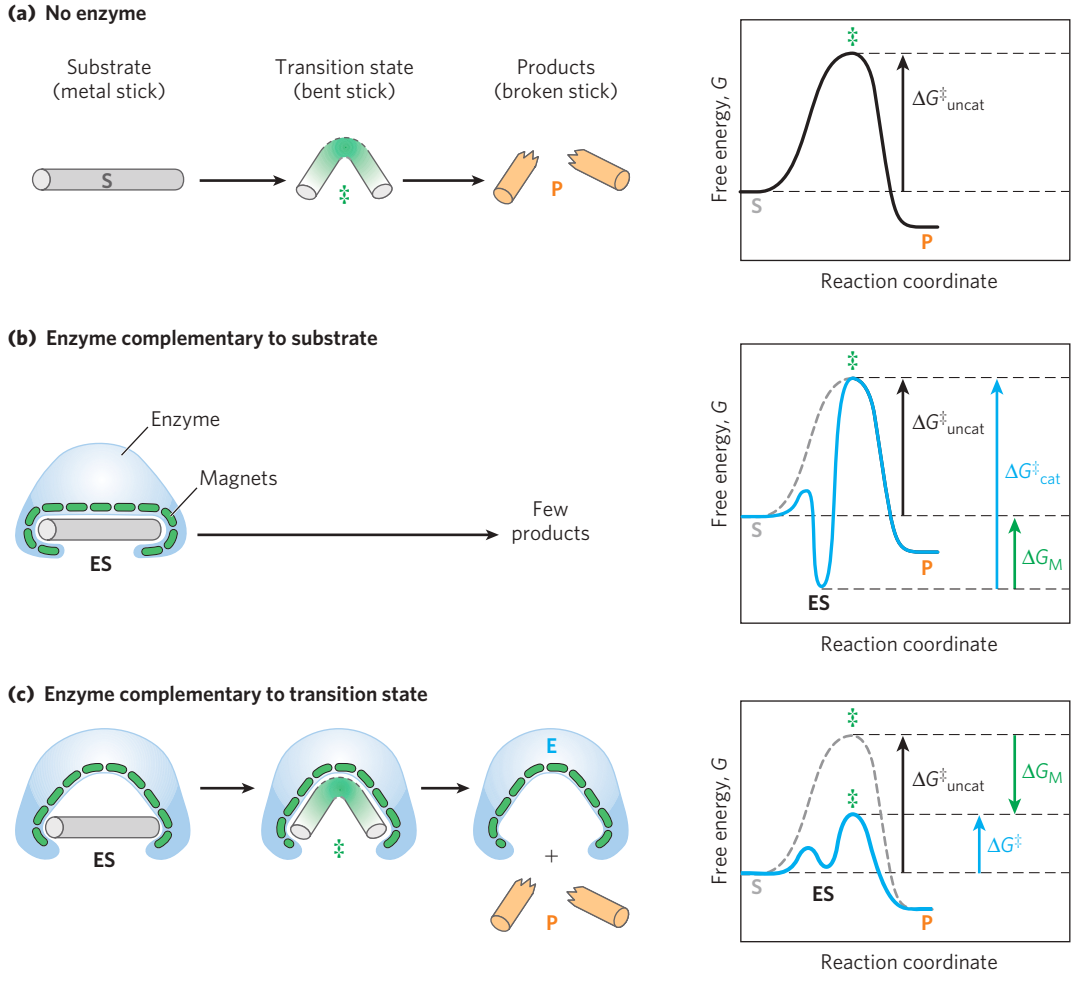


FIGURE 6–5 An imaginary enzyme (stickase) designed to catalyze breakage of a metal stick. (a) Before the stick is broken, it must first be bent (the transition state). In both stickase examples, magnetic interactions take the place of weak bonding interactions between enzyme and substrate. **(b)** A stickase with a magnet-lined pocket complementary in structure to the stick (the substrate) stabilizes the substrate. Bending is impeded by the magnetic attraction between stick and stickase. **(c)** An enzyme with a pocket complementary to the reaction transition state helps to destabilize the stick, contributing to catalysis of the reaction. The binding energy of the magnetic interactions compensates for the increase in free energy required to bend the stick. Reaction coordinate diagrams (right) show the energy consequences of complementarity to substrate versus complementarity to transition state (EP complexes are omitted). ΔG_M, the difference between the transition-state energies of the uncatalyzed and catalyzed reactions, is contributed by the magnetic interactions between the stick and stickase. When the enzyme is complementary to the substrate (b), the ES complex is more stable and has less free energy in the ground state than substrate alone. The result is an *increase* in the activation energy.

William P. Jencks in the 1970s: in order to catalyze reactions, an enzyme must be complementary to the *reaction transition state*. This means that optimal interactions between substrate and enzyme occur only in the transition state. Figure 6–5c demonstrates how such an enzyme can work. The metal stick binds to the stickase, but only a subset of the possible magnetic interactions are used in forming the ES complex. The bound substrate must still undergo the increase in free energy needed to reach the transition state. Now, however, the increase in free energy required to draw the stick into a bent and partially broken conformation is offset, or "paid for," by the magnetic interactions (binding energy) that form between the enzyme and substrate in the transition state. Many of these interactions involve parts of the stick that are distant from the point of breakage; thus interactions between the stickase and nonreacting parts of the stick provide some of the energy needed to catalyze stick breakage. This "energy payment" translates into a lower net activation energy and a faster reaction rate.

Real enzymes work on an analogous principle. Some weak interactions are formed in the ES complex, but the full complement of such interactions between substrate and enzyme is formed only when the substrate reaches the transition state. The free energy (binding energy) released by the formation of these interactions partially offsets the energy required to reach the top of the energy hill. The summation of the unfavorable (positive) activation energy $\Delta G^{\ddagger}$ and the favorable (negative) binding energy ΔG_B results in a lower *net* activation energy **(Fig. 6–6)**. Even on the enzyme, the transition state is not a stable species but a brief point in time that the substrate spends atop an energy hill. The enzyme-catalyzed reaction is much faster than the uncatalyzed process, however, because the hill is much smaller. The important principle is that *weak binding interactions between the enzyme and the substrate provide a substantial driving force for enzymatic catalysis*. The groups on the substrate that are involved in these weak interactions can be at some distance from the bonds that are broken or changed. The weak interactions formed only in the transition state are those that make the primary contribution to catalysis.

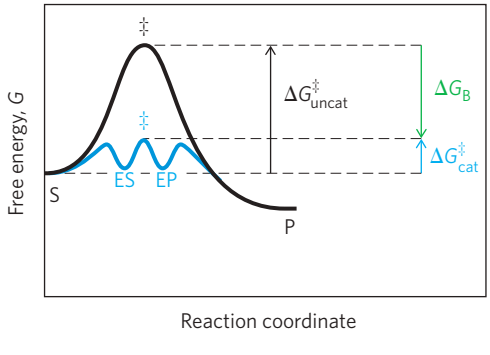


FIGURE 6–6 Role of binding energy in catalysis. To lower the activation energy for a reaction, the system must acquire an amount of energy equivalent to the amount by which $\Delta G^{\ddagger}$ is lowered. Much of this energy comes from binding energy (ΔG_B) contributed by formation of weak noncovalent interactions between substrate and enzyme in the transition state. The role of ΔG_B is analogous to that of ΔG_M in Figure 6–5.

The requirement for multiple weak interactions to drive catalysis is one reason why enzymes (and some coenzymes) are so large. An enzyme must provide functional groups for ionic, hydrogen-bond, and other interactions, and also must precisely position these groups so that binding energy is optimized in the transition state. Adequate binding is accomplished most readily by positioning a substrate in a cavity (the active site) where it is effectively removed from water. The size of proteins reflects the need for superstructure to keep interacting groups properly positioned and to keep the cavity from collapsing.

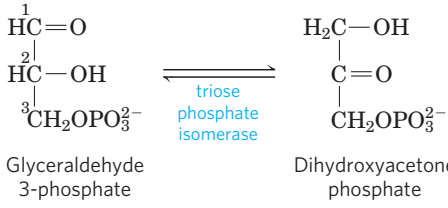
Binding Energy Contributes to Reaction Specificity and Catalysis

Can we demonstrate quantitatively that binding energy accounts for the huge rate accelerations brought about by enzymes? Yes. As a point of reference, Equation 6–6 allows us to calculate that $\Delta G^{\ddagger}$ must be lowered by about 5.7 kJ/mol to accelerate a first-order reaction by a factor of 10, under conditions commonly found in cells. The energy available from formation of a single weak interaction is generally estimated to be 4 to 30 kJ/mol. The overall energy available from a number of such interactions is therefore sufficient to lower activation energies by the 60 to 100 kJ/mol required to explain the large rate enhancements observed for many enzymes.

The same binding energy that provides energy for catalysis also gives an enzyme its **specificity**, the ability to discriminate between a substrate and a competing molecule. Conceptually, specificity is easy to distinguish from catalysis, but this distinction is much more difficult to make experimentally, because catalysis and specificity arise from the same phenomenon. If an enzyme active site has functional groups arranged optimally to form a variety of weak interactions with a particular substrate in the transition state, the enzyme will not be able to interact to the same degree with any other molecule. For example, if the substrate has a hydroxyl group that forms a hydrogen bond with a specific Glu residue on the enzyme, any molecule lacking a hydroxyl group at that particular position will be a poorer substrate for the enzyme. In addition, any molecule with an extra functional group for which the enzyme has no pocket or binding site is likely to be excluded from the enzyme. In general, *specificity* is derived from the formation of many weak interactions between the enzyme and its specific substrate molecule.

The importance of binding energy to catalysis can be readily demonstrated. For example, the glycolytic enzyme triose phosphate isomerase catalyzes the interconversion

of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate:



This reaction rearranges the carbonyl and hydroxyl groups on carbons 1 and 2. However, more than 80% of the enzymatic rate acceleration has been traced to enzyme-substrate interactions involving the phosphate group on carbon 3 of the substrate. This was determined by comparing the enzyme-catalyzed reactions with glyceraldehyde 3-phosphate and with glyceraldehyde (no phosphate group at position 3) as substrate.

The general principles outlined above can be illustrated by a variety of recognized catalytic mechanisms. These mechanisms are not mutually exclusive, and a given enzyme might incorporate several types in its overall mechanism of action.

Consider what needs to occur for a reaction to take place. Prominent physical and thermodynamic factors contributing to $\Delta G^{\ddagger}$, the barrier to reaction, might include: (1) the entropy (freedom of motion) of molecules in solution, which reduces the possibility that they will react together, (2) the solvation shell of hydrogen-bonded water that surrounds and helps to stabilize most biomolecules in aqueous solution, (3) the distortion of substrates that must occur in many reactions, and (4) the need for proper alignment of catalytic functional groups on the enzyme. Binding energy can be used to overcome all these barriers.

First, a large restriction in the relative motions of two substrates that are to react, or **entropy reduction**, is one obvious benefit of binding them to an enzyme. Binding energy holds the substrates in the proper orientation to react—a substantial contribution to catalysis, because productive collisions between molecules in solution can be exceedingly rare. Substrates can be precisely aligned on the enzyme, with many weak interactions between each substrate and strategically located groups on the enzyme clamping the substrate molecules into the proper positions. Studies have shown that constraining the motion of two reactants can produce rate enhancements of many orders of magnitude **(Fig. 6–7)**.

Second, formation of weak bonds between substrate and enzyme results in **desolvation** of the substrate. Enzyme-substrate interactions replace most or all of the hydrogen bonds between the substrate and water that would otherwise impede reaction. Third, binding energy involving weak interactions formed only in the reaction transition state helps to compensate thermodynamically for any distortion, primarily electron redistribution, that the substrate must undergo to react.

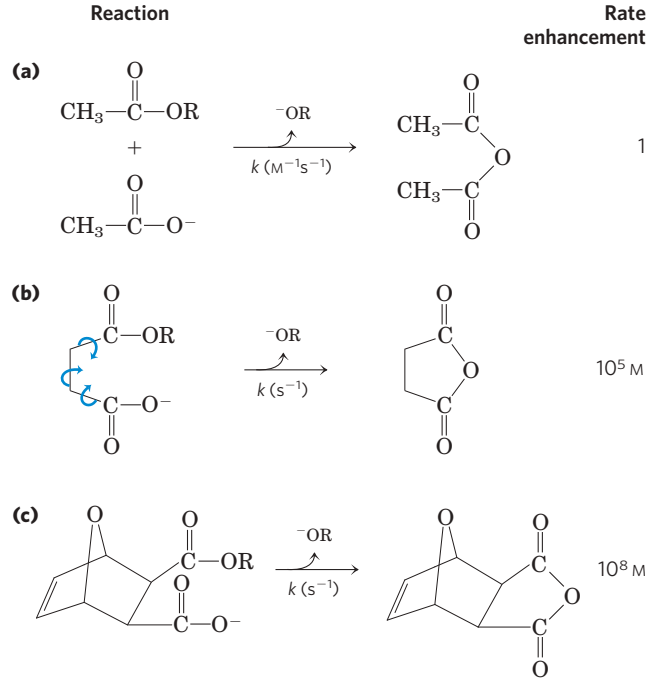


FIGURE 6–7 Rate enhancement by entropy reduction. Shown here are reactions of an ester with a carboxylate group to form an anhydride. The R group is the same in each case. **(a)** For this bimolecular reaction, the rate constant k is second-order, with units of $M^{-1}s^{-1}$. **(b)** When the two reacting groups are in a single molecule, and thus have less freedom of motion, the reaction is much faster. For this unimolecular reaction, k has units of s^{-1}. Dividing the rate constant for (b) by the rate constant for (a) gives a rate enhancement of about 10^5 M. (The enhancement has units of molarity because we are comparing a unimolecular and a bimolecular reaction.) Put another way, if the reactant in (b) were present at a concentration of 1 M, the reacting groups would *behave* as though they were present at a concentration of 10^5 M. Note that the reactant in (b) has freedom of rotation about three bonds (shown with curved arrows), but this still represents a substantial reduction of entropy over (a). If the bonds that rotate in (b) are constrained as in **(c)**, the entropy is reduced further and the reaction exhibits a rate enhancement of 10^8 M relative to (a).

Finally, the enzyme itself usually undergoes a change in conformation when the substrate binds, induced by multiple weak interactions with the substrate. This is referred to as **induced fit**, a mechanism postulated by Daniel Koshland in 1958. The motions can affect a small part of the enzyme near the active site or can involve changes in the positioning of entire domains. Typically, a network of coupled motions occurs throughout the enzyme that ultimately brings about the required changes in the active site. Induced fit serves to bring specific functional groups on the enzyme into the proper position to catalyze the reaction. The conformational change also permits formation of additional weak bonding interactions in the transition state. In either case, the new enzyme conformation has enhanced catalytic properties. As we have seen, induced fit is a common feature of the reversible binding of ligands to proteins (Chapter 5). Induced fit is also important in the interaction of almost every enzyme with its substrate.

Specific Catalytic Groups Contribute to Catalysis

In most enzymes, the binding energy used to form the ES complex is just one of several contributors to the overall catalytic mechanism. Once a substrate is bound to an enzyme, properly positioned catalytic functional groups aid in the cleavage and formation of bonds by a variety of mechanisms, including general acid-base catalysis, covalent catalysis, and metal ion catalysis. These are distinct from mechanisms based on binding energy, because they generally involve transient *covalent* interaction with a substrate or group transfer to or from a substrate.

General Acid-Base Catalysis A proton transfer is the single most common reaction in biochemistry. One or, often, many proton transfers occur in the course of most reactions that take place in cells. Many biochemical reactions involve the formation of unstable charged intermediates that tend to break down rapidly to their constituent reactant species, thus impeding the reaction **(Fig. 6–8)**. Charged intermediates can often be stabilized by the transfer of protons to or from the substrate or intermediate to form a species that breaks down more readily to products. The effects of catalysis by acids and bases are often studied using nonenzymatic model reactions, in which the proton transfers can involve either the constituents of water alone or other weak proton donors or acceptors. Catalysis of the type that uses only the H^+ (H_3O^+) or OH^- ions present in water is referred to as **specific acid-base catalysis**. If protons are transferred between the intermediate and water faster than the intermediate breaks down to reactants, the intermediate is effectively stabilized every time it forms. No additional catalysis mediated by other proton acceptors or donors will occur. In many cases, however, water is not enough. The term **general acid-base catalysis** refers to proton transfers mediated by weak acids and bases other than water. For nonenzymatic reactions in aqueous solutions, added weak acids and bases provide an observed rate acceleration only when the unstable reaction intermediate breaks down to reactants faster than protons can be transferred to or from water alone. Many weak organic acids can supplement water as proton donors in this situation, or weak organic bases can serve as proton acceptors.

In the active site of an enzyme, where water may not be available as a proton donor or acceptor, general acid-base catalysis becomes crucial. A number of amino acid side chains can and do take on the role of proton donors and acceptors **(Fig. 6–9)**. These groups can be precisely positioned in an enzyme active site to allow proton transfers, providing rate enhancements of the order of 10^2 to 10^5. This type of catalysis occurs on the vast majority of enzymes.

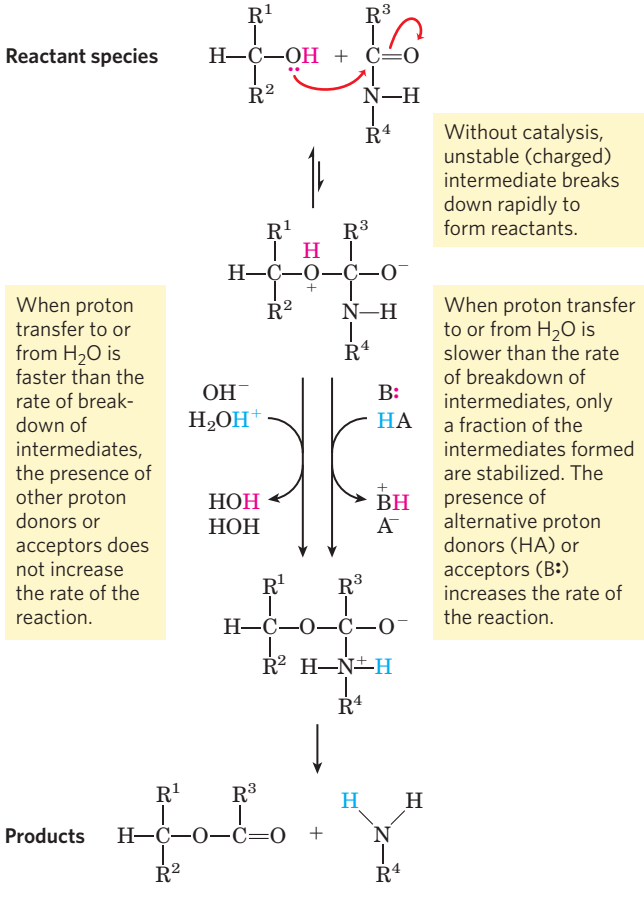


FIGURE 6–8 How a catalyst circumvents unfavorable charge development during cleavage of an amide. The hydrolysis of an amide bond, shown here, is the same reaction as that catalyzed by chymotrypsin and other proteases. Charge development is unfavorable and can be circumvented by donation of a proton by H_3O^+ (specific acid catalysis) or HA (general acid catalysis), where HA represents any acid. Similarly, charge can be neutralized by proton abstraction by OH^- (specific base catalysis) or B: (general base catalysis), where B: represents any base.

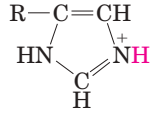
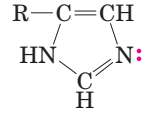
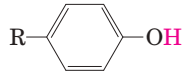
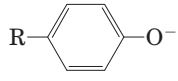
Amino acid residues	General acid form (proton donor)	General base form (proton acceptor)
Glu, Asp	R—COOH	R—COO^-
Lys, Arg	R—$\overset{+}{N}H_3$	R—NH_2
Cys	R—SH	R—S^-
His		
Ser	R—OH	R—O^-
Tyr		

FIGURE 6–9 Amino acids in general acid-base catalysis. Many organic reactions that are used to model biochemical processes are promoted by proton donors (general acids) or proton acceptors (general bases). The active sites of some enzymes contain amino acid functional groups, such as those shown here, that can participate in the catalytic process as proton donors or proton acceptors.

Covalent Catalysis In covalent catalysis, a transient covalent bond is formed between the enzyme and the substrate. Consider the hydrolysis of a bond between groups A and B:

$$A—B \xrightarrow{H_2O} A + B$$

In the presence of a covalent catalyst (an enzyme with a nucleophilic group X:) the reaction becomes

$$A—B + X: \longrightarrow A—X + B \xrightarrow{H_2O} A + X: + B$$

This alters the pathway of the reaction, and it results in catalysis *only* when the new pathway has a lower activation energy than the uncatalyzed pathway. Both of the new steps must be faster than the uncatalyzed reaction. A number of amino acid side chains, including all those in Figure 6–9, and the functional groups of some enzyme cofactors can serve as nucleophiles in the formation of covalent bonds with substrates. These covalent complexes always undergo further reaction to regenerate the free enzyme. The covalent bond formed between the enzyme and the substrate can activate a substrate for further reaction in a manner that is usually specific to the particular group or coenzyme.

Metal Ion Catalysis Metals, whether tightly bound to the enzyme or taken up from solution along with the substrate, can participate in catalysis in several ways. Ionic interactions between an enzyme-bound metal and a substrate can help orient the substrate for reaction or stabilize charged reaction transition states. This use of weak bonding interactions between metal and substrate is similar to some of the uses of enzyme-substrate binding energy described earlier. Metals can also mediate oxidation-reduction reactions by reversible changes in the metal ion's oxidation state. Nearly a third of all known enzymes require one or more metal ions for catalytic activity.

Most enzymes combine several catalytic strategies to bring about a rate enhancement. A good example is the use of covalent catalysis, general acid-base catalysis, and transition-state stabilization in the reaction catalyzed by chymotrypsin, detailed in Section 6.4.

SUMMARY 6.2 How Enzymes Work

- Enzymes are highly effective catalysts, commonly enhancing reaction rates by a factor of 10^5 to 10^{17}.
- Enzyme-catalyzed reactions are characterized by the formation of a complex between substrate and enzyme (an ES complex). Substrate binding occurs in a pocket on the enzyme called the active site.
- The function of enzymes and other catalysts is to lower the activation energy, $\Delta G^{\ddagger}$, for a reaction and thereby enhance the reaction rate. The equilibrium of a reaction is unaffected by the enzyme.
- A significant part of the energy used for enzymatic rate enhancements is derived from weak interactions (hydrogen bonds and hydrophobic and ionic interactions) between substrate and enzyme. The enzyme active site is structured so that some of these weak interactions occur preferentially in the reaction transition state, thus stabilizing the transition state. The need for multiple interactions is one reason for the large size of enzymes. The binding energy, ΔG_B, is used to offset the energy required for activation, $\Delta G^{\ddagger}$, in several ways. It can be used, for example, to lower substrate entropy, for substrate desolvation, or to cause a conformational change in the enzyme (induced fit). Binding energy also accounts for the exquisite specificity of enzymes for their substrates.
- Additional catalytic mechanisms employed by enzymes include general acid-base catalysis, covalent catalysis, and metal ion catalysis. Catalysis often involves transient covalent interactions between the substrate and the enzyme, or group transfers to and from the enzyme, so as to provide a new, lower-energy reaction path. In all cases, the enzyme reverts to the unbound state once the reaction is complete.

6.3 Enzyme Kinetics as an Approach to Understanding Mechanism

Biochemists commonly use several approaches to study the mechanism of action of purified enzymes. The three-dimensional structure of the protein provides important information, which is enhanced by classical protein chemistry and modern methods of site-directed mutagenesis (changing the amino acid sequence of a protein by genetic engineering; see Fig. 9–10). These technologies permit enzymologists to examine the role of individual amino acids in enzyme structure and action. However, the oldest approach to understanding enzyme mechanisms, and the one that remains most important, is to determine the *rate* of a reaction and how it changes in response to changes in experimental parameters, a discipline known as **enzyme kinetics**. We provide here a basic introduction to the kinetics of enzyme-catalyzed reactions. More advanced treatments are available in the sources cited at the end of the chapter.

Substrate Concentration Affects the Rate of Enzyme-Catalyzed Reactions

A key factor affecting the rate of a reaction catalyzed by an enzyme is the concentration of substrate, [S]. However, studying the effects of substrate concentration is complicated by the fact that [S] changes during the course of an in vitro reaction as substrate is converted to product. One simplifying approach in kinetics experiments is to measure the **initial rate** (or **initial**

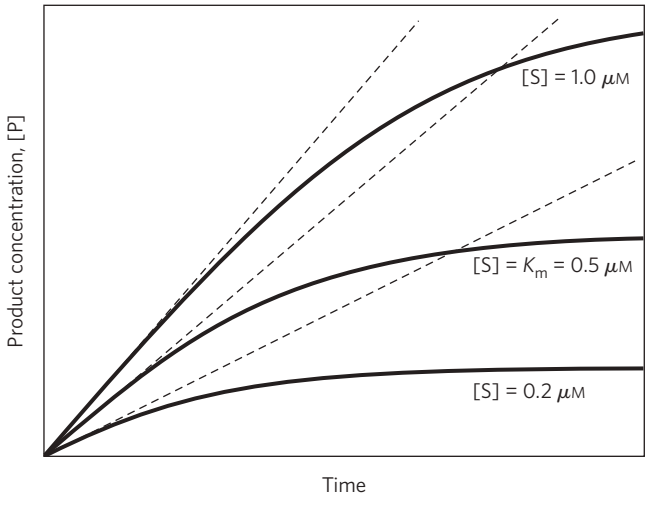


FIGURE 6–10 Initial velocities of enzyme-catalyzed reactions. A theoretical enzyme catalyzes the reaction S ⇌ P, and is present at a concentration sufficient to catalyze the reaction at a maximum velocity, V_{max}, of 1 μM/min. The Michaelis constant, K_m (explained in the text), is 0.5 μM. Progress curves are shown for substrate concentrations below, at, and above the K_m. The rate of an enzyme-catalyzed reaction declines as substrate is converted to product. A tangent to each curve taken at time = 0 defines the initial velocity, V_0, of each reaction.

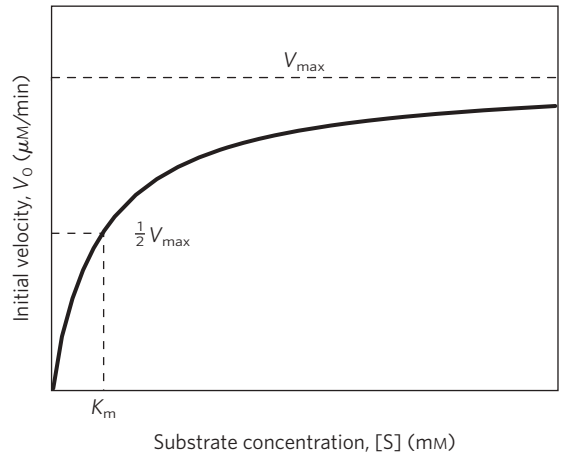


FIGURE 6–11 Effect of substrate concentration on the initial velocity of an enzyme-catalyzed reaction. The maximum velocity, V_{max}, is extrapolated from the plot, because V_0 approaches but never quite reaches V_{max}. The substrate concentration at which V_0 is half maximal is K_m, the Michaelis constant. The concentration of enzyme in an experiment such as this is generally so low that [S] >> [E] even when [S] is described as low or relatively low. The units shown are typical for enzyme-catalyzed reactions and are given only to help illustrate the meaning of V_0 and [S]. (Note that the curve describes *part* of a rectangular hyperbola, with one asymptote at V_{max}. If the curve were continued below [S] = 0, it would approach a vertical asymptote at [S] = $-K_m$.)

velocity), designated $\boldsymbol{V_0}$ **(Fig. 6–10)**. In a typical reaction, the enzyme may be present in nanomolar quantities, whereas [S] may be five or six orders of magnitude higher. If only the beginning of the reaction is monitored (often the first 60 seconds or less), changes in [S] can be limited to a few percent, and [S] can be regarded as constant. V_0 can then be explored as a function of [S], which is adjusted by the investigator. The effect on V_0 of varying [S] when the enzyme concentration is held constant is shown in **Figure 6–11**. At relatively low concentrations of substrate, V_0 increases almost linearly with an increase in [S]. At higher substrate concentrations, V_0 increases by smaller and smaller amounts in response to increases in [S]. Finally, a point is reached beyond which increases in V_0 are vanishingly small as [S] increases. This plateau-like V_0 region is close to the **maximum velocity**, $\boldsymbol{V_{max}}$.

The ES complex is the key to understanding this kinetic behavior, just as it was a starting point for our discussion of catalysis. The kinetic pattern in Figure 6–11 led Victor Henri, following the lead of Wurtz, to propose in 1903 that the combination of an enzyme with its substrate molecule to form an ES complex is a necessary step in enzymatic catalysis. This idea was expanded into a general theory of enzyme action, particularly by Leonor Michaelis and Maud Menten in 1913. They postulated that the enzyme first combines reversibly with its substrate to form an enzyme-substrate complex in a relatively fast reversible step:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \qquad (6–7)$$

The ES complex then breaks down in a slower second step to yield the free enzyme and the reaction product P:

$$ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P \qquad (6–8)$$

Because the slower second reaction (Eqn 6–8) must limit the rate of the overall reaction, the overall rate must be proportional to the concentration of the species that reacts in the second step, that is, ES.

At any given instant in an enzyme-catalyzed reaction, the enzyme exists in two forms, the free or uncombined form E and the combined form ES. At low [S], most of the enzyme is in the uncombined form E. Here, the rate is proportional to [S] because the equilibrium of Equation 6–7 is pushed toward formation of more ES as [S] increases. The maximum initial rate of the catalyzed reaction (V_{max}) is observed when virtually all the enzyme is present as the ES complex and [E] is vanishingly

Leonor Michaelis, 1875–1949

Maud Menten, 1879–1960

small. Under these conditions, the enzyme is "saturated" with its substrate, so that further increases in [S] have no effect on rate. This condition exists when [S] is sufficiently high that essentially all the free enzyme has been converted to the ES form. After the ES complex breaks down to yield the product P, the enzyme is free to catalyze reaction of another molecule of substrate (and will do so rapidly under saturating conditions). The saturation effect is a distinguishing characteristic of enzymatic catalysts and is responsible for the plateau observed in Figure 6–11. The pattern seen in Figure 6–11 is sometimes referred to as saturation kinetics.

When the enzyme is first mixed with a large excess of substrate, there is an initial period, the **pre–steady state**, during which the concentration of ES builds up. This period is usually too short to be easily observed, lasting just microseconds, and is not evident in Figure 6–10. The reaction quickly achieves a **steady state** in which [ES] (and the concentrations of any other intermediates) remains approximately constant over time. The concept of a steady state was introduced by G. E. Briggs and Haldane in 1925. The measured V_0 generally reflects the steady state, even though V_0 is limited to the early part of the reaction, and analysis of these initial rates is referred to as **steady-state kinetics**.

The Relationship between Substrate Concentration and Reaction Rate Can Be Expressed Quantitatively

The curve expressing the relationship between [S] and V_0 (Fig. 6–11) has the same general shape for most enzymes (it approaches a rectangular hyperbola), which can be expressed algebraically by the Michaelis-Menten equation. Michaelis and Menten derived this equation starting from their basic hypothesis that the rate-limiting step in enzymatic reactions is the breakdown of the ES complex to product and free enzyme. The equation is

$$V_0 = \frac{V_{max}[S]}{K_m + [S]} \qquad (6\text{–}9)$$

All these terms—[S], V_0, V_{max}, and a constant called the Michaelis constant, K_m—are readily measured experimentally.

Here we develop the basic logic and the algebraic steps in a modern derivation of the Michaelis-Menten equation, which includes the steady-state assumption introduced by Briggs and Haldane. The derivation starts with the two basic steps of the formation and breakdown of ES (Eqns 6–7 and 6–8). Early in the reaction, the concentration of the product, [P], is negligible, and we make the simplifying assumption that the reverse reaction, P → S (described by k_{-2}), can be ignored. This assumption is not critical but it simplifies our task. The overall reaction then reduces to

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \qquad (6\text{–}10)$$

V_0 is determined by the breakdown of ES to form product, which is determined by [ES]:

$$V_0 = k_2[ES] \qquad (6\text{–}11)$$

Because [ES] in Equation 6–11 is not easily measured experimentally, we must begin by finding an alternative expression for this term. First, we introduce the term $[E_t]$, representing the total enzyme concentration (the sum of free and substrate-bound enzyme). Free or unbound enzyme [E] can then be represented by $[E_t] - [ES]$. Also, because [S] is ordinarily far greater than $[E_t]$, the amount of substrate bound by the enzyme at any given time is negligible compared with the total [S]. With these conditions in mind, the following steps lead us to an expression for V_0 in terms of easily measurable parameters.

Step 1 The rates of formation and breakdown of ES are determined by the steps governed by the rate constants k_1 (formation) and $k_{-1} + k_2$ (breakdown to reactants and products, respectively), according to the expressions

$$\text{Rate of ES formation} = k_1([E_t] - [ES])[S] \qquad (6\text{–}12)$$

$$\text{Rate of ES breakdown} = k_{-1}[ES] + k_2[ES] \qquad (6\text{–}13)$$

Step 2 We now make an important assumption: that the initial rate of reaction reflects a steady state in which [ES] is constant—that is, the rate of formation of ES is equal to the rate of its breakdown. This is called the **steady-state assumption**. The expressions in Equations 6–12 and 6–13 can be equated for the steady state, giving

$$k_1([E_t] - [ES])[S] = k_{-1}[ES] + k_2[ES] \qquad (6\text{–}14)$$

Step 3 In a series of algebraic steps, we now solve Equation 6–14 for [ES]. First, the left side is multiplied out and the right side simplified to give

$$k_1[E_t][S] - k_1[ES][S] = (k_{-1} + k_2)[ES] \qquad (6\text{–}15)$$

Adding the term $k_1[ES][S]$ to both sides of the equation and simplifying gives

$$k_1[E_t][S] = (k_1[S] + k_{-1} + k_2)[ES] \qquad (6\text{–}16)$$

We then solve this equation for [ES]:

$$[ES] = \frac{k_1[E_t][S]}{k_1[S] + k_{-1} + k_2} \qquad (6\text{–}17)$$

This can now be simplified further, combining the rate constants into one expression:

$$[ES] = \frac{[E_t][S]}{[S] + (k_{-1} + k_2)/k_1} \qquad (6\text{–}18)$$

The term $(k_{-1} + k_2)/k_1$ is defined as the **Michaelis constant**, $\boldsymbol{K_m}$. Substituting this into Equation 6–18 simplifies the expression to

$$[ES] = \frac{[E_t][S]}{K_m + [S]} \qquad (6\text{–}19)$$

Step 4 We can now express V_0 in terms of [ES]. Substituting the right side of Equation 6–19 for [ES] in Equation 6–11 gives

$$V_0 = \frac{k_2[E_t][S]}{K_m + [S]} \qquad (6\text{–}20)$$

This equation can be further simplified. Because the maximum velocity occurs when the enzyme is saturated (that is, with $[ES] = [E_t]$), V_{max} can be defined as $k_2[E_t]$. Substituting this in Equation 6–20 gives Equation 6–9:

$$V_0 = \frac{V_{max}[S]}{K_m + [S]}$$

This is the **Michaelis-Menten equation**, the **rate equation** for a one-substrate enzyme-catalyzed reaction. It is a statement of the quantitative relationship between the initial velocity V_0, the maximum velocity V_{max}, and the initial substrate concentration [S], all related through the Michaelis constant K_m. Note that K_m has units of molar concentration. Does the equation fit experimental observations? Yes; we can confirm this by considering the limiting situations where [S] is very high or very low, as shown in **Figure 6–12**.

An important numerical relationship emerges from the Michaelis-Menten equation in the special case when V_0 is exactly one-half V_{max} (Fig. 6–12). Then

$$\frac{V_{max}}{2} = \frac{V_{max}[S]}{K_m + [S]} \qquad (6\text{–}21)$$

On dividing by V_{max}, we obtain

$$\frac{1}{2} = \frac{[S]}{K_m + [S]} \qquad (6\text{–}22)$$

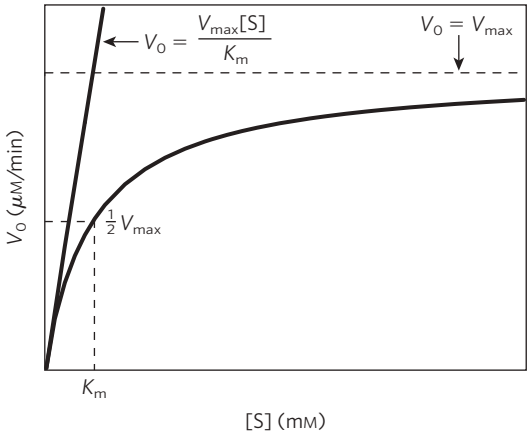


FIGURE 6–12 Dependence of initial velocity on substrate concentration. This graph shows the kinetic parameters that define the limits of the curve at high and low [S]. At low [S], $K_m >> [S]$ and the [S] term in the denominator of the Michaelis-Menten equation (Eqn 6–9) becomes insignificant. The equation simplifies to $V_0 = V_{max}[S]/K_m$ and V_0 exhibits a linear dependence on [S], as observed here. At high [S], where $[S] >> K_m$, the K_m term in the denominator of the Michaelis-Menten equation becomes insignificant and the equation simplifies to $V_0 = V_{max}$; this is consistent with the plateau observed at high [S]. The Michaelis-Menten equation is therefore consistent with the observed dependence of V_0 on [S], and the shape of the curve is defined by the terms V_{max}/K_m at low [S] and V_{max} at high [S].

Solving for K_m, we get $K_m + [S] = 2[S]$, or

$$K_m = [S], \quad \text{when} \quad V_0 = \tfrac{1}{2}V_{max} \qquad (6\text{–}23)$$

This is a very useful, practical definition of K_m: K_m is equivalent to the substrate concentration at which V_0 is one-half V_{max}.

The Michaelis-Menten equation (Eqn 6–9) can be algebraically transformed into versions that are useful in the practical determination of K_m and V_{max} (Box 6–1) and, as we describe later, in the analysis of inhibitor action (see Box 6–2 on p. 209).

Kinetic Parameters Are Used to Compare Enzyme Activities

It is important to distinguish between the Michaelis-Menten equation and the specific kinetic mechanism on which it was originally based. The equation describes the kinetic behavior of a great many enzymes, and all enzymes that exhibit a hyperbolic dependence of V_0 on [S] are said to follow **Michaelis-Menten kinetics**. The practical rule that $K_m = [S]$ when $V_0 = \frac{1}{2}V_{max}$ (Eqn 6–23) holds for all enzymes that follow Michaelis-Menten kinetics. (The most important exceptions to Michaelis-Menten kinetics are the regulatory enzymes, discussed in Section 6.5.) However, the Michaelis-Menten equation does not depend on the relatively simple two-step reaction mechanism proposed by Michaelis and Menten (Eqn 6–10). Many enzymes that follow Michaelis-Menten kinetics have quite different reaction mechanisms, and enzymes that catalyze reactions with six or eight identifiable steps often exhibit the same steady-state kinetic behavior. Even though Equation 6–23 holds true for many enzymes, both the magnitude and the real meaning of V_{max} and K_m can differ from one enzyme to the next. This is an important limitation of the steady-state approach to enzyme kinetics. The parameters V_{max} and K_m can be obtained experimentally for any given enzyme, but by themselves they provide little information about the number, rates, or chemical nature of discrete steps in the reaction. Steady-state kinetics nevertheless is the standard language by which biochemists compare and characterize the catalytic efficiencies of enzymes.

Interpreting V_{max} and K_m Figure 6–12 shows a simple graphical method for obtaining an approximate value for K_m. A more convenient procedure, using a **double-reciprocal plot**, is presented in Box 6–1. The K_m can vary greatly from enzyme to enzyme, and even for different substrates of the same enzyme (Table 6–6). The term is sometimes used (often inappropriately) as an indicator of the affinity of an enzyme for its substrate. The actual meaning of K_m depends on specific aspects of the reaction mechanism such as the number and relative rates of the individual steps. For reactions with two steps,

BOX 6–1 Transformations of the Michaelis-Menten Equation: The Double-Reciprocal Plot

The Michaelis-Menten equation

$$V_0 = \frac{V_{max}[S]}{K_m + [S]}$$

can be algebraically transformed into equations that are more useful in plotting experimental data. One common transformation is derived simply by taking the reciprocal of both sides of the Michaelis-Menten equation:

$$\frac{1}{V_0} = \frac{K_m + [S]}{V_{max}[S]}$$

Separating the components of the numerator on the right side of the equation gives

$$\frac{1}{V_0} = \frac{K_m}{V_{max}[S]} + \frac{[S]}{V_{max}[S]}$$

which simplifies to

$$\frac{1}{V_0} = \frac{K_m}{V_{max}[S]} + \frac{1}{V_{max}}$$

This form of the Michaelis-Menten equation is called the **Lineweaver-Burk equation**. For enzymes obeying the Michaelis-Menten relationship, a plot of $1/V_0$ versus 1/[S] (the "double reciprocal" of the V_0 versus [S] plot we have been using to this point) yields a straight line (Fig. 1). This line has a slope of K_m/V_{max}, an intercept of $1/V_{max}$ on the $1/V_0$ axis, and an intercept of $-1/K_m$ on the 1/[S] axis. The double-reciprocal presentation, also called a Lineweaver-Burk plot, has the great advantage of allowing a more accurate determination of V_{max}, which can only be *approximated* from a simple plot of V_0 versus [S] (see Fig. 6–12).

Other transformations of the Michaelis-Menten equation have been derived, each with some particular advantage in analyzing enzyme kinetic data. (See Problem 14 at the end of this chapter.)

The double-reciprocal plot of enzyme reaction rates is very useful in distinguishing between certain types of enzymatic reaction mechanisms (see Fig. 6–14) and in analyzing enzyme inhibition (see Box 6–2).

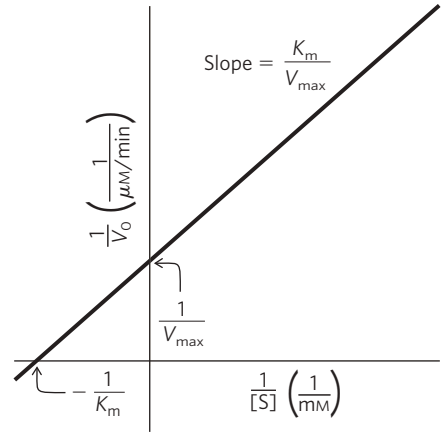


FIGURE 1 A double-reciprocal or Lineweaver-Burk plot.

$$K_m = \frac{k_2 + k_{-1}}{k_1} \qquad (6\text{–}24)$$

When k_2 is rate-limiting, $k_2 << k_{-1}$ and K_m reduces to k_{-1}/k_1, which is defined as the **dissociation constant**, K_d, of the ES complex. Where these conditions hold, K_m does represent a measure of the affinity of the enzyme for its substrate in the ES complex. However, this scenario does not apply for most enzymes. Sometimes $k_2 >> k_{-1}$, and then $K_m = k_2/k_1$. In other cases, k_2 and k_{-1} are comparable and K_m remains a more complex function of all three rate constants (Eqn 6–24). The Michaelis-Menten equation and the characteristic saturation behavior of the enzyme still apply, but K_m cannot be considered a simple measure of substrate affinity. Even more common are cases in which the reaction goes through several steps after formation of ES; K_m can then become a very complex function of many rate constants.

The quantity V_{max} also varies greatly from one enzyme to the next. If an enzyme reacts by the two-step Michaelis-Menten mechanism, $V_{max} = k_2[E_t]$, where k_2 is rate limiting. However, the number of reaction steps and the identity of the rate-limiting step(s) can vary from enzyme to enzyme. For example, consider the quite common situation where product release, EP → E + P, is rate-limiting. Early in the reaction (when [P] is low), the overall reaction can be described by the scheme

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EP \overset{k_3}{\rightleftharpoons} E + P \qquad (6\text{–}25)$$

In this case, most of the enzyme is in the EP form at saturation, and $V_{max} = k_3[E_t]$. It is useful to define a more general rate constant,

TABLE 6–6 K_m for Some Enzymes and Substrates

Enzyme	Substrate	K_m (mM)
Hexokinase (brain)	ATP	0.4
	D-Glucose	0.05
	D-Fructose	1.5
Carbonic anhydrase	HCO_3^-	26
Chymotrypsin	Glycyltyrosinylglycine	108
	N-Benzoyltyrosinamide	2.5
β-Galactosidase	D-Lactose	4.0
Threonine dehydratase	L-Threonine	5.0

k_{cat}, to describe the limiting rate of any enzyme-catalyzed reaction at saturation. If the reaction has several steps and one is clearly rate limiting, k_{cat} is equivalent to the rate constant for that limiting step. For the simple reaction of Equation 6–10, $k_{cat} = k_2$. For the reaction of Equation 6–25, when product release is clearly rate-limiting, $k_{cat} = k_3$. When several steps are partially rate-limiting, k_{cat} can become a complex function of several of the rate constants that define each individual reaction step. In the Michaelis-Menten equation, $k_{cat} = V_{max}/[E_t]$, and Equation 6–9 becomes

$$V_0 = \frac{k_{cat}[E_t][S]}{K_m + [S]} \qquad (6\text{–}26)$$

The constant k_{cat} is a first-order rate constant and hence has units of reciprocal time. It is also called the **turnover number**. It is equivalent to the number of substrate molecules converted to product in a given unit of time on a single enzyme molecule when the enzyme is saturated with substrate. The turnover numbers of several enzymes are given in Table 6–7.

TABLE 6–7 Turnover Number, k_{cat}, of Some Enzymes

Enzyme	Substrate	k_{cat} (s^{-1})
Catalase	H_2O_2	40,000,000
Carbonic anhydrase	HCO_3^-	400,000
Acetylcholinesterase	Acetylcholine	14,000
β-Lactamase	Benzylpenicillin	2,000
Fumarase	Fumarate	800
RecA protein (an ATPase)	ATP	0.5

Comparing Catalytic Mechanisms and Efficiencies The kinetic parameters k_{cat} and K_m are useful for the study and comparison of different enzymes, whether their reaction mechanisms are simple or complex. Each enzyme has values of k_{cat} and K_m that reflect the cellular environment, the concentration of substrate normally encountered in vivo by the enzyme, and the chemistry of the reaction being catalyzed.

The parameters k_{cat} and K_m also allow us to evaluate the kinetic efficiency of enzymes, but either parameter alone is insufficient for this task. Two enzymes catalyzing different reactions may have the same k_{cat} (turnover number), yet the rates of the uncatalyzed reactions may be different and thus the rate enhancements brought about by the enzymes may differ greatly. Experimentally, the K_m for an enzyme tends to be similar to the cellular concentration of its substrate. An enzyme that acts on a substrate present at a very low concentration in the cell usually has a lower K_m than an enzyme that acts on a substrate that is more abundant.

The best way to compare the catalytic efficiencies of different enzymes or the turnover of different substrates by the same enzyme is to compare the ratio k_{cat}/K_m for the two reactions. This parameter, sometimes called the **specificity constant**, is the rate constant for the conversion of E + S to E + P. When $[S] << K_m$, Equation 6–26 reduces to the form

$$V_0 = \frac{k_{cat}}{K_m}[E_t][S] \qquad (6\text{–}27)$$

V_0 in this case depends on the concentration of two reactants, $[E_t]$ and [S]; therefore this is a second-order rate equation and the constant k_{cat}/K_m is a second-order rate constant with units of $M^{-1}s^{-1}$. There is an upper limit to k_{cat}/K_m, imposed by the rate at which E and S can diffuse together in an aqueous solution. This diffusion-controlled limit is 10^8 to 10^9 $M^{-1}s^{-1}$, and many enzymes have a k_{cat}/K_m near this range (Table 6–8). Such enzymes are said to have achieved catalytic perfection. Note that different values of k_{cat} and K_m can produce the maximum ratio.

TABLE 6–8 Enzymes for Which k_{cat}/K_m Is Close to the Diffusion-Controlled Limit (10^8 to 10^9 $M^{-1}s^{-1}$)

Enzyme	Substrate	K_{cat} (s^{-1})	K_m (M)	K_{cat}/K_m ($M^{-1}s^{-1}$)
Acetylcholinesterase	Acetylcholine	1.4×10^4	9×10^{-5}	1.6×10^8
Carbonic anhydrase	CO_2	1×10^6	1.2×10^{-2}	8.3×10^7
	HCO_3^-	4×10^5	2.6×10^{-2}	1.5×10^7
Catalase	H_2O_2	4×10^7	1.1×10^0	4×10^7
Crotonase	Crotonyl-CoA	5.7×10^3	2×10^{-5}	2.8×10^8
Fumarase	Fumarate	8×10^2	5×10^{-6}	1.6×10^8
	Malate	9×10^2	2.5×10^{-5}	3.6×10^7
β-Lactamase	Benzylpenicillin	2.0×10^3	2×10^{-5}	1×10^8

Source: Fersht, A. (1999) *Structure and Mechanism in Protein Science*, p. 166, W. H. Freeman and Company, New York.

WORKED EXAMPLE 6–1 Determination of K_m

An enzyme is discovered that catalyzes the chemical reaction

$$\text{SAD} \rightleftharpoons \text{HAPPY}$$

A team of motivated researchers sets out to study the enzyme, which they call happyase. They find that the k_{cat} for happyase is 600 s^{-1}. They carry out several experiments.

When $[E_t]$ = 20 nM and [SAD] = 40 μM, the reaction velocity, V_0, is 9.6 $\mu M\ s^{-1}$. Calculate K_m for the substrate SAD.

Solution: We know k_{cat}, $[E_t]$, [S], and V_0. We want to solve for K_m. Equation 6–26, in which we substitute $k_{cat}[E_t]$ for V_{max} in the Michaelis-Menten equation, is most useful here. Substituting our known values in Equation 6–26 allows us to solve for K_m.

$$V_0 = \frac{k_{cat}[E_t][S]}{K_m + [S]}$$

$$9.6\ \mu M\ s^{-1} = \frac{(600\ s^{-1})(0.020\ \mu M)(40\ \mu M)}{K_m + 40\ \mu M}$$

$$9.6\ \mu M\ s^{-1} = \frac{480\ \mu M^2\ s^{-1}}{K_m + 40\ \mu M}$$

$$9.6\ \mu M\ s^{-1}(K_m + 40\ \mu M) = 480\ \mu M^2\ s^{-1}$$

$$K_m + 40\ \mu M = \frac{480\ \mu M^2\ s^{-1}}{9.6\ \mu M\ s^{-1}}$$

$$K_m + 40\ \mu M = 50\ \mu M$$

$$K_m = 50\ \mu M - 40\ \mu M$$

$$K_m = 10\ \mu M$$

Once you have worked with this equation, you will recognize shortcuts to solve problems like this. For example, one can calculate V_{max} knowing that $k_{cat}[E_t] = V_{max}$ (600 s^{-1} × 0.020 μM = 12 $\mu M\ s^{-1}$ in this case). A simple rearrangement of Equation 6–26 by dividing both sides by V_{max} gives

$$\frac{V_0}{V_{max}} = \frac{[S]}{K_m + [S]}$$

Thus, the ratio V_0/V_{max} = 9.6 $\mu M\ s^{-1}$/12 $\mu M\ s^{-1}$ = $[S]/(K_m + [S])$. This simplifies the process of solving for K_m, giving 0.25[S] or 10 μM.

WORKED EXAMPLE 6–2 Determination of [S]

In a separate happyase experiment using $[E_t]$ = 10 nM, the reaction velocity, V_0, is measured as 3 $\mu M\ s^{-1}$. What is the [S] used in this experiment?

Solution: Using the same logic as in Worked Example 6–1, we see that the V_{max} for this enzyme concentration is 6 $\mu M\ s^{-1}$. Note that the V_0 is exactly half of the V_{max}. Recall that K_m is by definition equal to the [S] where $V_0 = \frac{1}{2}V_{max}$. Thus, the [S] in this problem must be the same as the K_m, or 10 μM. If V_0 were anything other than $\frac{1}{2}V_{max}$, it would be simplest to use the expression $V_0/V_{max} = [S]/(K_m + [S])$ to solve for [S].

Many Enzymes Catalyze Reactions with Two or More Substrates

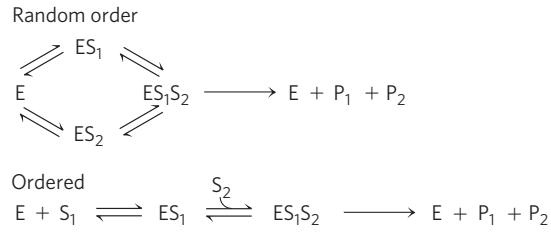
We have seen how [S] affects the rate of a simple enzymatic reaction (S → P) with only one substrate molecule. In most enzymatic reactions, however, two (and sometimes more) different substrate molecules bind to the enzyme and participate in the reaction. For example, in the reaction catalyzed by hexokinase, ATP and glucose are the substrate molecules, and ADP and glucose 6-phosphate are the products:

$$\text{ATP} + \text{glucose} \longrightarrow \text{ADP} + \text{glucose 6-phosphate}$$

The rates of such bisubstrate reactions can also be analyzed by the Michaelis-Menten approach. Hexokinase has a characteristic K_m for each of its substrates (Table 6–6).

Enzymatic reactions with two substrates usually involve transfer of an atom or a functional group from one substrate to the other. These reactions proceed by one of several different pathways. In some cases, both substrates are bound to the enzyme concurrently at some point in the course of the reaction, forming a noncovalent ternary complex **(Fig. 6–13a)**; the substrates bind in a random sequence or in a specific order. In

(a) Enzyme reaction involving a ternary complex



(b) Enzyme reaction in which no ternary complex is formed

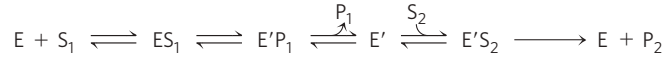


FIGURE 6–13 Common mechanisms for enzyme-catalyzed bisubstrate reactions. (a) The enzyme and both substrates come together to form a ternary complex. In ordered binding, substrate 1 must bind before substrate 2 can bind productively. In random binding, the substrates can bind in either order. **(b)** An enzyme-substrate complex forms, a product leaves the complex, the altered enzyme forms a second complex with another substrate molecule, and the second product leaves, regenerating the enzyme. Substrate 1 may transfer a functional group to the enzyme (to form the covalently modified E′), which is subsequently transferred to substrate 2. This is called a Ping-Pong or double-displacement mechanism.

other cases, the first substrate is converted to product and dissociates before the second substrate binds, so no ternary complex is formed. An example of this is the Ping-Pong, or double-displacement, mechanism (Fig. 6–13b). Steady-state kinetics can often help distinguish among these possibilities **(Fig. 6–14)**.

Pre–Steady State Kinetics Can Provide Evidence for Specific Reaction Steps

We have introduced kinetics as the primary method for studying the steps in an enzymatic reaction, and we have also outlined the limitations of the most common kinetic parameters in providing such information. The two most important experimental parameters obtained from steady-state kinetics are k_{cat} and k_{cat}/K_m. Variation in k_{cat} and k_{cat}/K_m with changes in pH or temperature can provide additional information about steps in a reaction pathway. In the case of bisubstrate reactions, steady-state kinetics can help determine whether a ternary complex is formed during the reaction (Fig. 6–14). A more complete picture generally requires more sophisticated kinetic methods that go beyond the scope of an introductory text. Here, we briefly introduce one of the most important kinetic approaches for studying reaction mechanisms, pre–steady state kinetics.

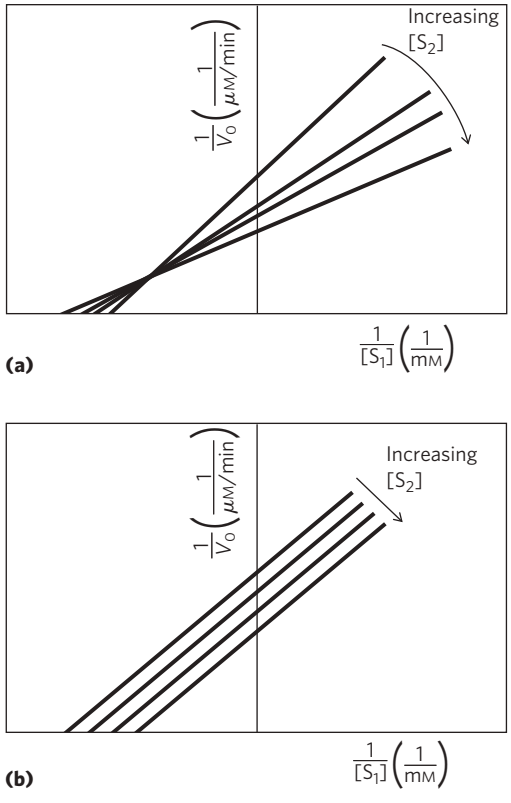


FIGURE 6–14 Steady-state kinetic analysis of bisubstrate reactions. In these double-reciprocal plots (see Box 6–1), the concentration of substrate 1 is varied while the concentration of substrate 2 is held constant. This is repeated for several values of $[S_2]$, generating several separate lines. **(a)** Intersecting lines indicate that a ternary complex is formed in the reaction; **(b)** parallel lines indicate a Ping-Pong (double-displacement) pathway.

A complete description of an enzyme-catalyzed reaction requires direct measurement of the rates of individual reaction steps—for example, the association of enzyme and substrate to form the ES complex. It is during the pre–steady state (p. 202) that the rates of many reaction steps can be measured independently and events during reaction of a single substrate molecule can be observed. Because the pre–steady state phase is generally very short, the experiments often require specialized techniques for very rapid mixing and sampling. One objective is to gain a complete and quantitative picture of the energy changes during the reaction. As we have already noted, reaction rates and equilibria are related to the free-energy changes during a reaction. Another objective is to measure the rate of individual reaction steps. In a number of cases, investigators have been able to record the rates of every individual step in a multistep enzymatic reaction. Some examples of the application of pre–steady state kinetics are included in the descriptions of specific enzymes in Section 6.4.

Enzymes Are Subject to Reversible or Irreversible Inhibition

Enzyme inhibitors are molecules that interfere with catalysis, slowing or halting enzymatic reactions. Enzymes catalyze virtually all cellular processes, so it should not be surprising that enzyme inhibitors are among the most important pharmaceutical agents known. For example, aspirin (acetylsalicylate) inhibits the enzyme that catalyzes the first step in the synthesis of prostaglandins, compounds involved in many processes, including some that produce pain. The study of enzyme inhibitors also has provided valuable information about enzyme mechanisms and has helped define some metabolic pathways. There are two broad classes of enzyme inhibitors: reversible and irreversible.

Reversible Inhibition One common type of **reversible inhibition** is called competitive **(Fig. 6–15a)**. A **competitive inhibitor** competes with the substrate for the active site of an enzyme. While the inhibitor (I) occupies the active site, it prevents binding of the substrate to the enzyme. Many competitive inhibitors are structurally similar to the substrate and combine with the enzyme to form an EI complex, but without leading to catalysis. Even fleeting combinations of this type will reduce the efficiency of the enzyme. By taking into account the molecular geometry of inhibitors, we can reach conclusions about which parts of the normal substrate bind to the enzyme. Competitive inhibition can be analyzed quantitatively by steady-state kinetics. In the presence of a competitive inhibitor, the Michaelis-Menten equation (Eqn 6–9) becomes

$$V_0 = \frac{V_{max}[S]}{\alpha K_m + [S]} \qquad (6\text{–}28)$$

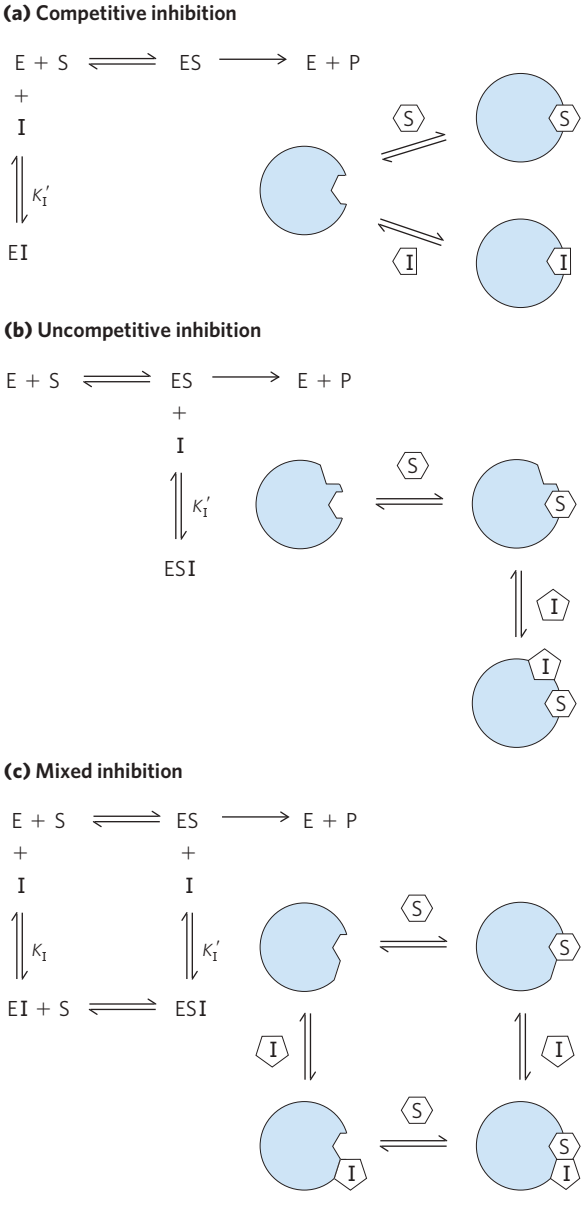


FIGURE 6–15 Three types of reversible inhibition. (a) Competitive inhibitors bind to the enzyme's active site; K_I is the equilibrium constant for inhibitor binding to E. **(b)** Uncompetitive inhibitors bind at a separate site, but bind only to the ES complex; K_I' is the equilibrium constant for inhibitor binding to ES. **(c)** Mixed inhibitors bind at a separate site, but may bind to either E or ES.

where

$$\alpha = 1 + \frac{[I]}{K_I} \quad \text{and} \quad K_I = \frac{[E][I]}{[EI]}$$

Equation 6–28 describes the important features of competitive inhibition. The experimentally determined variable αK_m, the K_m observed in the presence of the inhibitor, is often called the "apparent" K_m.

Because the inhibitor binds reversibly to the enzyme, the competition can be biased to favor the substrate simply by adding more substrate. When [S] far exceeds [I], the probability that an inhibitor molecule will bind to the enzyme is minimized and the reaction exhibits a normal V_{max}. However, the [S] at which $V_0 = \frac{1}{2}V_{max}$, the apparent K_m, increases in the presence of inhibitor by the factor α. This effect on apparent K_m, combined with the absence of an effect on V_{max}, is diagnostic of competitive inhibition and is readily revealed in a double-reciprocal plot (Box 6–2). The equilibrium constant for inhibitor binding, K_I, can be obtained from the same plot.

A medical therapy based on competition at the active site is used to treat patients who have ingested methanol, a solvent found in gas-line antifreeze. The liver enzyme alcohol dehydrogenase converts methanol to formaldehyde, which is damaging to many tissues. Blindness is a common result of methanol ingestion, because the eyes are particularly sensitive to formaldehyde. Ethanol competes effectively with methanol as an alternative substrate for alcohol dehydrogenase. The effect of ethanol is much like that of a competitive inhibitor, with the distinction that ethanol is also a substrate for alcohol dehydrogenase and its concentration will decrease over time as the enzyme converts it to acetaldehyde. The therapy for methanol poisoning is slow intravenous infusion of ethanol, at a rate that maintains a controlled concentration in the bloodstream for several hours. This slows the formation of formaldehyde, lessening the danger while the kidneys filter out the methanol to be excreted harmlessly in the urine. ■

Two other types of reversible inhibition, uncompetitive and mixed, can be defined in terms of one-substrate enzymes, but are in practice observed only with enzymes having two or more substrates. An **uncompetitive inhibitor** (Fig. 6–15b) binds at a site distinct from the substrate active site and, unlike a competitive inhibitor, binds only to the ES complex. In the presence of an uncompetitive inhibitor, the Michaelis-Menten equation is altered to

$$V_0 = \frac{V_{max}[S]}{K_m + \alpha'[S]} \qquad (6\text{–}29)$$

where

$$\alpha' = 1 + \frac{[I]}{K_I'} \quad \text{and} \quad K_I' = \frac{[ES][I]}{[ESI]}$$

As described by Equation 6–29, at high concentrations of substrate, V_0 approaches V_{max}/α'. Thus, an uncompetitive inhibitor lowers the measured V_{max}. Apparent K_m also decreases, because the [S] required to reach one-half V_{max} decreases by the factor α'.

A **mixed inhibitor** (Fig. 6–15c) also binds at a site distinct from the substrate active site, but it binds to either E or ES. The rate equation describing mixed inhibition is

$$V_0 = \frac{V_{max}[S]}{\alpha K_m + \alpha'[S]} \qquad (6\text{–}30)$$

where α and α' are defined as above. A mixed inhibitor usually affects both K_m and V_{max}. The special case of $\alpha = \alpha'$, rarely encountered in experiments, classically has been defined as **noncompetitive inhibition**.

BOX 6–2 Kinetic Tests for Determining Inhibition Mechanisms

The double-reciprocal plot (see Box 6–1) offers an easy way of determining whether an enzyme inhibitor is competitive, uncompetitive, or mixed. Two sets of rate experiments are carried out, with the enzyme concentration held constant in each set. In the first set, [S] is also held constant, permitting measurement of the effect of increasing inhibitor concentration [I] on the initial rate V_0 (not shown). In the second set, [I] is held constant but [S] is varied. The results are plotted as $1/V_0$ versus 1/[S].

Figure 1 shows a set of double-reciprocal plots, one obtained in the absence of inhibitor and two at different concentrations of a competitive inhibitor. Increasing [I] results in a family of lines with a common intercept on the $1/V_0$ axis but with different slopes. Because the intercept on the $1/V_0$ axis equals $1/V_{max}$, we know that V_{max} is unchanged by the presence of a competitive inhibitor. That is, regardless of the concentration of a competitive inhibitor, a sufficiently high substrate concentration will always displace the inhibitor from the enzyme's active site. Above the graph is the rearrangement of Equation 6–28 on which the plot is based. The value of α can be calculated from the change in slope at any given [I]. Knowing [I] and α, we can calculate K_I from the expression

$$\alpha = 1 + \frac{[I]}{K_I}$$

For uncompetitive and mixed inhibition, similar plots of rate data give the families of lines shown in Figures 2 and 3. Changes in axis intercepts signal changes in V_{max} and K_m.

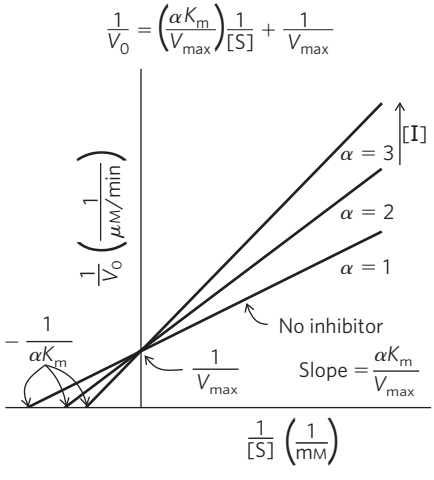


FIGURE 1 Competitive inhibition.

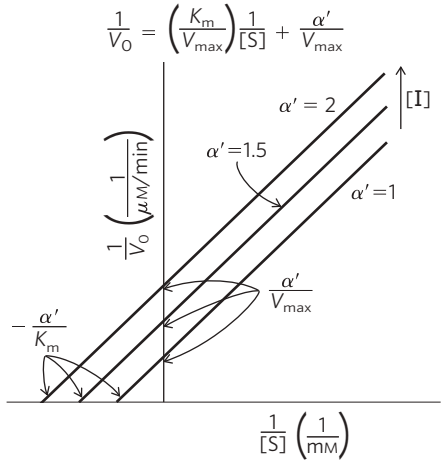


FIGURE 2 Uncompetitive inhibition.

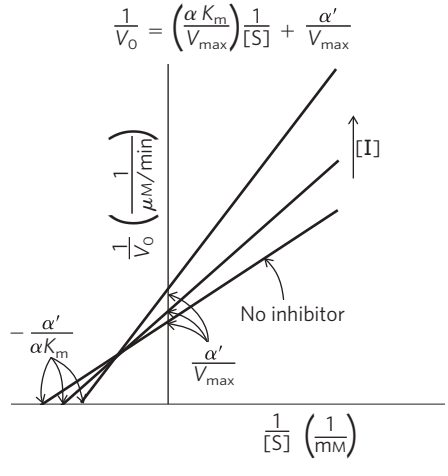


FIGURE 3 Mixed inhibition.

Examine Equation 6–30 to see why a noncompetitive inhibitor would affect the V_{max} but not the K_m.

Equation 6–30 is a general expression for the effects of reversible inhibitors, simplifying to the expressions for competitive and uncompetitive inhibition when $\alpha' = 1.0$ or $\alpha = 1.0$, respectively. From this expression we can summarize the effects of inhibitors on individual kinetic parameters. For all reversible inhibitors, the apparent $V_{max} = V_{max}/\alpha'$, because the right side of Equation 6–30 always simplifies to V_{max}/α' at sufficiently high substrate concentrations. For competitive inhibitors, $\alpha' = 1.0$ and can thus be ignored. Taking this expression for apparent V_{max}, we can also derive a general expression for apparent K_m to show how this parameter changes in the presence of reversible inhibitors. Apparent K_m, as always, equals the [S] at which V_0 is one-half apparent V_{max} or, more generally, when $V_0 = V_{max}/2\alpha'$. This condition is met when [S] = $\alpha K_m/\alpha'$. Thus, apparent $K_m = \alpha K_m/\alpha'$. This expression is simpler when either α or α' is 1.0 (for uncompetitive or competitive inhibitors), as summarized in Table 6–9.

In practice, uncompetitive and mixed inhibition are observed only for enzymes with two or more substrates—

TABLE 6–9 Effects of Reversible Inhibitors on Apparent V_{max} and Apparent K_m

Inhibitor type	Apparent V_{max}	Apparent K_m
None	V_{max}	K_m
Competitive	V_{max}	αK_m
Uncompetitive	V_{max}/α'	K_m/α'
Mixed	V_{max}/α'	$\alpha K_m/\alpha'$

say, S_1 and S_2—and are very important in the experimental analysis of such enzymes. If an inhibitor binds to the site normally occupied by S_1, it may act as a competitive inhibitor in experiments in which $[S_1]$ is varied. If an inhibitor binds to the site normally occupied by S_2, it may act as a mixed or uncompetitive inhibitor of S_1. The actual inhibition patterns observed depend on whether the S_1- and S_2-binding events are ordered or random, and thus the order in which substrates bind and products leave the active site can be determined. Use of one of the reaction products as an inhibitor is often particularly informative. If only one of two reaction products is present, no reverse reaction can take place. However, a product generally binds to some part of the active site, thus serving as an inhibitor. Enzymologists can use elaborate kinetic studies involving different combinations and amounts of products and inhibitors to develop a detailed picture of the mechanism of a bisubstrate reaction.

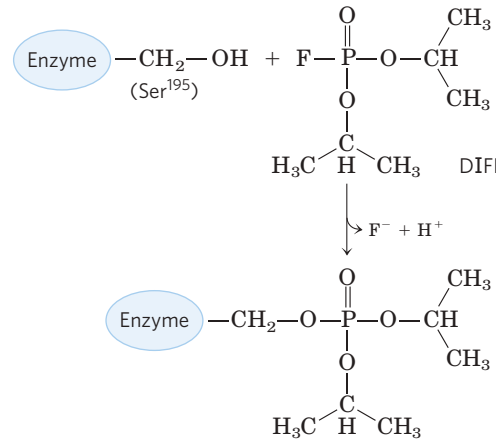


FIGURE 6–16 Irreversible inhibition. Reaction of chymotrypsin with diisopropylfluorophosphate (DIFP), which modifies Ser^{195}, irreversibly inhibits the enzyme. This has led to the conclusion that Ser^{195} is the key active-site Ser residue in chymotrypsin.

WORKED EXAMPLE 6–3 Effect of Inhibitor on K_m

The researchers working on happyase (see Worked Examples 6–1 and 6–2) discover that the compound STRESS is a potent competitive inhibitor of happyase. Addition of 1 nM STRESS increases the measured K_m for SAD by a factor of 2. What are the values for α and α' under these conditions?

Solution: Recall that the apparent K_m, the K_m measured in the presence of a competitive inhibitor, is defined as αK_m. Because K_m for SAD increases by a factor of 2 in the presence of 1 nM STRESS, the value of α must be 2. The value of α' for a competitive inhibitor is 1, by definition.

Irreversible Inhibition The **irreversible inhibitors** bind covalently with or destroy a functional group on an enzyme that is essential for the enzyme's activity, or form a particularly stable noncovalent association. Formation of a covalent link between an irreversible inhibitor and an enzyme is one approach. Irreversible inhibitors are another useful tool for studying reaction mechanisms. Amino acids with key catalytic functions in the active site can sometimes be identified by determining which residue is covalently linked to an inhibitor after the enzyme is inactivated. An example is shown in **Figure 6–16**.

A special class of irreversible inhibitors is the **suicide inactivators**. These compounds are relatively unreactive until they bind to the active site of a specific enzyme. A suicide inactivator undergoes the first few chemical steps of the normal enzymatic reaction, but instead of being transformed into the normal product, the inactivator is converted to a very reactive compound that combines irreversibly with the enzyme. These compounds are also called **mechanism-based inactivators**, because they hijack the normal enzyme reaction mechanism to inactivate the enzyme. Suicide inactivators play a significant role in *rational drug design*, a modern approach to obtaining new pharmaceutical agents in which chemists synthesize novel substrates based on knowledge of substrates and reaction mechanisms. A well-designed suicide inactivator is specific for a single enzyme and is unreactive until it is within that enzyme's active site, so drugs based on this approach can offer the important advantage of few side effects (Box 6–3). Some additional examples of irreversible inhibitors of medical importance are described in Section 6.4.

An irreversible inhibitor need not bind covalently to the enzyme. Noncovalent binding is enough, if that binding is so tight that the inhibitor dissociates only rarely. How does one develop a tight-binding inhibitor? Recall that enzymes evolve to bind most tightly to the transition states of the reactions that they catalyze. In principle, if one can design a molecule that looks like that reaction transition state, it should bind tightly to the enzyme. Even though transition states cannot be observed directly, chemists can often predict the approximate structure of a transition state based on accumulated knowledge about reaction mechanisms. The transition state is by definition transient and so unstable that direct measurement of the binding interaction between this species and the enzyme is impossible. In some cases, however, stable molecules can be designed that resemble transition states. These are called **transition-state analogs**. They bind to an enzyme more tightly than does the substrate in the ES complex, because they fit into the active site better (that is, form a greater number of weak interactions) than the substrate itself. The idea of transition-state analogs was suggested by Pauling in the 1940s, and it has been explored using a number of enzymes. For example, transition-state analogs designed to inhibit the glycolytic enzyme aldolase bind to that enzyme more than four orders of magnitude more

BOX 6–3 MEDICINE Curing African Sleeping Sickness with a Biochemical Trojan Horse

African sleeping sickness, or African trypanosomiasis, is caused by protists (single-celled eukaryotes) called trypanosomes (Fig. 1). This disease (and related trypanosome-caused diseases) is medically and economically significant in many developing nations. Until the late twentieth century, the disease was virtually incurable. Vaccines are ineffective because the parasite has a novel mechanism to evade the host immune system.

The cell coat of trypanosomes is covered with a single protein, which is the antigen to which the immune system responds. Every so often, however, by a process of genetic recombination (see Table 28–1), a few cells in the population of infecting trypanosomes switch to a new protein coat, not recognized by the immune system. This process of "changing coats" can occur hundreds of times. The result is a chronic cyclic infection: the human host develops a fever, which subsides as the immune system beats back the first infection; trypanosomes with changed coats then become the seed for a second infection, and the fever recurs. This cycle can repeat for weeks, and the weakened person eventually dies.

Some modern approaches to treating African sleeping sickness have been based on an understanding of enzymology and metabolism. In at least one such approach, this involves pharmaceutical agents designed as mechanism-based enzyme inactivators (suicide inactivators). A vulnerable point in trypanosome metabolism is the pathway of polyamine biosynthesis. The polyamines spermine and spermidine, involved in DNA packaging, are required in large amounts in rapidly dividing cells. The first step in their synthesis is catalyzed by ornithine decarboxylase, an enzyme that requires a coenzyme called pyridoxal

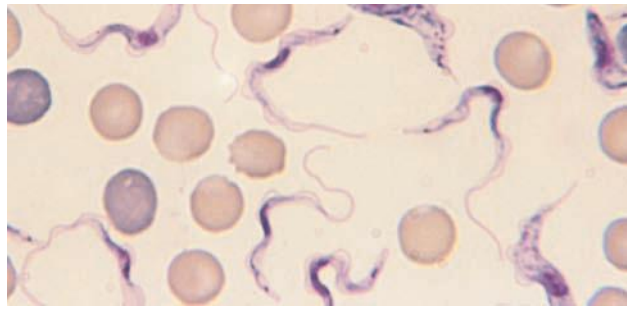


FIGURE 1 *Trypanosoma brucei rhodesiense*, one of several trypanosomes known to cause African sleeping sickness.

phosphate for its function. Pyridoxal phosphate (PLP), derived from vitamin B_6, makes a covalent bond to the amino acid substrates of the reactions it is involved in and acts as an electron sink to facilitate a variety of reactions (see Fig. 22–32). In mammalian cells, ornithine decarboxylase undergoes rapid turnover—that is, a constant round of enzyme degradation and synthesis. In some trypanosomes, however, the enzyme (for reasons not well understood) is stable, not readily replaced by newly synthesized enzyme. An inhibitor of ornithine decarboxylase that binds permanently to the enzyme would thus have little effect on human cells, which could rapidly replace inactivated enzyme, but would adversely affect the parasite.

The first few steps of the normal reaction catalyzed by ornithine decarboxylase are shown in Figure 2. Once CO_2 is released, the electron movement is reversed and putrescine is produced (see Fig. 22–32). Based on this mechanism, several suicide inactivators have been designed, one of which is difluoromethylornithine (DFMO). DFMO is relatively inert in solution. When it

(*continued on next page*)

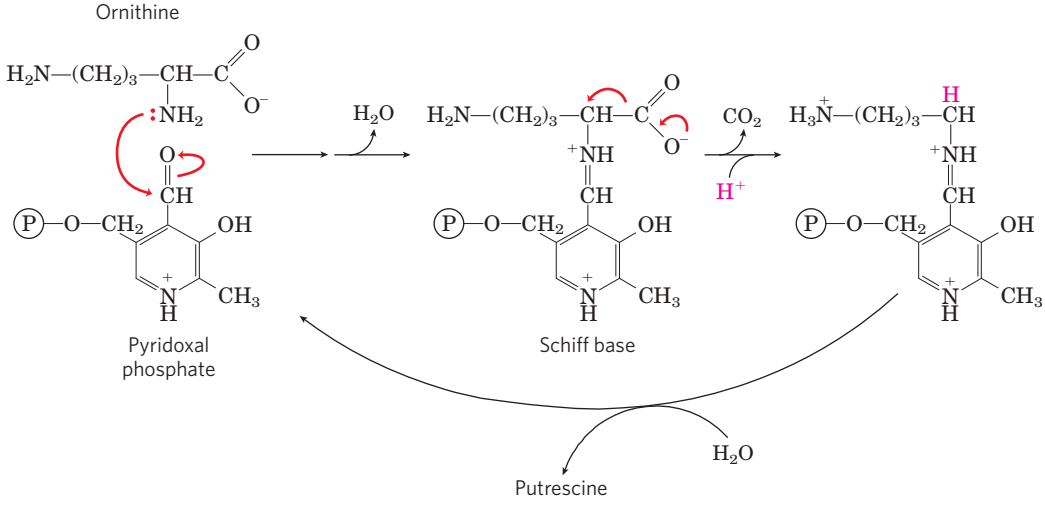


FIGURE 2 Mechanism of ornithine decarboxylase reaction.

BOX 6–3 MEDICINE Curing African Sleeping Sickness with a Biochemical Trojan Horse (*Continued*)

binds to ornithine decarboxylase, however, the enzyme is quickly inactivated (Fig. 3). The inhibitor acts by providing an alternative electron sink in the form of two strategically placed fluorine atoms, which are excellent leaving groups. Instead of electrons moving into the ring structure of PLP, the reaction results in displacement of a fluorine atom. The —S of a Cys residue at the enzyme's active site then forms a covalent complex with the highly reactive PLP-inhibitor adduct in an essentially irreversible reaction. In this way, the inhibitor makes use of the enzyme's own reaction mechanisms to kill it.

DFMO has proved highly effective against African sleeping sickness in clinical trials and is now used to treat African sleeping sickness caused by *Trypanosoma brucei gambiense*. Approaches such as this show great promise for treating a wide range of diseases. The design of drugs based on enzyme mechanism and structure can complement the more traditional trial-and-error methods of developing pharmaceuticals.

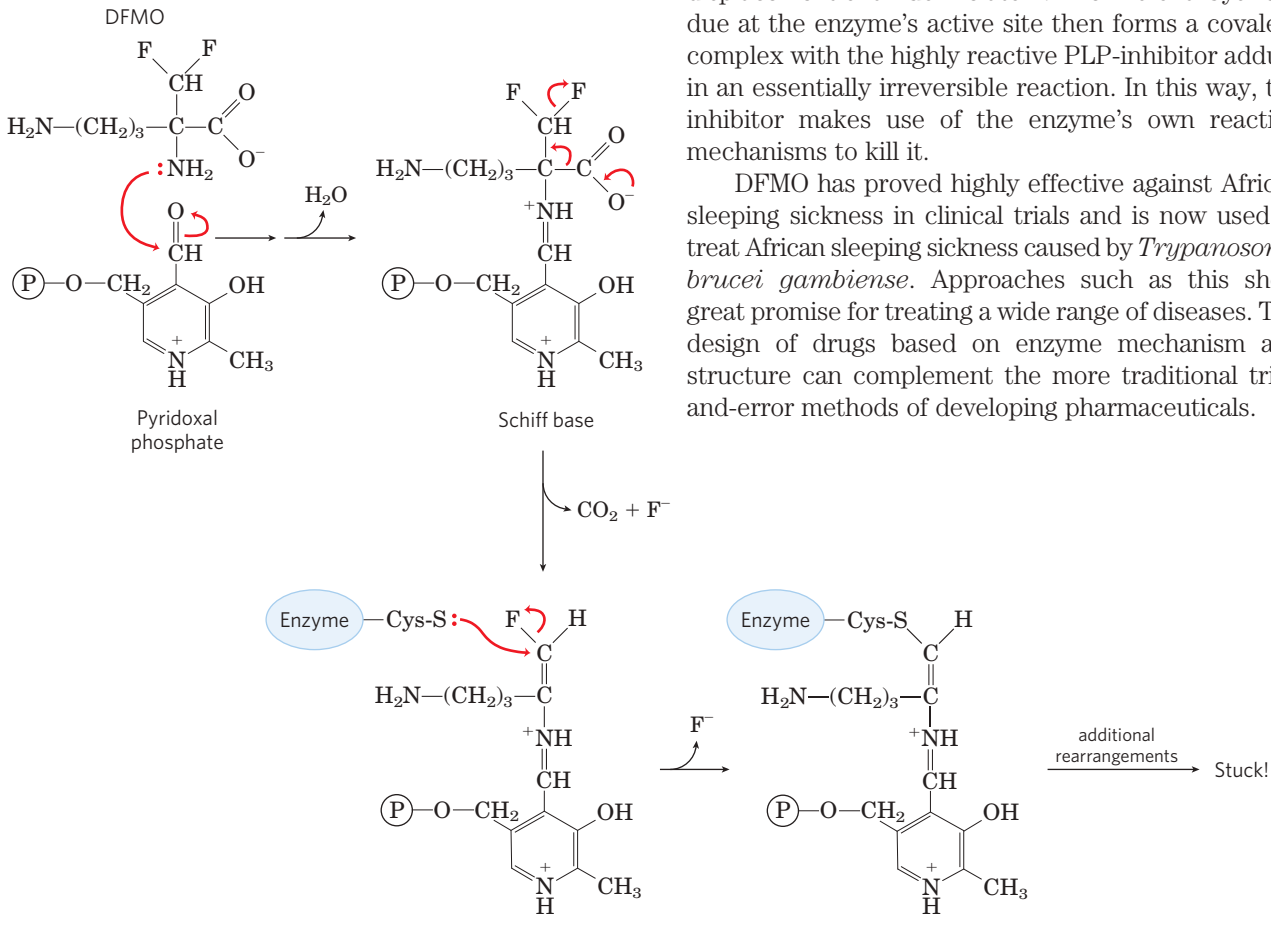


FIGURE 3 Inhibition of ornithine decarboxylase by DFMO.

tightly than substrates **(Fig. 6–17)**. These experiments have the limitation that a transition-state analog cannot perfectly mimic a transition state. Some analogs, however, bind to a target enzyme 10^2 to 10^8 times more tightly than does the normal substrate, providing good evidence that enzyme active sites are indeed complementary to transition states. The concept of transition-state analogs is important to the design of new pharmaceutical agents. As we shall see, the powerful anti-HIV drugs called protease inhibitors were designed in part as tight-binding transition-state analogs.

Enzyme Activity Depends on pH

Enzymes have an optimum pH (or pH range) at which their activity is maximal **(Fig. 6–18)**; at higher or lower pH, activity decreases. This is not surprising. Amino acid side chains in the active site may act as weak acids and bases with critical functions that depend on their maintaining a certain state of ionization, and elsewhere in the protein ionized side chains may play an essential role in the interactions that maintain protein structure. Removing a proton from a His residue, for example, might eliminate an ionic interaction essential for stabilizing the active conformation of the enzyme. A less common cause of pH sensitivity is titration of a group on the substrate.

The pH range over which an enzyme undergoes changes in activity can provide a clue to the type of amino acid residue involved (see Table 3–1). A change in activity near pH 7.0, for example, often reflects titration of a His residue. The effects of pH must be interpreted with some caution, however. In the closely packed environment of a protein, the pK_a of amino acid side chains can be significantly altered. For example, a nearby positive charge can lower the pK_a of a Lys residue, and a nearby negative charge can increase it. Such effects

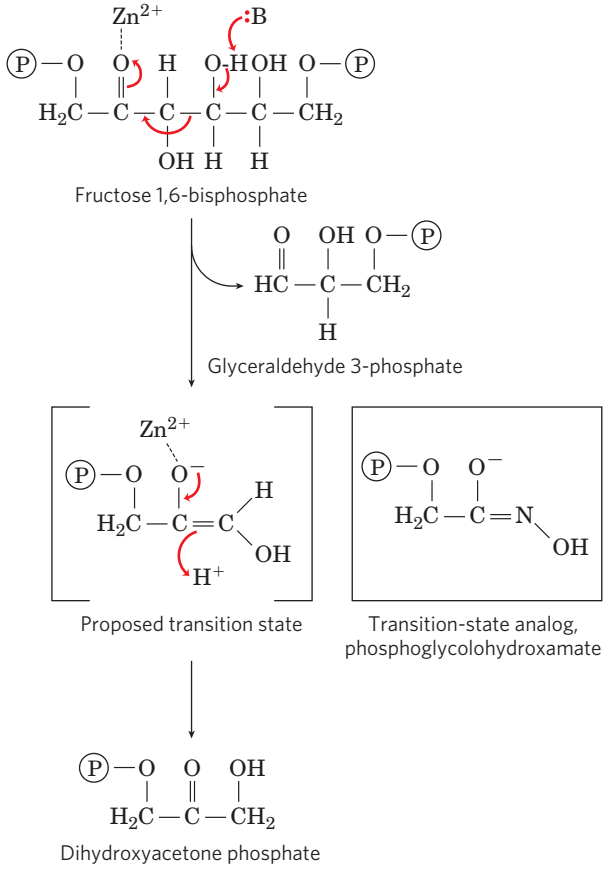


FIGURE 6–17 A transition-state analog. In glycolysis, a class II aldolase (found in bacteria and fungi) catalyzes the cleavage of fructose 1,6-bisphosphate to form glyceraldehyde 3-phosphate and dihydroxyacetone phosphate (see Fig. 14–6 for an example of a class I aldolase, found in animals and higher plants). The reaction proceeds via a reverse aldol condensation–like mechanism. The compound phosphoglycolohydroxamate resembles the proposed enediolate transition state, and binds to the enzyme nearly 10,000 times better than does the dihydroxyacetone phosphate product.

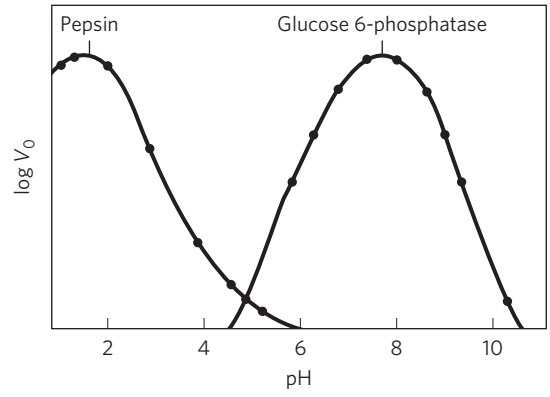


FIGURE 6–18 The pH-activity profiles of two enzymes. These curves are constructed from measurements of initial velocities when the reaction is carried out in buffers of different pH. Because pH is a logarithmic scale reflecting 10-fold changes in $[H^+]$, the changes in V_0 are also plotted on a logarithmic scale. The pH optimum for the activity of an enzyme is generally close to the pH of the environment in which the enzyme is normally found. Pepsin, a peptidase found in the stomach, has a pH optimum of about 1.6. The pH of gastric juice is between 1 and 2. Glucose 6-phosphatase of hepatocytes (liver cells), with a pH optimum of about 7.8, is responsible for releasing glucose into the blood. The normal pH of the cytosol of hepatocytes is about 7.2.

sometimes result in a pK_a that is shifted by several pH units from its value in the free amino acid. In the enzyme acetoacetate decarboxylase, for example, one Lys residue has a pK_a of 6.6 (compared with 10.5 in free lysine) due to electrostatic effects of nearby positive charges.

SUMMARY 6.3 Enzyme Kinetics as an Approach to Understanding Mechanism

- Most enzymes have certain kinetic properties in common. When substrate is added to an enzyme, the reaction rapidly achieves a steady state in which the rate at which the ES complex forms balances the rate at which it breaks down. As [S] increases, the steady-state activity of a fixed concentration of enzyme increases in a hyperbolic fashion to approach a characteristic maximum rate, V_{max}, at which essentially all the enzyme has formed a complex with substrate.
- The substrate concentration that results in a reaction rate equal to one-half V_{max} is the Michaelis constant K_m, which is characteristic for each enzyme acting on a given substrate. The Michaelis-Menten equation

$$V_0 = \frac{V_{max}[S]}{K_m + [S]}$$

relates initial velocity to [S] and V_{max} through the constant K_m. Michaelis-Menten kinetics is also called steady-state kinetics.
- K_m and V_{max} have different meanings for different enzymes. The limiting rate of an enzyme-catalyzed reaction at saturation is described by the constant k_{cat}, the turnover number. The ratio k_{cat}/K_m provides a good measure of catalytic efficiency. The Michaelis-Menten equation is also applicable to bisubstrate reactions, which occur by ternary complex or Ping-Pong (double-displacement) pathways.
- Reversible inhibition of an enzyme may be competitive, uncompetitive, or mixed. Competitive inhibitors compete with substrate by binding reversibly to the active site, but they are not transformed by the enzyme. Uncompetitive inhibitors bind only to the ES complex, at a site distinct from the active site. Mixed inhibitors bind to either E or ES, again at a site distinct from the active site. In irreversible inhibition an inhibitor binds permanently to an active site by forming a covalent bond or a very stable noncovalent interaction.
- Every enzyme has an optimum pH (or pH range) at which it has maximal activity.

6.4 Examples of Enzymatic Reactions

Thus far we have focused on the general principles of catalysis and on introducing some of the kinetic parameters used to describe enzyme action. We now turn to several examples of specific enzyme reaction mechanisms.

An understanding of the complete mechanism of action of a purified enzyme requires identification of all substrates, cofactors, products, and regulators. Moreover, it requires a knowledge of (1) the temporal sequence in which enzyme-bound reaction intermediates form, (2) the structure of each intermediate and each transition state, (3) the rates of interconversion between intermediates, (4) the structural relationship of the enzyme to each intermediate, and (5) the energy contributed by all reacting and interacting groups to intermediate complexes and transition states. As yet, there is probably no enzyme for which we have an understanding that meets all these requirements.

We present here the mechanisms for four enzymes: chymotrypsin, hexokinase, enolase, and lysozyme. These examples are not intended to cover all possible classes of enzyme chemistry. They are chosen in part because they are among the best-understood enzymes, and in part because they clearly illustrate some general principles outlined in this chapter. The discussion concentrates on selected principles, along with some key experiments that have helped to bring these principles into focus. We use the chymotrypsin example to review some of the conventions used to depict enzyme mechanisms. Much mechanistic detail and experimental evidence is necessarily omitted; no one book could completely document the rich experimental history of these enzymes. In addition, the special contribution of coenzymes to the catalytic activity of many enzymes is considered only briefly. The function of coenzymes is chemically varied, and we describe each coenzyme in detail as it is encountered in Part II.

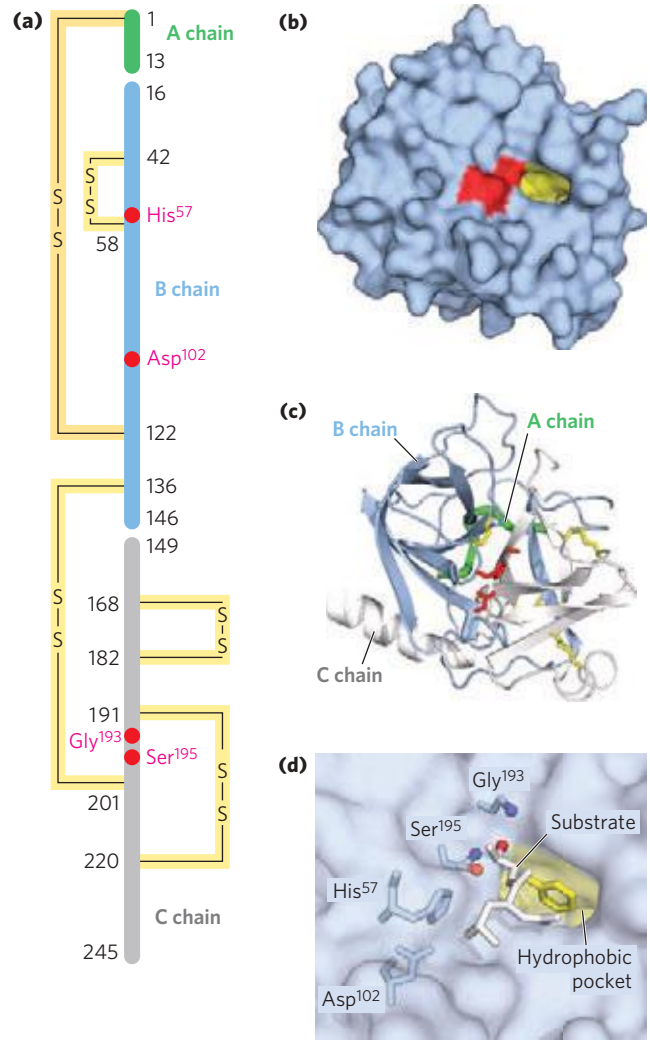


FIGURE 6–19 Structure of chymotrypsin. (PDB ID 7GCH) **(a)** A representation of primary structure, showing disulfide bonds and the amino acid residues crucial to catalysis. The protein consists of three polypeptide chains linked by disulfide bonds. (The numbering of residues in chymotrypsin, with "missing" residues 14, 15, 147, and 148, is explained in Fig. 6–38.) The active-site amino acid residues are grouped together in the three-dimensional structure. **(b)** A depiction of the enzyme emphasizing its surface. The hydrophobic pocket in which the aromatic amino acid side chain of the substrate is bound is shown in yellow. Key active-site residues, including Ser^{195}, His^{57}, and Asp^{102}, are red. The roles of these residues in catalysis are illustrated in Figure 6–22. **(c)** The polypeptide backbone as a ribbon structure. Disulfide bonds are yellow; the three chains are colored as in part (a). **(d)** A close-up of the active site with a substrate (white and yellow) bound. The hydroxyl of Ser^{195} attacks the carbonyl group of the substrate (the oxygens are red); the developing negative charge on the oxygen is stabilized by the oxyanion hole (amide nitrogens from Ser^{195} and Gly^{193}, in blue), as explained in Figure 6–22. The aromatic amino acid side chain of the substrate (yellow) sits in the hydrophobic pocket. The amide nitrogen of the peptide bond to be cleaved (protruding toward the viewer and projecting the path of the rest of the substrate polypeptide chain) is shown in white.

The Chymotrypsin Mechanism Involves Acylation and Deacylation of a Ser Residue

Bovine pancreatic chymotrypsin (M_r 25,191) is a protease, an enzyme that catalyzes the hydrolytic cleavage of peptide bonds. This protease is specific for peptide bonds adjacent to aromatic amino acid residues (Trp, Phe, Tyr). The three-dimensional structure of chymotrypsin is shown in **Figure 6–19**, with functional groups in the active site emphasized. The reaction catalyzed by this enzyme illustrates the principle of transition-state stabilization and also provides a classic example of general acid-base catalysis and covalent catalysis.

Chymotrypsin enhances the rate of peptide bond hydrolysis by a factor of at least 10^9. It does not catalyze a direct attack of water on the peptide bond; instead, a transient covalent acyl-enzyme intermediate is formed. The reaction thus has two distinct phases. In the acylation phase, the peptide bond is cleaved and an ester linkage is formed between the peptide carbonyl carbon and the enzyme. In the deacylation phase, the ester linkage is hydrolyzed and the nonacylated enzyme regenerated.

The first evidence for a covalent acyl-enzyme intermediate came from a classic application of pre–steady state kinetics. In addition to its action on polypeptides, chymotrypsin also catalyzes the hydrolysis of small esters and amides. These reactions are much slower than hydrolysis of peptides because less binding energy is available with smaller substrates, and they are therefore easier to study. Investigations by B. S. Hartley and B. A. Kilby in 1954 found that chymotrypsin hydrolysis of the ester *p*-nitrophenylacetate, as measured by release of *p*-nitrophenol, proceeds with a rapid burst before leveling off to a slower rate **(Fig. 6–20)**. By extrapolating back to zero time, they concluded that the burst phase corresponded to just under one molecule of *p*-nitrophenol released for every enzyme molecule present. Hartley and Kilby suggested that this reflected a rapid acylation of all the enzyme molecules (with release of *p*-nitrophenol), with the rate for subsequent turnover of the enzyme limited by a slow deacylation step. Similar results have since been obtained with many other enzymes. The observation of a burst phase provides yet another example of the use of kinetics to break down a reaction into its constituent steps.

Additional features of the chymotrypsin mechanism have been discovered by analyzing the dependence of the reaction on pH. The rate of chymotrypsin-catalyzed cleavage generally exhibits a bell-shaped pH-rate profile **(Fig. 6–21)**. The rates plotted in Figure 6–21a are obtained at low (subsaturating) substrate concentrations and therefore represent k_{cat}/K_m (see Eqn 6–27, p. 205). A more complete analysis of the rates at different substrate concentrations at each pH allows researchers to determine the individual contributions of the k_{cat} and K_m terms. After obtaining the maximum rates at each pH, one can plot the k_{cat} alone versus pH (Fig. 6–21b); after obtaining the K_m at each pH, researchers can then plot $1/K_m$ (Fig. 6–21c). Kinetic and structural analyses have revealed that the change in k_{cat} reflects the ionization state of His^{57}. The decline in k_{cat} at low pH results from protonation of His^{57} (so that it cannot extract a proton from Ser^{195} in step ❷ of the reaction; see **Fig. 6–22**). This rate reduction illustrates the

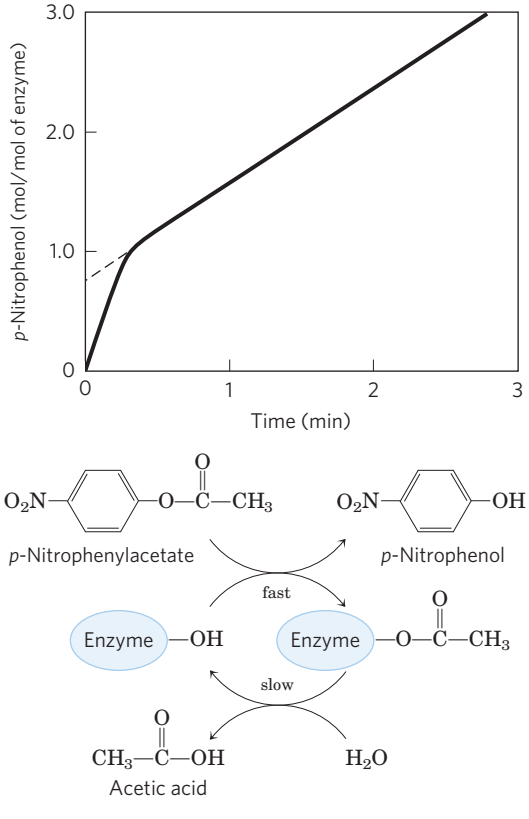


FIGURE 6–20 Pre-steady state kinetic evidence for an acyl-enzyme intermediate. The hydrolysis of *p*-nitrophenylacetate by chymotrypsin is measured by release of *p*-nitrophenol (a colored product). Initially, the reaction releases a rapid burst of *p*-nitrophenol nearly stoichiometric with the amount of enzyme present. This reflects the fast acylation phase of the reaction. The subsequent rate is slower, because enzyme turnover is limited by the rate of the slower deacylation phase.

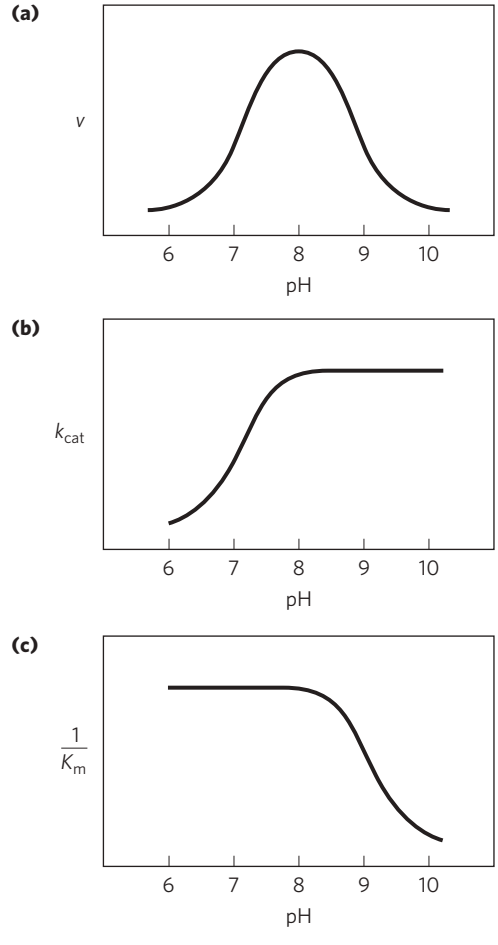


FIGURE 6–21 The pH dependence of chymotrypsin-catalyzed reactions. **(a)** The rates of chymotrypsin-mediated cleavage produce a bell-shaped pH-rate profile with an optimum at pH 8.0. The rate (v) being plotted is that at low substrate concentrations and thus reflects the term k_{cat}/K_m. The plot can be broken down to its components by using kinetic methods to determine the terms k_{cat} and K_m separately at each pH. When this is done **(b, c)**, it becomes clear that the transition just above pH 7 is due to changes in k_{cat}, whereas the transition above pH 8.5 is due to changes in $1/K_m$. Kinetic and structural studies have shown that the transitions illustrated in (b) and (c) reflect the ionization states of the His^{57} side chain (when substrate is not bound) and the α-amino group of Ile^{16} (at the amino terminus of the B chain), respectively. For optimal activity, His^{57} must be unprotonated and Ile^{16} must be protonated.

How to Read Reaction Mechanisms—A Refresher

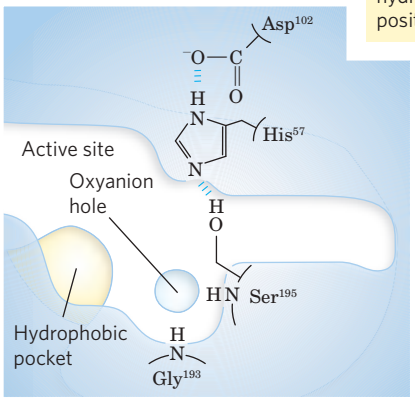
Chemical reaction mechanisms, which trace the formation and breakage of covalent bonds, are communicated with dots and curved arrows, a convention known informally as "electron pushing." A covalent bond consists of a shared pair of electrons. Nonbonded electrons important to the reaction mechanism are designated by dots ($^{-}\ddot{O}H$). Curved arrows (↷) represent the movement of electron pairs. For movement of a single electron (as in a free radical reaction), a single-headed (fishhook-type) arrow is used (↼). Most reaction steps involve an unshared electron pair (as in the chymotrypsin mechanism).

Some atoms are more electronegative than others; that is, they more strongly attract electrons. The relative electronegativities of atoms encountered in this text are F > O > N > C≈S > P≈H. For example, the two electron pairs making up a C=O (carbonyl) bond are not shared equally; the carbon is relatively electron-deficient as the oxygen draws away the electrons. Many reactions involve an electron-rich atom (a nucleophile) reacting with an electron-deficient atom (an electrophile). Some common nucleophiles and electrophiles in biochemistry are shown at right.

In general, a reaction mechanism is initiated at an unshared electron pair of a nucleophile. In mechanism diagrams, the base of the electron-pushing arrow originates near the electron-pair dots, and the head of the arrow points directly at the electrophilic center being attacked. Where the unshared electron pair confers a formal negative charge on the nucleophile, the negative charge symbol itself can represent the unshared electron pair and serves as the base of the arrow. In the chymotrypsin mechanism, the nucleophilic electron pair in the ES complex between steps ❶ and ❷ is provided by the oxygen of the Ser^{195} hydroxyl group. This electron pair (2 of the 8 valence electrons of the hydroxyl oxygen) provides the base of the curved arrow. The electrophilic center under attack is the carbonyl carbon of the peptide bond to be cleaved. The C, O, and N atoms have a maximum of 8 valence electrons, and H has a maximum of 2. These atoms are occasionally found in unstable states with less than their maximum allotment of electrons, but C, O, and N cannot have more than 8. Thus, when the electron pair from chymotrypsin's Ser^{195} attacks the substrate's carbonyl carbon, an electron pair is displaced from the carbon valence shell (you cannot have 5 bonds to carbon!). These electrons move toward the more electronegative carbonyl oxygen. The oxygen has 8 valence electrons both before and after this chemical process, but the number shared with the carbon is reduced from 4 to 2, and the carbonyl oxygen acquires a negative charge. In the next step, the electron pair conferring the negative charge on the oxygen moves back to re-form a bond with carbon and reestablish the carbonyl linkage. Again, an electron pair must be displaced from the carbon, and this time it is the electron pair shared with the amino group of the peptide linkage. This breaks the peptide bond. The remaining steps follow a similar pattern.

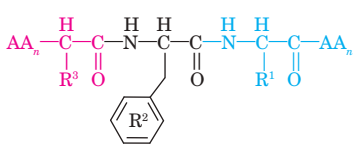
When substrate binds, the side chain of the residue adjacent to the peptide bond to be cleaved nestles in a hydrophobic pocket on the enzyme, positioning the peptide bond for attack.

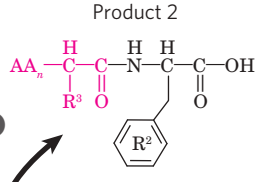
Chymotrypsin (free enzyme)



❶

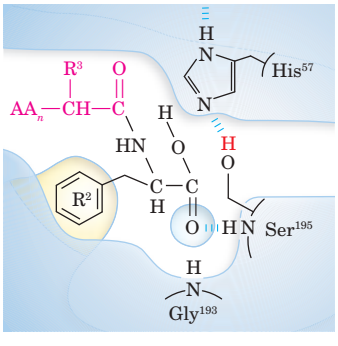
Substrate (a polypeptide)





❼

Enzyme-product 2 complex

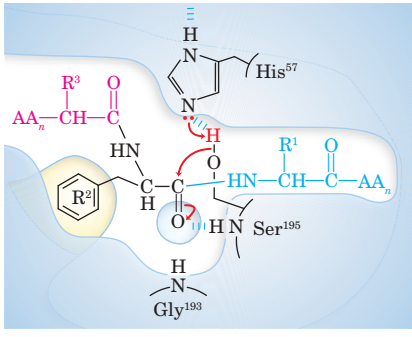


Dissociation of the second product from the active site regenerates free enzyme.

Nucleophiles	Electrophiles
Negatively charged oxygen (as in an unprotonated hydroxyl group or an ionized carboxylic acid)	Carbon atom of a carbonyl group (the more electronegative oxygen of the carbonyl group pulls electrons away from the carbon)
Negatively charged sulfhydryl	Protonated imine group (activated for nucleophilic attack at the carbon by protonation of the imine)
Carbanion	Phosphorus of a phosphate group
Uncharged amine group	Proton
Imidazole	
Hydroxide ion	

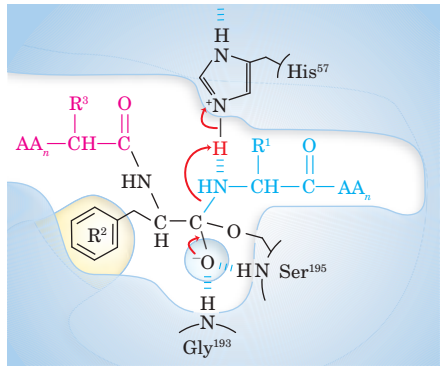
Enzyme-substrate complex

Interaction of Ser^{195} and His^{57} generates a strongly nucleophilic alkoxide ion on Ser^{195}; the ion attacks the peptide carbonyl group, forming a tetrahedral acyl-enzyme. This is accompanied by formation of a short-lived negative charge on the carbonyl oxygen of the substrate, which is stabilized by hydrogen bonding in the oxyanion hole.



Short-lived intermediate* (acylation)

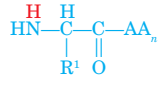
Instability of the negative charge on the substrate carbonyl oxygen leads to collapse of the tetrahedral intermediate; re-formation of a double bond with carbon displaces the bond between carbon and the amino group of the peptide linkage, breaking the peptide bond. The amino leaving group is protonated by His^{57}, facilitating its displacement.



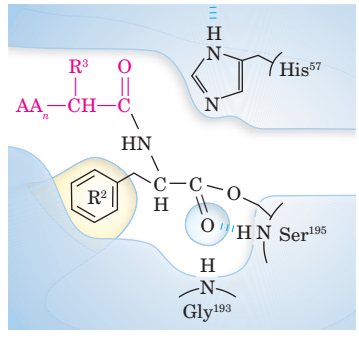
MECHANISM FIGURE 6–22 **Hydrolytic cleavage of a peptide bond by chymotrypsin.** The reaction has two phases. In the acylation phase (steps ❶ to ❹), formation of a covalent acyl-enzyme intermediate is coupled to cleavage of the peptide bond. In the deacylation phase (steps ❺ to ❼), deacylation regenerates the free enzyme; this is essentially the reverse of the acylation phase, with water mirroring, in reverse, the role of the amine component of the substrate.

Chymotrypsin Mechanism

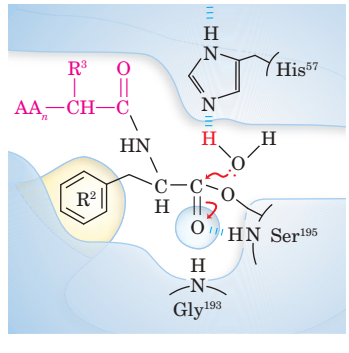
Product 1



Acyl-enzyme intermediate

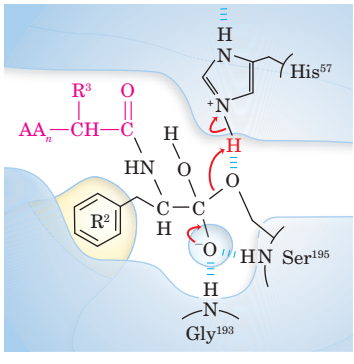


Acyl-enzyme intermediate



An incoming water molecule is deprotonated by general base catalysis, generating a strongly nucleophilic hydroxide ion. Attack of hydroxide on the ester linkage of the acyl-enzyme generates a second tetrahedral intermediate, with oxygen in the oxyanion hole again taking on a negative charge.

Short-lived intermediate* (deacylation)



Collapse of the tetrahedral intermediate forms the second product, a carboxylate anion, and displaces Ser^{195}.

*The tetrahedral intermediate in the chymotrypsin reaction pathway, and the second tetrahedral intermediate that forms later, are sometimes referred to as transition states, which can lead to confusion. An *intermediate* is any chemical species with a finite lifetime, "finite" being defined as longer than the time required for a molecular vibration (~10^{-13} second). A *transition state* is simply the maximum-energy species formed on the reaction coordinate and does not have a finite lifetime. The tetrahedral intermediates formed in the chymotrypsin reaction closely resemble, both energetically and structurally, the transition states leading to their formation and breakdown. However, the intermediate represents a committed stage of completed bond formation, whereas the transition state is part of the process of reaction. In the case of chymotrypsin, given the close relationship between the intermediate and the actual transition state the distinction between them is routinely glossed over. Furthermore, the interaction of the negatively charged oxygen with the amide nitrogens in the oxyanion hole, often referred to as transition-state stabilization, also serves to stabilize the intermediate in this case. Not all intermediates are so short-lived that they resemble transition states. The chymotrypsin acyl-enzyme intermediate is much more stable and more readily detected and studied, and it is never confused with a transition state.

importance of general acid and general base catalysis in the mechanism for chymotrypsin. The changes in the $1/K_m$ term reflect the ionization of the α-amino group of Ile^{16} (at the amino-terminal end of one of chymotrypsin's three polypeptide chains). This group forms a salt bridge to Asp^{194}, stabilizing the active conformation of the enzyme. When this group loses its proton at high pH, the salt bridge is eliminated and a conformational change closes the hydrophobic pocket where the aromatic amino acid side chain of the substrate inserts (Fig. 6–19). Substrates can no longer bind properly, which is measured kinetically as an increase in K_m.

As can be seen in Figure 6-22, the nucleophile in the acylation phase is the oxygen of Ser^{195}. (Proteases with a Ser residue that plays this role in reaction mechanisms are called **serine proteases**.) The pK_a of a Ser hydroxyl group is generally too high for the unprotonated form to be present in significant concentrations at physiological pH. However, in chymotrypsin, Ser^{195} is linked to His^{57} and Asp^{102} in a hydrogen-bonding network referred to as the **catalytic triad**. When a peptide substrate binds to chymotrypsin, a subtle change in conformation compresses the hydrogen bond between His^{57} and Asp^{102}, resulting in a stronger interaction, called a low-barrier hydrogen bond. This enhanced interaction increases the pK_a of His^{57} from ~7 (for free histidine) to >12, allowing the His residue to act as an enhanced general base that can remove the proton from the Ser^{195} hydroxyl group. Deprotonation prevents development of a very unstable positive charge on the Ser^{195} hydroxyl and makes the Ser side chain a stronger nucleophile. At later reaction stages, His^{57} also acts as a proton donor, protonating the amino group in the displaced portion of the substrate (the leaving group).

As the Ser^{195} oxygen attacks the carbonyl group of the substrate, a very short-lived tetrahedral intermediate* is formed in which the carbonyl oxygen acquires a negative charge (Fig. 6–22, step ❷). This charge, forming within a pocket on the enzyme called the oxyanion hole, is stabilized by hydrogen bonds contributed by the amide groups of two peptide bonds in the chymotrypsin backbone. One of these hydrogen bonds (contributed by Gly^{193}) is present only in this intermediate and in the transition states for its formation and breakdown; it reduces the energy required to reach these states. This is an example of the use of binding energy in catalysis.

An Understanding of Protease Mechanisms Leads to New Treatments for HIV Infections

New pharmaceutical agents are almost always designed to inhibit an enzyme. The very successful therapies developed to treat HIV infections provide a case in point. The human immunodeficiency virus (HIV) is the causative agent of acquired immune deficiency syndrome (AIDS). In 2005, an estimated 37 to 45 million people worldwide were living with HIV infections, with 3.9 to 6.6 million new infections that year and more than 2.4 million fatalities. AIDS first surfaced as a world epidemic in the 1980s; HIV was discovered soon after and identified as a **retrovirus**. Retroviruses possess an RNA genome and an enzyme, reverse transcriptase, capable of using RNA to direct the synthesis of a complementary DNA. Efforts to understand HIV and develop therapies for HIV infection benefited from decades of basic research, both on enzyme mechanisms and on the properties of other retroviruses. A retrovirus such as HIV has a relatively simple life cycle (see Fig. 26–32). Its RNA genome is converted to duplex DNA in several steps catalyzed by a reverse transcriptase (described in Chapter 26). The duplex DNA is then inserted into a chromosome in the nucleus of the host cell by the enzyme integrase (described in Chapter 25). The integrated copy of the viral genome can remain dormant indefinitely. Alternatively, it can be transcribed back into RNA, which can then be translated into proteins to construct new virus particles. Most of the viral genes are translated into large polyproteins, which are cut by the HIV protease into the individual proteins needed to make the virus (see Fig. 26–33). There are only three key enzymes in this cycle—the reverse transcriptase, the integrase, and the protease. These enzymes thus represent the most promising drug targets.

There are four major subclasses of proteases. Serine proteases, such as chymotrypsin and trypsin, and cysteine proteases (in which Cys serves a catalytic role similar to that of Ser in the active site) feature covalent enzyme-substrate complexes; aspartyl proteases and metalloproteases do not. The HIV protease is an aspartyl protease. Two active-site Asp residues facilitate a direct attack of water on the peptide bond to be cleaved **(Fig. 6–23)**. The initial product of the attack of water on the carbonyl group of the peptide bond is an unstable tetrahedral intermediate, much as we have seen for the chymotrypsin reaction. This intermediate is close in structure and energy to the reaction transition state. The drugs that have been developed as HIV protease inhibitors form noncovalent complexes with the enzyme, but they bind to it so tightly that they can be considered irreversible inhibitors. The tight binding is derived in part from their design as transition-state analogs. The success of these drugs makes a point worth emphasizing. The catalytic principles we have studied in this chapter are not simply abstruse ideas to be memorized—their application saves lives.

The HIV protease cleaves peptide bonds between Phe and Pro residues most efficiently. The active site thus has a pocket to bind aromatic groups next to the bond to be cleaved. The structures of several HIV protease inhibitors are shown in **Figure 6–24**. Although the structures appear varied, they all share a core structure—a main chain with a hydroxyl group positioned next to a branch containing a benzyl group. This arrangement targets the benzyl group to an aromatic (hydrophobic) binding pocket. The adjacent hydroxyl group mimics the negatively charged oxygen in the

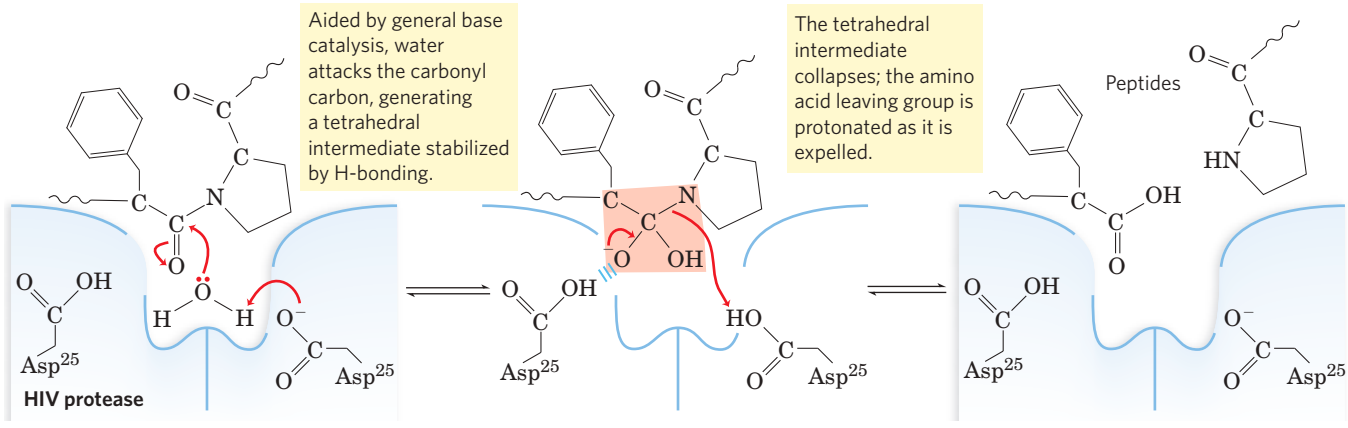
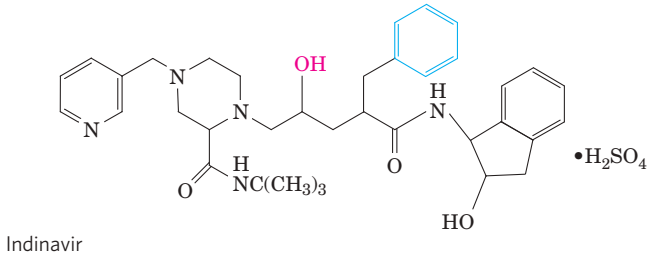
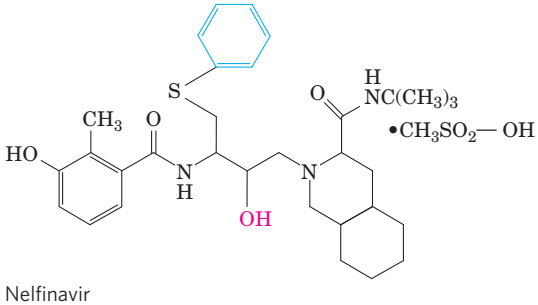


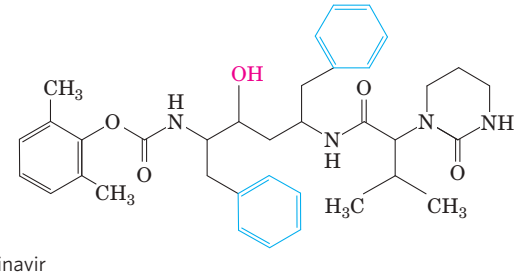
FIGURE 6–23 Mechanism of action of HIV protease. Two active-site Asp residues (from different subunits) act as general acid-base catalysts, facilitating the attack of water on the peptide bond. The unstable tetrahedral intermediate in the reaction pathway is shaded light red.



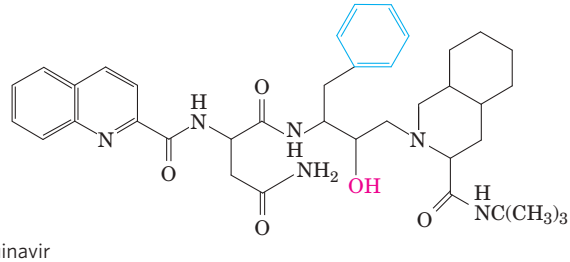
Indinavir



Nelfinavir



Lopinavir



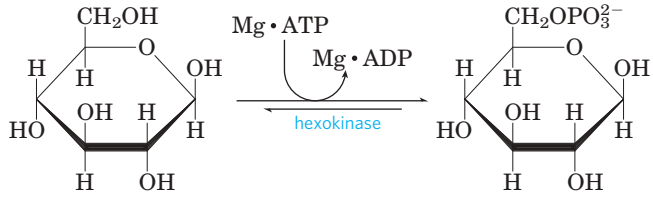
Saquinavir

FIGURE 6–24 HIV protease inhibitors. The hydroxyl group (red) acts as a transition-state analog, mimicking the oxygen of the tetrahedral intermediate. The adjacent benzyl group (blue) helps to properly position the drug in the active site.

tetrahedral intermediate in the normal reaction, providing a transition-state analog. The remainder of each inhibitor structure was designed to fit into and bind to various crevices along the surface of the enzyme, enhancing overall binding. Availability of these effective drugs has vastly increased the lifespan and quality of life of millions of people with HIV and AIDS. ■

Hexokinase Undergoes Induced Fit on Substrate Binding

Yeast hexokinase (M_r 107,862) is a bisubstrate enzyme that catalyzes the reversible reaction



β-D-Glucose Glucose 6-phosphate

ATP and ADP always bind to enzymes as a complex with the metal ion Mg^{2+}.

The hydroxyl at C-6 of glucose (to which the γ-phosphoryl of ATP is transferred in the hexokinase reaction) is similar in chemical reactivity to water, and water freely enters the enzyme active site. Yet hexokinase favors the reaction with glucose by a factor of 10^6. The enzyme can discriminate between glucose and water because of a conformational change in the enzyme when the correct substrates binds **(Fig. 6–25)**. Hexokinase thus provides a good example of induced fit. When glucose is not present, the enzyme is in an inactive conformation, with the active-site amino acid side chains out of position for reaction. When glucose (but not water) and Mg • ATP bind, the binding energy derived from this interaction induces a conformational change in hexokinase to the catalytically active form.

This model has been reinforced by kinetic studies. The five-carbon sugar xylose, stereochemically similar to glucose but one carbon shorter, binds to hexokinase

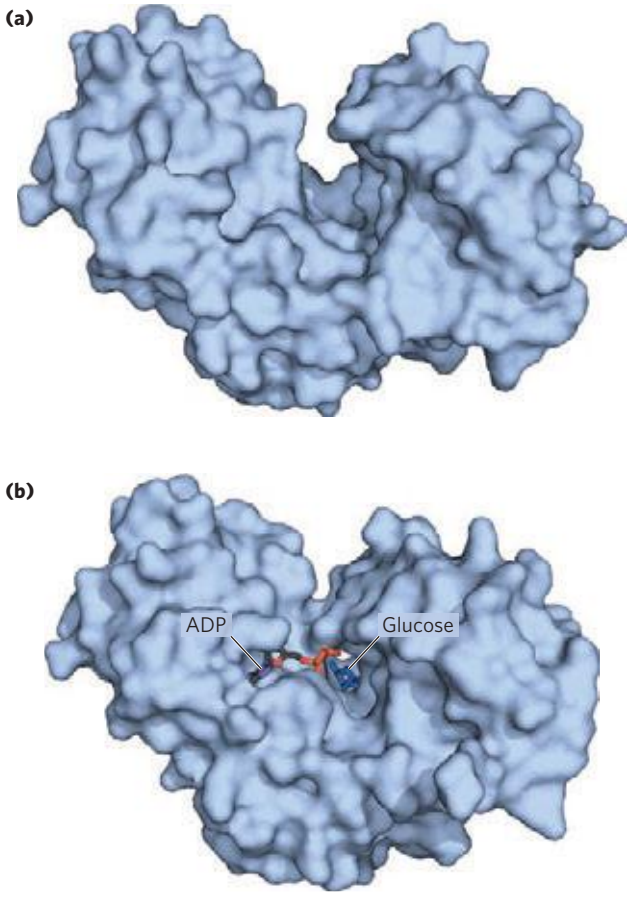
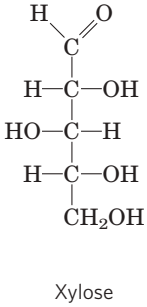
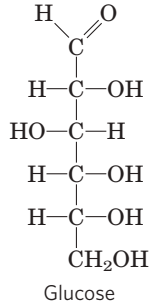


FIGURE 6–25 Induced fit in hexokinase. (a) Hexokinase has a U-shaped structure (PDB ID 2YHX). **(b)** The ends pinch toward each other in a conformational change induced by binding of D-glucose (derived from PDB ID 1HKG and PDB ID 1GLK).

but in a position where it cannot be phosphorylated. Nevertheless, addition of xylose to the reaction mixture increases the rate of ATP hydrolysis. Evidently, the binding of xylose is sufficient to induce a change in hexokinase to its active conformation, and the enzyme is thereby "tricked" into phosphorylating water. The hexokinase reaction also illustrates that enzyme specificity is not always a simple matter of binding one compound but not another. In the case of hexokinase, specificity is observed not in the formation of the ES complex but in the relative rates of subsequent catalytic steps. Water is not excluded from the active site, but reaction rates increase greatly in the presence of the functional phosphoryl group acceptor (glucose).

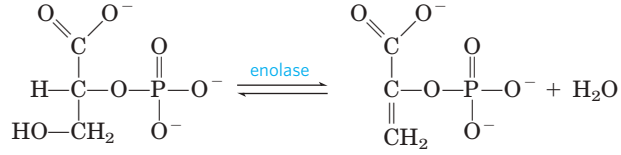

Xylose


Glucose

Induced fit is only one aspect of the catalytic mechanism of hexokinase—like chymotrypsin, hexokinase uses several catalytic strategies. For example, the active-site amino acid residues (those brought into position by the conformational change that follows substrate binding) participate in general acid-base catalysis and transition-state stabilization.

The Enolase Reaction Mechanism Requires Metal Ions

Another glycolytic enzyme, enolase, catalyzes the reversible dehydration of 2-phosphoglycerate to phosphoenolpyruvate:

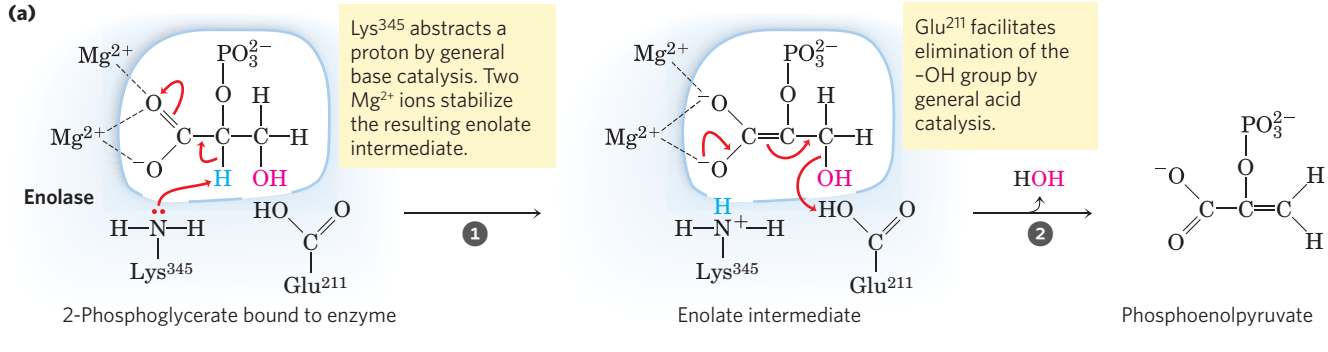


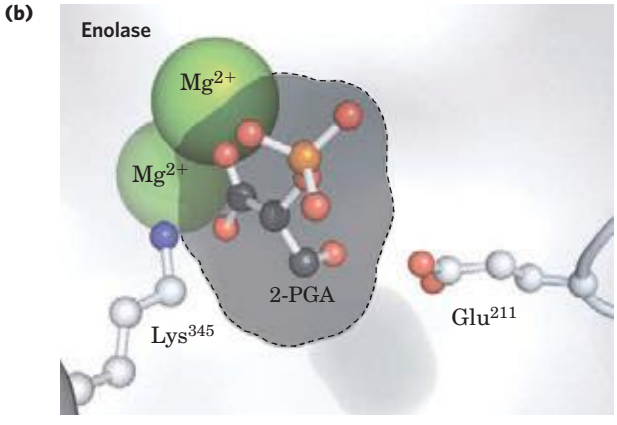
2-Phosphoglycerate Phosphoenolpyruvate

The reaction provides an example of the use of an enzymatic cofactor, in this case a metal ion (an example of coenzyme function is provided in Box 6–3). Yeast enolase (M_r 93,316) is a dimer with 436 amino acid residues per subunit. The enolase reaction illustrates one type of metal ion catalysis and provides an additional example of general acid-base catalysis and transition-state stabilization. The reaction occurs in two steps **(Fig. 6–26a)**. First, Lys^{345} acts as a general base catalyst, abstracting a proton from C-2 of 2-phosphoglycerate; then Glu^{211} acts as a general acid catalyst, donating a proton to the —OH leaving group. The proton at C-2 of 2-phosphoglycerate is not acidic and thus is quite resistant to its removal by Lys^{345}. However, the electronegative oxygen atoms of the adjacent carboxyl group pull electrons away from C-2, making the attached protons somewhat more labile. In the active site, the carboxyl group of 2-phosphoglycerate undergoes strong ionic interactions with two bound Mg^{2+} ions (Fig. 6–26b), strongly enhancing the electron withdrawal by the carboxyl. Together, these effects render the C-2 protons sufficiently acidic (lowering the pK_a) so that one can be abstracted to initiate the reaction. As the unstable enolate intermediate is formed, the metal ions further act to shield the two negative charges (on the carboxyl oxygen atoms) that transiently exist in close proximity to each other. Hydrogen bonding to other active-site amino acid residues also contributes to the overall mechanism. The various interactions effectively stabilize both the enolate intermediate and the transition state preceding its formation.

Lysozyme Uses Two Successive Nucleophilic Displacement Reactions

Lysozyme is a natural antibacterial agent found in tears and egg whites. The hen egg white lysozyme (M_r 14,296) is a monomer with 129 amino acid residues.





MECHANISM FIGURE 6–26 Two-step reaction catalyzed by enolase. (a) The mechanism by which enolase converts 2-phosphoglycerate (2-PGA) to phosphoenolpyruvate. The carboxyl group of 2-PGA is coordinated by two magnesium ions at the active site. **(b)** The substrate, 2-PGA, in relation to the Mg^{2+}, Lys^{345}, and Glu^{211} in the enolase active site (gray outline). Nitrogen is shown in blue, phosphorus in orange; hydrogen atoms are not shown (PDB ID 1ONE).

This was the first enzyme to have its three-dimensional structure determined, by David Phillips and colleagues in 1965. The structure revealed four stabilizing disulfide bonds and a cleft containing the active site **(Fig. 6–27a)**. More than five decades of investigations have provided a detailed picture of the structure and activity of the enzyme, and an interesting story of how biochemical science progresses.

The substrate of lysozyme is peptidoglycan, a carbohydrate found in many bacterial cell walls (see Fig. 20–30). Lysozyme cleaves the (β1→4) glycosidic C—O bond (p. 252) between the two types of sugar residues in the molecule, *N*-acetylmuramic acid (Mur2Ac) and *N*-acetylglucosamine (GlcNAc) (Fig. 6–27b), often referred to as NAM and NAG, respectively, in the research literature on enzymology. Six residues of the alternating Mur2Ac and GlcNAc in peptidoglycan bind in the active site, in binding sites labeled A through F. Model building has shown that the lactyl side chain of Mur2Ac cannot be accommodated in sites C and E, restricting Mur2Ac binding to sites B, D, and F. Only one of the bound glycosidic bonds is cleaved, that between a Mur2Ac residue in site D and a GlcNAc residue in site E. The key catalytic amino acid residues in the active site are Glu^{35} and Asp^{52} **(Fig. 6–28a)**. The reaction is a nucleophilic substitution, with —OH from water replacing the GlcNAc at C-1 of Mur2Ac.

With the active-site residues identified and a detailed structure of the enzyme available, the path to understanding the reaction mechanism seemed open in the 1960s. However, definitive evidence for a particular mechanism eluded investigators for nearly four decades.

There are two chemically reasonable mechanisms that could generate the observed product of lysozyme-mediated cleavage of the glycosidic bond. Phillips and colleagues proposed a dissociative (S_N1-type) mechanism (Fig. 6–28a, left), in which the GlcNAc initially dissociates in step ① to leave behind a glycosyl cation (a carbocation) intermediate. In this mechanism, the departing GlcNAc is protonated by general acid catalysis by Glu^{35}, located in a hydrophobic pocket that gives its carboxyl group an unusually high pK_a. The carbocation is stabilized by resonance involving the adjacent ring oxygen, as well as by electrostatic interaction with the negative charge on the nearby Asp^{52}. In step ② water attacks at C-1 of Mur2Ac to yield the product. The alternative mechanism (Fig. 6–28a, right) involves two consecutive direct displacement (S_N2-type) steps. In step ❶, Asp^{52} attacks C-1 of Mur2Ac to displace the GlcNAc. As in the first mechanism, Glu^{35} acts as a general acid to protonate the departing GlcNAc. In step ❷, water attacks at C-1 of Mur2Ac to displace the Asp^{52} and generate product.

The Phillips mechanism (S_N1), was widely accepted for more than three decades. However, some controversy persisted and tests continued. The scientific method sometimes advances an issue slowly, and a truly insightful experiment can be difficult to design. Some early arguments against the Phillips mechanism were suggestive but not completely persuasive. For example, the half-life of the proposed glycosyl cation was estimated to be 10^{-12} second, just longer than a molecular vibration and not long enough for the needed diffusion of other molecules. More important, lysozyme is a member

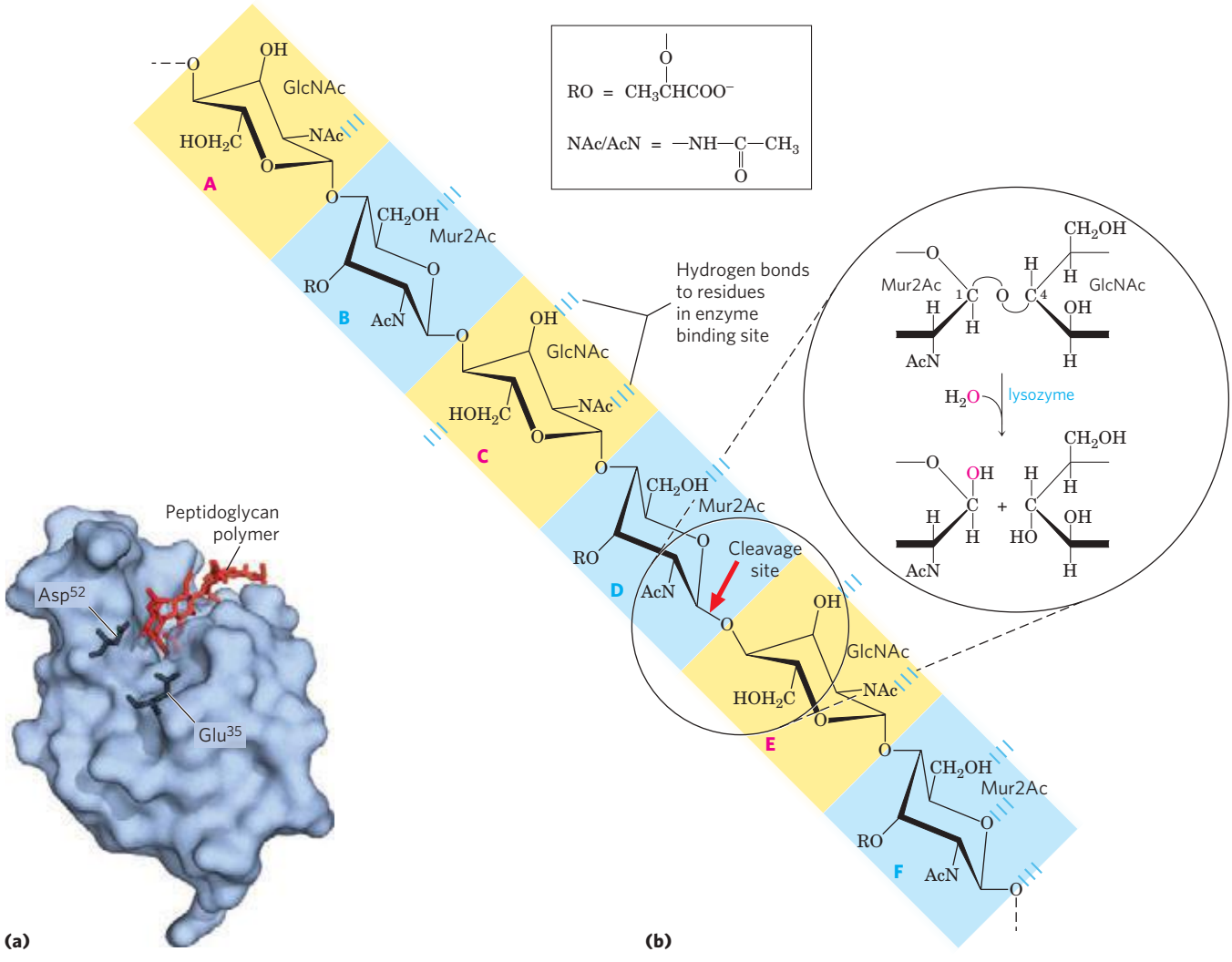
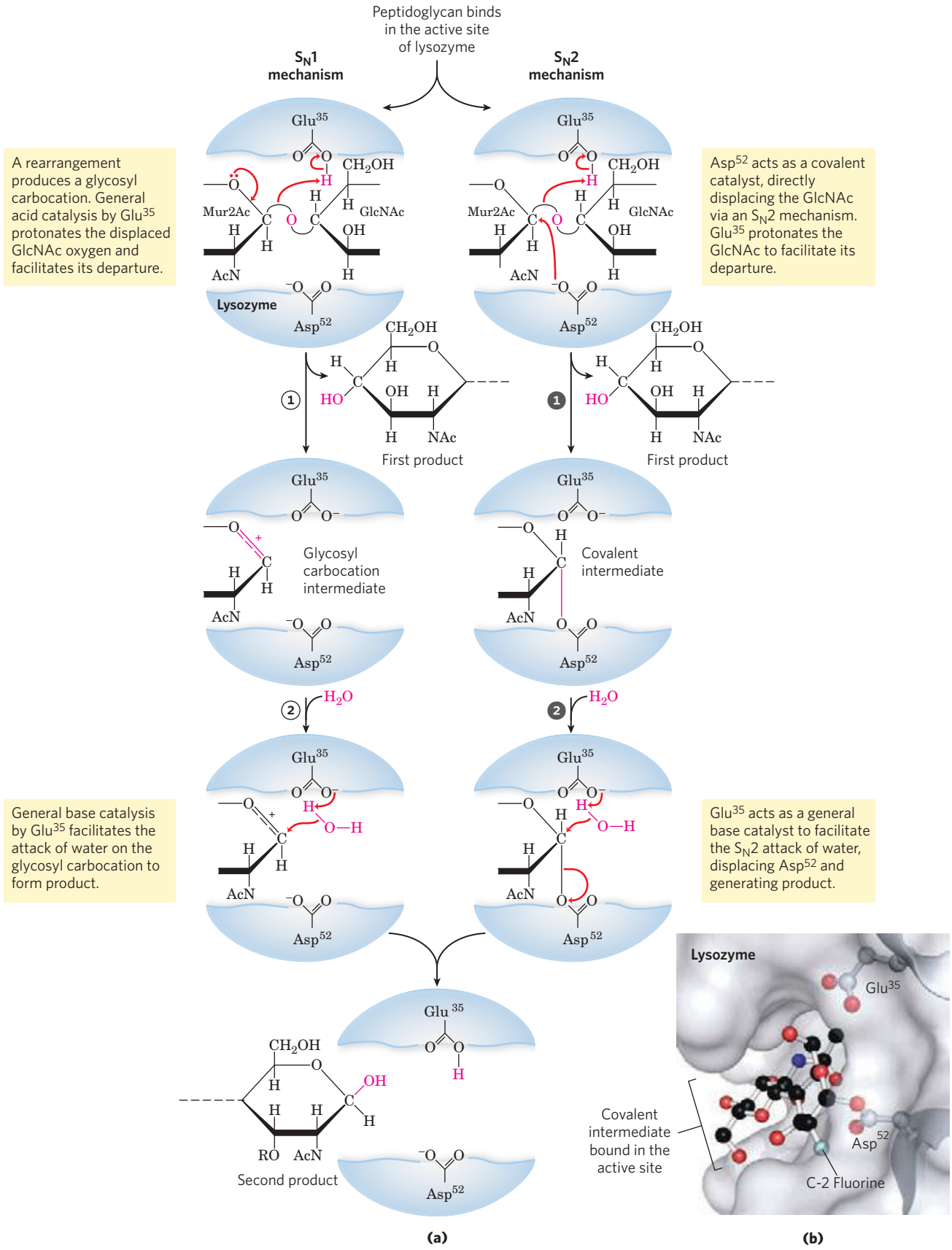


FIGURE 6–27 Hen egg white lysozyme and the reaction it catalyzes. **(a)** Surface representation of the enzyme, with the active-site residues Glu^{35} and Asp^{52} shown as black stick structures and bound substrate shown as a red stick structure (PDB ID 1LZE). **(b)** Reaction catalyzed by hen egg white lysozyme. A segment of a peptidoglycan polymer is shown, with the lysozyme binding sites A through F shaded. The glycosidic C—O bond between sugar residues bound to sites D and E is cleaved, as indicated by the red arrow. The hydrolytic reaction is shown in the inset, with the fate of the oxygen in the H_2O traced in red. Mur2Ac is *N*-acetylmuramic acid; GlcNAc, *N*-acetylglucosamine. RO— represents a lactyl (lactic acid) group; —NAc and AcN—, an *N*-acetyl group (see key).

of a family of enzymes called "retaining glycosidases," all of which catalyze reactions in which the product has the same anomeric configuration as the substrate (anomeric configurations of carbohydrates are examined in Chapter 7), and all of which are known to have reactive covalent intermediates like that envisioned in the alternative (S_N2) pathway. Hence, the Phillips mechanism ran counter to experimental findings for closely related enzymes.

A compelling experiment tipped the scales decidedly in favor of the S_N2 pathway, as reported by Stephen Withers and colleagues in 2001. Making use of a mutant enzyme (with residue 35 changed from Glu to Gln) and artificial substrates, which combined to slow the rate of key steps in the reaction, these workers were able to stabilize the elusive covalent intermediate. This in turn allowed them to observe the intermediate directly, using both mass spectrometry and x-ray crystallography (Fig. 6–28b).

Is the lysozyme mechanism now proven? No. A key feature of the scientific method, as Albert Einstein once summarized it, is "No amount of experimentation can ever prove me right; a single experiment can prove me wrong." In the case of the lysozyme mechanism, one might argue (and some have) that the artificial substrates, with fluorine substitutions at C-1 and C-2, as were used to stabilize the covalent intermediate, might have altered the reaction pathway. The highly electronegative fluorine could destabilize an already electron-deficient oxocarbenium ion in the glycosyl cation intermediate that might occur in an S_N1 pathway. However, the S_N2 pathway is now the mechanism most in concert with available data.



MECHANISM FIGURE 6–28 Lysozyme reaction. In this reaction (described in the text), the water introduced into the product at C-1 of Mur2Ac is in the same configuration as the original glycosidic bond. The reaction is thus a molecular substitution with retention of configuration. **(a)** Two proposed pathways potentially explain the overall reaction and its properties. The S_N1 pathway (left) is the original Phillips mechanism. The S_N2 pathway (right) is the mechanism most consistent with current data. **(b)** A surface rendering of the lysozyme active site, with the covalent enzyme-substrate intermediate shown as a ball-and-stick structure. Side chains of active-site residues are shown as ball-and-stick structures protruding from ribbons (PDB ID 1H6M).

An Understanding of Enzyme Mechanism Produces Useful Antibiotics

Penicillin was discovered in 1928 by Alexander Fleming, but it took another 15 years before this relatively unstable compound was understood well enough to use it as a pharmaceutical agent to treat bacterial infections. Penicillin interferes with the synthesis of peptidoglycan (described in Chapter 20, Fig. 20–31), the major component of the rigid cell wall that protects bacteria from osmotic lysis. Peptidoglycan consists of polysaccharides and peptides cross-linked in several steps that include a transpeptidase reaction **(Fig. 6–29)**. It is this reaction that is inhibited by penicillin and related compounds **(Fig. 6–30a)**, all of which mimic one conformation of the D-Ala–D-Ala segment of the peptidoglycan precursor. The peptide bond in the precursor is replaced by a highly reactive β-lactam ring. When penicillin binds to the transpeptidase, an active-site Ser attacks the carbonyl of the β-lactam ring and generates a covalent adduct between penicillin and the enzyme. However, the leaving group remains attached because it is linked by the remnant of the β-lactam ring (Fig. 6–30b). The covalent complex irreversibly inactivates the enzyme. This, in turn, blocks synthesis of the bacterial cell wall, and most bacteria die as the fragile inner membrane bursts under osmotic pressure.

Human use of penicillin and its derivatives has led to the evolution of strains of pathogenic bacteria that express **β-lactamases (Fig. 6–31a)**, enzymes that cleave β-lactam antibiotics, rendering them inactive. The bacteria thereby become resistant to the antibiotics. The genes for these enzymes have spread rapidly through bacterial populations under the selective pressure imposed by the use (and often overuse) of β-lactam antibiotics. Human medicine responded with the development of compounds such as clavulanic acid, a suicide inactivator, which irreversibly inactivates the β-lactamases (Fig. 6–31b). Clavulanic acid mimics the structure of a β-lactam antibiotic, and forms a covalent adduct with a Ser in the β-lactamase active site. This leads to a rearrangement that creates a much more reactive derivative, which is subsequently attacked by another nucleophile in the active site to irreversibly acylate the enzyme and inactivate it. Amoxicillin and clavulanic acid are combined in a widely used pharmaceutical formulation with the trade name Augmentin. The cycle of chemical warfare between humans and bacteria continues unabated. Strains of disease-causing bacteria that are resistant to both amoxicillin and clavulanic acid (reflecting mutations in β-lactamase that render it unreactive to clavulanic acid) have been discovered. The development of new antibiotics promises to be a growth industry for the foreseeable future. ■

SUMMARY 6.4 Examples of Enzymatic Reactions

- Chymotrypsin is a serine protease with a well-understood mechanism, featuring general acid-base catalysis, covalent catalysis, and transition-state stabilization.

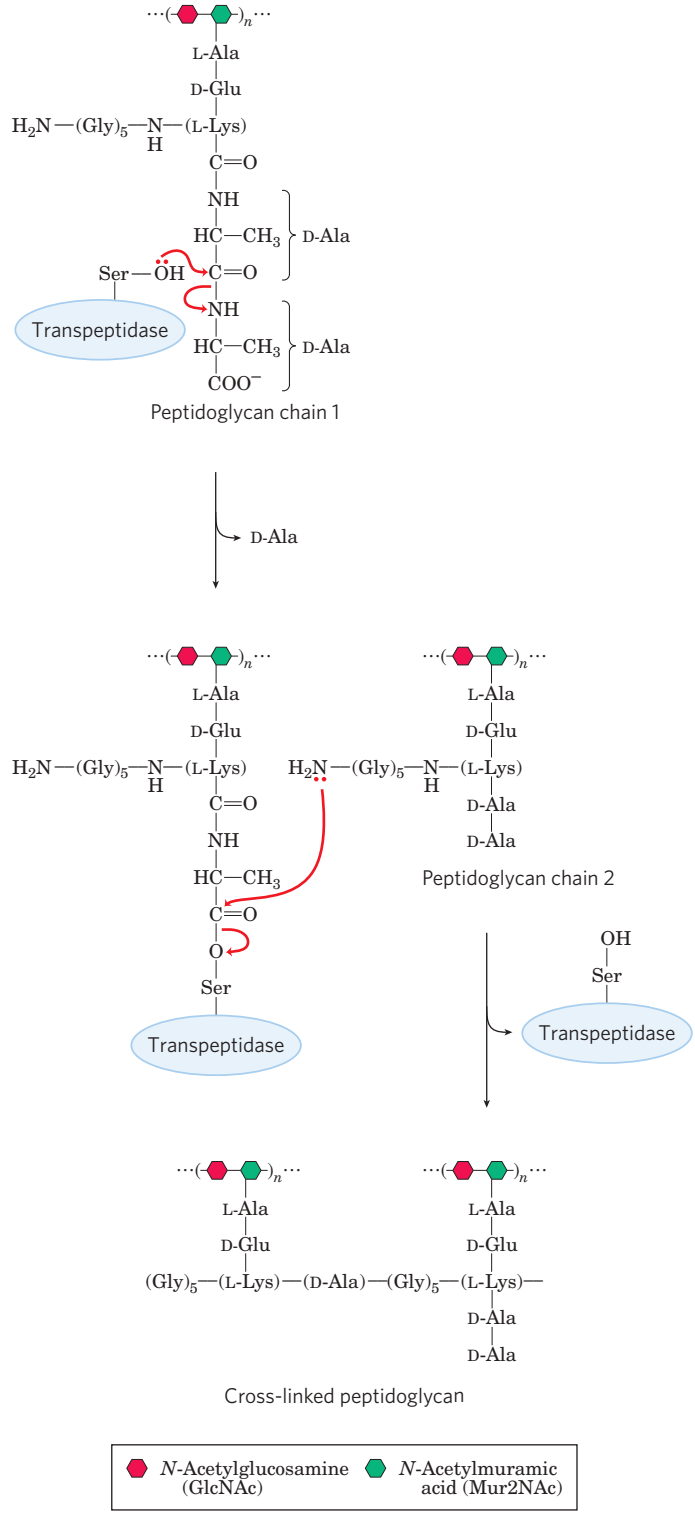
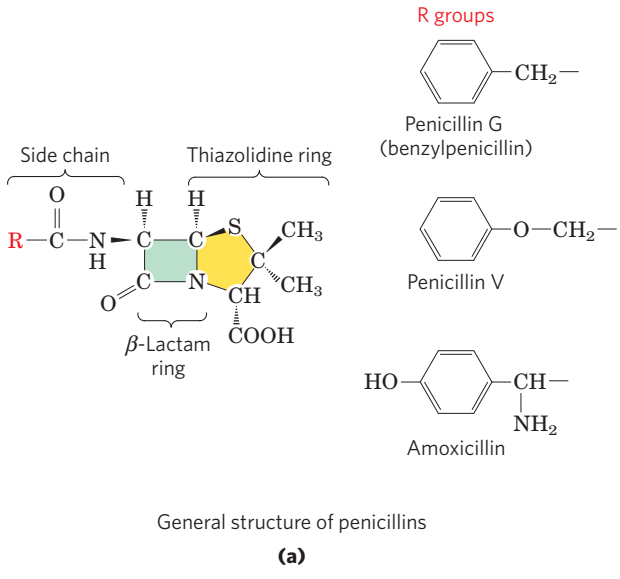


FIGURE 6–29 The transpeptidase reaction. This reaction, which links two peptidoglycan precursors into a larger polymer, is facilitated by an active-site Ser and a covalent catalytic mechanism similar to that of chymotrypsin. Note that peptidoglycan is one of the few places in nature where D-amino acid residues are found. The active-site Ser attacks the carbonyl of the peptide bond between the two D-Ala residues, creating a covalent ester linkage between the substrate and the enzyme, with release of the terminal D-Ala residue. An amino group from the second peptidoglycan precursor then attacks the ester linkage, displacing the enzyme and cross-linking the two precursors.



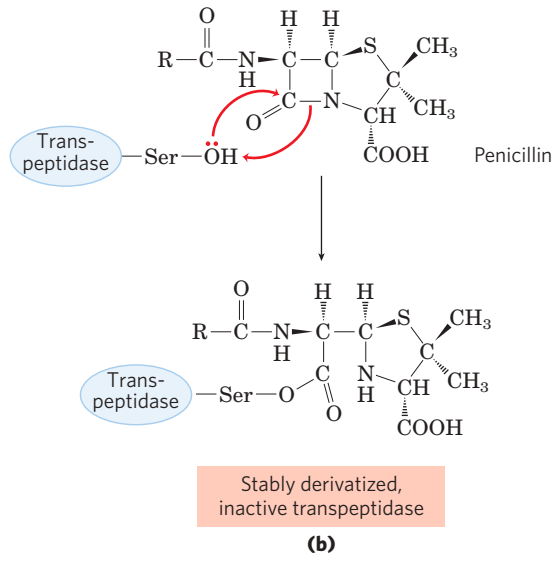
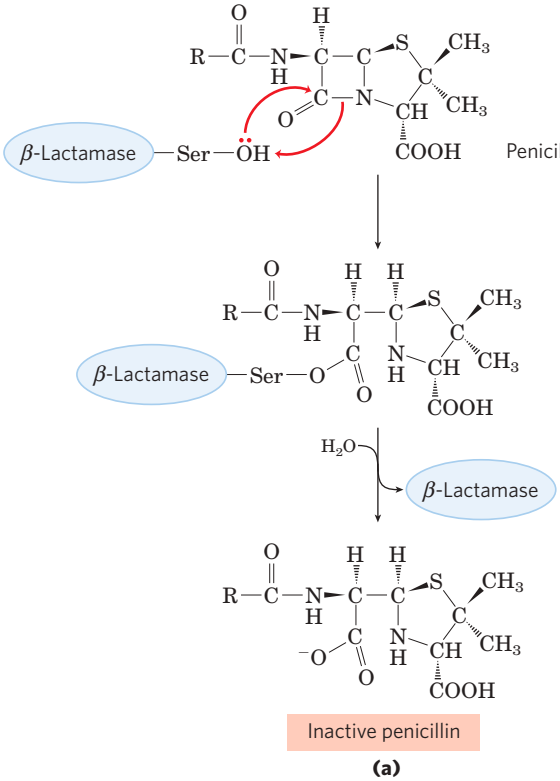


FIGURE 6–30 Transpeptidase inhibition by β-lactam antibiotics. **(a)** β-Lactam antibiotics feature a five-membered thiazolidine ring fused to a four-membered β-lactam ring. The latter ring is strained and includes an amide moiety that plays a critical role in the inactivation of peptidoglycan synthesis. The R group differs in different penicillins. Penicillin G was the first to be isolated and remains one of the most effective, but it is degraded by stomach acid and must be administered by injection. Penicillin V is nearly as effective and is acid stable, so it can be administered orally. Amoxicillin has a broad range of effectiveness, is readily administered orally, and is thus the most widely prescribed β-lactam antibiotic. **(b)** Attack on the amide moiety of the β-lactam ring by a transpeptidase active-site Ser results in a covalent acyl-enzyme product. This is hydrolyzed so slowly that adduct formation is practically irreversible, and the transpeptidase is inactivated.



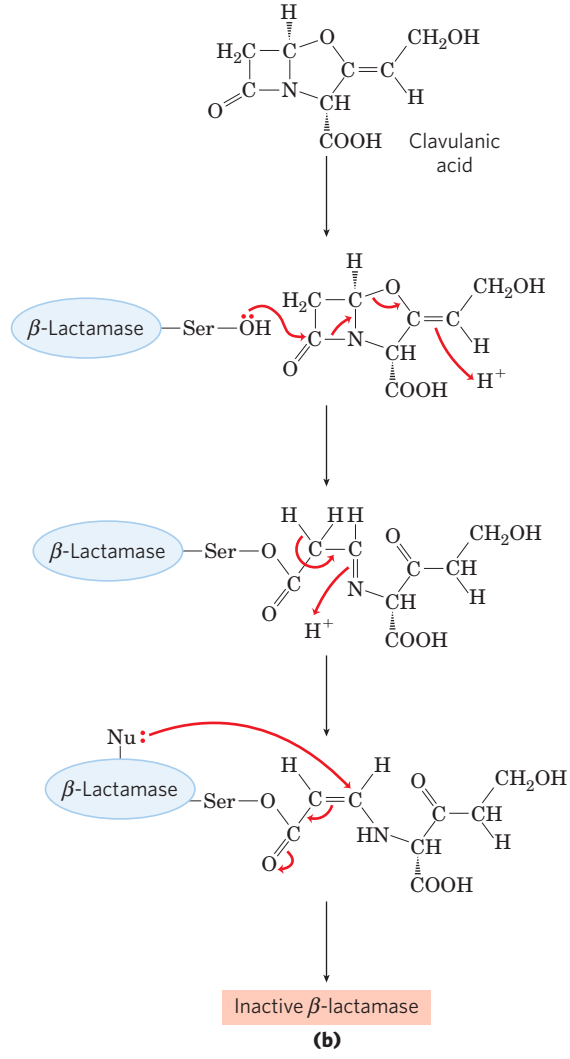


FIGURE 6–31 β-Lactamases and β-lactamase inhibition. **(a)** β-Lactamases promote cleavage of the β-lactam ring in β-lactam antibiotics, inactivating them. **(b)** Clavulanic acid is a suicide inhibitor, making use of the normal chemical mechanism of β-lactamases to create a reactive species at the active site. This reactive species is attacked by a nucleophilic group (Nu:) in the active site to irreversibly acylate the enzyme.

- Hexokinase provides an excellent example of induced fit as a means of using substrate binding energy.
- The enolase reaction proceeds via metal ion catalysis.
- Lysozyme makes use of covalent catalysis and general acid catalysis as it promotes two successive nucleophilic displacement reactions.
- Understanding enzyme mechanism allows the development of drugs to inhibit enzyme action.

6.5 Regulatory Enzymes

In cellular metabolism, groups of enzymes work together in sequential pathways to carry out a given metabolic process, such as the multireaction breakdown of glucose to lactate or the multireaction synthesis of an amino acid from simpler precursors. In such enzyme systems, the reaction product of one enzyme becomes the substrate of the next.

Most of the enzymes in each metabolic pathway follow the kinetic patterns we have already described. Each pathway, however, includes one or more enzymes that have a greater effect on the rate of the overall sequence. These **regulatory enzymes** exhibit increased or decreased catalytic activity in response to certain signals. Adjustments in the rate of reactions catalyzed by regulatory enzymes, and therefore in the rate of entire metabolic sequences, allow the cell to meet changing needs for energy and for biomolecules required in growth and repair.

The activities of regulatory enzymes are modulated in a variety of ways. **Allosteric enzymes** function through reversible, noncovalent binding of regulatory compounds called **allosteric modulators** or **allosteric effectors**, which are generally small metabolites or cofactors. Other enzymes are regulated by reversible **covalent modification**. Both classes of regulatory enzymes tend to be multisubunit proteins, and in some cases the regulatory site(s) and the active site are on separate subunits. Metabolic systems have at least two other mechanisms of enzyme regulation. Some enzymes are stimulated or inhibited when they are bound by separate **regulatory proteins**. Others are activated when peptide segments are removed by **proteolytic cleavage**; unlike effector-mediated regulation, regulation by proteolytic cleavage is irreversible. Important examples of both mechanisms are found in physiological processes such as digestion, blood clotting, hormone action, and vision.

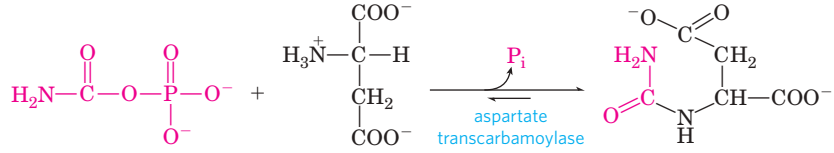
Cell growth and survival depend on efficient use of resources, and this efficiency is made possible by regulatory enzymes. No single rule governs the occurrence of different types of regulation in different systems. To a degree, allosteric (noncovalent) regulation may permit fine-tuning of metabolic pathways that are required continuously but at different levels of activity as cellular conditions change. Regulation by covalent modification may be all or none—usually the case with proteolytic cleavage—or it may allow for subtle changes in activity. Several types of regulation may occur in a single regulatory enzyme. The remainder of this chapter is devoted to a discussion of these methods of enzyme regulation.

Allosteric Enzymes Undergo Conformational Changes in Response to Modulator Binding

As we saw in Chapter 5, allosteric proteins are those having "other shapes" or conformations induced by the binding of modulators. The same concept applies to certain regulatory enzymes, as conformational changes induced by one or more modulators interconvert more-active and less-active forms of the enzyme. The modulators for allosteric enzymes may be inhibitory or stimulatory. Often the modulator is the substrate itself; regulation in which substrate and modulator are identical is called homotropic. The effect is similar to that of O_2 binding to hemoglobin (Chapter 5): binding of the ligand—or substrate, in the case of enzymes—causes conformational changes that affect the subsequent activity of other sites on the protein. In most cases, the conformational change converts a relatively inactive conformation (often referred to as a T state) to a more active conformation (an R state). When the modulator is a molecule other than the substrate, the enzyme is said to be heterotropic. Note that allosteric modulators should not be confused with uncompetitive and mixed inhibitors. Although the latter bind at a second site on the enzyme, they do not necessarily mediate conformational changes between active and inactive forms, and the kinetic effects are distinct.

The properties of allosteric enzymes are significantly different from those of simple nonregulatory enzymes. Some of the differences are structural. In addition to active sites, allosteric enzymes generally have one or more regulatory, or allosteric, sites for binding the modulator **(Fig. 6–32)**. Just as an enzyme's active site is specific for its substrate, each regulatory site is specific for its modulator. Enzymes with several modulators generally have different specific binding sites for each. In homotropic enzymes, the active site and regulatory site are the same.

Allosteric enzymes are typically larger and more complex than nonallosteric enzymes, with two or more subunits. A classic example is aspartate transcarbamoylase (often abbreviated ATCase), which catalyzes an early step in the biosynthesis of pyrimidine nucleotides, the reaction of carbamoyl phosphate and aspartate to form carbamoyl aspartate.



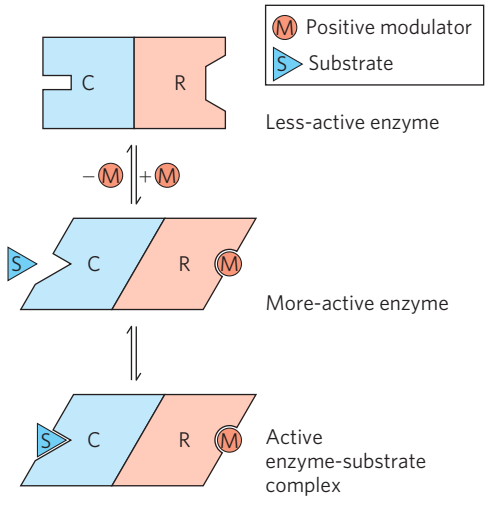


FIGURE 6–32 Subunit interactions in an allosteric enzyme, and interactions with inhibitors and activators. In many allosteric enzymes the substrate-binding site and the modulator-binding site(s) are on different subunits, the catalytic (C) and regulatory (R) subunits, respectively. Binding of the positive (stimulatory) modulator (M) to its specific site on the regulatory subunit is communicated to the catalytic subunit through a conformational change. This change renders the catalytic subunit active and capable of binding the substrate (S) with higher affinity. On dissociation of the modulator from the regulatory subunit, the enzyme reverts to its inactive or less active form.

ATCase has 12 polypeptide chains organized into 6 catalytic (organized as 2 trimeric complexes) and 6 regulatory (organized as 3 dimeric complexes) subunits. **Figure 6–33** shows the quaternary structure of this enzyme, deduced from x-ray analysis. The enzyme exhibits allosteric behavior as detailed below, as the catalytic subunits function cooperatively. The regulatory subunits have binding sites for ATP and CTP, which function as positive and negative regulators, respectively. CTP is one of the end products of the pathway, and negative regulation by CTP serves to limit ATCase action under conditions when CTP is abundant. On the other hand, high concentrations of ATP indicate that cellular metabolism is robust, the cell is growing, and additional pyrimidine nucleotides may be needed to support RNA transcription and DNA replication.

The Kinetic Properties of Allosteric Enzymes Diverge from Michaelis-Menten Behavior

Allosteric enzymes show relationships between V_0 and [S] that differ from Michaelis-Menten kinetics. They do exhibit saturation with the substrate when [S] is sufficiently high, but for allosteric enzymes, plots of V_0 versus [S] **(Fig. 6–34)** usually produce a sigmoid saturation curve, rather than the hyperbolic curve typical of nonregulatory enzymes. On the sigmoid saturation curve we can find a value of [S] at which V_0 is half-maximal, but we cannot refer to it with the designation K_m, because the enzyme does not follow the hyperbolic Michaelis-Menten relationship. Instead, the symbol $[S]_{0.5}$ or $K_{0.5}$ is often used to represent the substrate concentration giving half-maximal velocity of the reaction catalyzed by an allosteric enzyme (Fig. 6–34).

Sigmoid kinetic behavior generally reflects cooperative interactions between multiple protein subunits. In other words, changes in the structure of one subunit are translated into structural changes in adjacent subunits, an effect mediated by noncovalent interactions at the

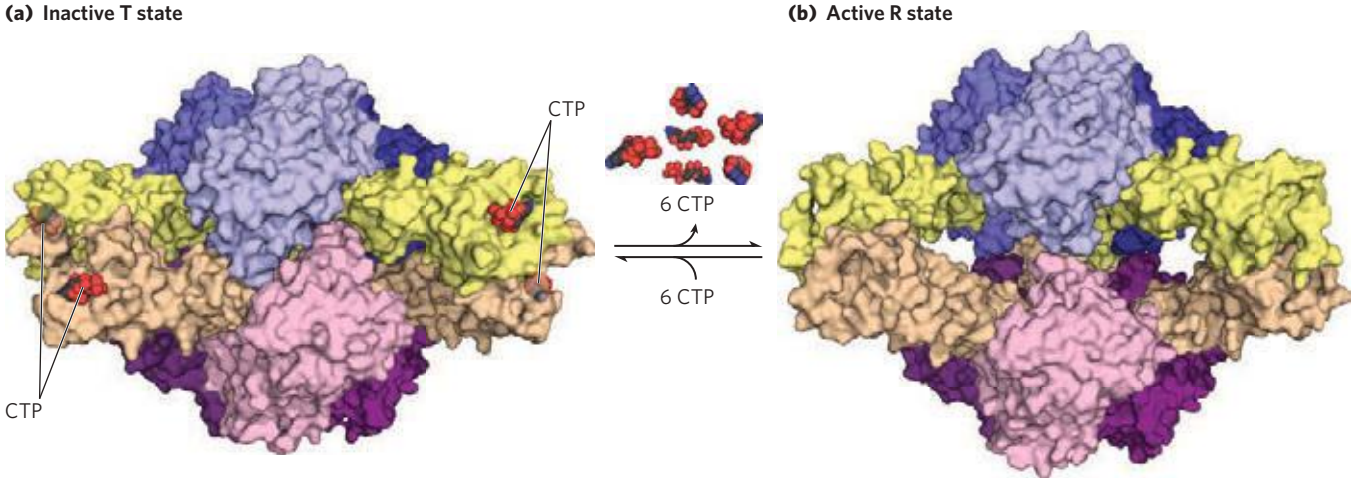


FIGURE 6–33 The regulatory enzyme aspartate transcarbamoylase. **(a)** The inactive T state (PDB ID 1RAB) and **(b)** active R state (PDB ID 1F1B) of the enzyme are shown. This allosteric regulatory enzyme has two stacked catalytic clusters, each with three catalytic polypeptide chains (in shades of blue and purple), and three regulatory clusters, each with two regulatory polypeptide chains (in beige and yellow). The regulatory clusters form the points of a triangle (not evident in this side view) surrounding the catalytic subunits. Binding sites for allosteric modulators (including CTP) are on the regulatory subunits. Modulator binding produces large changes in enzyme conformation and activity. The role of this enzyme in nucleotide synthesis, and details of its regulation, are discussed in Chapter 22.

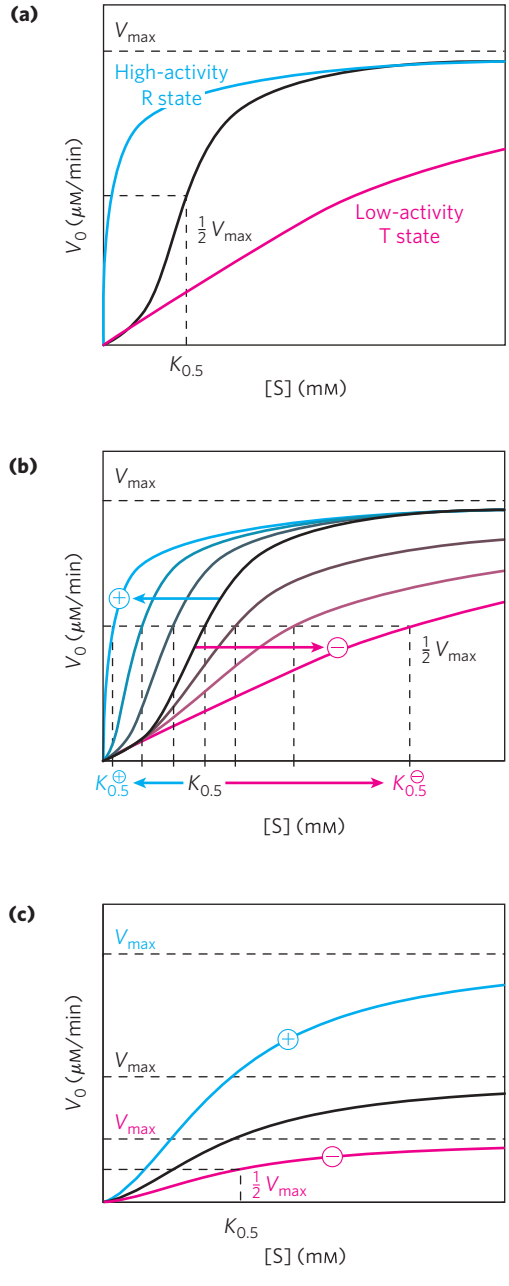


FIGURE 6–34 Substrate-activity curves for representative allosteric enzymes. Three examples of complex responses of allosteric enzymes to their modulators. **(a)** The sigmoid curve of a homotropic enzyme, in which the substrate also serves as a positive (stimulatory) modulator, or activator. Note the resemblance to the oxygen-saturation curve of hemoglobin (see Fig. 5–12). The sigmoidal curve is a hybrid curve in which the enzyme is present primarily in the relatively inactive T state at low substrate concentration, and primarily in the more active R state at high substrate concentration. The curves for the pure T and R states are plotted separately in color. ATCase exhibits a kinetic pattern similar to this. **(b)** The effects of several different concentrations of a positive modulator (+) or a negative modulator (−) on an allosteric enzyme in which $K_{0.5}$ is altered without a change in V_{max}. The central curve shows the substrate-activity relationship without a modulator. For ATCase, CTP is a negative modulator and ATP is a positive modulator. **(c)** A less common type of modulation, in which V_{max} is altered and $K_{0.5}$ is nearly constant.

interface between subunits. The principles are particularly well illustrated by a nonenzyme: O_2 binding to hemoglobin. Sigmoid kinetic behavior is explained by the concerted and sequential models for subunit interactions (see Fig. 5–15).

ATCase effectively illustrates both homotropic and heterotropic allosteric kinetic behavior. The binding of the substrates, aspartate and carbamoyl phosphate, to the enzyme gradually bring about a transition from the relatively inactive T state to the more active R state. This accounts for the sigmoid rather than hyperbolic change in V_0 with increasing [S]. One characteristic of sigmoid kinetics is that small changes in the concentration of a modulator can be associated with large changes in activity. As exemplified in Figure 6–34a, a relatively small increase in [S] in the steep part of the curve causes a comparatively large increase in V_0.

The heterotropic allosteric regulation of ATCase is brought about by its interactions with ATP and CTP. For heterotropic allosteric enzymes, an activator may cause the curve to become more nearly hyperbolic, with a decrease in $K_{0.5}$ but no change in V_{max}, resulting in an increased reaction velocity at a fixed substrate concentration. For ATCase, the interaction with ATP brings this about, and the enzyme exhibits a V_0 versus [S] curve that is characteristic of the active R state at sufficiently high ATP concentrations (V_0 is higher for any value of [S]; Fig. 6–34b). A negative modulator (an inhibitor) may produce a *more* sigmoid substrate-saturation curve, with an increase in $K_{0.5}$, as illustrated by the effects of CTP on ATCase kinetics (see curves for negative modulater, Fig. 6–34b). Other heterotropic allosteric enzymes respond to an activator by an increase in V_{max} with little change in $K_{0.5}$ (Fig. 6–34c). Heterotropic allosteric enzymes therefore show different kinds of responses in their substrate-activity curves because some have inhibitory modulators, some have activating modulators, and some (like ATCase) have both.

Some Enzymes Are Regulated by Reversible Covalent Modification

In another important class of regulatory enzymes, activity is modulated by covalent modification of one or more of the amino acid residues in the enzyme molecule. Over 500 different types of covalent modification have been found in proteins. Common modifying groups include phosphoryl, acetyl, adenylyl, uridylyl, methyl, amide, carboxyl, myristoyl, palmitoyl, prenyl, hydroxyl, sulfate, and adenosine diphosphate ribosyl groups **(Fig. 6–35)**. There are even entire proteins that are used as specialized modifying groups, including ubiquitin and sumo. These varied groups are generally linked to and removed from a regulated enzyme by separate enzymes. When an amino acid residue in an enzyme is modified, a novel amino acid with altered properties has effectively been introduced into the enzyme. Introduction of a charge can alter the local properties of the enzyme and induce

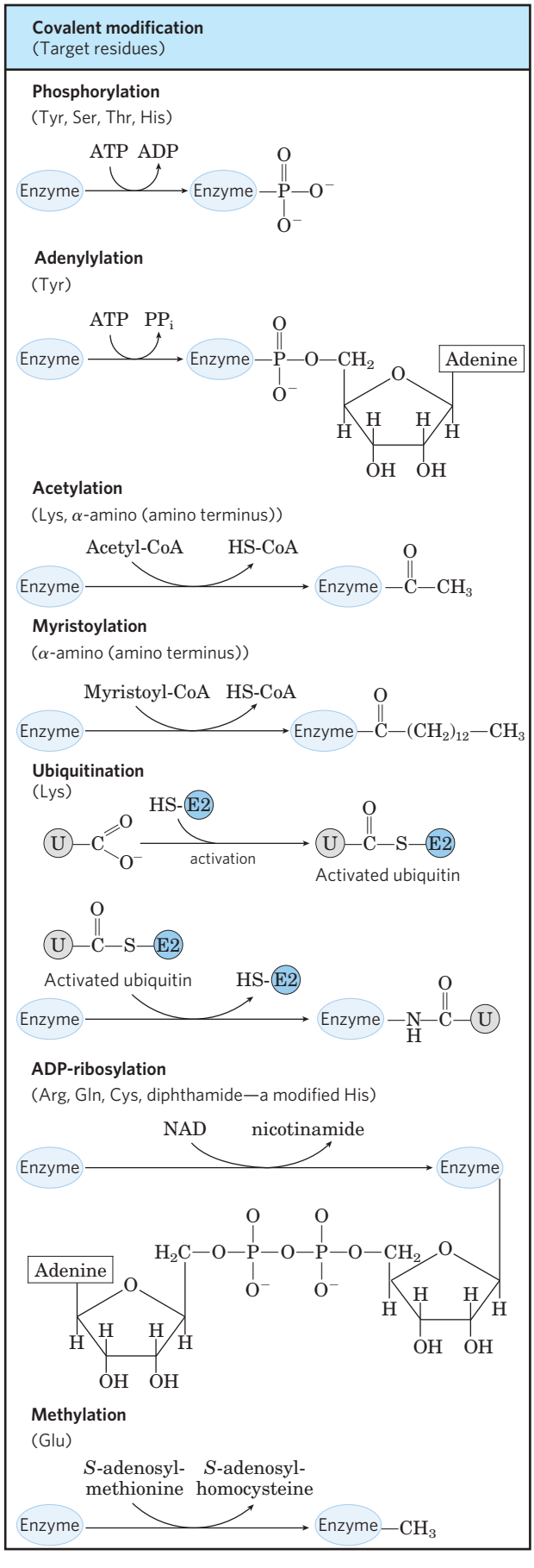


FIGURE 6–35 Some enzyme modification reactions.

a change in conformation. Introduction of a hydrophobic group can trigger association with a membrane. The changes are often substantial and can be critical to the function of the altered enzyme.

The variety of enzyme modifications is too great to cover in detail, but some examples can be offered. An example of an enzyme regulated by methylation is the methyl-accepting chemotaxis protein of bacteria. This protein is part of a system that permits a bacterium to swim toward an attractant (such as a sugar) in solution and away from repellent chemicals. The methylating agent is *S*-adenosylmethionine (adoMet) (see Fig. 18–18). Acetylation is another common modification, with approximately 80% of the soluble proteins in eukaryotes, including many enzymes, acetylated at their amino terminus. Ubiquitin is added to proteins as a tag that predestines them for proteolytic degradation (see Fig. 27–47). Ubiquitination can also have a regulatory function. Sumo is found attached to many eukaryotic nuclear proteins with roles in the regulation of transcription, chromatin structure, and DNA repair.

ADP-ribosylation is an especially interesting reaction, observed in a number of proteins; the ADP-ribose is derived from nicotinamide adenine dinucleotide (NAD) (see Fig. 8–38). This type of modification occurs for the bacterial enzyme dinitrogenase reductase, resulting in regulation of the important process of biological nitrogen fixation. Diphtheria toxin and cholera toxin are enzymes that catalyze the ADP-ribosylation (and inactivation) of key cellular enzymes or proteins.

Phosphorylation is the most important type of regulatory modification. It is estimated that one-third of all proteins in a eukaryotic cell are phosphorylated, and one or (often) many phosphorylation events are part of virtually every regulatory process. Some proteins have only one phosphorylated residue, others have several, and a few have dozens of sites for phosphorylation. This mode of covalent modification is central to a large number of regulatory pathways, and we therefore discuss it in some detail here, and again in Chapter 12.

We will encounter all of these types of modification again in later chapters.

Phosphoryl Groups Affect the Structure and Catalytic Activity of Enzymes

The attachment of phosphoryl groups to specific amino acid residues of a protein is catalyzed by **protein kinases**. In the reactions, the γ-phosphoryl group derived from a nucleoside triphosphate (usually ATP) is transferred to a particular Ser, Thr, or Tyr residue (occasionally His as well) on the target protein. This introduces a bulky, charged group into a region of the target protein that was only moderately polar. The oxygen atoms of a phosphoryl group can hydrogen-bond with one or several groups in a protein, commonly the amide groups of the peptide backbone at the start of an α helix or the charged guanidinium group of an Arg

residue. The two negative charges on a phosphorylated side chain can also repel neighboring negatively charged (Asp or Glu) residues. When the modified side chain is located in a region of an enzyme critical to its three-dimensional structure, phosphorylation can have dramatic effects on enzyme conformation and thus on substrate binding and catalysis. Removal of phosphoryl groups from these same target proteins is catalyzed by **protein phosphatases**.

An important example of enzyme regulation by phosphorylation is seen in glycogen phosphorylase (M_r 94,500) of muscle and liver (Chapter 15), which catalyzes the reaction

$$\underset{\text{Glycogen}}{(\text{Glucose})_n} + P_i \longrightarrow \underset{\substack{\text{Shortened}\\ \text{glycogen}\\ \text{chain}}}{(\text{glucose})_{n-1}} + \text{glucose 1-phosphate}$$

The glucose 1-phosphate so formed can be used for ATP synthesis in muscle or converted to free glucose in the liver. Note that glycogen phosphorylase, though it adds a phosphate to a substrate, is not itself a kinase, because it does not utilize ATP or any other nucleotide triphosphate as a phosphoryl donor in its catalyzed reaction. It is, however, the substrate for a protein kinase that phosphorylates it. In the discussion below, the phosphoryl groups we are concerned with are those involved in regulation of the enzyme, as distinguished from its catalytic function.

Glycogen phosphorylase occurs in two forms: the more active phosphorylase *a* and the less active phosphorylase *b* **(Fig. 6–36)**. Phosphorylase *a* has two subunits, each with a specific Ser residue that is phosphorylated at its hydroxyl group. These serine phosphate residues are required for maximal activity of the enzyme. The phosphoryl groups can be hydrolytically removed by a separate enzyme called phosphorylase phosphatase:

$$\underset{\text{(more active)}}{\text{Phosphorylase } a} + 2H_2O \longrightarrow \underset{\text{(less active)}}{\text{Phosphorylase } b} + 2P_i$$

In this reaction, phosphorylase *a* is converted to phosphorylase *b* by the cleavage of two serine phosphate covalent bonds, one on each subunit of glycogen phosphorylase.

Phosphorylase *b* can in turn be reactivated—covalently transformed back into active phosphorylase *a*—by another enzyme, phosphorylase kinase, which catalyzes the transfer of phosphoryl groups from ATP to the hydroxyl groups of the two specific Ser residues in phosphorylase *b*:

$$2\text{ATP} + \underset{\text{(less active)}}{\text{phosphorylase } b} \longrightarrow 2\text{ADP} + \underset{\text{(more active)}}{\text{phosphorylase } a}$$

The breakdown of glycogen in skeletal muscles and the liver is regulated by variations in the ratio of the two forms of glycogen phosphorylase. The *a* and *b* forms differ in their secondary, tertiary, and quaternary structures; the active site undergoes changes in structure and, consequently, changes in catalytic activity as the two forms are interconverted.

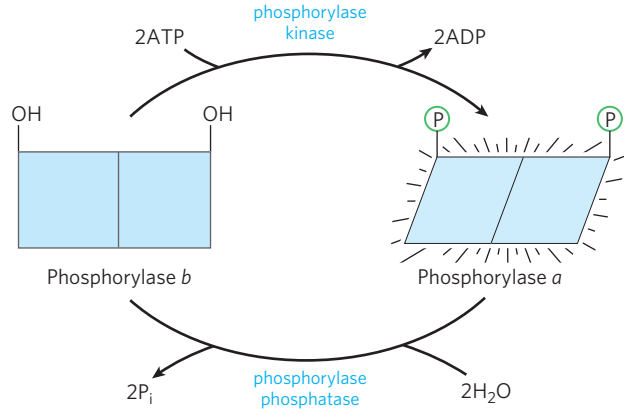


FIGURE 6–36 Regulation of muscle glycogen phosphorylase activity by phosphorylation. In the more active form of the enzyme, phosphorylase *a*, specific Ser residues, one on each subunit, are phosphorylated. Phosphorylase *a* is converted to the less active phosphorylase *b* by enzymatic loss of these phosphoryl groups, promoted by phosphoprotein phosphatase 1 (PP1). Phosphorylase *b* can be reconverted (reactivated) to phosphorylase *a* by the action of phosphorylase kinase.

The regulation of glycogen phosphorylase by phosphorylation illustrates the effects on both structure and catalytic activity of adding a phosphoryl group. In the unphosphorylated state, each subunit of this enzyme is folded so as to bring the 20 residues at its amino terminus, including a number of basic residues, into a region containing several acidic amino acids; this produces an electrostatic interaction that stabilizes the conformation. Phosphorylation of Ser^{14} interferes with this interaction, forcing the amino-terminal domain out of the acidic environment and into a conformation that allows interaction between the Ⓟ-Ser and several Arg side chains. In this conformation, the enzyme is much more active.

Phosphorylation of an enzyme can affect catalysis in another way: by altering substrate-binding affinity. For example, when isocitrate dehydrogenase (an enzyme of the citric acid cycle; Chapter 16) is phosphorylated, electrostatic repulsion by the phosphoryl group inhibits the binding of citrate (a tricarboxylic acid) at the active site.

Multiple Phosphorylations Allow Exquisite Regulatory Control

The Ser, Thr, or Tyr residues that are typically phosphorylated in regulated proteins occur within common structural motifs, called consensus sequences, that are recognized by specific protein kinases (Table 6–10). Some kinases are basophilic, preferentially phosphorylating a residue having basic neighbors; others have different substrate preferences, such as for a residue near a Pro residue. Amino acid sequence is not the only important factor in determining whether a given residue will be phosphorylated, however. Protein folding brings together

TABLE 6–10 Consensus Sequences for Protein Kinases

Protein kinase	Consensus sequence and phosphorylated residue
Protein kinase A	-x-R-[RK]-x-[ST]-B-
Protein kinase G	-x-R-[RK]-x-[ST]-x-
Protein kinase C	-[RK](2)-x-[ST]-B-[RK](2)-
Protein kinase B	-x-R-x-[ST]-x-K-
Ca^{2+}/calmodulin kinase I	-B-x-R-x(2)-[ST]-x(3)-B-
Ca^{2+}/calmodulin kinase II	-B-x-[RK]-x(2)-[ST]-x(2)-
Myosin light chain kinase (smooth muscle)	-K(2)-R-x(2)-S-x-B(2)-
Phosphorylase *b* kinase	*-K-R-K-Q-I-S-V-R-*
Extracellular signal–regulated kinase (ERK)	-P-x-[ST]-P(2)-
Cyclin-dependent protein kinase (cdc2)	-x-[ST]-P-x-[KR]-
Casein kinase I	-[SpTp]-x(2)-[ST]-B*
Casein kinase II	-x-[ST]-x(2)-[ED]-x-
β-Adrenergic receptor kinase	-[DE](*n*)-[ST]-x(3)
Rhodopsin kinase	-x(2)-[ST]-E(*n*)-
Insulin receptor kinase	-x-E(3)-Y-M(4)-*K(2)-S-R-G-D-Y-M-T-M-Q-I-G-K(3)-L-P-A-T-G-D-Y-M-N-M-S-P-V-G-D-*
Epidermal growth factor (EGF) receptor kinase	*-E(4)-Y-F-E-L-V-*

Sources: Pinna, L.A. & Ruzzene, M.H. (1996) How do protein kinases recognize their substrates? *Biochim. Biophys. Acta* 1314, 191–225; Kemp, B.E. & Pearson, R.B. (1990) Protein kinase recognition sequence motifs. *Trends Biochem. Sci.* 15, 342–346; Kennelly, P.J. & Krebs, E.G. (1991) Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Biol. Chem.* 266, 15,555–15,558.

Note: Shown here are deduced consensus sequences (in roman type) and actual sequences from known substrates (italic). The Ser (S), Thr (T), or Tyr (Y) residue that undergoes phosphorylation is in red; all amino acid residues are shown as their one-letter abbreviations (see Table 3–1). x represents any amino acid; B, any hydrophobic amino acid. Sp, Tp, and Yp are Ser, Thr, and Tyr residues that must already be phosphorylated for the kinase to recognize the site.

*The best target site has two amino acid residues separating the phosphorylated and target Ser/Thr residues; target sites with one or three intervening residues function at a reduced level.

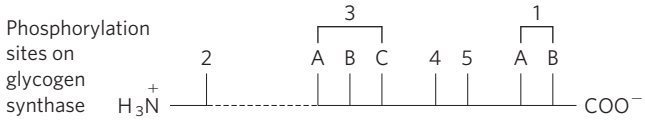
residues that are distant in the primary sequence; the resulting three-dimensional structure can determine whether a protein kinase has access to a given residue and can recognize it as a substrate. Another factor influencing the substrate specificity of certain protein kinases is the proximity of other phosphorylated residues.

Regulation by phosphorylation is often complicated. Some proteins have consensus sequences recognized by several different protein kinases, each of which can phosphorylate the protein and alter its enzymatic activity. In some cases, phosphorylation is hierarchical: a certain residue can be phosphorylated only if a neighboring residue has already been phosphorylated. For example, glycogen synthase, the enzyme that catalyzes the condensation of glucose monomers to form glycogen (Chapter 15), is inactivated by phosphorylation of specific Ser residues and is also modulated by at least four other protein kinases that phosphorylate four other sites in the enzyme **(Fig. 6–37)**. The enzyme is not a substrate for glycogen synthase kinase 3, for example, until one site has been phosphorylated by casein kinase II. Some phosphorylations inhibit glycogen synthase more than others, and some combinations of phosphorylations are cumulative. These multiple regulatory phosphorylations provide the potential for extremely subtle modulation of enzyme activity.

To serve as an effective regulatory mechanism, phosphorylation must be reversible. In general, phosphoryl groups are added and removed by different enzymes, and the processes can therefore be separately regulated. Cells contain a family of phosphoprotein phosphatases that hydrolyze specific Ⓟ–Ser, Ⓟ–Thr, and Ⓟ–Tyr esters, releasing P_i. The phosphoprotein phosphatases we know of thus far act only on a subset of phosphoproteins, but they show less substrate specificity than protein kinases.

Some Enzymes and Other Proteins Are Regulated by Proteolytic Cleavage of an Enzyme Precursor

For some enzymes, an inactive precursor called a **zymogen** is cleaved to form the active enzyme. Many proteolytic enzymes (proteases) of the stomach and pancreas are regulated in this way. Chymotrypsin and trypsin are initially synthesized as chymotrypsinogen and trypsinogen **(Fig. 6–38)**. Specific cleavage causes conformational changes that expose the enzyme active site.



Kinase	Phosphorylation sites	Degree of synthase inactivation
Protein kinase A	1A, 1B, 2, 4	+
Protein kinase G	1A, 1B, 2	+
Protein kinase C	1A	+
Ca^{2+}/calmodulin kinase	1B, 2	+
Phosphorylase *b* kinase	2	+
Casein kinase I	At least nine	+ + + +
Casein kinase II	5	0
Glycogen synthase kinase 3	3A, 3B, 3C	+ + +
Glycogen synthase kinase 4	2	+

FIGURE 6–37 Multiple regulatory phosphorylations. The enzyme glycogen synthase has at least nine separate sites in five designated regions susceptible to phosphorylation by one of the cellular protein kinases. Thus, regulation of this enzyme is a matter not of binary (on/off) switching but of finely tuned modulation of activity over a wide range in response to a variety of signals.

Because this type of activation is irreversible, other mechanisms are needed to inactivate these enzymes. Proteases are inactivated by inhibitor proteins that bind very tightly to the enzyme active site. For example, pancreatic trypsin inhibitor (M_r 6,000) binds to and inhibits trypsin. α_1-Antiproteinase (M_r 53,000) primarily inhibits neutrophil elastase (neutrophils are a type of leukocyte, or white blood cell; elastase is a protease acting on elastin, a component of some connective tissues). An insufficiency of α_1-antiproteinase, which can be caused by exposure to cigarette smoke, has been associated with lung damage, including emphysema.

Proteases are not the only proteins activated by proteolysis. In other cases, however, the precursors are called not zymogens but, more generally, **proproteins** or **proenzymes**, as appropriate. For example, the connective tissue protein collagen is initially synthesized as the soluble precursor procollagen.

A Cascade of Proteolytically Activated Zymogens Leads to Blood Coagulation

A blood clot is an aggregate of cell fragments called platelets, cross-linked and stabilized by proteinaceous fibers consisting mainly of fibrin (**Fig. 6–39a**). Fibrin is derived from a soluble zymogen called fibrinogen. After albumins and globulins, fibrinogen is generally the third most abundant type of protein in blood plasma. The formation of a blood clot provides a well-studied example of a **regulatory cascade**, a mechanism that allows a very sensitive response to—and amplification of—a molecular signal. The pathways also bring together several other types of regulation.

In a regulatory cascade, a signal leads to the activation of protein X. Protein X catalyzes the activation of protein Y. Protein Y catalyzes the activation of protein Z, and so on. Since proteins X, Y, and Z are catalysts and activate multiple copies of the next protein in the chain, the signal is amplified in each step. In some cases, the

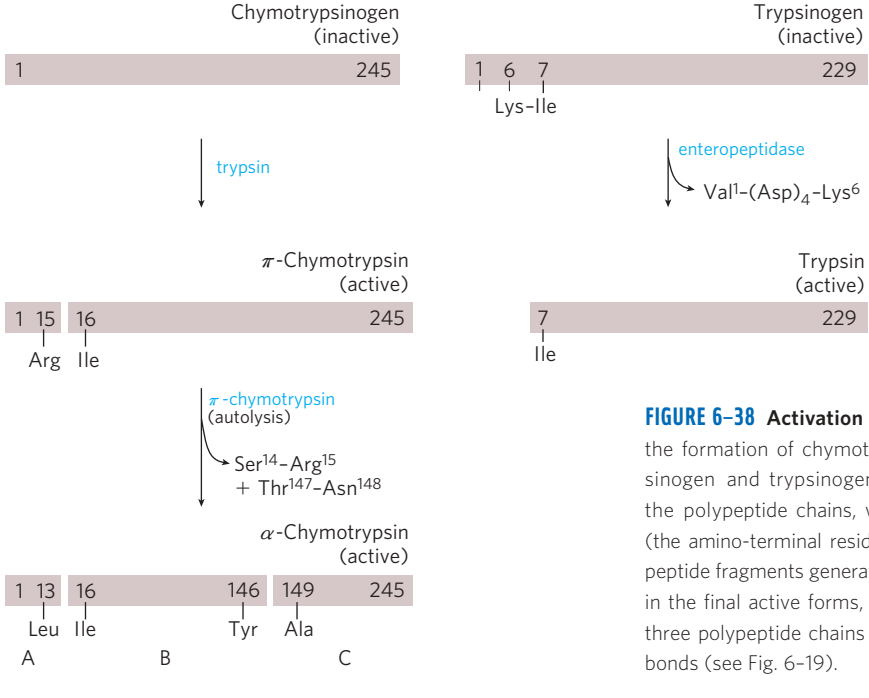


FIGURE 6–38 Activation of zymogens by proteolytic cleavage. Shown here is the formation of chymotrypsin and trypsin from their zymogens, chymotrypsinogen and trypsinogen. The bars represent the amino acid sequences of the polypeptide chains, with numbers indicating the positions of the residues (the amino-terminal residue is number 1). Residues at the termini of the polypeptide fragments generated by cleavage are indicated below the bars. Note that in the final active forms, some numbered residues are missing. Recall that the three polypeptide chains (A, B, and C) of chymotrypsin are linked by disulfide bonds (see Fig. 6–19).

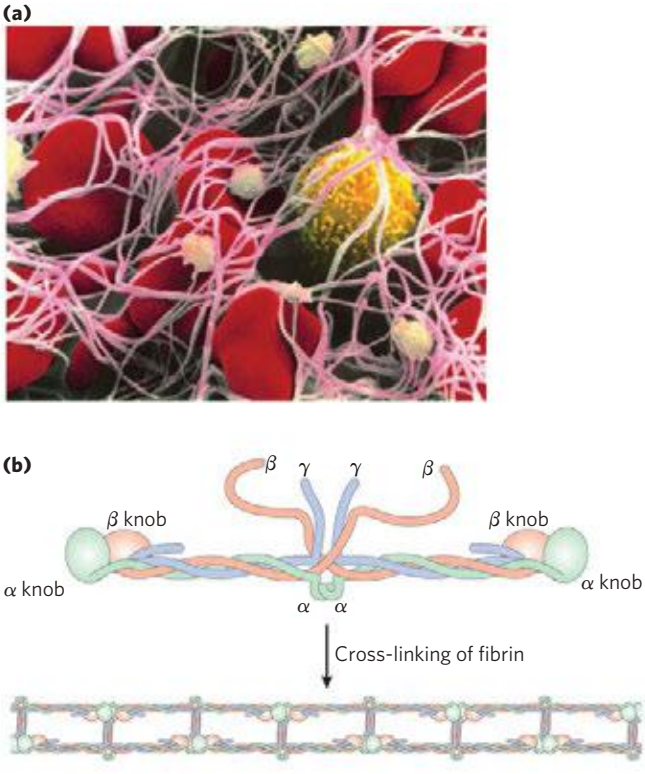


FIGURE 6–39 The function of fibrin in blood clots. (a) A blood clot consists of aggregated platelets (small, light-colored cells), tied together with strands of cross-linked fibrin. Erythrocytes (red) are also trapped in the matrix. **(b)** The soluble plasma protein fibrinogen consists of two complexes of α, β, and γ subunits ($\alpha_2\beta_2\gamma_2$). The removal of amino-terminal peptides from the α and β subunits (not shown) leads to the formation of higher-order complexes and eventual covalent cross-linking that results in the formation of fibrin fibers. The "knobs" are globular domains at the ends of the proteolyzed subunits.

activation steps involve proteolytic cleavage and are thus effectively irreversible. In others, activation entails protein modification steps such as phosphorylation, which is readily reversible. Regulatory cascades govern a wide range of biological processes, including some aspects of cell fate determination during development, the detection of light by retinal rods, programmed cell death (apoptosis), and blood coagulation.

Fibrinogen is a dimer of heterotrimers ($\alpha_2\beta_2\gamma_2$), with three different but evolutionarily related types of subunits (Fig. 6–39b). Fibrinogen is converted to fibrin, and thereby activated for blood clotting, by the proteolytic removal of 16 amino acid residues from the amino-terminal end of each α subunit and 14 amino acid residues from the amino terminus of each β subunit. Peptide removal is catalyzed by the serine protease **thrombin**. The newly exposed amino termini of the α and β subunits fit neatly into binding sites in the carboxyl-terminal globular portions of the γ and β subunits, respectively. The interactions are stabilized by covalent cross-links generated by the condensation of particular Lys residues in one subunit with Gln residues in another, catalyzed by a transglutaminase, **factor XIIIa**. The resulting cross-linked fiber, **fibrin**, helps to tie together a blood clot.

Fibrinogen activation to produce fibrin is the end point of not one but two parallel but intertwined regulatory cascades **(Fig. 6–40)**. One of these is referred to as the contact activation pathway ("contact" refers to interaction of key components of this system with anionic phospholipids presented on the surface of platelets at the site of a wound). As all components of this pathway are found in the blood plasma, it is also called the **intrinsic pathway**. The second path is the tissue factor or **extrinsic pathway**. A major component of this pathway, the protein **tissue factor (TF),** is not present in the bloodstream. Most of the protein factors in both pathways are designated by Roman numerals. Many of those factors are chymotrypsin-like serine proteases, with zymogen precursors that are synthesized in the liver and exported to the blood. Other factors are regulatory proteins that bind to the serine proteases and help to activate them.

Blood clotting begins with the activation of circulating **platelets**—specialized cell fragments that lack nuclei—at the site of a wound. Tissue damage causes

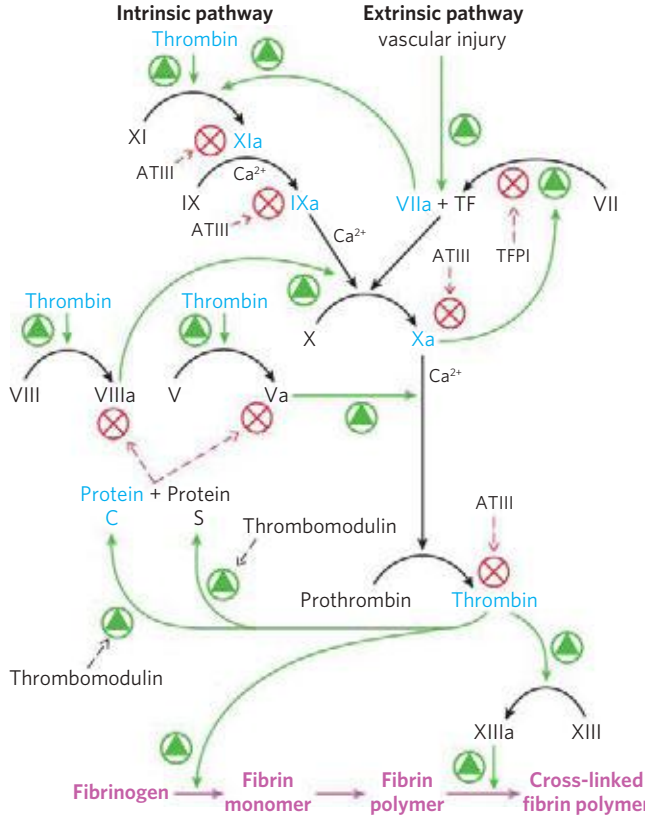


FIGURE 6–40 The coagulation cascades. The interlinked intrinsic and extrinsic pathways leading to the cleavage of fibrinogen to form active fibrin are shown. Active serine proteases in the pathways are shown in blue. Green arrows denote activating steps, and red arrows indicate inhibitory processes.

collagen molecules present beneath the epithelial cell layer that lines each blood vessel to become exposed to the blood. Platelet activation is primarily triggered by interaction with this collagen. Activation leads to the presentation of anionic phospholipids on the surface of each platelet and the release of signaling molecules such as **thromboxanes** (p. 371) that help stimulate the activation of additional platelets. The activated platelets aggregate at the site of a wound, forming a loose clot. Stabilization of the clot now requires the fibrin generated by the coagulation cascades.

The extrinsic pathway comes into play first. Tissue damage exposes the blood plasma to TF embedded largely in the membranes of fibroblasts and smooth muscle cells beneath the endothelial layer. An initiating complex is formed between TF and factor VII, present in the blood plasma. **Factor VII** is a zymogen of a serine protease, and TF is a regulatory protein that is required for its function. Factor VII is converted to its active form, **factor VIIa**, by proteolytic cleavage carried out by **factor Xa** (another serine protease). The TF-VIIa complex then cleaves **factor X**, creating the active form, factor Xa.

If TF-VIIa is needed to cleave X, and Xa is needed to cleave TF-VII, how does the process ever get started? A very small amount of factor VIIa is present in the blood at all times, enough to form a small amount of the active TF-VIIa complex immediately after tissue is damaged. This allows formation of factor Xa and establishes the initiating feedback loop. Once levels of factor Xa begin to build up, Xa (in a complex with regulatory protein factor Va) cleaves prothrombin to form active thrombin, and thrombin cleaves fibrinogen.

The extrinsic pathway thus provides a burst of thrombin. However, the TF-VIIa complex is rather quickly shut down by the protein **tissue factor protein inhibitor (TFPI)**. Clot formation is sustained by the activation of components of the intrinsic pathway. **Factor IX** is converted to the active serine protease **factor IXa** by the TF-VIIa protease during initiation of the clotting sequence. Factor IXa, in a complex with the regulatory protein **VIIIa**, is relatively stable and provides an alternative enzyme for the proteolytic conversion of factor X to Xa. Activated IXa can also be produced by the serine protease factor XIa. Most of the XIa is generated by cleavage of **factor XI** zymogen by thrombin in a feedback loop.

Left uncontrolled, blood coagulation could eventually lead to blockage of blood vessels, causing heart attacks or strokes. More regulation is thus needed. As a hard clot forms, regulatory pathways are already acting to limit the time during which the coagulation cascade is active. In addition to cleaving fibrinogen, thrombin also forms a complex with a protein embedded in the vascular surface of endothelial cells, **thrombomodulin**. The thrombin-thrombomodulin complex cleaves the serine protease zymogen **protein C**. Activated protein C, in a complex with the regulatory **protein S**, cleaves and inactivates factors Va and VIIIa, leading to suppression of the overall cascade. Another protein, **antithrombin III (ATIII)**, is a serine protease inhibitor. ATIII makes a covalent 1:1 complex between an Arg residue on ATIII and the active-site Ser residue of serine proteases, particularly thrombin and factor Xa. These two regulatory systems, in concert with TFPI, help to establish a threshold or level of exposure to TF that is needed to activate the coagulation cascade. Individuals with genetic defects that eliminate or decrease levels of protein C or ATIII in the blood have a greatly elevated risk of thrombosis (inappropriate formation of blood clots).

The control of blood coagulation has important roles in medicine, particularly in the prevention of blood clotting during surgery and in patients at risk for heart attacks or strokes. Several different medical approaches to anticoagulation are available. The first takes advantage of another feature of several proteins in the coagulation cascade that we have not yet considered. The factors VII, IX, X, and prothrombin, along with proteins C and S, have calcium-binding sites that are critical to their function. In each case, the calcium-binding sites are formed by modification of multiple Glu residues near the amino terminus of each protein to **γ-carboxyglutamate** residues (abbreviated **Gla**; p. 81). The Glu-to-Gla modifications are carried out by enzymes that depend on the function of the fat-soluble vitamin K (p. 374). Bound calcium functions to adhere these proteins to the anionic phospholipids that appear on the surface of activated platelets, effectively localizing the coagulation factors to the areas where the clot is to form. Vitamin K antagonists such as **warfarin** (Coumadin) have proven highly effective as anticoagulants. A second approach to anticoagulation is the administration of heparins. **Heparins** are highly sulfated polysaccharides (see Figs 7–22 and 7–23). They act as anticoagulants by increasing the affinity of ATIII for factor Xa and thrombin, thus facilitating the inactivation of key cascade elements. Finally, **aspirin** (acetylsalicylic acid; p. 845) is effective as an anticoagulant. Aspirin inhibits the enzyme cyclooxygenase, required for the production of thromboxanes. As aspirin reduces thromboxane release from platelets, the capacity of the platelets to aggregate declines.

Humans born with a deficiency in most components of the clotting cascade exhibit a tendency to bleed that varies from mild to essentially uncontrollable, a fatal condition. Genetic defects in genes encoding proteins required for blood clotting result in diseases referred to as hemophilias. Hemophilia A is a sex-linked trait resulting from a deficiency in factor VIII. This is the most common human hemophilia, affecting about one in 5,000 males worldwide. The most famous example of hemophilia A occurred among European royalty. Queen Victoria (1819–1901) was evidently a carrier. Prince Leopold, her eighth child, suffered from hemophilia A and died at the age of 31 after a minor fall. At least two of her daughters were carriers and passed the defective gene to other royal families of Europe (**Fig. 6–41**). ■

Some Regulatory Enzymes Use Several Regulatory Mechanisms

Glycogen phosphorylase catalyzes the first reaction in a pathway that feeds stored glucose into energy-yielding carbohydrate metabolism (Chapters 14 and 15). This is an important metabolic pathway, and its regulation is correspondingly complex. Although the primary regulation of glycogen phosphorylase is through covalent modification, as outlined in Figure 6–36, glycogen phosphorylase is also modulated allosterically by AMP, which is an activator of phosphorylase *b*, and by glucose 6-phosphate and ATP, both inhibitors. In addition, the enzymes that add and remove the phosphoryl groups are themselves regulated by—and so the entire system is sensitive to—the levels of hormones that regulate blood sugar (**Fig. 6–42**; see also Chapters 15 and 23).

Other complex regulatory enzymes are found at key metabolic crossroads. Bacterial glutamine synthetase, which catalyzes a reaction that introduces reduced nitrogen into cellular metabolism (Chapter 22), is among the most complex regulatory enzymes known. It is regulated allosterically (with at least eight different modulators); by reversible covalent modification; and by the association of other regulatory proteins, a mechanism examined in detail when we consider the regulation of specific metabolic pathways.

What is the advantage of such complexity in the regulation of enzymatic activity? We began this chapter by stressing the central importance of catalysis to the very existence of life. The *control* of catalysis is also critical to life. If all possible reactions in a cell were catalyzed simultaneously, macromolecules and metabolites would quickly be broken down to much simpler chemical forms. Instead, cells catalyze only the reactions they need at a given moment. When chemical resources are plentiful, cells synthesize and store glucose and other metabolites. When chemical resources are scarce, cells use these stores to fuel cellular metabolism. Chemical energy is used economically, parceled out to various metabolic pathways as cellular needs dictate. The availability of powerful catalysts, each specific for a given reaction, makes the regulation of these reactions possible. This in turn gives rise to the complex, highly regulated symphony we call life.

SUMMARY 6.5 Regulatory Enzymes

- The activities of metabolic pathways in cells are regulated by control of the activities of certain enzymes.
- The activity of an allosteric enzyme is adjusted by reversible binding of a specific modulator to a regulatory site. A modulator may be the substrate itself or some other metabolite, and the effect of the modulator may be inhibitory or stimulatory. The kinetic behavior of allosteric enzymes reflects cooperative interactions among enzyme subunits.
- Other regulatory enzymes are modulated by covalent modification of a specific functional group necessary for activity. The phosphorylation of specific amino acid residues is a particularly common way to regulate enzyme activity.

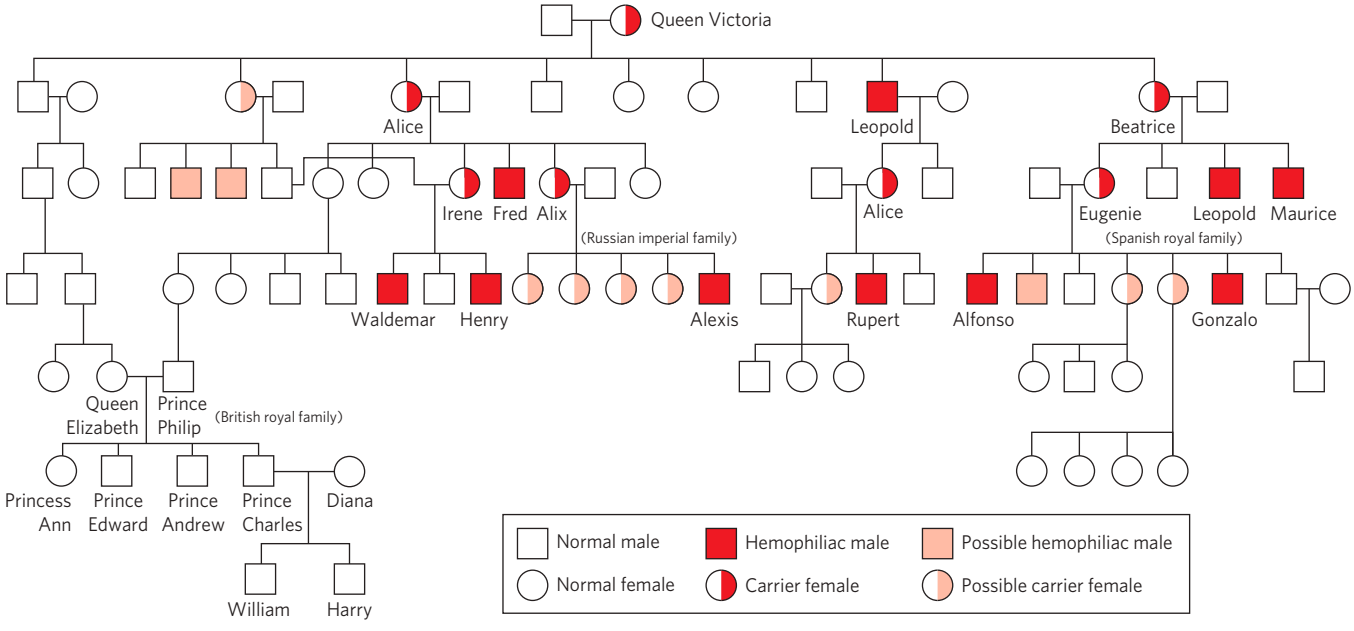


FIGURE 6–41 The royal families of Europe and inheritance of hemophilia A. Males are indicated by squares and females by circles. Males who suffered from hemophilia are represented by red squares, and presumed female carriers by half-red circles.

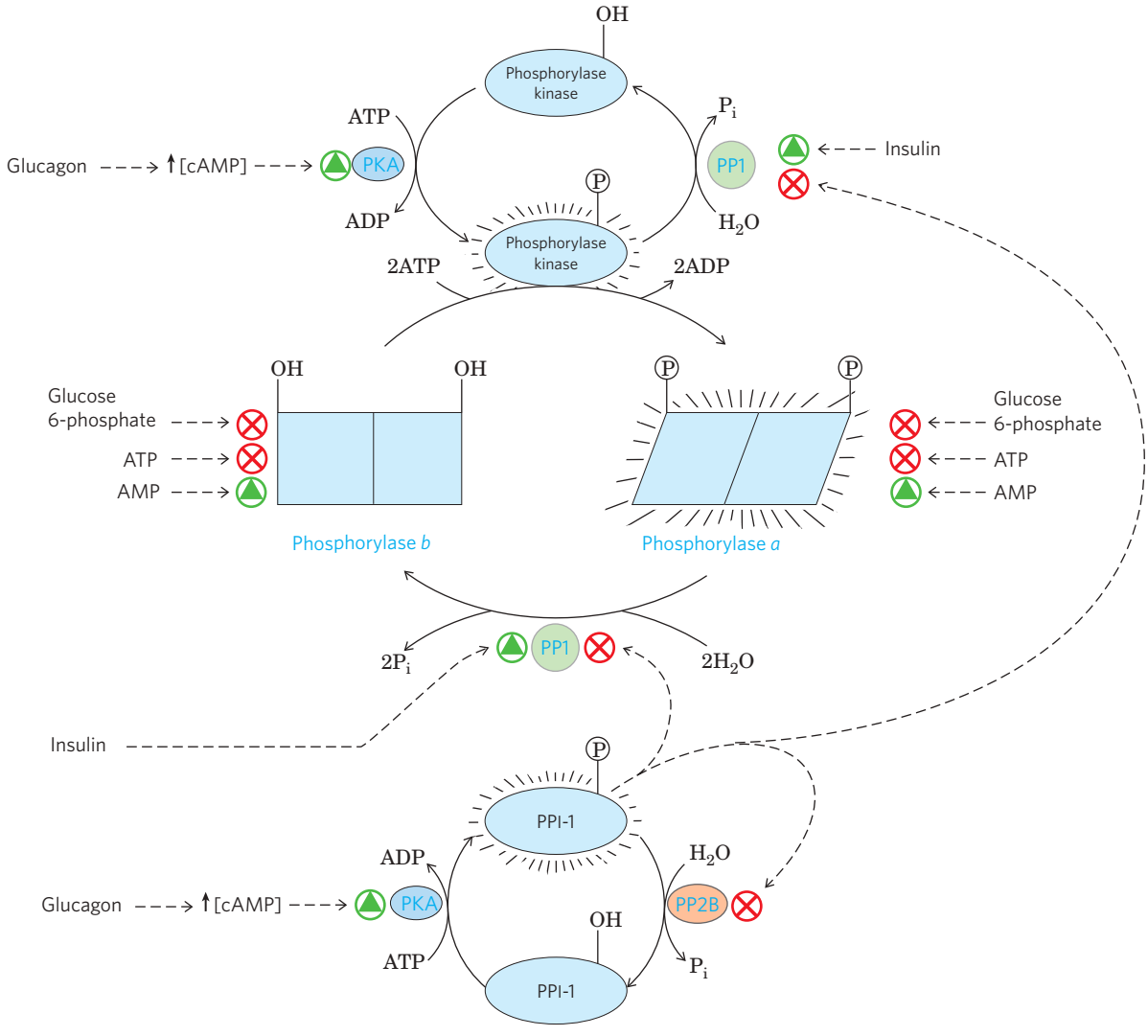


FIGURE 6–42 Regulation of muscle glycogen phosphorylase activity by phosphorylation. The activity of glycogen phosphorylase in muscle is subjected to a multilevel system of regulation involving much more than the covalent modification (phosphorylation) shown in Figure 6–36. Allosteric regulation, and a regulatory cascade sensitive to hormonal status that acts on the enzymes involved in phosphorylation and dephosphorylation, also play important roles. The activity of both forms of the enzyme is allosterically regulated by an activator (AMP) and by inhibitors (glucose 6-phosphate and ATP) that bind to separate sites on the enzyme. The activities of phosphorylase kinase and phosphorylase phosphatase 1 (PP1) are also regulated by covalent modification, via a short pathway that responds to the hormones glucagon and epinephrine. One path leads to the phosphorylation of phosphorylase kinase and phosphoprotein phosphatase inhibitor 1 (PPI-1). The phosphorylated phosphorylase kinase is activated and in turn phosphorylates and activates glycogen phosphorylase. At the same time, the phosphorylated PPI-1 interacts with and inhibits PP1. PPI-1 also keeps itself active (phosphorylated) by inhibiting phosphoprotein phosphatase 2B (PP2B), the enzyme that dephosphorylates (inactivates) it. In this way, the equilibrium between the *a* and *b* forms of glycogen phosphorylase is shifted decisively toward the more active glycogen phosphorylase *a*. Note that the two forms of phosphorylase kinase are both activated to a degree by Ca^{2+} ion (not shown). This pathway is discussed in more detail in Chapters 14, 15, and 23.

- Many proteolytic enzymes are synthesized as inactive precursors called zymogens, which are activated by cleavage of small peptide fragments.
- Blood clotting is mediated by two interlinked regulatory cascades of proteolytically activated zymogens.
- Enzymes at important metabolic intersections may be regulated by complex combinations of effectors, allowing coordination of the activities of interconnected pathways.

Key Terms

Terms in bold are defined in the glossary.

enzyme 190
cofactor 190
coenzyme 190
prosthetic group 190
holoenzyme 190
apoenzyme 190
apoprotein 190
active site 192
substrate 192
ground state 192
transition state 193
activation energy ($\Delta G^{\ddagger}$) 193
reaction intermediate 193

Further Reading

General

Evolution of Catalytic Function. (1987) *Cold Spring Harb. Symp. Quant. Biol.* **52**.

A collection of excellent papers on fundamentals; continues to be very useful.

Fersht, A. (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, W. H. Freeman and Company, New York.

A clearly written, concise introduction. More advanced.

Frey, P.A. & Hegeman, A.D. (2006) *Enzymatic Reaction Mechanisms*, Oxford University Press, New York.

An authoritative and up-to-date resource on the reactions that occur in living systems.

Jencks, W.P. (1987) *Catalysis in Chemistry and Enzymology*, Dover Publications, Inc., New York.

An outstanding book on the subject. More advanced.

Kornberg, A. (1989) *For the Love of Enzymes: The Odyssey of a Biochemist*, Harvard University Press, Cambridge, MA.

Principles of Catalysis

Amyes, T.L., O'Donoghue, A.C., & Richard, J.P. (2001) Contribution of phosphate intrinsic binding energy to the enzymatic rate acceleration for triosephosphate isomerase. *J. Am. Chem. Soc.* **123**, 11,325–11,326.

Gutteridge, A. & Thornton, J.M. (2005) Understanding nature's catalytic toolkit. *Trends Biochem. Sci.* **30**, 622–629.

A nice discussion of where binding energy comes from and how it is used.

Hammes-Schiffer, S. & Benkovic, S.J. (2006) Relating protein motion to catalysis. *Annu. Rev. Biochem.* **75**, 519–541.

A good description of the importance of protein motions in catalysis.

Hansen, D.E. & Raines, R.T. (1990) Binding energy and enzymatic catalysis. *J. Chem. Educ.* **67**, 483–489.

A good place for the beginning student to acquire a better understanding of principles.

Harris, T.K. & Turner, G.J. (2002) Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. *IUBMB Life* **53**, 85–98.

Kraut, D.A., Carroll, K.S., & Herschlag, D. (2003) Challenges in enzyme mechanism and energetics. *Annu. Rev. Biochem.* **72**, 517–571.

A good summary of the principles of enzymatic catalysis as currently understood, and of what we still do not understand.

Schramm, V.L. (2011) Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annu. Rev. Biochem.* **80**, 703–732.

Williams, D.H. (2010) Enzyme catalysis from improved packing in their transition-state structures. *Curr. Opin. Chem. Biol.* **14**, 666–670.

Wolfenden, R. (2011) Benchmark reaction rates, the stability of biological molecules in water, and the evolution of catalytic power in enzymes. *Annu. Rev. Biochem.* **80**, 645–667.

Kinetics

Cleland, W.W. (2002) Enzyme kinetics: steady state. In *Encyclopedia of Life Sciences*, John Wiley & Sons, Inc./Wiley Interscience, www.els.net.

A clear and concise presentation of the basics.

Raines, R.T. & Hansen, D.E. (1988) An intuitive approach to steady-state kinetics. *J. Chem. Educ.* **65**, 757–759.

Enzyme Examples

Babbitt, P.C. & Gerlt, J.A. (1997) Understanding enzyme superfamilies: chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **27**, 30,591–30,594.

An interesting description of the evolution of enzymes with different catalytic specificities, and the use of a limited repertoire of protein structural motifs.

Babbitt, P.C., Hasson, M.S., Wedekind, J.E., Palmer, D.R.J., Barrett, W.C., Reed, G.H., Rayment, I., Ringe, D., Kenyon, G.L., & Gerlt, J.A. (1996) The enolase superfamily: a general strategy for enzyme-catalyzed abstraction of the α-protons of carboxylic acids. *Biochemistry* **35**, 16,489–16,501.

Kirby, A.J. (2001) The lysozyme mechanism sorted—after 50 years. *Nat. Struct. Biol.* **8**, 737–739.

A nice discussion of the catalytic power of enzymes and the principles underlying it.

Regulatory Enzymes

Changeux, J.-P. (2012) Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu. Rev. Biophys.* **41**, 83–113.

Ehrmann, M. & Clausen, T. (2004) Proteolysis as a regulatory mechanism. *Annu. Rev. Genet.* **38**, 709–724.

Hunter, T. & Plowman, G.D. (1997) The protein kinases of budding yeast: six score and more. *Trends Biochem. Sci.* **22**, 18–22.

Details of the variety of these important enzymes in a model eukaryote.

Johnson, L.N. & Barford, D. (1993) The effects of phosphorylation on the structure and function of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 199–232.

Problems

1. Keeping the Sweet Taste of Corn The sweet taste of freshly picked corn (maize) is due to the high level of sugar in the kernels. Store-bought corn (several days after picking) is not as sweet, because about 50% of the free sugar is converted to starch within one day of picking. To preserve the sweetness of fresh corn, the husked ears can be immersed in boiling water for a few minutes ("blanched"), then cooled in cold water. Corn processed in this way and stored in a freezer maintains its sweetness. What is the biochemical basis for this procedure?

2. Intracellular Concentration of Enzymes To approximate the actual concentration of enzymes in a bacterial cell, assume that the cell contains equal concentrations of 1,000 different enzymes in solution in the cytosol and that each protein has a molecular weight of 100,000. Assume also that the bacterial cell is a cylinder (diameter 1.0 μm, height 2.0 μm), that the cytosol (specific gravity 1.20) is 20% soluble protein by weight, and that the soluble protein consists entirely of enzymes. Calculate the *average* molar concentration of each enzyme in this hypothetical cell.

3. Rate Enhancement by Urease The enzyme urease enhances the rate of urea hydrolysis at pH 8.0 and 20°C by a factor of 10^{14}. If a given quantity of urease can completely hydrolyze a given quantity of urea in 5.0 min at 20°C and pH 8.0, how long would it take for this amount of urea to be hydrolyzed under the same conditions in the absence of urease? Assume that both reactions take place in sterile systems so that bacteria cannot attack the urea.

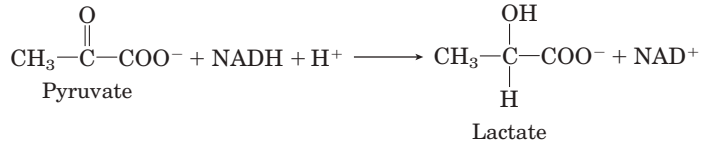
4. Protection of an Enzyme against Denaturation by Heat When enzyme solutions are heated, there is a progressive loss of catalytic activity over time due to denaturation of the enzyme. A solution of the enzyme hexokinase incubated at 45°C lost 50% of its activity in 12 min, but when incubated at 45°C in the presence of a very large concentration of one of its substrates, it lost only 3% of its activity in 12 min. Suggest why thermal denaturation of hexokinase was retarded in the presence of one of its substrates.

5. Requirements of Active Sites in Enzymes Carboxypeptidase, which sequentially removes carboxyl-terminal amino acid residues from its peptide substrates, is a single polypeptide of 307 amino acids. The two essential catalytic groups in the active site are furnished by Arg^{145} and Glu^{270}.

(a) If the carboxypeptidase chain were a perfect α helix, how far apart (in Å) would Arg^{145} and Glu^{270} be? (Hint: See Fig. 4–4a.)

(b) Explain how the two amino acid residues can catalyze a reaction occurring in the space of a few angstroms.

6. Quantitative Assay for Lactate Dehydrogenase The muscle enzyme lactate dehydrogenase catalyzes the reaction



NADH and NAD^+ are the reduced and oxidized forms, respectively, of the coenzyme NAD. Solutions of NADH, but *not* NAD^+, absorb light at 340 nm. This property is used to determine the concentration of NADH in solution by measuring spectrophotometrically the amount of light absorbed at 340 nm by the solution. Explain how these properties of NADH can be used to design a quantitative assay for lactate dehydrogenase.

7. Effect of Enzymes on Reactions Which of the listed effects would be brought about by any enzyme catalyzing the following simple reaction?

$$S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} P \qquad \text{where} \qquad K'_{eq} = \frac{[P]}{[S]}$$

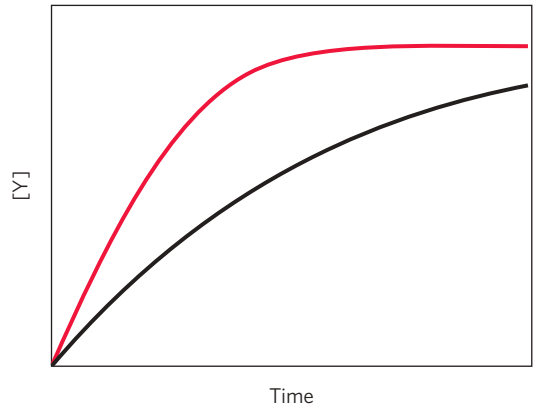
(a) Decreased K'_{eq}; (b) increased k_1; (c) increased K'_{eq}; (d) increased $\Delta G^{\ddagger}$; (e) decreased $\Delta G^{\ddagger}$; (f) more negative $\Delta G'^{\circ}$; (g) increased k_2.

8. Relation between Reaction Velocity and Substrate Concentration: Michaelis-Menten Equation

(a) At what substrate concentration would an enzyme with a k_{cat} of 30.0 s^{-1} and a K_m of 0.0050 M operate at one-quarter of its maximum rate?

(b) Determine the fraction of V_{max} that would be obtained at the following substrate concentrations [S]: $\frac{1}{2}K_m$, $2K_m$, and $10K_m$.

(c) An enzyme that catalyzes the reaction X ⇌ Y is isolated from two bacterial species. The enzymes have the same V_{max}, but different K_m values for the substrate X. Enzyme A has a K_m of 2.0 μM, while enzyme B has a K_m of 0.5 μM. The plot below shows the kinetics of reactions carried out with the same concentration of each enzyme and with [X] = 1 μM. Which curve corresponds to which enzyme?



9. Applying the Michaelis-Menten Equation I A research group discovers a new version of happyase, which they call happyase*, that catalyzes the chemical reaction

$$\text{HAPPY} \rightleftharpoons \text{SAD}$$

The researchers begin to characterize the enzyme.

(a) In the first experiment, with $[E_t]$ at 4 nM, they find that the V_{max} is 1.6 μM s^{-1}. Based on this experiment, what is the k_{cat} for happyase*? (Include appropriate units.)

(b) In another experiment, with $[E_t]$ at 1 nM and [HAPPY] at 30 μM, the researchers find that $V_0 = 300$ nM s^{-1}. What is the measured K_m of happyase* for its substrate HAPPY? (Include appropriate units.)

(c) Further research shows that the purified happyase* used in the first two experiments was actually contaminated with a reversible inhibitor called ANGER. When ANGER is carefully removed from the happyase* preparation and the two experiments repeated, the measured V_{max} in (a) is increased to 4.8 μM s^{-1}, and the measured K_m in (b) is now 15 μM. For the inhibitor ANGER, calculate the values of α and α'.

(d) Based on the information given above, what type of inhibitor is ANGER?

10. Applying the Michaelis-Menten Equation II An enzyme is found that catalyzes the reaction

$$\text{A} \rightleftharpoons \text{B}$$

Researchers find that the K_m for the substrate A is 4 μM, and the k_{cat} is 20 min^{-1}.

(a) In an experiment, [A] = 6 mM, and $V_0 = 480$ nM min^{-1}. What was the $[E_t]$ used in the experiment?

(b) In another experiment, $[E_t] = 0.5$ μM, and the measured $V_0 = 5$ μM min^{-1}. What was the [A] used in the experiment?

(c) The compound Z is found to be a very strong competitive inhibitor of the enzyme, with an α of 10. In an experiment with the same $[E_t]$ as in (a), but a different [A], an amount of Z is added that reduces V_0 to 240 nM min^{-1}. What is the [A] in this experiment?

(d) Based on the kinetic parameters given above, has this enzyme evolved to achieve catalytic perfection? Explain your answer briefly, using the kinetic parameter(s) that define catalytic perfection.

11. Estimation of V_{max} and K_m by Inspection Although graphical methods are available for accurate determination of the V_{max} and K_m of an enzyme-catalyzed reaction (see Box 6–1), sometimes these quantities can be quickly estimated by inspecting values of V_0 at increasing [S]. Estimate the V_{max} and K_m of the enzyme-catalyzed reaction for which the following data were obtained.

[S] (M)	V_0 (μM/min)
2.5×10^{-6}	28
4.0×10^{-6}	40
1×10^{-5}	70
2×10^{-5}	95
4×10^{-5}	112
1×10^{-4}	128
2×10^{-3}	139
1×10^{-2}	140

12. Properties of an Enzyme of Prostaglandin Synthesis Prostaglandins are a class of eicosanoids, fatty acid derivatives with a variety of extremely potent actions on vertebrate tissues. They are responsible for producing fever and inflammation and its associated pain. Prostaglandins are derived from the 20-carbon fatty acid arachidonic acid in a reaction catalyzed by the enzyme prostaglandin endoperoxide synthase. This enzyme, a cyclooxygenase, uses oxygen to convert arachidonic acid to PGG_2, the immediate precursor of many different prostaglandins (prostaglandin synthesis is described in Chapter 21).

(a) The kinetic data given below are for the reaction catalyzed by prostaglandin endoperoxide synthase. Focusing here on the first two columns, determine the V_{max} and K_m of the enzyme.

[Arachidonic acid] (mM)	Rate of formation of PGG_2 (mM/min)	Rate of formation of PGG_2 with 10 mg/mL ibuprofen (mM/min)
0.5	23.5	16.67
1.0	32.2	25.25
1.5	36.9	30.49
2.5	41.8	37.04
3.5	44.0	38.91

(b) Ibuprofen is an inhibitor of prostaglandin endoperoxide synthase. By inhibiting the synthesis of prostaglandins, ibuprofen reduces inflammation and pain. Using the data in the first and third columns of the table, determine the type of inhibition that ibuprofen exerts on prostaglandin endoperoxide synthase.

13. Graphical Analysis of V_{max} and K_m The following experimental data were collected during a study of the catalytic activity of an intestinal peptidase with the substrate glycylglycine:

$$\text{Glycylglycine} + H_2O \longrightarrow 2\ \text{glycine}$$

[S] (mM)	Product formed (μmol/min)
1.5	0.21
2.0	0.24
3.0	0.28
4.0	0.33
8.0	0.40
16.0	0.45

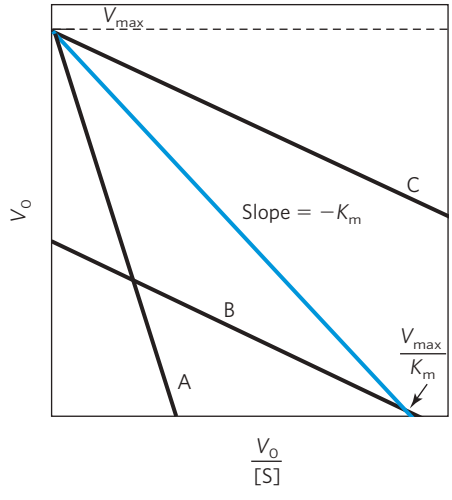
Use graphical analysis (see Box 6–1 and its associated Living Graph) to determine the K_m and V_{max} for this enzyme preparation and substrate.

14. The Eadie-Hofstee Equation There are several ways to transform the Michaelis-Menten equation so as to plot data and derive kinetic parameters, each with different advantages depending on the data set being analyzed. One transformation of the Michaelis-Menten equation is the Lineweaver-Burk, or

double-reciprocal, equation. Multiplying both sides of the Lineweaver-Burk equation by V_{max} and rearranging gives the Eadie-Hofstee equation:

$$V_0 = (-K_m)\frac{V_0}{[S]} + V_{max}$$

A plot of V_0 versus $V_0/[S]$ for an enzyme-catalyzed reaction is shown below. The blue curve was obtained in the absence of inhibitor. Which of the other curves (A, B, or C) shows the enzyme activity when a competitive inhibitor is added to the reaction mixture? Hint: See Equation 6–30.



15. The Turnover Number of Carbonic Anhydrase Carbonic anhydrase of erythrocytes (M_r 30,000) has one of the highest turnover numbers known. It catalyzes the reversible hydration of CO_2:

$$H_2O + CO_2 \rightleftharpoons H_2CO_3$$

This is an important process in the transport of CO_2 from the tissues to the lungs. If 10.0 μg of pure carbonic anhydrase catalyzes the hydration of 0.30 g of CO_2 in 1 min at 37 °C at V_{max}, what is the turnover number (k_{cat}) of carbonic anhydrase (in units of min^{-1})?

16. Deriving a Rate Equation for Competitive Inhibition The rate equation for an enzyme subject to competitive inhibition is

$$V_0 = \frac{V_{max}[S]}{\alpha K_m + [S]}$$

Beginning with a new definition of total enzyme as

$$[E_t] = [E] + [ES] + [EI]$$

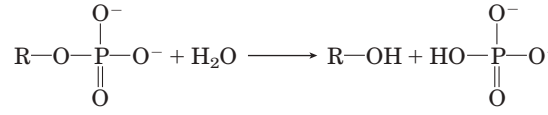
and the definitions of α and K_I provided in the text, derive the rate equation above. Use the derivation of the Michaelis-Menten equation as a guide.

17. Irreversible Inhibition of an Enzyme Many enzymes are inhibited irreversibly by heavy metal ions such as Hg^{2+}, Cu^{2+}, or Ag^+, which can react with essential sulfhydryl groups to form mercaptides:

$$\text{Enz–SH} + Ag^+ \longrightarrow \text{Enz–S–Ag} + H^+$$

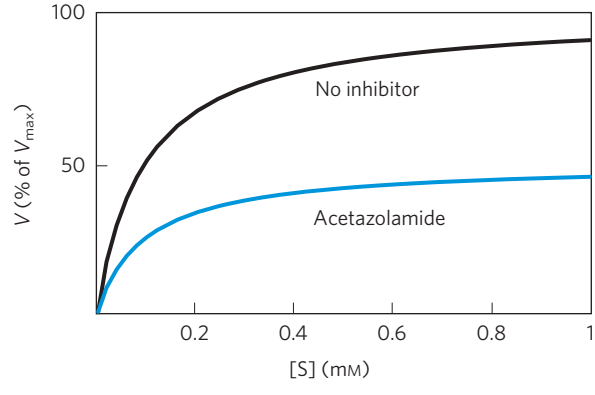
The affinity of Ag^+ for sulfhydryl groups is so great that Ag^+ can be used to titrate —SH groups quantitatively. To 10.0 mL of a solution containing 1.0 mg/mL of a pure enzyme, an investigator added just enough $AgNO_3$ to completely inactivate the enzyme. A total of 0.342 μmol of $AgNO_3$ was required. Calculate the minimum molecular weight of the enzyme. Why does the value obtained in this way give only the *minimum* molecular weight?

18. Clinical Application of Differential Enzyme Inhibition Human blood serum contains a class of enzymes known as acid phosphatases, which hydrolyze biological phosphate esters under slightly acidic conditions (pH 5.0):



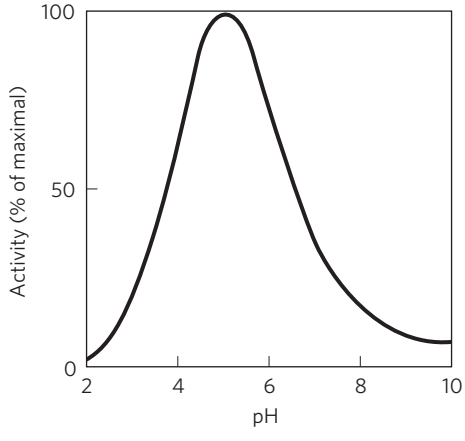
Acid phosphatases are produced by erythrocytes, the liver, kidney, spleen, and prostate gland. The enzyme of the prostate gland is clinically important, because its increased activity in the blood can be an indication of prostate cancer. The phosphatase from the prostate gland is strongly inhibited by tartrate ion, but acid phosphatases from other tissues are not. How can this information be used to develop a specific procedure for measuring the activity of the acid phosphatase of the prostate gland in human blood serum?

19. Inhibition of Carbonic Anhydrase by Acetazolamide Carbonic anhydrase is strongly inhibited by the drug acetazolamide, which is used as a diuretic (i.e., to increase the production of urine) and to lower excessively high pressure in the eye (due to accumulation of intraocular fluid) in glaucoma. Carbonic anhydrase plays an important role in these and other secretory processes, because it participates in regulating the pH and bicarbonate content of several body fluids. The experimental curve of initial reaction velocity (as percentage of V_{max}) versus [S] for the carbonic anhydrase reaction is illustrated below (upper curve). When the experiment is repeated in the presence of acetazolamide, the lower curve is obtained. From an inspection of the curves and your knowledge of the kinetic properties of competitive and mixed enzyme inhibitors, determine the nature of the inhibition by acetazolamide. Explain your reasoning.



20. The Effects of Reversible Inhibitors Derive the expression for the effect of a reversible inhibitor on observed K_m (apparent $K_m = \alpha K_m/\alpha'$). Start with Equation 6–30 and the statement that apparent K_m is equivalent to the [S] at which $V_0 = V_{max}/2\alpha'$.

21. pH Optimum of Lysozyme The active site of lysozyme contains two amino acid residues essential for catalysis: Glu^{35} and Asp^{52}. The pK_a values of the carboxyl side chains of these residues are 5.9 and 4.5, respectively. What is the ionization state (protonated or deprotonated) of each residue at pH 5.2, the pH optimum of lysozyme? How can the ionization states of these residues explain the pH-activity profile of lysozyme shown below?



22. Working with Kinetics Go to the Living Graphs for Chapter 6.

(a) Using the Living Graph for Equation 6–9, create a V versus [S] plot. Use $V_{max} = 100\ \mu M\ s^{-1}$, and $K_m = 10\ \mu M$. How much does V_0 increase when [S] is doubled, from 0.2 to 0.4 μM? What is V_0 when [S] = 10 μM? How much does the V_0 increase when [S] increases from 100 to 200 μM? Observe how the graph changes when the values for V_{max} or K_m are halved or doubled.

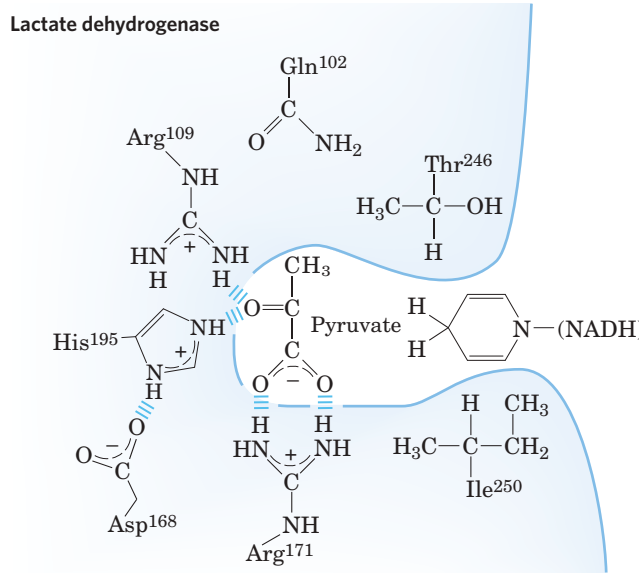
(b) Using the Living Graph for Equation 6–30 and the kinetic parameters in (a), create a plot in which both α and α' are 1.0. Now observe how the plot changes when $\alpha = 2.0$, when $\alpha' = 3.0$, and when $\alpha = 2.0$ and $\alpha' = 3.0$.

(c) Using the Living Graphs for Equation 6–30 and the Lineweaver-Burk equation in Box 6–1, create Lineweaver-Burk (double-reciprocal) plots for all the cases in (a) and (b). When $\alpha = 2.0$, does the x intercept move to the right or to the left? If $\alpha = 2.0$ and $\alpha' = 3.0$, does the x intercept move to the right or to the left?

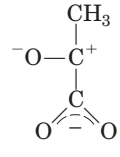
Data Analysis Problem

23. Exploring and Engineering Lactate Dehydrogenase Examining the structure of an enzyme results in hypotheses about the relationship between different amino acids in the protein's structure and the protein's function. One way to test these hypotheses is to use recombinant DNA technology to generate mutant versions of the enzyme and then examine the structure and function of these altered forms. The technology used to do this is described in Chapter 9.

One example of this kind of analysis is the work of A. R. Clarke and colleagues on the enzyme lactate dehydrogenase, published in 1989. Lactate dehydrogenase (LDH) catalyzes the reduction of pyruvate with NADH to form lactate (see Section 14.3). A schematic of the enzyme's active site is shown below; the pyruvate is in the center:



The reaction mechanism is similar to that of many NADH reductions (see Fig. 13–24); it is approximately the reverse of steps ❷ and ❸ of Figure 14–8. The transition state involves a strongly polarized carbonyl group of the pyruvate molecule, as shown below:



(a) A mutant form of LDH in which Arg^{109} is replaced with Gln shows only 5% of the pyruvate binding and 0.07% of the activity of wild-type enzyme. Provide a plausible explanation for the effects of this mutation.

(b) A mutant form of LDH in which Arg^{171} is replaced with Lys shows only 0.05% of the wild-type level of substrate binding. Why is this dramatic effect surprising?

(c) In the crystal structure of LDH, the guanidinium group of Arg^{171} and the carboxyl group of pyruvate are aligned as shown above in a co-planar "forked" configuration. Based on this, explain the dramatic effect of substituting Arg^{171} with Lys.

(d) A mutant form of LDH in which Ile^{250} is replaced with Gln shows reduced binding of NADH. Provide a plausible explanation for this result.

Clarke and colleagues also set out to engineer a mutant version of LDH that would bind and reduce oxaloacetate rather than pyruvate. They made a single substitution, replacing Gln^{102} with Arg; the resulting enzyme would reduce oxaloacetate to

malate and would no longer reduce pyruvate to lactate. They had therefore converted LDH to malate dehydrogenase.

(e) Sketch the active site of this mutant LDH with oxaloacetate bound.

(f) Why does this mutant enzyme now use oxaloacetate as a substrate instead of pyruvate?

(g) The authors were surprised that substituting a larger amino acid in the active site allowed a larger substrate to bind. Explain this result.

References

Clarke, A.R., Atkinson, T., & Holbrook, J.J. (1989) From analysis to synthesis: new ligand binding sites on the lactate dehydrogenase framework, Part I. *Trends Biochem. Sci.* **14**, 101–105.

Clarke, A.R., Atkinson, T., & Holbrook, J.J. (1989) From analysis to synthesis: new ligand binding sites on the lactate dehydrogenase framework, Part II. *Trends Biochem. Sci.* **14**, 145–148.

7

Carbohydrates and Glycobiology

Carbohydrates are the most abundant biomolecules on Earth. Each year, photosynthesis converts more than 100 billion metric tons of CO_2 and H_2O into cellulose and other plant products. Certain carbohydrates (sugar and starch) are a dietary staple in most parts of the world, and the oxidation of carbohydrates is the central energy-yielding pathway in most nonphotosynthetic cells. Carbohydrate polymers (also called glycans) serve as structural and protective elements in the cell walls of bacteria and plants and in the connective tissues of animals. Other carbohydrate polymers lubricate skeletal joints and participate in recognition and adhesion between cells. Complex carbohydrate polymers covalently attached to proteins or lipids act as signals that determine the intracellular destination or metabolic fate of these hybrid molecules, called **glycoconjugates**. This chapter introduces the major classes of carbohydrates and glycoconjugates and provides a few examples of their many structural and functional roles.

Carbohydrates are polyhydroxy aldehydes or ketones, or substances that yield such compounds on hydrolysis. Many, but not all, carbohydrates have the empirical formula $(CH_2O)_n$; some also contain nitrogen, phosphorus, or sulfur. There are three major size classes of carbohydrates: monosaccharides, oligosaccharides, and polysaccharides (the word "saccharide" is derived from the Greek *sakcharon*, meaning "sugar"). **Monosaccharides**, or simple sugars, consist of a single polyhydroxy aldehyde or ketone unit. The most abundant monosaccharide in nature is the six-carbon sugar D-glucose, sometimes referred to as dextrose. Monosaccharides of four or more carbons tend to have cyclic structures.

Oligosaccharides consist of short chains of monosaccharide units, or residues, joined by characteristic linkages called glycosidic bonds. The most abundant are the **disaccharides**, with two monosaccharide units. Typical is sucrose (cane sugar), which consists of the six-carbon sugars D-glucose and D-fructose. All common monosaccharides and disaccharides have names ending with the suffix "-ose." In cells, most oligosaccharides consisting of three or more units do not occur as free entities but are joined to nonsugar molecules (lipids or proteins) in glycoconjugates.

The **polysaccharides** are sugar polymers containing more than 20 or so monosaccharide units; some have hundreds or thousands of units. Some polysaccharides, such as cellulose, are linear chains; others, such as glycogen, are branched. Both glycogen and cellulose consist of recurring units of D-glucose, but they differ in the type of glycosidic linkage and consequently have strikingly different properties and biological roles.

7.1 Monosaccharides and Disaccharides

The simplest of the carbohydrates, the monosaccharides, are either aldehydes or ketones with two or more hydroxyl groups; the six-carbon monosaccharides glucose and fructose have five hydroxyl groups. Many of the carbon atoms to which hydroxyl groups are attached are chiral centers, which give rise to the many sugar stereoisomers found in nature. Stereoisomerism in sugars is biologically significant because the enzymes that act on sugars are strictly stereospecific, typically preferring one stereoisomer to another by three or more orders of magnitude, as reflected in K_m values or binding constants. It is as difficult to fit the wrong sugar stereoisomer into an enzyme's binding site as it is to put your left glove on your right hand.

We begin by describing the families of monosaccharides with backbones of three to seven carbons—their structure and stereoisomeric forms, and the means of representing their three-dimensional structures on paper. We then discuss several chemical reactions of the carbonyl groups of monosaccharides. One such reaction, the addition of a hydroxyl group from within the same molecule, generates cyclic forms having four or more backbone carbons (the forms that predominate in aqueous solution). This ring closure creates a new chiral center, adding further stereochemical complexity to this class of compounds. The nomenclature for unambiguously specifying the configuration about each carbon atom in a cyclic form and the means of representing these structures on paper are therefore described in some detail; this information will be useful as we discuss the metabolism of monosaccharides in Part II. We also introduce here some important monosaccharide derivatives encountered in later chapters.

The Two Families of Monosaccharides Are Aldoses and Ketoses

Monosaccharides are colorless, crystalline solids that are freely soluble in water but insoluble in nonpolar solvents. Most have a sweet taste (see Box 7–2, p. 254). The backbones of common monosaccharides are unbranched carbon chains in which all the carbon atoms are linked by single bonds. In this open-chain form, one of the carbon atoms is double-bonded to an oxygen atom to form a carbonyl group; each of the other carbon atoms has a hydroxyl group. If the carbonyl group is at an end of the carbon chain (that is, in an aldehyde group) the monosaccharide is an **aldose**; if the carbonyl group is at any other position (in a ketone group) the monosaccharide is a **ketose**. The simplest monosaccharides are the two three-carbon trioses: glyceraldehyde, an aldotriose, and dihydroxyacetone, a ketotriose **(Fig. 7–1a)**.

Monosaccharides with four, five, six, and seven carbon atoms in their backbones are called, respectively, tetroses, pentoses, hexoses, and heptoses. There are aldoses and ketoses of each of these chain lengths: aldotetroses and ketotetroses, aldopentoses and ketopentoses, and so on. The hexoses, which include the aldohexose D-glucose and the ketohexose D-fructose (Fig. 7–1b), are the most common monosaccharides in nature—the products of photosynthesis, and key intermediates in the central energy-yielding reaction sequence in most organisms. The aldopentoses D-ribose and 2-deoxy-D-ribose (Fig. 7–1c) are components of nucleotides and nucleic acids (Chapter 8).

Monosaccharides Have Asymmetric Centers

All the monosaccharides except dihydroxyacetone contain one or more asymmetric (chiral) carbon atoms and thus occur in optically active isomeric forms (pp. 17–18). The simplest aldose, glyceraldehyde, contains one chiral center (the middle carbon atom) and therefore has two different optical isomers, or **enantiomers (Fig. 7–2)**.

KEY CONVENTION: One of the two enantiomers of glyceraldehyde is, by convention, designated the D isomer, the other the L isomer. As for other biomolecules with chiral centers, the absolute configurations of sugars are known from x-ray crystallography. To represent three-dimensional sugar structures on paper, we often use **Fischer projection formulas** (Fig. 7–2). In Fischer projection formulas, horizontal bonds project out of the plane of the paper, toward the reader; vertical bonds project behind the plane of the paper, away from the reader. ■

In general, a molecule with n chiral centers can have 2^n stereoisomers. Glyceraldehyde has $2^1 = 2$; the aldohexoses, with four chiral centers, have $2^4 = 16$. The stereoisomers of monosaccharides of each carbon-chain

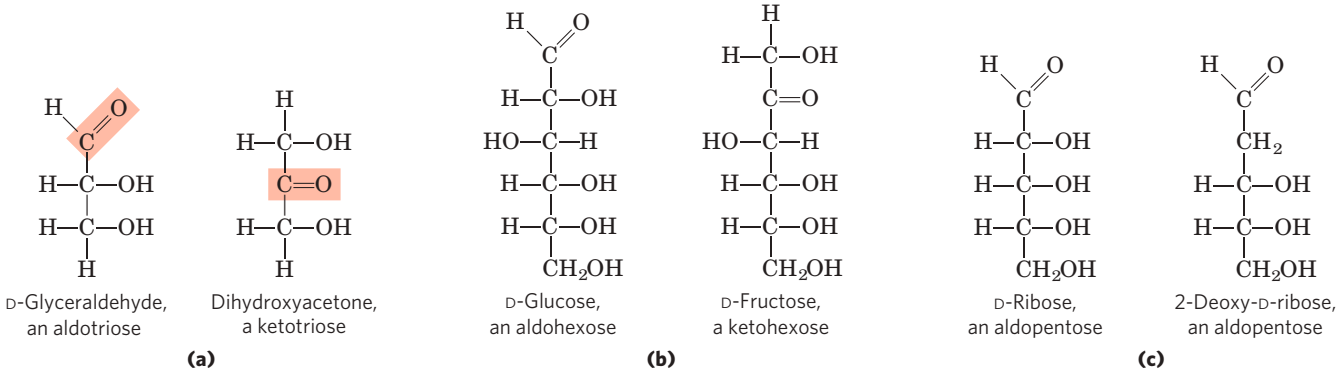


FIGURE 7–1 Representative monosaccharides. (a) Two trioses, an aldose and a ketose. The carbonyl group in each is shaded. **(b)** Two common hexoses. **(c)** The pentose components of nucleic acids. D-Ribose is a component of ribonucleic acid (RNA), and 2-deoxy-D-ribose is a component of deoxyribonucleic acid (DNA).

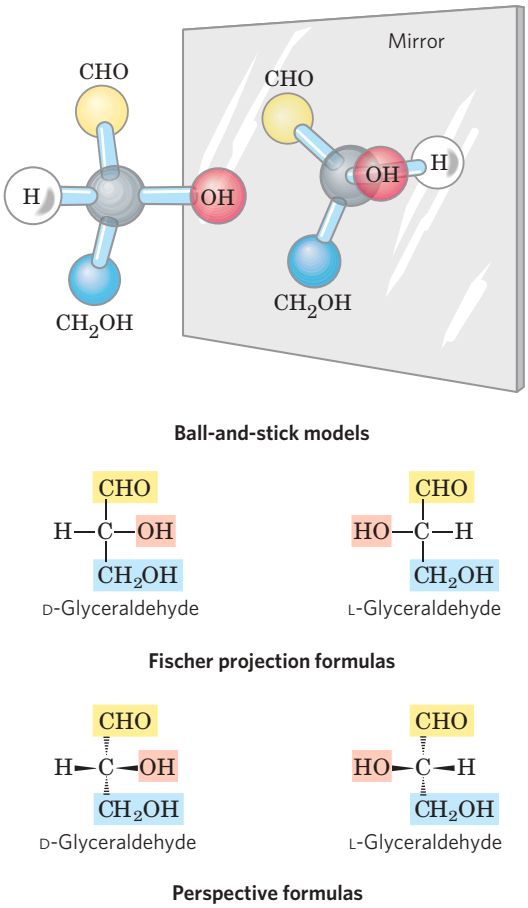
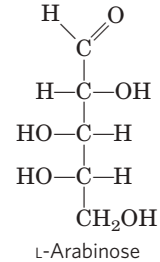


FIGURE 7–2 Three ways to represent the two enantiomers of glyceraldehyde. The enantiomers are mirror images of each other. Ball-and-stick models show the actual configuration of molecules. Recall (see Fig. 1–18) that in perspective formulas, the wide end of a solid wedge projects out of the plane of the paper, toward the reader; a dashed wedge extends behind.

length can be divided into two groups that differ in the configuration about the chiral center *most distant* from the carbonyl carbon. Those in which the configuration at this reference carbon is the same as that of D-glyceraldehyde are designated D isomers, and those with the same configuration as L-glyceraldehyde are L isomers. In other words, when the hydroxyl group on the reference carbon is on the right (*dextro*) in a projection formula that has the carbonyl carbon at the top, the sugar is the D isomer; when on the left (*levo*), it is the L isomer. Of the 16 possible aldohexoses, eight are D forms and eight are L. Most of the hexoses of living organisms are D isomers. Why D isomers? An interesting and unanswered question. Recall that all of the amino acids found in protein are exclusively one of two possible stereoisomers, L. The basis for this initial preference for one isomer during evolution is also unknown; however, once one isomer had been selected, it was likely that evolving enzymes would retain their preference for that stereoisomer (p. 78).

Figure 7–3 shows the structures of the D stereoisomers of all the aldoses and ketoses having three to six carbon atoms. The carbons of a sugar are numbered beginning at the end of the chain nearest the carbonyl group. Each of the eight D-aldohexoses, which differ in the stereochemistry at C-2, C-3, or C-4, has its own name: D-glucose, D-galactose, D-mannose, and so forth (Fig. 7–3a). The four- and five-carbon ketoses are designated by inserting "ul" into the name of a corresponding aldose; for example, D-ribulose is the ketopentose corresponding to the aldopentose D-ribose. (We will see the importance of ribulose when we discuss the fixation of atmospheric CO_2 by green plants, in Chapter 20.) The ketohexoses are named otherwise: for example, fructose (from the Latin *fructus*, "fruit"; fruits are one source of this sugar) and sorbose (from *Sorbus*, the genus of mountain ash, which has berries rich in the related sugar alcohol sorbitol). Two sugars that differ only in the configuration around one carbon atom are called **epimers**; D-glucose and D-mannose, which differ only in the stereochemistry at C-2, are epimers, as are D-glucose and D-galactose (which differ at C-4) **(Fig. 7–4)**.

Some sugars occur naturally in their L form; examples are L-arabinose and the L isomers of some sugar derivatives that are common components of glycoconjugates (Section 7.3).



L-Arabinose

The Common Monosaccharides Have Cyclic Structures

For simplicity, we have thus far represented the structures of aldoses and ketoses as straight-chain molecules (Figs 7–3, 7–4). In fact, in aqueous solution, aldotetroses and all monosaccharides with five or more carbon atoms in the backbone occur predominantly as cyclic (ring) structures in which the carbonyl group has formed a covalent bond with the oxygen of a hydroxyl group along the chain. The formation of these ring structures is the result of a general reaction between alcohols and aldehydes or ketones to form derivatives called **hemiacetals** or **hemiketals**. Two molecules of an alcohol can add to a carbonyl carbon; the product of the first addition is a hemiacetal (for addition to an aldose) or a hemiketal (for addition to a ketose). If the —OH and carbonyl groups are from the same molecule,

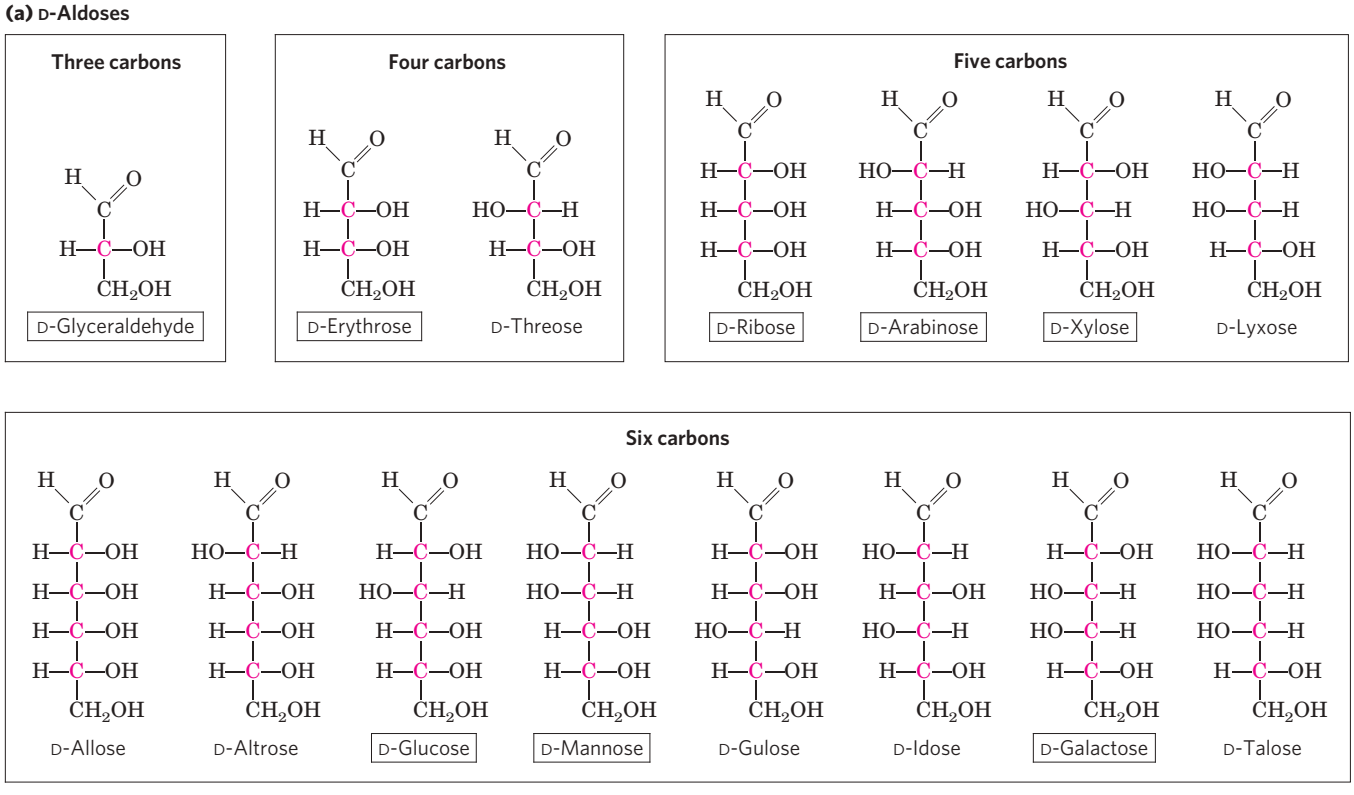


FIGURE 7–3 Aldoses and ketoses. The series of **(a)** D-aldoses and **(b)** D-ketoses having from three to six carbon atoms, shown as projection formulas. The carbon atoms in red are chiral centers. In all these D isomers, the chiral carbon *most distant from the carbonyl carbon* has the same configuration as the chiral carbon in D-glyceraldehyde. The sugars named in boxes are the most common in nature; you will encounter these again in this and later chapters.

a five- or six-membered ring results. Addition of the second molecule of alcohol produces the full acetal or ketal **(Fig. 7–5)**, and the bond formed is a glycosidic linkage. When the two molecules that react are both monosaccharides, the acetal or ketal formed is a disaccharide.

The reaction with the first molecule of alcohol creates an additional chiral center (the carbonyl carbon). Because the alcohol can add in either of two ways, attacking either the "front" or the "back" of the carbonyl carbon, the reaction can produce either of two stereoisomeric configurations, denoted α and β. For example, D-glucose exists in solution as an intramolecular hemiacetal in which the free hydroxyl group at C-5 has reacted with the aldehydic C-1, rendering the latter carbon asymmetric and producing two possible stereoisomers, designated α and β **(Fig. 7–6)**. Isomeric forms of monosaccharides that differ only in their configuration about the hemiacetal or hemiketal carbon atom are called **anomers**, and the carbonyl carbon atom is called the **anomeric carbon**.

Six-membered ring compounds are called **pyranoses** because they resemble the six-membered ring compound

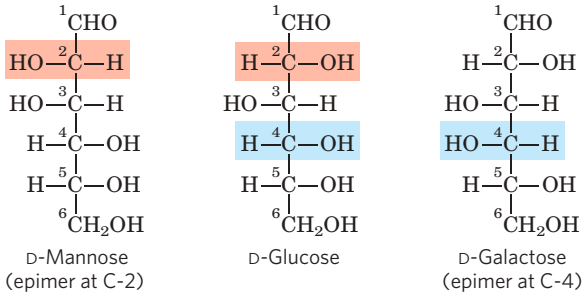


FIGURE 7–4 Epimers. D-Glucose and two of its epimers are shown as projection formulas. Each epimer differs from D-glucose in the configuration at one chiral center (shaded light red or blue).

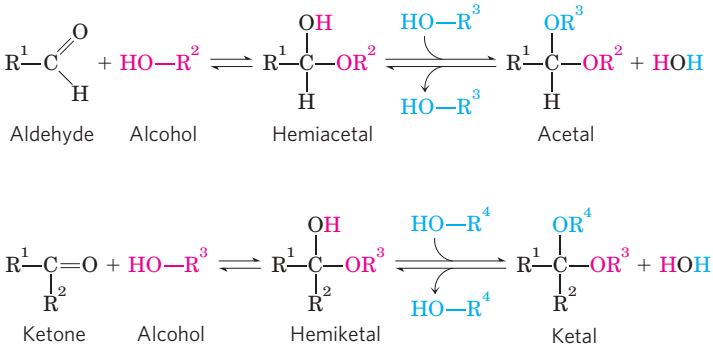


FIGURE 7–5 Formation of hemiacetals and hemiketals. An aldehyde or ketone can react with an alcohol in a 1:1 ratio to yield a hemiacetal or hemiketal, respectively, creating a new chiral center at the carbonyl carbon. Substitution of a second alcohol molecule produces an acetal or ketal. When the second alcohol is part of another sugar molecule, the bond produced is a glycosidic bond (p. 252).

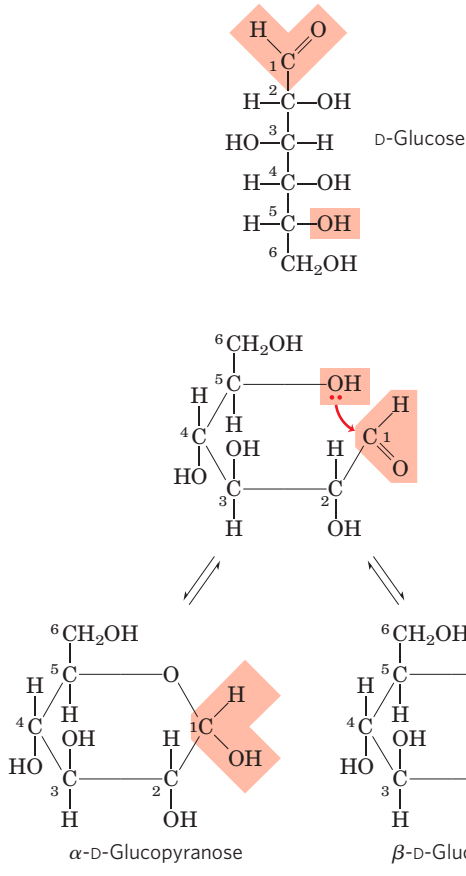
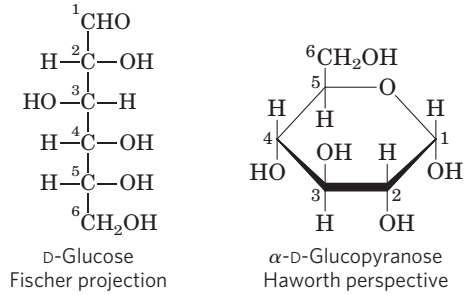


FIGURE 7–6 Formation of the two cyclic forms of D-glucose. Reaction between the aldehyde group at C-1 and the hydroxyl group at C-5 forms a hemiacetal linkage, producing either of two stereoisomers, the α and β anomers, which differ only in the stereochemistry around the hemiacetal carbon. This reaction is reversible. The interconversion of α and β anomers is called mutarotation.

pyran **(Fig. 7–7)**. The systematic names for the two ring forms of D-glucose are therefore α-D-glucopyranose and β-D-glucopyranose. Ketohexoses (such as fructose) also occur as cyclic compounds with α and β anomeric forms. In these compounds the hydroxyl group at C-5 (or C-6) reacts with the keto group at C-2, forming a **furanose** (or pyranose) ring containing a hemiketal linkage (Fig. 7–5). D-Fructose readily forms the furanose ring (Fig. 7–7); the more common anomer of this sugar in combined forms or in derivatives is β-D-fructofuranose.

Cyclic sugar structures are more accurately represented in **Haworth perspective formulas** than in the Fischer projections commonly used for linear sugar structures. In Haworth projections the six-membered ring is tilted to make its plane almost perpendicular to that of the paper, with the bonds closest to the reader drawn thicker than those farther away, as in Figure 7–7.

KEY CONVENTION: To convert the Fischer projection formula of any linear D-hexose to a Haworth perspective formula showing the molecule's cyclic structure, draw the six-membered ring (five carbons and one oxygen, at the upper right), number the carbons in a clockwise direction beginning with the anomeric carbon, then place the hydroxyl groups. If a hydroxyl group is to the right in the Fischer projection, it is placed pointing down (i.e., below the plane of the ring) in the Haworth perspective; if it is to the left in the Fischer projection, it is placed pointing up (i.e., above the plane) in the Haworth perspective. The terminal —CH_2OH group projects upward for the D-enantiomer, downward for the L-enantiomer. The hydroxyl on the anomeric carbon can point up or down. When the anomeric hydroxyl of a D-hexose is on the same side of the ring as C-6, the structure is by definition β; when it is on the opposite side from C-6, the structure is α. ■



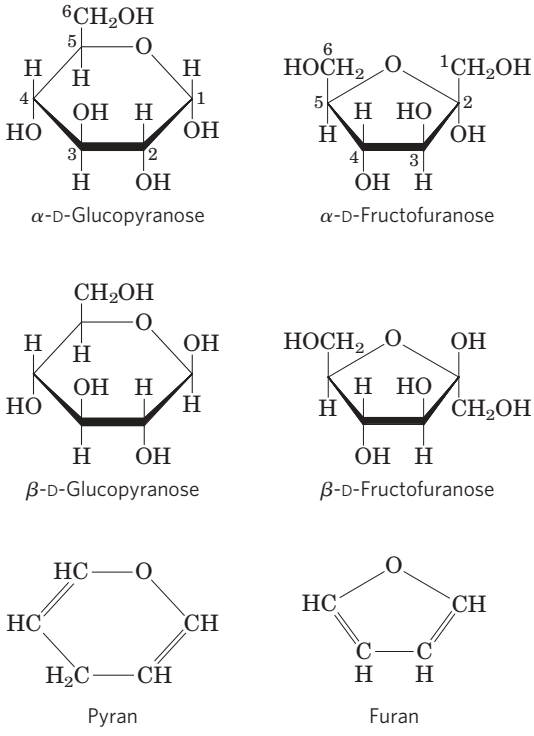
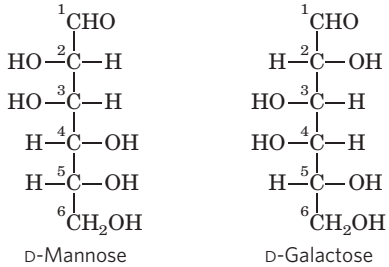


FIGURE 7–7 Pyranoses and furanoses. The pyranose forms of D-glucose and the furanose forms of D-fructose are shown here as Haworth perspective formulas. The edges of the ring nearest the reader are represented by bold lines. Hydroxyl groups below the plane of the ring in these Haworth perspectives would appear at the right side of a Fischer projection (compare with Fig. 7–6). Pyran and furan are shown for comparison.

WORKED EXAMPLE 7–1 Conversion of Fischer Projection to Haworth Perspective Formulas

Draw the Haworth perspective formulas for D-mannose and D-galactose.



Solution: Pyranoses are six-membered rings, so start with six-membered Haworth structures with the oxygen atom at the top right. Number the carbon atoms clockwise, starting with the aldose carbon. For mannose, place the hydroxyls on C-2, C-3, and C-4 above, above, and below the ring, respectively (because in the Fischer projection they are on the left, left, and right sides of the mannose structure). For D-galactose, the hydroxyls are oriented below, above, and above for C-2, C-3, and C-4, respectively. The hydroxyl at C-1 can be either up or down; there are two possible configurations, α and β, at this carbon.

WORKED EXAMPLE 7–2 Drawing Haworth Perspective Formulas of Sugar Isomers

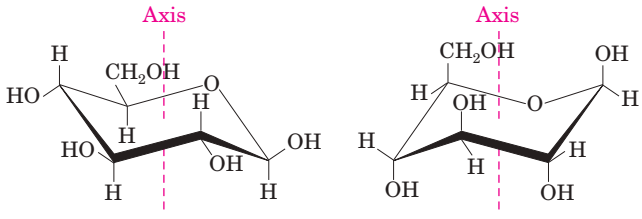
Draw the Haworth perspective formulas for α-D-mannose and β-L-galactose.

Solution: The Haworth perspective formula of D-mannose from Worked Example 7–1 can have the hydroxyl group at C-1 pointing either up or down. According to the Key Convention, for the α form, the C-1 hydroxyl is pointing down when C-6 is up, as it is in D-mannose.

For β-L-galactose, use the Fischer representation of D-galactose (see Worked Example 7–1) to draw the correct Fischer representation of L-galactose, which is its mirror image: the hydroxyls at C-2, C-3, C-4, and C-5 are on the left, right, right, and left sides, respectively. Now draw the Haworth perspective, a six-membered ring in which the —OH groups on C-2, C-3, and C-4 are oriented up, down, and down, respectively, because in the Fischer representation they are on the left, right, and right sides. Because it is the β form, the —OH on the anomeric carbon points down (same side as C-5).

The α and β anomers of D-glucose interconvert in aqueous solution by a process called **mutarotation**, in which one ring form (say, the α anomer) opens briefly into the linear form, then closes again to produce the β anomer (Fig. 7–6). Thus, a solution of β-D-glucose and a solution of α-D-glucose eventually form identical equilibrium mixtures having identical optical properties. This mixture consists of about one-third α-D-glucose, two-thirds β-D-glucose, and very small amounts of the linear and five-membered ring (glucofuranose) forms.

Haworth perspective formulas like those in Figure 7–7 are commonly used to show the stereochemistry of ring forms of monosaccharides. However, the six-membered pyranose ring is not planar, as Haworth perspectives suggest, but tends to assume either of two "chair" conformations **(Fig. 7–8)**. Recall from Chapter 1 (pp. 18–19) that two *conformations* of a molecule are interconvertible without the breakage of covalent bonds, whereas two *configurations* can be interconverted only by breaking a covalent bond. To interconvert α and β configurations, the bond involving the ring oxygen atom would have to be broken, but interconversion of the two chair forms (which are *conformers*) does not require bond breakage and does not change configurations at any of the ring carbons. The specific three-dimensional structures of the monosaccharide units are important in determining the biological properties and functions of some polysaccharides, as we shall see.



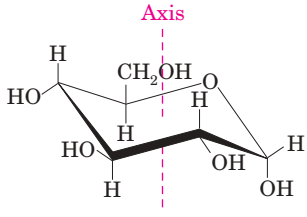


FIGURE 7–8 Conformational formulas of pyranoses. (a) Two chair forms of the pyranose ring of β-D-glucopyranose. Two *conformers* such as these are not readily interconvertible; an input of about 46 kJ of energy per mole of sugar is required to force the interconversion of chair forms. Another conformation, the "boat" (not shown), is seen only in derivatives with very bulky substituents. **(b)** The preferred chair conformation of α-D-glucopyranose.

Organisms Contain a Variety of Hexose Derivatives

In addition to simple hexoses such as glucose, galactose, and mannose, there are a number of sugar derivatives in which a hydroxyl group in the parent compound is replaced with another substituent, or a carbon atom is oxidized to a carboxyl group **(Fig. 7–9)**. In glucosamine, galactosamine, and mannosamine, the hydroxyl at C-2 of the parent compound is replaced with an amino group. The amino group is commonly condensed with acetic

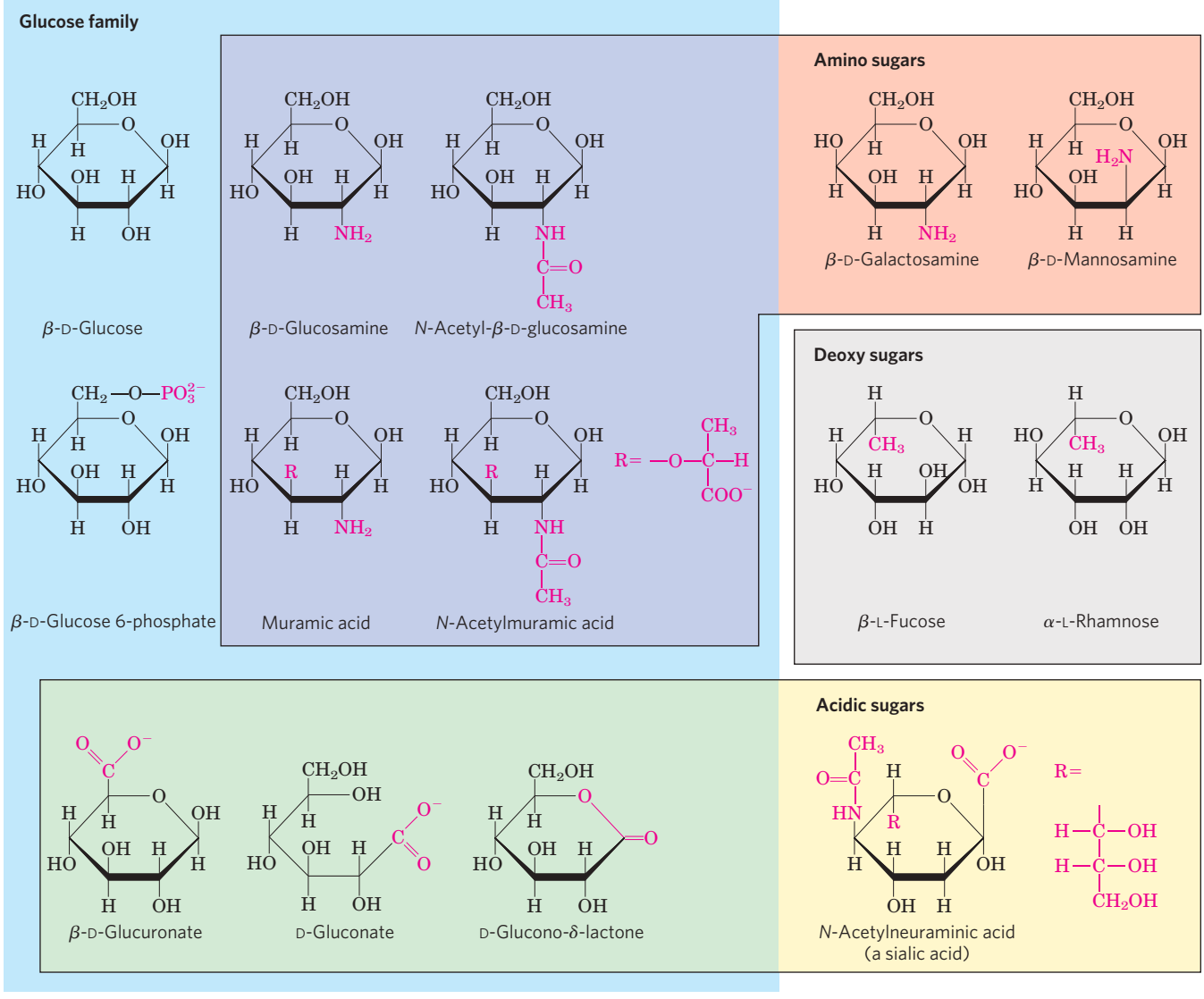


FIGURE 7–9 Some hexose derivatives important in biology. In amino sugars, an $—NH_2$ group replaces one of the —OH groups in the parent hexose. Substitution of —H for —OH produces a deoxy sugar; note that the deoxy sugars shown here occur in nature as the L isomers. The acidic sugars contain a carboxylate group, which confers a negative charge at neutral pH. D-Glucono-δ-lactone results from formation of an ester linkage between the C-1 carboxylate group and the C-5 (also known as the δ carbon) hydroxyl group of D-gluconate.

acid, as in *N*-acetylglucosamine. This glucosamine derivative is part of many structural polymers, including those of the bacterial cell wall. The substitution of a hydrogen for the hydroxyl group at C-6 of L-galactose or L-mannose produces L-fucose or L-rhamnose, respectively. L-Fucose is found in the complex oligosaccharide components of glycoproteins and glycolipids; L-rhamnose is found in plant polysaccharides.

Oxidation of the carbonyl (aldehyde) carbon of glucose to the carboxyl level produces gluconic acid, used in medicine as an innocuous counterion with which to administer positively charged drugs (such as quinine) or ions (such as Ca^{2+}). Other aldoses yield other **aldonic acids**. Oxidation of the carbon at the other end of the carbon chain—C-6 of glucose, galactose, or mannose—forms the corresponding **uronic acid**: glucuronic, galacturonic, or mannuronic acid. Both aldonic and uronic acids form stable intramolecular esters called lactones (Fig. 7–9, lower left). The sialic acids are a family of sugars with the same nine-carbon backbone. One of them, *N*-acetylneuraminic acid (often referred to simply as "sialic acid"), is a derivative of *N*-acetylmannosamine that occurs in many glycoproteins and glycolipids on animal cell surfaces, providing sites of recognition by other cells or extracellular carbohydrate-binding proteins. The carboxylic acid groups of the acidic sugar derivatives are ionized at pH 7, and the compounds are therefore correctly named as the carboxylates—glucuronate, galacturonate, and so forth.

In the synthesis and metabolism of carbohydrates, the intermediates are very often not the sugars

BOX 7–1 MEDICINE Blood Glucose Measurements in the Diagnosis and Treatment of Diabetes

Glucose is the principal fuel for the brain. When the amount of glucose reaching the brain is too low, the consequences can be dire: lethargy, coma, permanent brain damage, and death (see Fig. 23–24). Animals have evolved complex hormonal mechanisms to ensure that the concentration of glucose in the blood remains high enough (about 5 mM) to satisfy the brain's needs, but not too high, because elevated blood glucose can also have serious physiological consequences.

Individuals with insulin-dependent diabetes mellitus do not produce sufficient insulin, the hormone that normally serves to reduce blood glucose concentration, and if the diabetes is untreated their blood glucose levels may rise to severalfold higher than normal. These high glucose levels are believed to be at least one cause of the serious long-term consequences of untreated diabetes—kidney failure, cardiovascular disease, blindness, and impaired wound healing—so one goal of therapy is to provide just enough insulin (by injection) to keep blood glucose levels near normal. To maintain the correct balance of exercise, diet, and insulin for the individual, blood glucose concentration needs to be measured several times a day, and the amount of insulin injected adjusted appropriately.

The concentrations of glucose in blood and urine can be determined by a simple assay for reducing sugar, such as Fehling's reaction, which for many years was used as a diagnostic test for diabetes. Modern measurements require just a drop of blood, added to a test strip containing the enzyme glucose oxidase, which catalyzes the following reaction:

$$\text{D-Glucose} + O_2 \xrightarrow{\text{glucose oxidase}} \text{D-Glucono-}\delta\text{-lactone} + H_2O_2$$

A second enzyme, a peroxidase, catalyzes the reaction of the H_2O_2 with colorless compound to create a colored product, which is quantified with a simple photometer that reads out the blood glucose concentration.

Because blood glucose levels change with the timing of meals and exercise, single-time measurements do not reflect the *average* blood glucose over hours and days, so dangerous increases may go undetected. The average glucose concentration can be assessed by looking at its effect on hemoglobin, the oxygen-carrying protein in erythrocytes (p. 163). Transporters in the erythrocyte membrane equilibrate intracellular and plasma glucose concentrations, so hemoglobin is constantly exposed to glucose at whatever concentration is present in the blood. A nonenzymatic reaction occurs between glucose and primary amino groups in hemoglobin (either the amino-terminal Val or the ε-amino groups of Lys residues) (Fig. 1). The rate of this process is proportional to the concentration of glucose, so the reaction can be used as the basis for estimating the average blood glucose level over weeks. The amount of glycated hemoglobin (GHB) present at any time reflects the average blood glucose concentration over the circulating "lifetime" of the erythrocyte (about 120 days), although the concentration in the last two weeks is the most important in setting the level of GHB.

The extent of **hemoglobin glycation** (so named to distinguish it from glycosylation, the *enzymatic* transfer of glucose to a protein) is measured clinically by extracting hemoglobin from a small sample of blood and separating GHB from unmodified hemoglobin electrophoretically, taking advantage of the charge difference resulting from modification of the amino group(s). Normal GHB values are about 5% of total hemoglobin (corresponding to blood glucose of

themselves but their phosphorylated derivatives. Condensation of phosphoric acid with one of the hydroxyl groups of a sugar forms a phosphate ester, as in glucose 6-phosphate (Fig. 7–9), the first metabolite in the pathway by which most organisms oxidize glucose for energy. Sugar phosphates are relatively stable at neutral pH and bear a negative charge. One effect of sugar phosphorylation within cells is to trap the sugar inside the cell; most cells do not have plasma membrane transporters for phosphorylated sugars. Phosphorylation also activates sugars for subsequent chemical transformation. Several important phosphorylated derivatives of sugars are components of nucleotides (discussed in the next chapter).

Monosaccharides Are Reducing Agents

Monosaccharides can be oxidized by relatively mild oxidizing agents such as cupric (Cu^{2+}) ion. The carbonyl carbon is oxidized to a carboxyl group. Glucose and other sugars capable of reducing cupric ion are called **reducing sugars**. Cupric ion oxidizes glucose and certain other sugars to a complex mixture of carboxylic acids. This is the basis of Fehling's reaction, a semi-quantitative test for the presence of reducing sugar that for many years was used to detect and measure elevated glucose levels in people with diabetes mellitus. Today, more sensitive methods that involve an immobilized enzyme on a test strip are used; they require only a single drop of blood (Box 7–1). ■

120 mg/100 mL). In people with untreated diabetes, however, this value may be as high as 13%, indicating an average blood glucose level of about 300 mg/100 mL—dangerously high. One criterion for success in an individual program of insulin therapy (the timing, frequency, and amount of insulin injected) is maintaining GHB values at about 7%.

In the hemoglobin glycation reaction, the first step (formation of a Schiff base) is followed by a series of rearrangements, oxidations, and dehydrations of the carbohydrate moiety to produce a heterogeneous mixture of AGEs, *a*dvanced *g*lycation *e*nd products. These products can leave the erythrocyte and form covalent cross-links between proteins, interfering with normal protein function (Fig. 1). The accumulation of relatively high concentrations of AGEs in people with diabetes may, by cross-linking critical proteins, cause the damage to the kidneys, retinas, and cardiovascular system that characterizes the disease. This pathogenic process is a potential target for drug action.

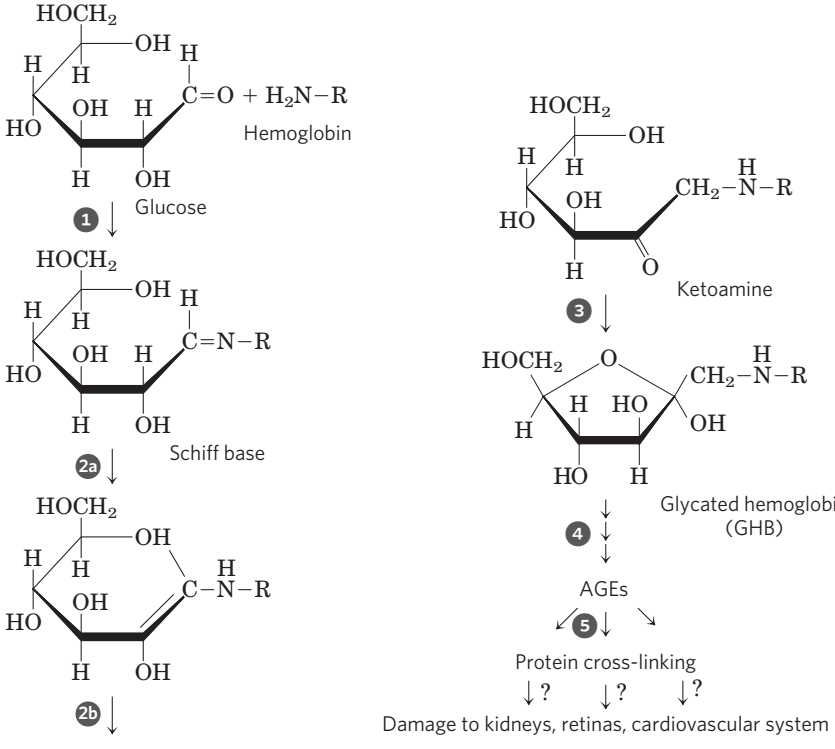


FIGURE 1 The nonenzymatic reaction of glucose with a primary amino group in hemoglobin begins with ❶ formation of a Schiff base, which ❷ undergoes a rearrangement to generate a stable product; ❸ this ketoamine can further cyclize to yield GHB. ❹ Subsequent reactions generate advanced glycation end products (AGEs), such as ε-*N*-carboxymethyllysine and methylglyoxal, compounds that ❺ can damage other proteins by cross-linking them, causing pathological changes.

Disaccharides Contain a Glycosidic Bond

Disaccharides (such as maltose, lactose, and sucrose) consist of two monosaccharides joined covalently by an ***O*-glycosidic bond**, which is formed when a hydroxyl group of one sugar molecule, typically cyclic, reacts with the anomeric carbon of the other **(Fig. 7–10)**. This reaction represents the formation of an acetal from a hemiacetal (such as glucopyranose) and an alcohol (a hydroxyl group of the second sugar molecule) (Fig. 7–5), and the resulting compound is called a glycoside. Glycosidic bonds are readily hydrolyzed by acid but resist cleavage by base. Thus disaccharides can be hydrolyzed to yield their free monosaccharide components by boiling with dilute acid. ***N*-glycosyl bonds** join the anomeric carbon of a sugar to a nitrogen atom in glycoproteins (see Fig. 7–30) and nucleotides (see Fig. 8–1).

The oxidation of a sugar by cupric ion (the reaction that defines a reducing sugar) occurs only with the linear form, which exists in equilibrium with the cyclic form(s). When the anomeric carbon is involved in a glycosidic bond (that is, when the compound is a full acetal or ketal; see Fig. 7–5), the easy interconversion of linear and cyclic forms shown in Figure 7–6 is prevented. Because the carbonyl carbon can be oxidized only when the sugar is in its linear form, formation of a glycosidic bond renders a sugar nonreducing. In describing disaccharides or polysaccharides, the end of a chain with a free anomeric carbon (one not involved in a glycosidic bond) is commonly called the **reducing end**.

The disaccharide maltose (Fig. 7–10) contains two D-glucose residues joined by a glycosidic linkage between C-1 (the anomeric carbon) of one glucose residue and C-4 of the other. Because the disaccharide retains a free anomeric carbon (C-1 of the glucose residue on the right in Fig. 7–10), maltose is a reducing sugar. The configuration of the anomeric carbon atom in the glycosidic linkage is α. The glucose residue with the free anomeric carbon is capable of existing in α- and β-pyranose forms.

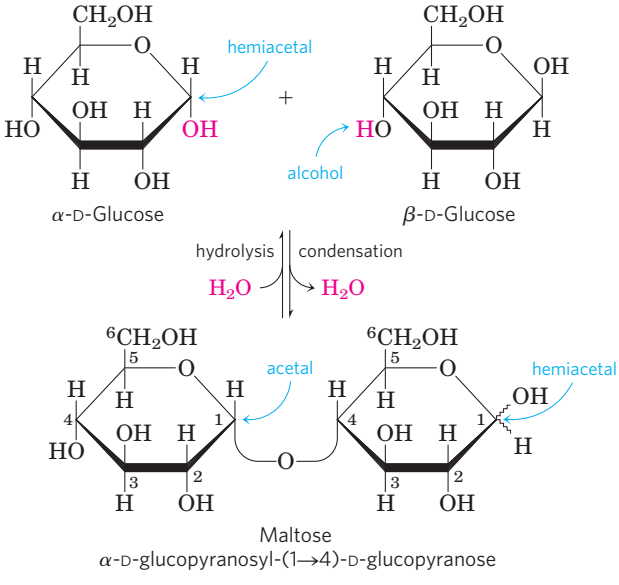


FIGURE 7–10 Formation of maltose. A disaccharide is formed from two monosaccharides (here, two molecules of D-glucose) when an —OH (alcohol) of one monosaccharide molecule (right) condenses with the intramolecular hemiacetal of the other (left), with elimination of H_2O and formation of a glycosidic bond. The reversal of this reaction is hydrolysis—attack by H_2O on the glycosidic bond. The maltose molecule shown here retains a reducing hemiacetal at the C-1 not involved in the glycosidic bond. Because mutarotation interconverts the α and β forms of the hemiacetal, the bonds at this position are sometimes depicted with wavy lines, as shown here, to indicate that the structure may be either α or β.

KEY CONVENTION: To name reducing disaccharides such as maltose unambiguously, and especially to name more complex oligosaccharides, several rules are followed. By convention, the name describes the compound written with its nonreducing end to the left, and we can "build up" the name in the following order. (1) Give the configuration (α or β) at the anomeric carbon joining the first monosaccharide unit (on the left) to the second. (2) Name the nonreducing residue; to distinguish five- and six-membered ring structures, insert "furano" or "pyrano" into the name. (3) Indicate in parentheses the two carbon atoms joined by the glycosidic bond, with an arrow connecting the two numbers; for example, (1→4) shows that C-1 of the first-named sugar residue is joined to C-4 of the second. (4) Name the second residue. If there is a third residue, describe the second glycosidic bond by the same conventions. (To shorten the description of complex polysaccharides, three-letter abbreviations or colored symbols for the monosaccharides are often used, as given in Table 7–1.) Following this convention for naming oligosaccharides, maltose is α-D-glucopyranosyl-(1→4)-D-glucopyranose. Because

TABLE 7–1 Symbols and Abbreviations for Common Monosaccharides and Some of Their Derivatives

Abequose	Abe	Glucuronic acid	GlcA
Arabinose	Ara	Galactosamine	GalN
Fructose	Fru	Glucosamine	GlcN
Fucose	Fuc	*N*-Acetylgalactosamine	GalNAc
Galactose	Gal	*N*-Acetylglucosamine	GlcNAc
Glucose	Glc	Iduronic acid	IdoA
Mannose	Man	Muramic acid	Mur
Rhamnose	Rha	*N*-Acetylmuramic acid	Mur2Ac
Ribose	Rib	*N*-Acetylneuraminic acid (a sialic acid)	Neu5Ac
Xylose	Xyl		

Note: In a commonly used convention, hexoses are represented as circles, *N*-acetylhexosamines as squares, and hexosamines as squares divided diagonally. All sugars with the "gluco" configuration are blue, those with the "galacto" configuration are yellow, and "manno" sugars are green. Other substituents can be added as needed: sulfate (S), phosphate (P), *O*-acetyl (OAc), or *O*-methyl (OMe).

most sugars encountered in this book are the D enantiomers and the pyranose form of hexoses predominates, we generally use a shortened version of the formal name of such compounds, giving the configuration of the anomeric carbon and naming the carbons joined by the glycosidic bond. In this abbreviated nomenclature, maltose is Glc(α1→4)Glc. ■

The disaccharide lactose **(Fig. 7–11)**, which yields D-galactose and D-glucose on hydrolysis, occurs naturally in milk. The anomeric carbon of the glucose residue is available for oxidation, and thus lactose is a reducing disaccharide. Its abbreviated name is Gal(β1→4)Glc. Sucrose (table sugar) is a disaccharide of glucose and fructose. It is formed by plants but not by animals. In contrast to maltose and lactose, sucrose contains no free anomeric carbon atom; the anomeric carbons of both monosaccharide units are involved in the glycosidic bond (Fig. 7–11). Sucrose is therefore a nonreducing sugar, and its stability toward oxidation makes it a suitable molecule for the storage and transport of energy in plants. In the abbreviated nomenclature, a double-headed arrow connects the symbols specifying the anomeric carbons and their configurations. For example, the abbreviated name of sucrose is either Glc(α1↔2β)Fru or Fru(β2↔1α)Glc. Sucrose is a major intermediate product of photosynthesis; in many plants it is the principal form in which sugar is transported from the leaves to other parts of the plant body. Trehalose, Glc(α1↔1α)Glc (Fig. 7–11)—a disaccharide of D-glucose that, like sucrose, is a nonreducing sugar—is a major constituent of the circulating fluid (hemolymph) of insects, serving as an energy-storage compound. Lactose gives milk its sweetness, and sucrose, of course, is table sugar. Trehalose is also used commercially as a sweetener. Box 7–2 explains how humans detect sweetness, and how artificial sweeteners such as aspartame act.

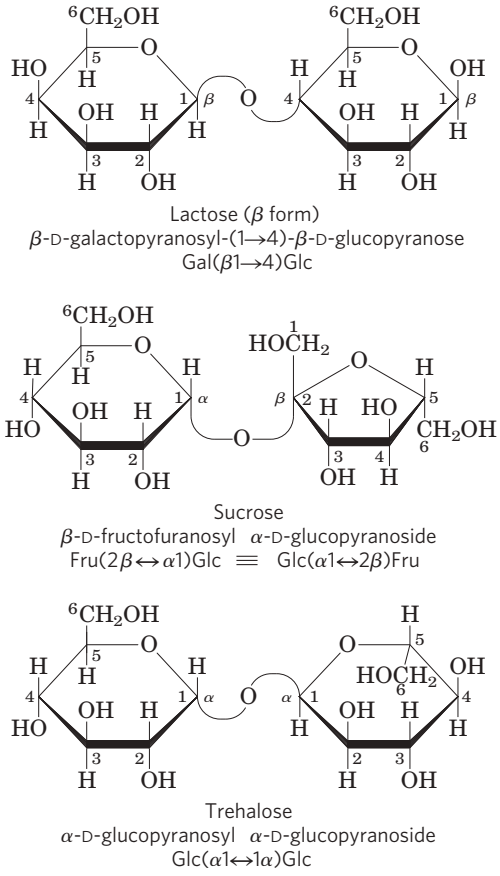


FIGURE 7–11 Two common disaccharides. Like maltose in Figure 7–10, these are shown as Haworth perspectives. The common name, full systematic name, and abbreviation are given for each disaccharide. Formal nomenclature for sucrose names glucose as the parent glycoside, although it is typically depicted as shown, with glucose on the left. The two abbreviated symbols shown for sucrose are equivalent (≡).

SUMMARY 7.1 Monosaccharides and Disaccharides

- Sugars (also called saccharides) are compounds containing an aldehyde or ketone group and two or more hydroxyl groups.
- Monosaccharides generally contain several chiral carbons and therefore exist in a variety of stereochemical forms, which may be represented on paper as Fischer projections. Epimers are sugars that differ in configuration at only one carbon atom.
- Monosaccharides commonly form internal hemiacetals or hemiketals, in which the aldehyde or ketone group joins with a hydroxyl group of the same molecule, creating a cyclic structure; this can be represented as a Haworth perspective formula. The carbon atom originally found in the aldehyde or ketone group (the anomeric carbon) can assume either of two configurations, α and β, which are interconvertible by mutarotation. In the linear form of the monosaccharide, which is in equilibrium with the cyclic forms, the anomeric carbon is easily oxidized, making the compound a reducing sugar.
- A hydroxyl group of one monosaccharide can add to the anomeric carbon of a second monosaccharide to form an acetal called a glycoside. In this disaccharide, the glycosidic bond protects the anomeric carbon from oxidation, making it a nonreducing sugar.
- Oligosaccharides are short polymers of several monosaccharides joined by glycosidic bonds. At one end of the chain, the reducing end, is a monosaccharide unit with its anomeric carbon not involved in a glycosidic bond.
- The common nomenclature for di- or oligosaccharides specifies the order of monosaccharide units, the configuration at each anomeric carbon, and the carbon atoms involved in the glycosidic linkage(s).

BOX 7–2 Sugar Is Sweet, and So Are . . . a Few Other Things

Sweetness is one of the five basic flavors that humans can taste (Fig. 1); the others are sour, bitter, salty, and umami. Sweet taste is detected by protein receptors in the plasma membranes of gustatory cells in the taste buds on the surface of the tongue. In humans, two closely related genes (*T1R2* and *T1R3*) encode sweetness receptors (Fig. 2). When a molecule with a compatible structure binds these receptors on a gustatory cell's extracellular domain, it triggers a series of events in the cell (including activation of a GTP-binding protein; see Fig. 12–42) that lead to an electrical signal being sent to the brain that is interpreted there as "sweet." During evolution, there has probably been selection for the ability to taste compounds found in foods containing important nutrients, such as the carbohydrates that are major fuels for most organisms. Most simple sugars, including sucrose, glucose, and fructose,

FIGURE 1 A strong stimulus for the sweetness receptors.

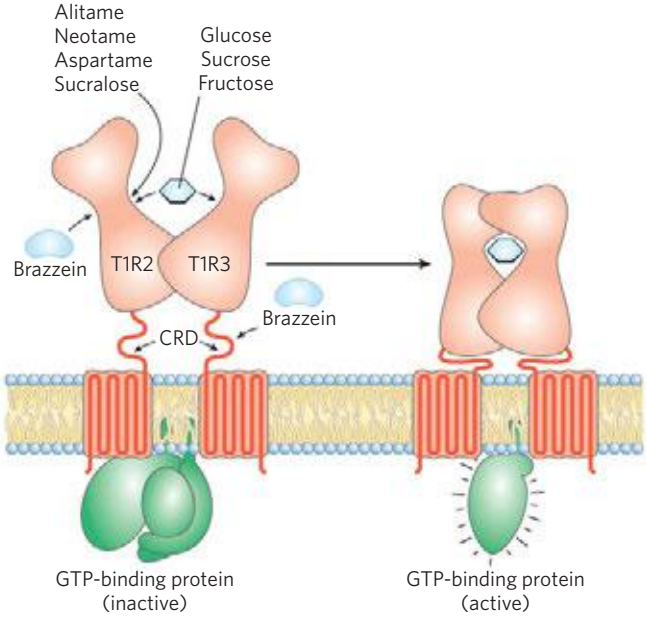


FIGURE 2 The receptor for sweet-tasting substances, with its regions of interaction (short arrows) with various sweet-tasting compounds indicated. Each receptor has an extracellular domain, a cysteine-rich domain (CRD), and a membrane domain with seven transmembrane helices, a common feature of signaling receptors. Artificial sweeteners bind to only one of the two receptor subunits; natural sugars bind to both. See Chapter 1, Problem 14, for the structures of many of these artificial sweeteners.

7.2 Polysaccharides

Most carbohydrates found in nature occur as polysaccharides, polymers of medium to high molecular weight (M_r >20,000). Polysaccharides, also called **glycans**, differ from each other in the identity of their recurring monosaccharide units, in the length of their chains, in the types of bonds linking the units, and in the degree of branching. **Homopolysaccharides** contain only a single monomeric species; **heteropolysaccharides** contain two or more different kinds **(Fig. 7–12)**. Some homopolysaccharides serve as storage forms of monosaccharides that are used as fuels; starch and glycogen are homopolysaccharides of this type. Other homopolysaccharides (cellulose and chitin, for example) serve as structural elements in plant cell walls and animal exoskeletons. Heteropolysaccharides provide extracellular support for organisms of all kingdoms. For example, the rigid layer of the bacterial cell envelope (the peptidoglycan) is composed in part of a heteropolysaccharide built from two alternating monosaccharide units (see Fig. 20–30). In animal tissues, the extracellular space is

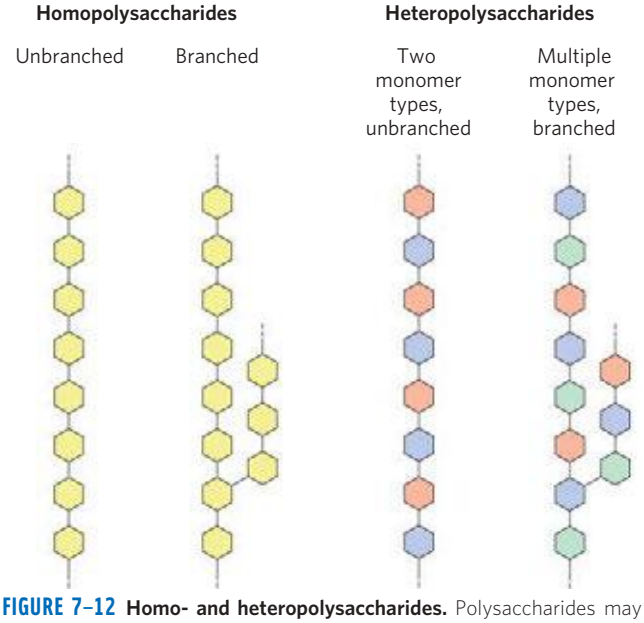


FIGURE 7–12 Homo- and heteropolysaccharides. Polysaccharides may be composed of one, two, or several different monosaccharides, in straight or branched chains of varying length.

taste sweet, but there are other classes of compounds that also bind the sweet receptors: the amino acids glycine, alanine, and serine are mildly sweet and harmless; nitrobenzene and ethylene glycol have a strong sweet taste, but are toxic. (See Box 18–2 for a remarkable medical mystery involving ethylene glycol poisoning.) Several natural products are extraordinarily sweet: stevioside, a sugar derivative isolated from the leaves of the stevia plant (*Stevia rebaudiana* Berton), is several hundred times sweeter than an equivalent amount of sucrose (table sugar), and the small (54 amino acids) protein brazzein, isolated from berries of the Oubli vine (*Pentadiplandra brazzeana* Baillon) in Gabon and Cameroon, is 17,000 times as sweet as sucrose on a molar basis. Presumably the sweet taste of the berries encourages their consumption by animals, which then disperse the seeds geographically so new plants may be established.

There is great interest in the development of artificial sweeteners as weight-reduction aids—compounds that give foods a sweet taste without adding the calories found in sugars. The artificial sweetener aspartame demonstrates the importance of stereochemistry in biology (Fig. 3). According to one simple model of sweetness receptor binding, binding involves three sites on the receptor: AH^+, B^-, and X. Site AH^+ contains some group (an alcohol or amine) that can form a hydrogen bond with a partial negative charge, such as a carbonyl oxygen, on the sweetener molecule; the carboxylic acid of aspartame contains such an oxygen. Site B^- contains a group with a partially negative oxygen available to hydrogen-bond with some partially positive atom on the sweetener molecule, such as the amine group of aspartame. Site X is oriented perpendicular to the other two groups and is capable of interacting with a hydrophobic patch on the sweetener molecule, such as the benzene ring of aspartame.

When the steric match is correct, as on the left in Figure 3, the sweet receptor is stimulated and the signal "sweet" is conducted to the brain. When the match is not correct, as on the right in Figure 3, the sweet receptor is not stimulated; in fact, in this case, another receptor (for bitterness) is stimulated by the "wrong" stereoisomer of aspartame. Stereoisomerism really matters!

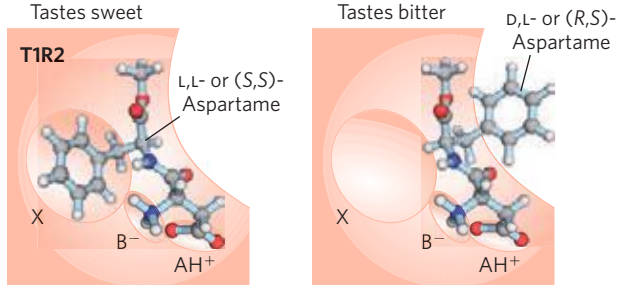


FIGURE 3 Stereochemical basis for the taste of two isomers of aspartame.

occupied by several types of heteropolysaccharides, which form a matrix that holds individual cells together and provides protection, shape, and support to cells, tissues, and organs.

Unlike proteins, polysaccharides generally do not have defining molecular weights. This difference is a consequence of the mechanisms of assembly of the two types of polymer. As we shall see in Chapter 27, proteins are synthesized on a template (messenger RNA) of defined sequence and length, by enzymes that follow the template exactly. For polysaccharide synthesis there is no template; rather, the program for polysaccharide synthesis is intrinsic to the enzymes that catalyze the polymerization of the monomeric units, and there is no specific stopping point in the synthetic process; the products thus vary in length.

Some Homopolysaccharides Are Stored Forms of Fuel

The most important storage polysaccharides are starch in plant cells and glycogen in animal cells. Both polysaccharides occur intracellularly as large clusters or granules. Starch and glycogen molecules are heavily hydrated, because they have many exposed hydroxyl groups available to hydrogen-bond with water. Most plant cells have the ability to form starch (see Fig. 20–2), and starch storage is especially abundant in tubers (underground stems), such as potatoes, and in seeds.

Starch contains two types of glucose polymer, amylose and amylopectin **(Fig. 7–13)**. Amylose consists of long, unbranched chains of D-glucose residues connected by (α1→4) linkages (as in maltose). Such chains vary in molecular weight from a few thousand to more than a million. Amylopectin also has a high molecular weight (up to 200 million) but unlike amylose is highly branched. The glycosidic linkages joining successive glucose residues in amylopectin chains are (α1→4); the branch points (occurring every 24 to 30 residues) are (α1→6) linkages.

Glycogen is the main storage polysaccharide of animal cells. Like amylopectin, glycogen is a polymer of (α1→4)-linked subunits of glucose, with (α1→6)-linked branches, but glycogen is more extensively branched (on average, every 8 to 12 residues) and more compact than

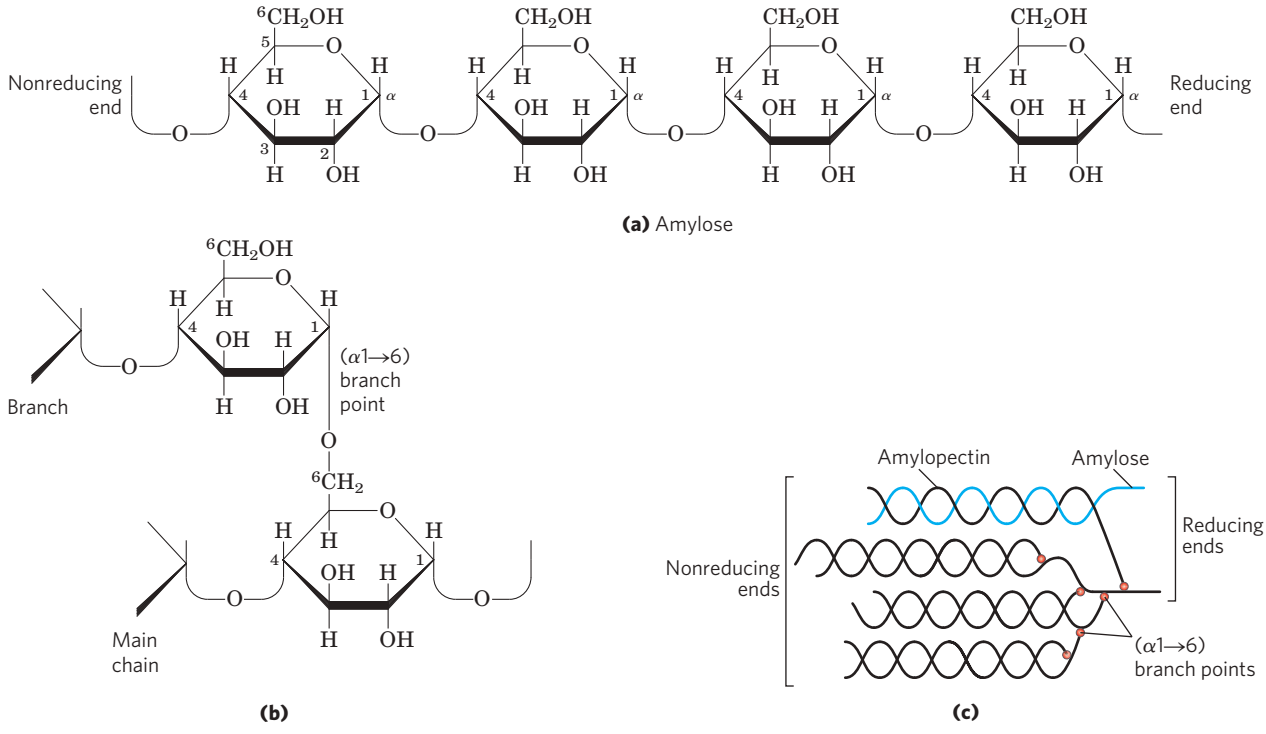


FIGURE 7–13 **Glycogen and starch.** **(a)** A short segment of amylose, a linear polymer of D-glucose residues in (α1⟶4) linkage. A single chain can contain several thousand glucose residues. Amylopectin has stretches of similarly linked residues between branch points. Glycogen has the same basic structure, but has more branching than amylopectin. **(b)** An (α1⟶6) branch point of glycogen or amylopectin. **(c)** A cluster of amylose and amylopectin like that believed to occur in starch granules. Strands of amylopectin (black) form double-helical structures with each other or with amylose strands (blue). Amylopectin has frequent (α1⟶6) branch points (red). Glucose residues at the nonreducing ends of the outer branches are removed enzymatically during the mobilization of starch for energy production. Glycogen has a similar structure but is more highly branched and more compact.

starch. Glycogen is especially abundant in the liver, where it may constitute as much as 7% of the wet weight; it is also present in skeletal muscle. In hepatocytes glycogen is found in large granules, which are themselves clusters of smaller granules composed of single, highly branched glycogen molecules with an average molecular weight of several million. Such glycogen granules also contain, in tightly bound form, the enzymes responsible for the synthesis and degradation of glycogen.

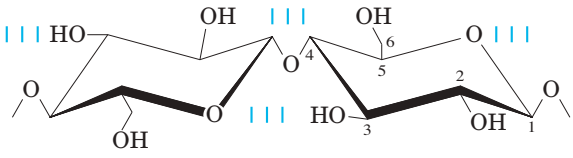
Because each branch in glycogen ends with a nonreducing sugar unit, a glycogen molecule with n branches has $n + 1$ nonreducing ends, but only one reducing end. When glycogen is used as an energy source, glucose units are removed one at a time from the nonreducing ends. Degradative enzymes that act only at nonreducing ends can work simultaneously on the many branches, speeding the conversion of the polymer to monosaccharides.

Why not store glucose in its monomeric form? It has been calculated that hepatocytes store glycogen equivalent to a glucose concentration of 0.4 M. The actual concentration of glycogen, which is insoluble and contributes little to the osmolarity of the cytosol, is about 0.01 μM. If the cytosol contained 0.4 M glucose, the osmolarity would be threateningly elevated, leading to osmotic entry of water that might rupture the cell (see Fig. 2–13). Furthermore, with an intracellular glucose concentration of 0.4 M and an external concentration of about 5 mM (the concentration in the blood of a mammal), the free-energy change for glucose uptake into cells against this very high concentration gradient would be prohibitively large.

Dextrans are bacterial and yeast polysaccharides made up of (α1⟶6)-linked poly-D-glucose; all have (α1⟶3) branches, and some also have (α1⟶2) or (α1⟶4) branches. Dental plaque, formed by bacteria growing on the surface of teeth, is rich in dextrans, which are adhesive and allow the bacteria to stick to teeth and to each other. Dextrans also provide a source of glucose for bacterial metabolism. Synthetic dextrans are used in several commercial products (for example, Sephadex) that serve in the fractionation of proteins by size-exclusion chromatography (see Fig. 3–17b). The dextrans in these products are chemically cross-linked to form insoluble materials of various sizes.

Some Homopolysaccharides Serve Structural Roles

Cellulose, a fibrous, tough, water-insoluble substance, is found in the cell walls of plants, particularly in stalks, stems, trunks, and all the woody portions of the plant body. Cellulose constitutes much of the mass of wood, and cotton is almost pure cellulose. Like amylose, the



(β1→4)-linked D-glucose units

FIGURE 7–14 Cellulose. Two units of a cellulose chain; the D-glucose residues are in (β1→4) linkage. The rigid chair structures can rotate relative to one another.

cellulose molecule is a linear, unbranched homopolysaccharide, consisting of 10,000 to 15,000 D-glucose units. But there is a very important difference: in cellulose the glucose residues have the β configuration **(Fig. 7–14)**, whereas in amylose the glucose is in the α configuration. The glucose residues in cellulose are linked by (β1→4) glycosidic bonds, in contrast to the (α1→4) bonds of amylose. This difference causes individual molecules of cellulose and amylose to fold differently in space, giving them very different macroscopic structures and physical properties (see below). The tough, fibrous nature of cellulose makes it useful in such commercial products as cardboard and insulation material, and it is a major constituent of cotton and linen fabrics. Cellulose is also the starting material for the commercial production of cellophane and rayon.

Glycogen and starch ingested in the diet are hydrolyzed by α-amylases and glycosidases, enzymes in saliva and the intestine that break (α1→4) glycosidic bonds between glucose units. Most vertebrate animals cannot use cellulose as a fuel source, because they lack an enzyme to hydrolyze the (β1→4) linkages. Termites readily digest cellulose (and therefore wood), but only because their intestinal tract harbors a symbiotic microorganism, *Trichonympha*, that secretes cellulase, which hydrolyzes the (β1→4) linkages **(Fig. 7–15)**. Molecular genetic studies have revealed that genes encoding cellulose-degrading enzymes are present in the genomes of a wide range of invertebrate animals, including arthropods and nematodes. There is one important exception to the absence of cellulase in vertebrates: ruminant animals such as cattle, sheep, and goats harbor symbiotic microorganisms in the rumen (the first of their four stomach compartments) that can hydrolyze cellulose, allowing the animal to degrade dietary cellulose from soft grasses, but not from woody plants. Fermentation in the rumen yields acetate, propionate, and β-hydroxybutyrate, which the animal uses to synthesize the sugars in milk (p. 560).

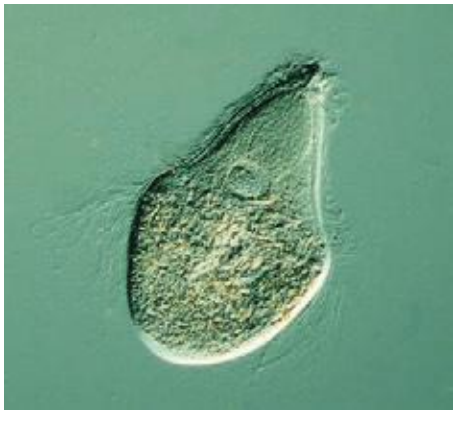


FIGURE 7–15 Cellulose breakdown by *Trichonympha*, a protist in the gut of a wood-eating termite. *Trichonympha* produces the enzyme cellulase, which breaks the (β1→4) glycosidic bonds in cellulose, making wood a source of metabolizable sugar (glucose) for the protist and the termite. A number of invertebrates can digest cellulose, but only a few vertebrates (the ruminants, such as cattle, sheep, and goats); the ruminants are able to use cellulose as food because the first of their four stomach compartments (rumen) teems with bacteria and protists that secrete cellulase.

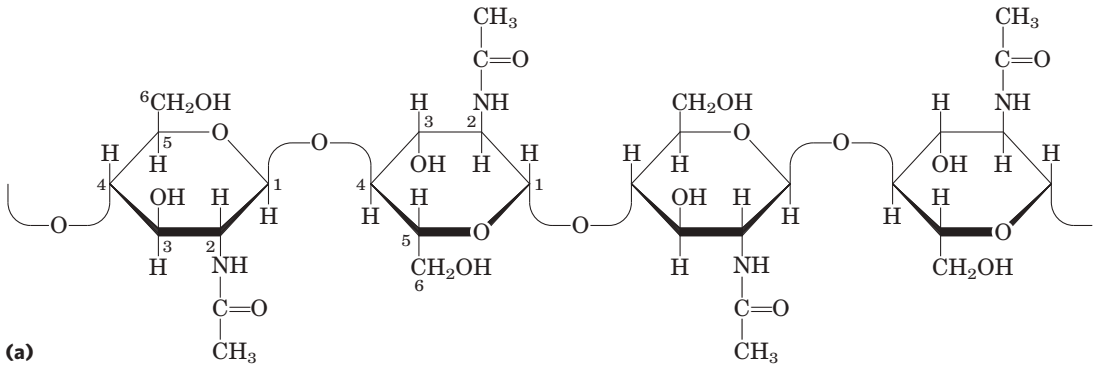
Biomass (such as switch grass) that is rich in cellulose can be used as starting material for the fermentation of carbohydrates to ethanol, to be used as a gasoline additive. The annual production of biomass on Earth (accomplished primarily by photosynthetic organisms) is the energetic equivalent of nearly a trillion barrels of crude oil, when converted to ethanol by fermentation. Because of their potential use in biomass conversion to bioenergy, cellulose-degrading enzymes such as cellulase are under vigorous investigation. Supramolecular complexes called cellulosomes, found on the outside surface of the bacterium *Clostridium cellulolyticum*, include the catalytic subunit of cellulase, along with proteins that hold one or more cellulase molecules to the bacterial surface, and a subunit that binds cellulose and positions it in the catalytic site.

A major fraction of photosynthetic biomass is the woody portion of plants and trees, which consists of cellulose plus several other polymers derived from carbohydrates that are not easily digestible, either chemically or biologically. Lignins, for example, make up some 30% of the mass of wood. Synthesized from precursors that include phenylalanine and glucose, lignins are complex polymers with covalent cross-links to cellulose that complicate the digestion of cellulose by cellulase. If woody plants are to be used in the production of ethanol from biomass, better means of digesting wood components will need to be found.

Chitin is a linear homopolysaccharide composed of *N*-acetylglucosamine residues in (β1→4) linkage **(Fig. 7–16)**. The only chemical difference from cellulose is the replacement of the hydroxyl group at C-2 with an acetylated amino group. Chitin forms extended fibers similar to those of cellulose, and like cellulose cannot be digested by vertebrates. Chitin is the principal component of the hard exoskeletons of nearly a million species of arthropods—insects, lobsters, and crabs, for example—and is probably the second most abundant polysaccharide, next to cellulose, in nature; an estimated 1 billion tons of chitin are produced each year in the biosphere.

Steric Factors and Hydrogen Bonding Influence Homopolysaccharide Folding

The folding of polysaccharides in three dimensions follows the same principles as those governing polypeptide



(a)

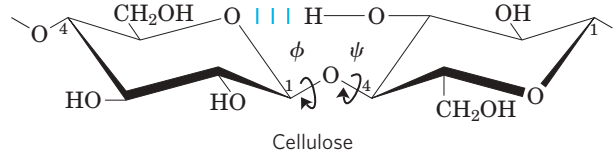
(b)

FIGURE 7–16 Chitin. (a) A short segment of chitin, a homopolymer of *N*-acetyl-D-glucosamine units in (β1⟶4) linkage. **(b)** A spotted June beetle (*Pelidnota punctata*), showing its surface armor (exoskeleton) of chitin.

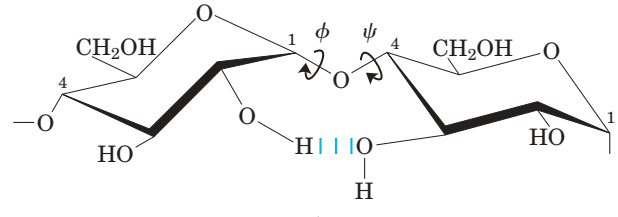
structure: subunits with a more-or-less rigid structure dictated by covalent bonds form three-dimensional macromolecular structures that are stabilized by weak interactions within or between molecules, such as hydrogen bonds and hydrophobic and van der Waals interactions, and, for polymers with charged subunits, electrostatic interactions. Because polysaccharides have so many hydroxyl groups, hydrogen bonding has an especially important influence on their structure. Glycogen, starch, and cellulose are composed of pyranoside subunits (having six-membered rings), as are the oligosaccharides of glycoproteins and glycolipids, to be discussed later. Such molecules can be represented as a series of rigid pyranose rings connected by an oxygen atom bridging two carbon atoms (the glycosidic bond). There is, in principle, free rotation about both C—O bonds linking the residues (Fig. 7–14), but as in polypeptides (see Figs 4–2, 4–9), rotation about each bond is limited by steric hindrance by substituents. The three-dimensional structures of these molecules can be described in terms of the dihedral angles, ϕ and ψ, about the glycosidic bond **(Fig. 7–17)**, analogous to angles ϕ and ψ made by the peptide bond (see Fig. 4–2).

The bulkiness of the pyranose ring and its substituents, and electronic effects at the anomeric carbon, place constraints on the angles ϕ and ψ; thus certain conformations are much more stable than others, as can be shown on a map of energy as a function of ϕ and ψ **(Fig. 7–18)**.

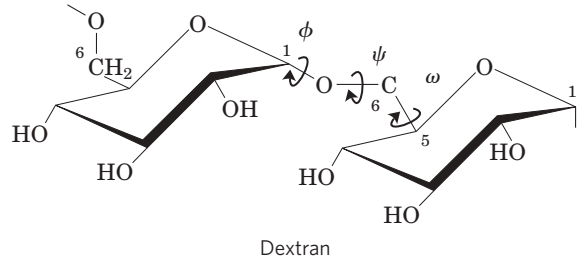
The most stable three-dimensional structure for the (α1⟶4)-linked chains of starch and glycogen is a tightly coiled helix **(Fig. 7–19)**, stabilized by interchain hydrogen bonds. In amylose (with no branches) this structure is regular enough to allow crystallization and thus determination of the structure by x-ray diffraction. The average plane of each residue along the amylose chain forms a 60° angle with the average plane of the



Cellulose
(β1⟶4)Glc repeats



Amylose
(α1⟶4)Glc repeats



Dextran
(α1⟶6)Glc repeats (with (α1⟶3) branches, not shown)

FIGURE 7–17 Conformation at the glycosidic bonds of cellulose, amylose, and dextran. The polymers are depicted as rigid pyranose rings joined by glycosidic bonds, with free rotation about these bonds. Note that in dextran there is also free rotation about the bond between C-5 and C-6 (torsion angle ω (omega)).

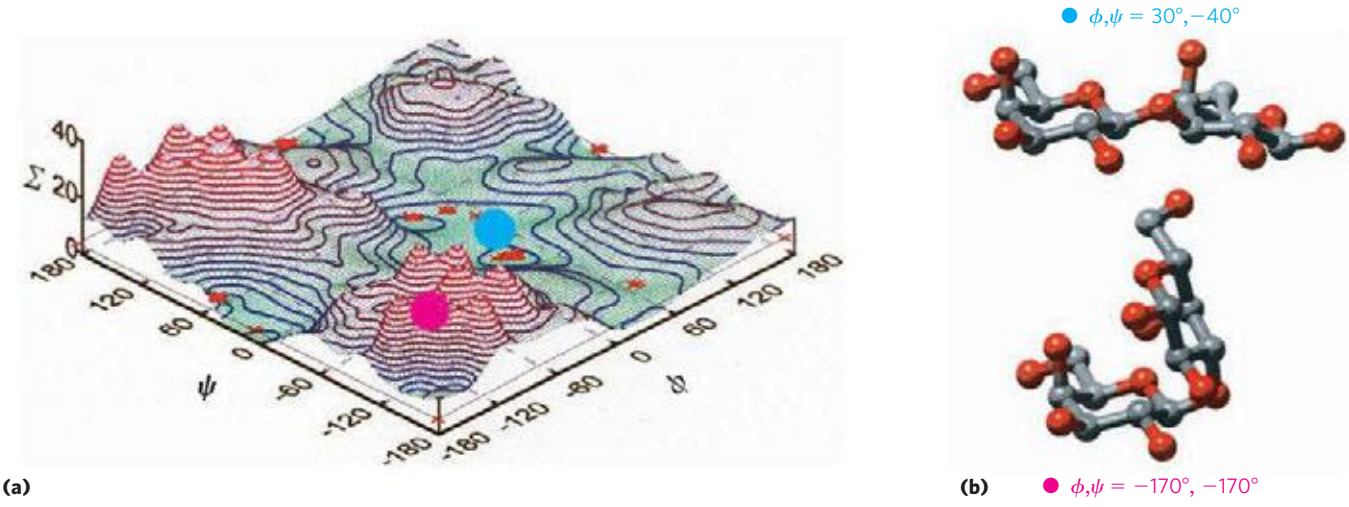


FIGURE 7–18 A map of favored conformations for oligosaccharides and polysaccharides. The torsion angles ψ and ϕ (see Fig. 7–17), which define the spatial relationship between adjacent rings, can in principle have any value from 0° to 360°. In fact, some of the torsion angles would give conformations that are sterically hindered, whereas others give conformations that maximize hydrogen bonding. **(a)** When the relative energy (Σ) is plotted for each value of ϕ and ψ, with isoenergy ("same energy") contours drawn at intervals of 1 kcal/mol above the minimum energy state, the result is a map of preferred conformations. This is analogous to the Ramachandran plot for peptides (see Figs 4–3, 4–9). **(b)** Two energetic extremes for the disaccharide Gal(β1⟶3)Gal; these values fall on the energy diagram (a) as shown by the red and blue dots. The red dot indicates the least favored conformation; the blue dot, the most favored conformation. The known conformations of the three polysaccharides shown in Figure 7–17 have been determined by x-ray crystallography, and all fall within the lowest-energy regions of the map.

preceding residue, so the helical structure has six residues per turn. For amylose, the core of the helix is of precisely the right dimensions to accommodate iodine as complex ions (I_3^- and I_5^-), giving an intensely blue complex. This interaction is a common qualitative test for amylose.

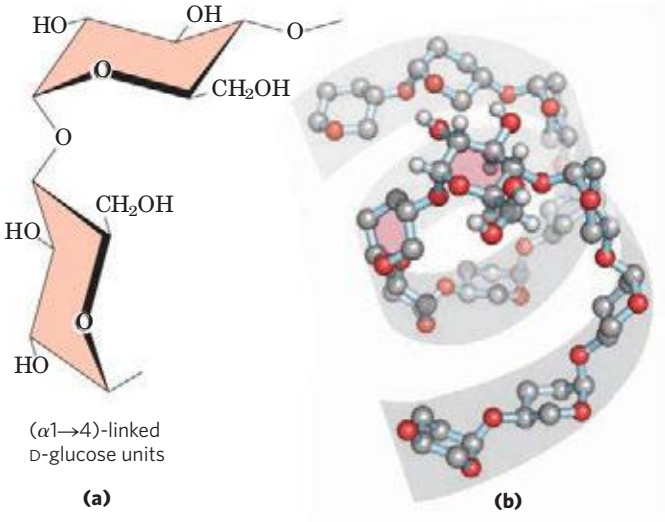


FIGURE 7–19 Helical structure of starch (amylose). (a) In the most stable conformation, with adjacent rigid chairs, the polysaccharide chain is curved, rather than linear as in cellulose (see Fig. 7–14). **(b)** A model of a segment of amylose; for clarity, the hydroxyl groups have been omitted from all but one of the glucose residues. Compare the two residues shaded in pink with the chemical structures in (a). The conformation of (α1⟶4) linkages in amylose, amylopectin, and glycogen causes these polymers to assume tightly coiled helical structures. These compact structures produce the dense granules of stored starch or glycogen seen in many cells (see Fig. 20–2).

For cellulose, the most stable conformation is that in which each chair is turned 180° relative to its neighbors, yielding a straight, extended chain. All —OH groups are available for hydrogen bonding with neighboring chains. With several chains lying side by side, a stabilizing network of interchain and intrachain hydrogen bonds produces straight, stable supramolecular fibers of great tensile strength **(Fig. 7–20)**. This property of cellulose has made it a useful substance to civilizations for millennia. Many manufactured products, including papyrus, paper, cardboard, rayon, insulating tiles, and a variety of other useful materials, are derived from cellulose. The water content of these materials is low because extensive interchain hydrogen bonding between cellulose molecules satisfies their capacity for hydrogen-bond formation.

Bacterial and Algal Cell Walls Contain Structural Heteropolysaccharides

The rigid component of bacterial cell walls (peptidoglycan) is a heteropolymer of alternating (β1⟶4)-linked *N*-acetylglucosamine and *N*-acetylmuramic acid residues (see Fig. 20–30). The linear polymers lie side by side in the cell wall, cross-linked by short peptides, the exact structure of which depends on the bacterial species. The peptide cross-links weld the polysaccharide chains into a strong sheath (peptidoglycan) that envelops the entire cell and prevents cellular swelling and lysis due to the osmotic entry of water. The enzyme

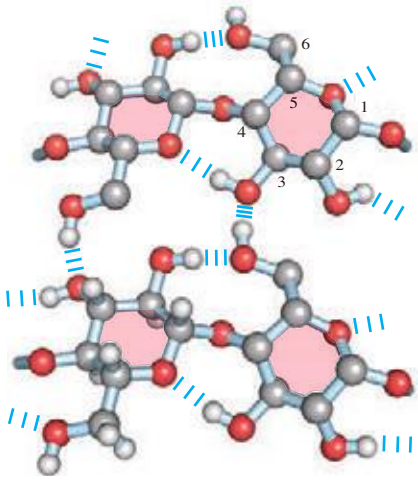
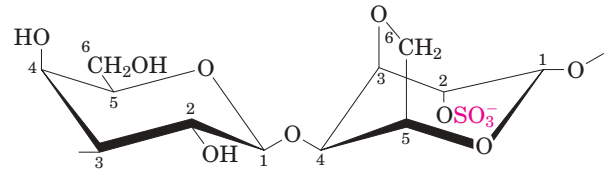


FIGURE 7–20 Cellulose chains. Scale drawing of segments of two parallel cellulose chains, showing the conformation of the D-glucose residues and the hydrogen-bond cross-links. In the hexose unit at the lower left, all hydrogen atoms are shown; in the other three hexose units, the hydrogens attached to carbon have been omitted for clarity, as they do not participate in hydrogen bonding.

lysozyme kills bacteria by hydrolyzing the (β1→4) glycosidic bond between *N*-acetylglucosamine and *N*-acetylmuramic acid (see Fig. 6–27); the enzyme is found in human tears, where it is presumably a defense against bacterial infections of the eye, and is also produced by certain bacterial viruses to ensure their release from the host bacterial cell, an essential step of the viral infection cycle. Penicillin and related antibiotics kill bacteria by preventing synthesis of the cross-links, leaving the cell wall too weak to resist osmotic lysis (p. 224).

Certain marine red algae, including some of the seaweeds, have cell walls that contain **agar**, a mixture of sulfated heteropolysaccharides made up of D-galactose and an L-galactose derivative ether-linked between C-3 and C-6. Agar is a complex mixture of polysaccharides, all with the same backbone structure but substituted to varying degrees with sulfate and pyruvate. **Agarose** (M_r ~150,000) is the agar component with the fewest charged groups (sulfates, pyruvates) **(Fig. 7–21)**. The remarkable gel-forming property of agarose makes it useful in the biochemistry laboratory. When a suspension of agarose in water is heated and cooled, the agarose forms a double helix: two molecules in parallel orientation twist together with a helix repeat of three residues; water molecules are trapped in the central cavity. These structures in turn associate with each other to form a gel—a three-dimensional matrix that traps large amounts of water. Agarose gels are used as inert supports for the electrophoretic separation of nucleic acids, an essential part of the DNA-sequencing process (p. 302). Agar is also used to form a surface for the growth of bacterial colonies. Another commercial use of agar is for the capsules in which some vitamins and drugs are packaged; the dried agar material dissolves readily in the stomach and is metabolically inert.



Agarose
3)D-Gal(β1→4)3,6-anhydro-L-Gal2S (α1
repeating units

FIGURE 7–21 Agarose. The repeating unit consists of D-galactose (β1⟶4)-linked to 3,6-anhydro-L-galactose (in which an ether bridge connects C-3 and C-6). These units are joined by (α1⟶3) glycosidic links to form a polymer 600 to 700 residues long. A small fraction of the 3,6-anhydrogalactose residues have a sulfate ester at C-2 (as shown here). The open parentheses in the systematic name indicate that the repeating unit extends from both ends.

Glycosaminoglycans Are Heteropolysaccharides of the Extracellular Matrix

The extracellular space in the tissues of multicellular animals is filled with a gel-like material, the **extracellular matrix (ECM)**, also called ground substance, which holds the cells together and provides a porous pathway for the diffusion of nutrients and oxygen to individual cells. The ECM that surrounds fibroblasts and other connective tissue cells is composed of an interlocking meshwork of heteropolysaccharides and fibrous proteins such as fibrillar collagens, elastins, and fibronectins. Basement membrane is a specialized ECM that underlies epithelial cells; it consists of specialized collagens, laminins, and heteropolysaccharides. These heteropolysaccharides, the **glycosaminoglycans**, are a family of linear polymers composed of repeating disaccharide units **(Fig. 7–22)**. They are unique to animals and bacteria and are not found in plants. One of the two monosaccharides is always either *N*-acetylglucosamine or *N*-acetylgalactosamine; the other is in most cases a uronic acid, usually D-glucuronic or L-iduronic acid. Some glycosaminoglycans contain esterified sulfate groups. The combination of sulfate groups and the carboxylate groups of the uronic acid residues gives glycosaminoglycans a very high density of negative charge. To minimize the repulsive forces among neighboring charged groups, these molecules assume an extended conformation in solution, forming a rodlike helix in which the negatively charged carboxylate groups occur on alternate sides of the helix (as shown for heparin in Fig. 7–22). The extended rod form also provides maximum separation between the negatively charged sulfate groups. The specific patterns of sulfated and nonsulfated sugar residues in glycosaminoglycans provide for specific recognition by a variety of protein ligands that bind electrostatically to these molecules. The sulfated glycosaminoglycans are attached to extracellular proteins to form proteoglycans (Section 7.3).

The glycosaminoglycan **hyaluronan** (hyaluronic acid) contains alternating residues of D-glucuronic acid and *N*-acetylglucosamine (Fig. 7–22). With up to 50,000 repeats of the basic disaccharide unit, hyaluronan has a

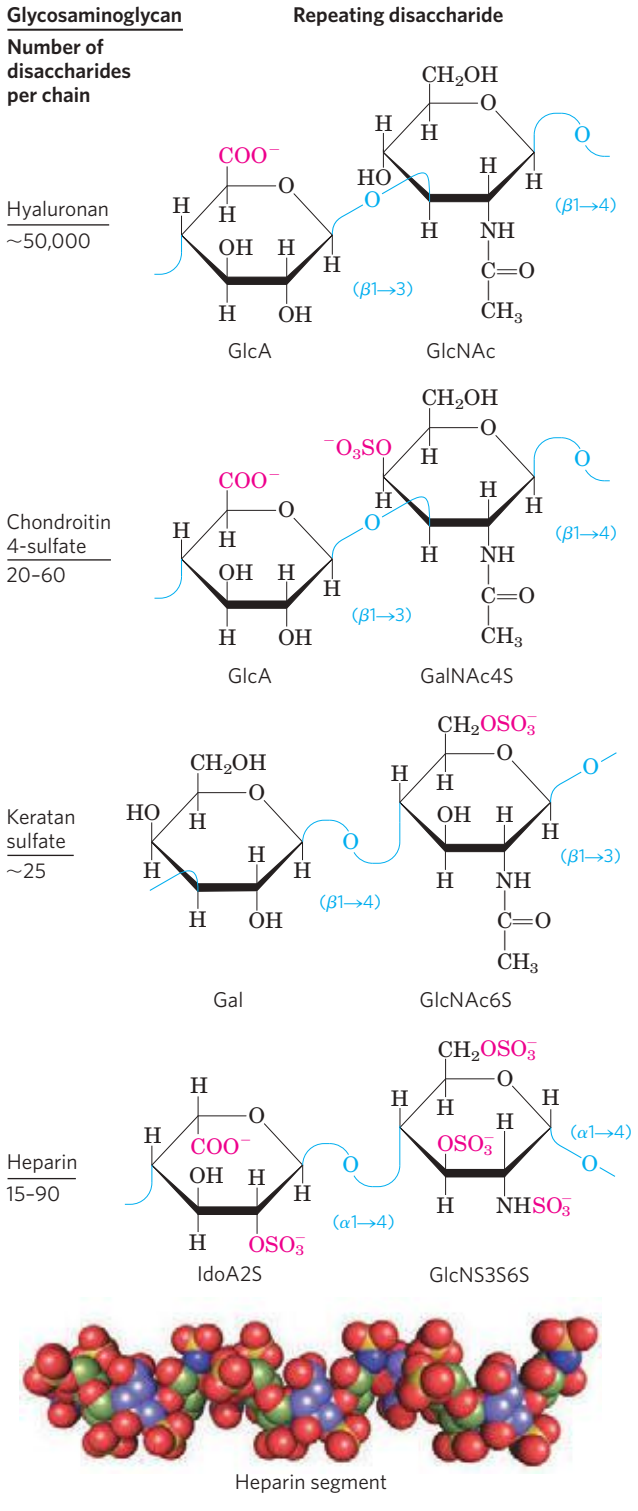
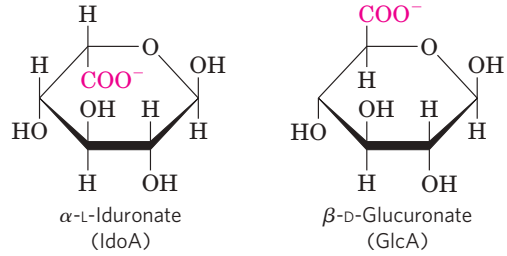


FIGURE 7–22 Repeating units of some common glycosaminoglycans of extracellular matrix. The molecules are copolymers of alternating uronic acid and amino sugar residues (keratan sulfate is the exception), with sulfate esters in any of several positions, except in hyaluronan. The ionized carboxylate and sulfate groups (red in the perspective formulas) give these polymers their characteristic high negative charge. Therapeutic heparin contains primarily iduronic acid (IdoA) and a smaller proportion of glucuronic acid (GlcA, not shown), and is generally highly sulfated and heterogeneous in length. The space-filling model shows a heparin segment as its solution structure, as determined by NMR spectroscopy (PDB ID 1HPN). The carbons in the iduronic acid sulfate are colored blue; those in glucosamine sulfate are green. Oxygen and sulfur atoms are shown in their standard colors of red and yellow, respectively. The hydrogen atoms are not shown (for clarity). Heparan sulfate (not shown) is similar to heparin but has a higher proportion of GlcA and fewer sulfate groups, arranged in a less regular pattern.

molecular weight of several million; it forms clear, highly viscous solutions that serve as lubricants in the synovial fluid of joints and give the vitreous humor of the vertebrate eye its jellylike consistency (the Greek *hyalos* means "glass"; hyaluronan can have a glassy or translucent appearance). Hyaluronan is also a component of the extracellular matrix of cartilage and tendons, to which it contributes tensile strength and elasticity as a result of its strong noncovalent interactions with other components of the matrix. Hyaluronidase, an enzyme secreted by some pathogenic bacteria, can hydrolyze the glycosidic linkages of hyaluronan, rendering tissues more susceptible to bacterial invasion. In many animal species, a similar enzyme in sperm hydrolyzes an outer glycosaminoglycan coat around the ovum, allowing sperm penetration.

Other glycosaminoglycans differ from hyaluronan in three respects: they are generally much shorter polymers, they are covalently linked to specific proteins (proteoglycans), and one or both monomeric units differ from those of hyaluronan. **Chondroitin sulfate** (Greek *chondros*, "cartilage") contributes to the tensile strength of cartilage, tendons, ligaments, and the walls of the aorta. Dermatan sulfate (Greek *derma*, "skin") contributes to the pliability of skin and is also present in blood vessels and heart valves. In this polymer, many of the glucuronate residues present in chondroitin sulfate are replaced by their 5-epimer, L-iduronate (IdoA).



Keratan sulfates (Greek *keras*, "horn") have no uronic acid and their sulfate content is variable. They are present in cornea, cartilage, bone, and a variety of horny structures formed of dead cells: horn, hair, hoofs, nails, and claws. **Heparan sulfate** (Greek *hēpar*, "liver"; it was originally isolated from dog liver) is produced by all animal cells and contains variable arrangements of sulfated and nonsulfated sugars. The sulfated segments of the chain allow it to interact with a large number of proteins, including growth factors and ECM components, as well as various enzymes and factors present in plasma. Heparin is a fractionated form of heparan sulfate derived mostly from mast cells (a type of leukocyte). Heparin is a therapeutic agent used to inhibit coagulation through its capacity to bind the protease inhibitor antithrombin.

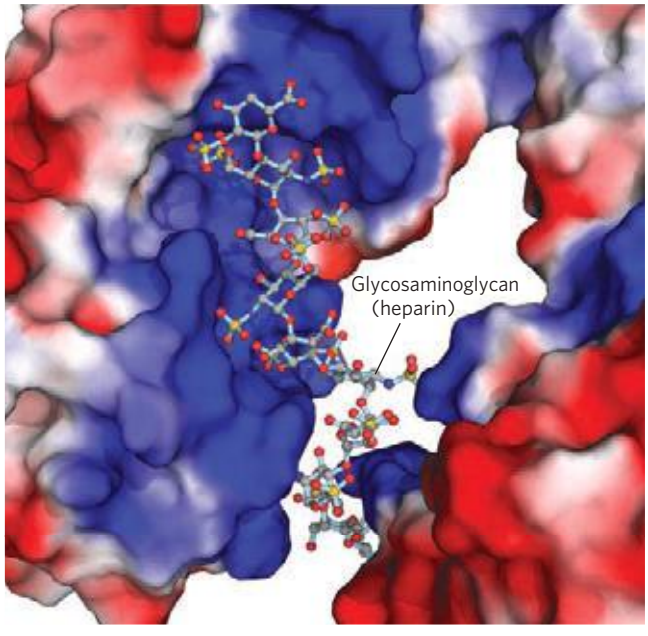


FIGURE 7–23 Interaction between a glycosaminoglycan and its binding protein. Fibroblast growth factor 1 (FGF1), its cell surface receptor (FGFR), and a short segment of a glycosaminoglycan (heparin) were co-crystallized to yield the structure shown here (PDB ID 1E0O). The proteins are represented as surface contour images, with color to represent surface electrostatic potential: red, predominantly negative charge; blue, predominantly positive charge. Heparin is shown in a ball-and-stick representation, with the negative charges ($—SO_3^-$ and $—COO^-$) attracted to the positive (blue) surface of the FGF1 protein. Heparin was used in this experiment, but the glycosaminoglycan that binds FGF1 in vivo is heparan sulfate on the cell surface.

Heparin binding causes antithrombin to bind to and inhibit thrombin, a protease essential to blood clotting. The interaction is strongly electrostatic; heparin has the highest negative charge density of any known biological macromolecule **(Fig. 7–23)**. Purified heparin is routinely added to blood samples obtained for clinical analysis, and to blood donated for transfusion, to prevent clotting.

Table 7–2 summarizes the composition, properties, roles, and occurrence of the polysaccharides described in Section 7.2.

TABLE 7–2 Structures and Roles of Some Polysaccharides

Polymer	Type*	Repeating unit†	Size (number of monosaccharide units)	Roles/significance
Starch				Energy storage: in plants
Amylose	Homo-	(α1→4)Glc, linear	50–5,000	
Amylopectin	Homo-	(α1→4)Glc, with (α1→6)Glc branches every 24–30 residues	Up to 10^6	
Glycogen	Homo-	(α1→4)Glc, with (α1→6)Glc branches every 8–12 residues	Up to 50,000	Energy storage: in bacteria and animal cells
Cellulose	Homo-	(β1→4)Glc	Up to 15,000	Structural: in plants, gives rigidity and strength to cell walls
Chitin	Homo-	(β1→4)GlcNAc	Very large	Structural: in insects, spiders, crustaceans, gives rigidity and strength to exoskeletons
Dextran	Homo-	(α1→6)Glc, with (α1→3) branches	Wide range	Structural: in bacteria, extracellular adhesive
Peptidoglycan	Hetero-; peptides attached	4)Mur2Ac(β1→4)GlcNAc(β1	Very large	Structural: in bacteria, gives rigidity and strength to cell envelope
Agarose	Hetero-	3)D-Gal(β1→4)3,6-anhydro-L-Gal(α1	1,000	Structural: in algae, cell wall material
Hyaluronan (a glycosaminoglycan)	Hetero-; acidic	4)GlcA(β1→3)GlcNAc(β1	Up to 100,000	Structural: in vertebrates, extracellular matrix of skin and connective tissue; viscosity and lubrication in joints

*Each polymer is classified as a homopolysaccharide (homo-) or heteropolysaccharide (hetero-).

†The abbreviated names for the peptidoglycan, agarose, and hyaluronan repeating units indicate that the polymer contains repeats of this disaccharide unit. For example, in peptidoglycan, the GlcNAc of one disaccharide unit is (β1→4)-linked to the first residue of the next disaccharide unit.

SUMMARY 7.2 Polysaccharides

- Polysaccharides (glycans) serve as stored fuel and as structural components of cell walls and extracellular matrix.
- The homopolysaccharides starch and glycogen are stored fuels in plant, animal, and bacterial cells. They consist of D-glucose with (α1→4) linkages, and both contain some branches.
- The homopolysaccharides cellulose, chitin, and dextran serve structural roles. Cellulose, composed of (β1→4)-linked D-glucose residues, lends strength and rigidity to plant cell walls. Chitin, a polymer of (β1→4)-linked *N*-acetylglucosamine, strengthens the exoskeletons of arthropods. Dextran forms an adhesive coat around certain bacteria.
- Homopolysaccharides fold in three dimensions. The chair form of the pyranose ring is essentially rigid, so the conformation of the polymers is determined by rotation about the bonds from the rings to the oxygen atom in the glycosidic linkage. Starch and glycogen form helical structures with intrachain hydrogen bonding; cellulose and chitin form long, straight strands that interact with neighboring strands.
- Bacterial and algal cell walls are strengthened by heteropolysaccharides—peptidoglycan in bacteria, agar in red algae. The repeating disaccharide in peptidoglycan is GlcNAc(β1→4)Mur2Ac; in agar, it is D-Gal(β1→4)3,6-anhydro-L-Gal.
- Glycosaminoglycans are extracellular heteropolysaccharides in which one of the two monosaccharide units is a uronic acid (keratin sulfate is an exception) and the other an *N*-acetylated amino sugar. Sulfate esters on some of the hydroxyl groups and on the amino group of some glucosamine residues in heparin and in heparan sulfate give these polymers a high density of negative charge, forcing them to assume extended conformations. These polymers (hyaluronan, chondroitin sulfate, dermatan sulfate, and keratan sulfate) provide viscosity, adhesiveness, and tensile strength to the extracellular matrix.

7.3 Glycoconjugates: Proteoglycans, Glycoproteins, and Glycosphingolipids

In addition to their important roles as stored fuels (starch, glycogen, dextran) and as structural materials (cellulose, chitin, peptidoglycans), polysaccharides and oligosaccharides are information carriers. Some provide communication between cells and their extracellular surroundings; others label proteins for transport to and localization in specific organelles, or for destruction when the protein is malformed or superfluous; and others serve as recognition sites for extracellular signal molecules (growth factors, for example) or extracellular parasites (bacteria or viruses). On almost every eukaryotic cell, specific oligosaccharide chains attached to components of the plasma membrane form a carbohydrate layer (the glycocalyx), several nanometers thick, that serves as an information-rich surface that the cell shows to its surroundings. These oligosaccharides are central players in cell-cell recognition and adhesion, cell migration during development, blood clotting, the immune response, wound healing, and other cellular processes. In most of these cases, the informational carbohydrate is covalently joined to a protein or a lipid to form a **glycoconjugate**, which is the biologically active molecule **(Fig. 7–24)**.

Proteoglycans are macromolecules of the cell surface or extracellular matrix in which one or more sulfated glycosaminoglycan chains are joined covalently to a membrane protein or a secreted protein. The glycosaminoglycan chain can bind to extracellular proteins through electrostatic interactions between the protein and the negatively charged sugar moieties on the proteoglycan. Proteoglycans are major components of all extracellular matrices.

Glycoproteins have one or several oligosaccharides of varying complexity joined covalently to a protein. They are usually found on the outer face of the plasma membrane (as part of the glycocalyx), in the extracellular matrix, and in the blood. Inside cells they are found in specific organelles such as Golgi complexes,

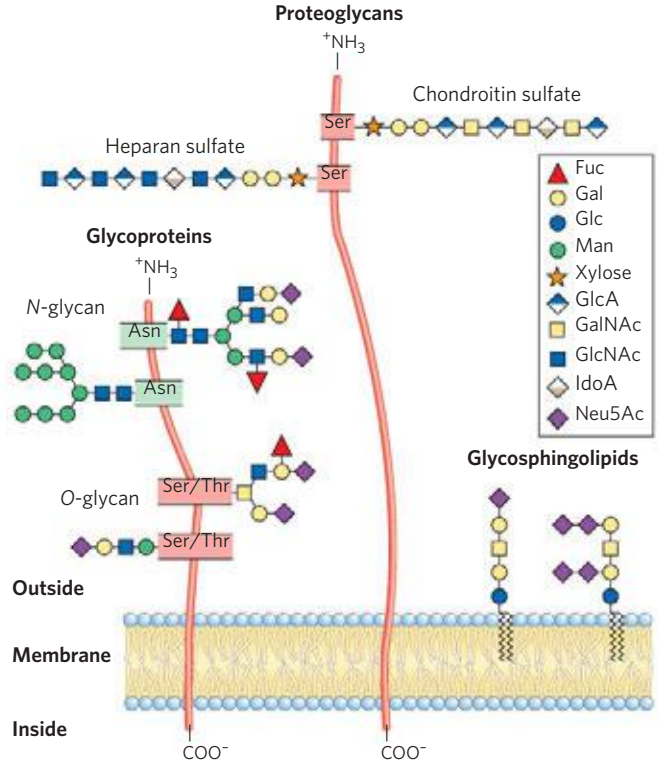


FIGURE 7–24 Glycoconjugates. The structures of some typical proteoglycans, glycoproteins, and glycosphingolipids described in the text.

secretory granules, and lysosomes. The oligosaccharide portions of glycoproteins are very heterogeneous and, like glycosaminoglycans, they are rich in information, forming highly specific sites for recognition and high-affinity binding by carbohydrate-binding proteins called lectins. Some cytosolic and nuclear proteins can be glycosylated as well.

Glycosphingolipids are plasma membrane components in which the hydrophilic head groups are oligosaccharides. As in glycoproteins, the oligosaccharides act as specific sites for recognition by lectins. The brain and neurons are rich in glycosphingolipids, which help in nerve conduction and myelin formation. Glycosphingolipids also play a role in signal transduction in cells. Sphingolipids are considered in more detail in Chapters 10 and 11.

Proteoglycans Are Glycosaminoglycan-Containing Macromolecules of the Cell Surface and Extracellular Matrix

Mammalian cells can produce 40 types of proteoglycans. These molecules act as tissue organizers, and they influence various cellular activities, such as growth factor activation and adhesion. The basic proteoglycan unit consists of a "core protein" with covalently attached glycosaminoglycan(s). The point of attachment is a Ser residue, to which the glycosaminoglycan is joined through a tetrasaccharide bridge **(Fig. 7–25)**. The Ser residue is generally in the sequence –Ser–Gly–X–Gly– (where X is any amino acid residue), although not every protein with this sequence has an attached glycosaminoglycan.

Many proteoglycans are secreted into the extracellular matrix, but some are integral membrane proteins (see Fig. 11–7). For example, the sheetlike extracellular matrix (basal lamina) that separates organized groups of cells from other groups contains a family of core proteins (M_r 20,000 to 40,000), each with several covalently attached heparan sulfate chains. There are two major families of membrane heparan sulfate proteoglycans. **Syndecans** have a single transmembrane domain and an extracellular domain bearing three to five chains of heparan sulfate and in some cases chondroitin sulfate **(Fig. 7–26a)**. **Glypicans** are attached to the membrane by a lipid anchor, a derivative of the membrane lipid phosphatidylinositol (see Fig. 11–15). Both syndecans and glypicans can be shed into the extracellular space. A protease in the ECM that cuts close to the membrane surface releases syndecan ectodomains (those domains outside the plasma membrane), and a phospholipase that breaks the connection

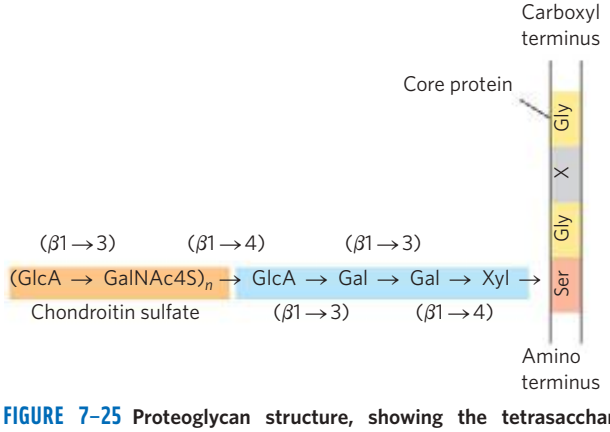


FIGURE 7–25 Proteoglycan structure, showing the tetrasaccharide bridge. A typical tetrasaccharide linker (blue) connects a glycosaminoglycan—in this case chondroitin 4-sulfate (orange)—to a Ser residue in the core protein. The xylose residue at the reducing end of the linker is joined by its anomeric carbon to the hydroxyl of the Ser residue.

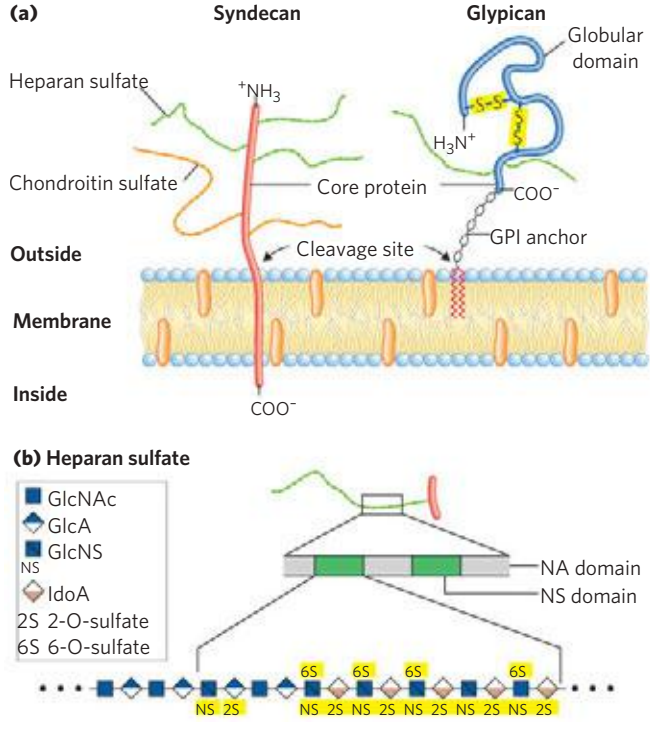


FIGURE 7–26 Two families of membrane proteoglycans. (a) Schematic diagrams of a syndecan and a glypican in the plasma membrane. Syndecans are held in the membrane by hydrophobic interactions between a sequence of nonpolar amino acid residues and plasma membrane lipids; they can be released by a single proteolytic cut near the membrane surface. In a typical syndecan, the extracellular amino-terminal domain is covalently attached (by tetrasaccharide linkers such as those in Fig. 7–25) to three heparan sulfate chains and two chondroitin sulfate chains. Glypicans are held in the membrane by a covalently attached membrane lipid (GPI anchor; see Fig. 11–15), but are shed if the bond between the lipid portion of the GPI anchor (phosphatidylinositol) and the oligosaccharide linked to the protein is cleaved by a phospholipase. All glypicans have 14 conserved Cys residues, which form disulfide bonds to stabilize the protein moiety, and either two or three glycosaminoglycan chains attached near the carboxyl terminus, close to the membrane surface. **(b)** Along a heparan sulfate chain, regions rich in sulfated sugars, the NS domains (green), alternate with regions with chiefly unmodified residues of GlcNAc and GlcA, the NA domains (gray). One of the NS domains is shown in more detail, revealing a high density of modified residues: GlcNS (*N*-sulfoglucosamine), with a sulfate ester at C-6; and both GlcA and IdoA, with a sulfate ester at C-2. The exact pattern of sulfation in the NS domain differs among proteoglycans.

to the membrane lipid releases glypicans. These mechanisms provide a way for a cell to change its surface features quickly. Shedding is highly regulated and is activated in proliferating cells, such as cancer cells. Proteoglycan shedding is involved in cell-cell recognition and adhesion, and in the proliferation and differentiation of cells. Numerous chondroitin sulfate and dermatan sulfate proteoglycans also exist, some as membrane-bound entities, others as secreted products in the ECM.

The glycosaminoglycan chains can bind to a variety of extracellular ligands and thereby modulate the ligands' interaction with specific receptors of the cell surface. Detailed studies of heparan sulfate demonstrate a domain structure that is not random; some domains (typically 3 to 8 disaccharide units long) differ from neighboring domains in sequence and in ability to bind to specific proteins. Highly sulfated domains (called NS domains) alternate with domains having unmodified GlcNAc and GlcA residues (*N*-acetylated, or NA, domains) (Fig. 7–26b). The exact pattern of sulfation in the NS domain depends on the particular proteoglycan; given the number of possible modifications of the GlcNAc–IdoA (iduronic acid) dimer, at least 32 different disaccharide units are possible. Furthermore, the same core protein can display different heparan sulfate structures when synthesized in different cell types.

Heparan sulfate molecules with precisely organized NS domains bind specifically to extracellular proteins and signaling molecules to alter their activities. The change in activity may result from a conformational change in the protein that is induced by the binding **(Fig. 7–27a)**, or it may be due to the ability of adjacent domains of heparan sulfate to bind to two different proteins, bringing them into close proximity and enhancing protein-protein interactions (Fig. 7–27b). A third general mechanism of action is the binding of extracellular signal molecules (growth factors, for example) to heparan sulfate, which increases their local concentrations and enhances their interaction with growth factor receptors in the cell surface; in this case, the heparan sulfate acts as a coreceptor (Fig. 7–27c). For example, fibroblast growth factor (FGF), an extracellular protein signal that stimulates cell division, first binds to heparan sulfate moieties of syndecan molecules in the target cell's plasma membrane. Syndecan presents FGF to the FGF plasma membrane receptor, and only then can FGF interact productively with its receptor to trigger cell division. Finally, in another type of mechanism, the NS domains interact—electrostatically and otherwise—with a variety of soluble molecules outside the cell, maintaining high local concentrations at the cell surface (Fig. 7–27d).

The importance of correctly synthesizing sulfated domains in heparan sulfate is demonstrated in mutant

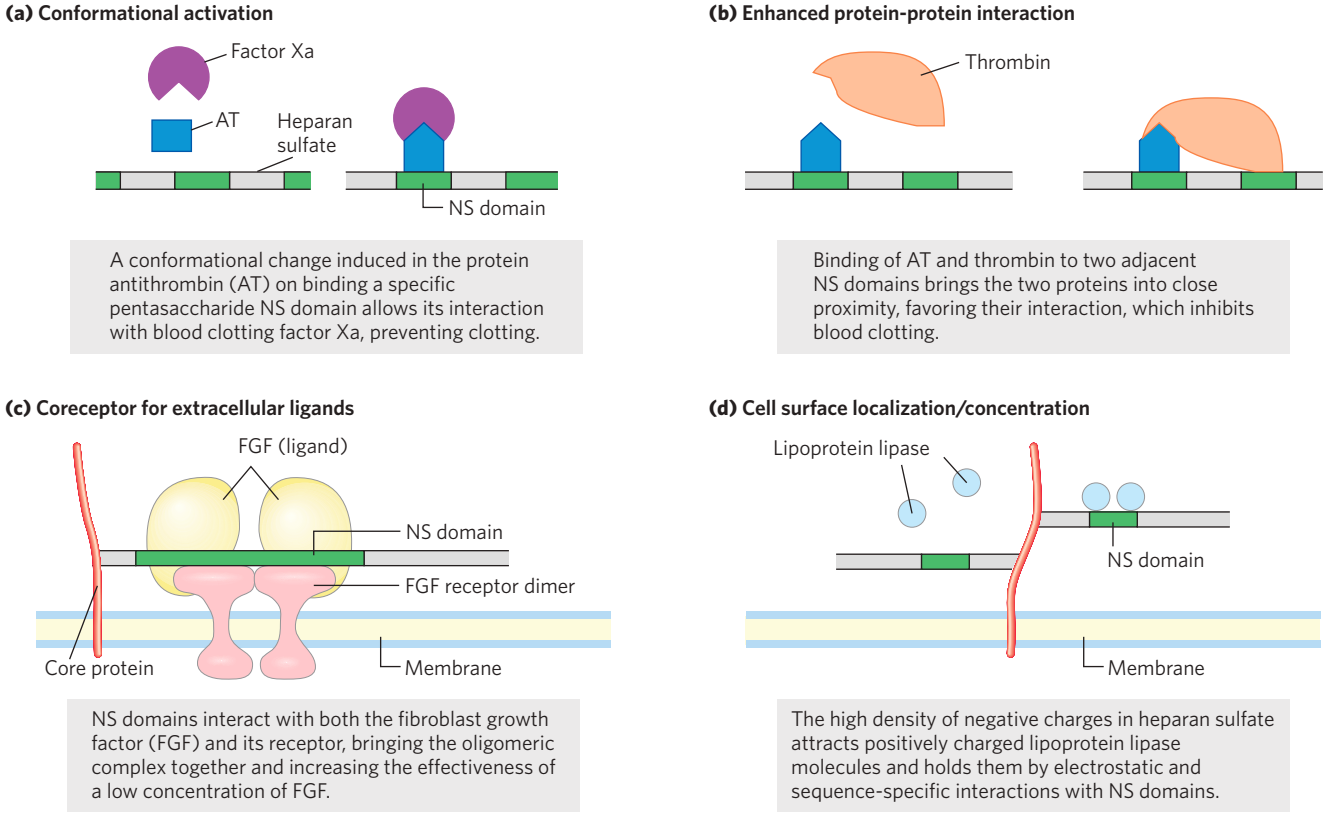


FIGURE 7–27 **Four types of protein interactions with NS domains of heparan sulfate.**

("knockout") mice lacking the enzyme that sulfates the C-2 hydroxyl of iduronate (IdoA). These animals are born without kidneys and with very severe developmental abnormalities of the skeleton and eyes. Other studies demonstrate that membrane proteoglycans are important in lipoprotein clearance in the liver. There is growing evidence that the path taken by developing axons in the nervous system, and thus the wiring circuitry, is influenced by proteoglycans containing heparan sulfates and chondroitin sulfate, which provide directional cues for axon outgrowth.

Some proteoglycans can form **proteoglycan aggregates**, enormous supramolecular assemblies of many core proteins all bound to a single molecule of hyaluronan. Aggrecan core protein (M_r ~250,000) has multiple chains of chondroitin sulfate and keratan sulfate, joined to Ser residues in the core protein through trisaccharide linkers, to give an aggrecan monomer of M_r ~2×10^6. When a hundred or more of these "decorated" core proteins bind a single, extended molecule of hyaluronate **(Fig. 7–28)**, the resulting proteoglycan aggregate (M_r >2×10^8) and its associated water of hydration occupy a volume about equal to that of a bacterial cell! Aggrecan interacts strongly with collagen in the extracellular matrix of cartilage, contributing to the development, tensile strength, and resilience of this connective tissue.

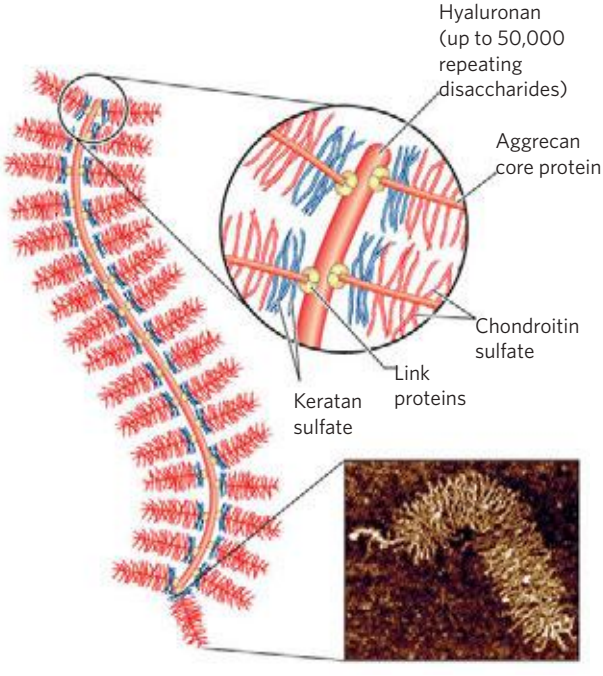


FIGURE 7–28 Proteoglycan aggregate of the extracellular matrix. Schematic drawing of a proteoglycan with many aggrecan molecules. One very long molecule of hyaluronan is associated noncovalently with about 100 molecules of the core protein aggrecan. Each aggrecan molecule contains many covalently bound chondroitin sulfate and keratan sulfate chains. Link proteins at the junction between each core protein and the hyaluronan backbone mediate the core protein–hyaluronan interaction. The micrograph shows a single molecule of aggrecan, viewed with the atomic force microscope (see Box 19-2).

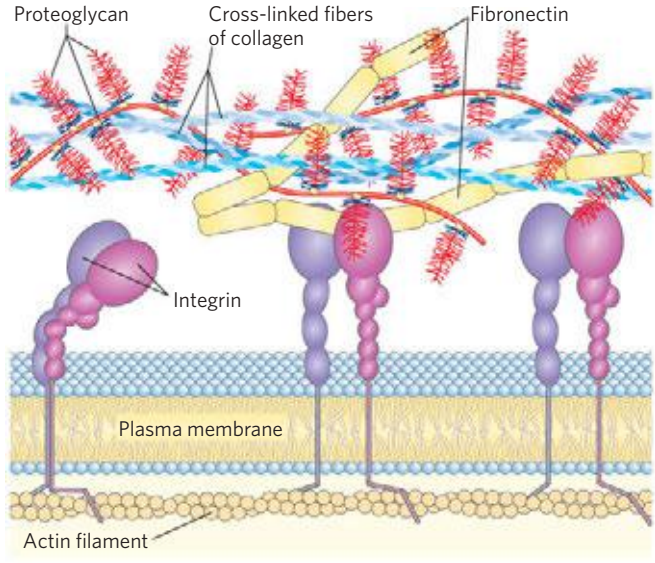


FIGURE 7–29 Interactions between cells and the extracellular matrix. The association between cells and the proteoglycan of the extracellular matrix is mediated by a membrane protein (integrin) and by an extracellular protein (fibronectin in this example) with binding sites for both integrin and the proteoglycan. Note the close association of collagen fibers with the fibronectin and proteoglycan.

Interwoven with these enormous extracellular proteoglycans are fibrous matrix proteins such as collagen, elastin, and fibronectin, forming a cross-linked meshwork that gives the whole extracellular matrix strength and resilience. Some of these proteins are multiadhesive, a single protein having binding sites for several different matrix molecules. Fibronectin, for example, has separate domains that bind fibrin, heparan sulfate, collagen, and a family of plasma membrane proteins called integrins that mediate signaling between the cell interior and the extracellular matrix (see Fig. 12–29). The overall picture of cell-matrix interactions that emerges **(Fig. 7–29)** shows an array of interactions between cellular and extracellular molecules. These interactions serve not merely to anchor cells to the extracellular matrix but also to provide paths that direct the migration of cells in developing tissue and to convey information in both directions across the plasma membrane.

Glycoproteins Have Covalently Attached Oligosaccharides

Glycoproteins are carbohydrate-protein conjugates in which the glycans are smaller, branched, and more structurally diverse than the huge glycosaminoglycans of proteoglycans. The carbohydrate is attached at its anomeric carbon through a glycosidic link to the —OH of a Ser or Thr residue (*O*-linked), or through

an *N*-glycosyl link to the amide nitrogen of an Asn residue (*N*-linked) **(Fig. 7–30)**. Some glycoproteins have a single oligosaccharide chain, but many have more than one; the carbohydrate may constitute from 1% to 70% or more of the glycoprotein by mass. About half of all proteins of mammals are glycosylated, and about 1% of all mammalian genes encode enzymes involved in the synthesis and attachment of these oligosaccharide chains. *N*-linked oligosaccharides are generally found in the consensus sequence N-{P}-[ST]; not all potential sites are used. (See Box 3–2 for the conventions on representing consensus sequences.) There appears to be no specific consensus sequence for *O*-linked oligosaccharides, although regions bearing *O*-linked chains tend to be rich in Gly, Val, and Pro residues.

One class of glycoproteins found in the cytoplasm and the nucleus is unique in that the glycosylated positions in the protein carry only single residues of *N*-acetylglucosamine, in *O*-glycosidic linkage to the hydroxyl group of Ser side chains. This modification is reversible and often occurs on the same Ser residues that are phosphorylated at some stage in the protein's activity. The two modifications are mutually exclusive, and this type of glycosylation is important in the regulation of protein activity. We discuss protein phosphorylation at length in Chapter 12.

As we shall see in Chapter 11, the external surface of the plasma membrane has many membrane glycoproteins with arrays of covalently attached oligosaccharides of varying complexity. **Mucins** are secreted or membrane glycoproteins that can contain large numbers of *O*-linked oligosaccharide chains. Mucins are present in most secretions; they are what gives mucus its characteristic slipperiness.

Glycomics is the systematic characterization of all of the carbohydrate components of a given cell or tissue, including those attached to proteins and to lipids. For glycoproteins, this also means determining which proteins are glycosylated and where in the amino acid sequence each oligosaccharide is attached. This is a challenging undertaking, but worthwhile because of the potential insights it offers into normal patterns of glycosylation and the ways in which they are altered during development or in genetic diseases or cancer. Current methods of characterizing the whole carbohydrate complement of cells depend heavily on sophisticated application of mass spectrometry (see Fig. 7–39).

The structures of a large number of *O*- and *N*-linked oligosaccharides from a variety of glycoproteins are known; Figures 7–24 and 7–30 show a few typical examples. We consider the mechanisms by which specific proteins acquire specific oligosaccharide moieties in Chapter 27.

Many of the proteins secreted by eukaryotic cells are glycoproteins, including most of the proteins of blood. For example, immunoglobulins (antibodies) and certain hormones, such as follicle-stimulating hormone, luteinizing hormone, and thyroid-stimulating hormone, are glycoproteins. Many milk proteins, including the major whey protein α-lactalbumin, and some of the proteins secreted by the pancreas (such as ribonuclease) are glycosylated, as are most of the proteins contained in lysosomes.

The biological advantages of adding oligosaccharides to proteins are slowly being uncovered. The very hydrophilic clusters of carbohydrate alter the polarity and solubility of the proteins with which they are conjugated. Oligosaccharide chains that are attached to newly synthesized proteins in the endoplasmic reticulum (ER) and elaborated in the Golgi complex serve as destination labels (see Fig. 27–39) and also act in protein quality control, targeting misfolded proteins for degradation (see Fig. 27–40). When numerous negatively charged oligosaccharide chains are clustered in a single region of a protein, the charge repulsion among them favors the formation of an extended,

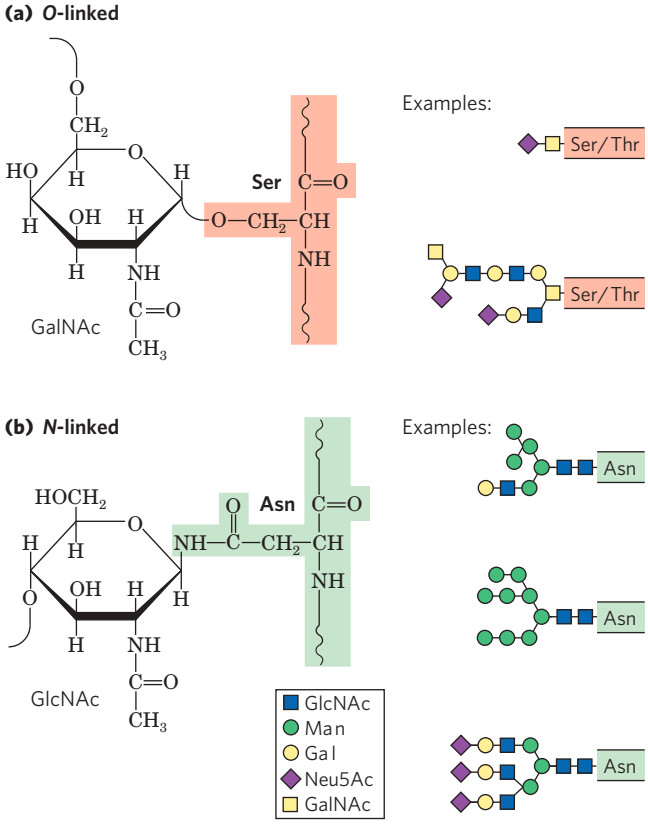


FIGURE 7–30 Oligosaccharide linkages in glycoproteins. (a) *O*-linked oligosaccharides have a glycosidic bond to the hydroxyl group of Ser or Thr residues (light red), illustrated here with GalNAc as the sugar at the reducing end of the oligosaccharide. One simple chain and one complex chain are shown. **(b)** *N*-linked oligosaccharides have an *N*-glycosyl bond to the amide nitrogen of an Asn residue (green), illustrated here with GlcNAc as the terminal sugar. Three common types of oligosaccharide chains that are *N*-linked in glycoproteins are shown. A complete description of oligosaccharide structure requires specification of the position and stereochemistry (α or β) of each glycosidic linkage.

rodlike structure in that region. The bulkiness and negative charge of oligosaccharide chains also protect some proteins from attack by proteolytic enzymes. Beyond these global physical effects on protein structure, there are also more specific biological effects of oligosaccharide chains in glycoproteins (Section 7.4). The importance of normal protein glycosylation is clear from the finding of at least 18 different genetic disorders of glycosylation in humans, all causing severely defective physical or mental development; some of these disorders are fatal.

Glycolipids and Lipopolysaccharides Are Membrane Components

Glycoproteins are not the only cellular components that bear complex oligosaccharide chains; some lipids, too, have covalently bound oligosaccharides. **Gangliosides** are membrane lipids of eukaryotic cells in which the polar head group, the part of the lipid that forms the outer surface of the membrane, is a complex oligosaccharide containing a sialic acid (Fig. 7–9) and other monosaccharide residues. Some of the oligosaccharide moieties of gangliosides, such as those that determine human blood groups (see Fig. 10–15), are identical with those found in certain glycoproteins, which therefore also contribute to blood group type. Like the oligosaccharide moieties of glycoproteins, those of membrane lipids are generally, perhaps always, found on the outer face of the plasma membrane.

Lipopolysaccharides are the dominant surface feature of the outer membrane of gram-negative bacteria such as *Escherichia coli* and *Salmonella typhimurium*. These molecules are prime targets of the antibodies produced by the vertebrate immune system in response to bacterial infection and are therefore important determinants of the serotype of bacterial strains (serotypes are strains that are distinguished on the basis of antigenic properties). The lipopolysaccharides of *S. typhimurium* contain six fatty acids bound to two glucosamine residues, one of which is the point of attachment for a complex oligosaccharide **(Fig. 7–31)**. *E. coli* has similar but unique lipopolysaccharides. The lipid A portion of the lipopolysaccharides of some bacteria is called endotoxin; its toxicity to humans and other animals is responsible for the dangerously lowered blood pressure that occurs in toxic shock syndrome resulting from gram-negative bacterial infections. ■

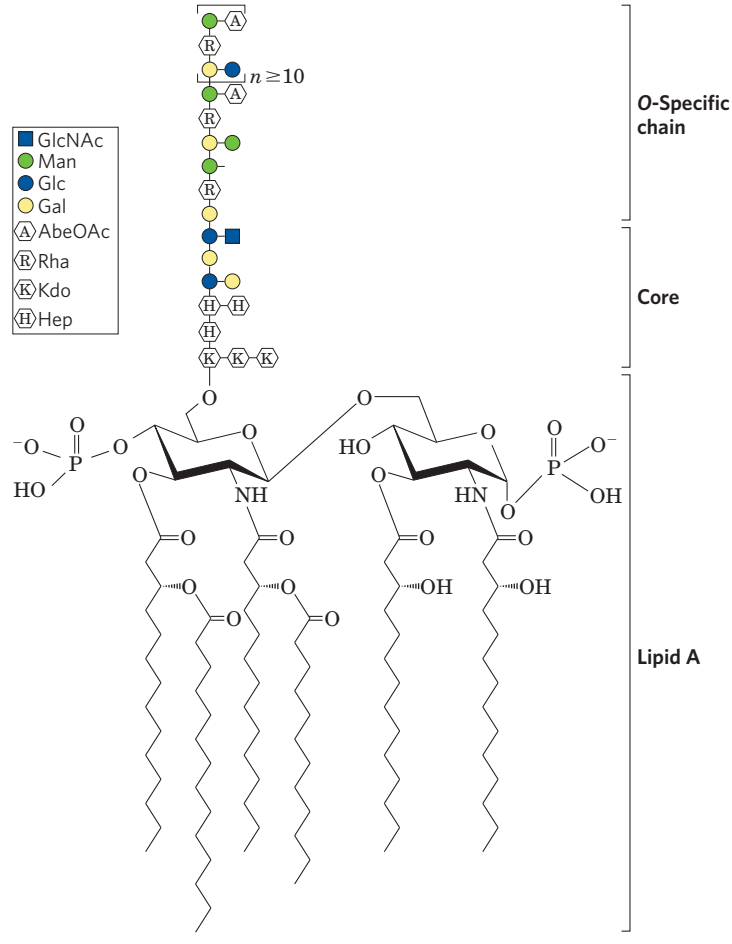


FIGURE 7–31 Bacterial lipopolysaccharides. Schematic diagram of the lipopolysaccharide of the outer membrane of *Salmonella typhimurium*. Kdo is 3-deoxy-D-*manno*-octulosonic acid (previously called ketodeoxyoctonic acid); Hep is L-*glycero*-D-*manno*-heptose; AbeOAc is abequose (a 3,6-dideoxyhexose) acetylated on one of its hydroxyls. There are six fatty acid residues in the lipid A portion of the molecule. Different bacterial species have subtly different lipopolysaccharide structures, but they have in common a lipid region (lipid A), a core oligosaccharide also known as endotoxin, and an "*O*-specific" chain, which is the principal determinant of the serotype (immunological reactivity) of the bacterium. The outer membranes of the gram-negative bacteria *S. typhimurium* and *E. coli* contain so many lipopolysaccharide molecules that the cell surface is virtually covered with *O*-specific chains.

SUMMARY 7.3 Glycoconjugates: Proteoglycans, Glycoproteins, and Glycosphingolipids

- Proteoglycans are glycoconjugates in which one or more large glycans, called sulfated glycosaminoglycans (heparan sulfate, chondroitin sulfate, dermatan sulfate, or keratan sulfate) are covalently attached to a core protein. Bound to the outside of the plasma membrane by a transmembrane peptide or a covalently attached lipid, proteoglycans provide points of adhesion, recognition, and information transfer between cells, or between the cell and the extracellular matrix.
- Glycoproteins contain oligosaccharides covalently linked to Asp or Ser/Thr residues. The glycans are typically branched and smaller than glycosaminoglycans. Many cell surface or

extracellular proteins are glycoproteins, as are most secreted proteins. The covalently attached oligosaccharides influence the folding and stability of the proteins, provide critical information about the targeting of newly synthesized proteins, and allow for specific recognition by other proteins.

- Glycomics is the determination of the full complement of sugar-containing molecules in a cell or tissue, and the determination of the function of each such molecule.
- Glycolipids and glycosphingolipids in plants and animals and lipopolysaccharides in bacteria are components of the cell envelope, with covalently attached oligosaccharide chains exposed on the cell's outer surface.

7.4 Carbohydrates as Informational Molecules: The Sugar Code

Glycobiology, the study of the structure and function of glycoconjugates, is one of the most active and exciting areas of biochemistry and cell biology. It is becoming increasingly clear that cells use specific oligosaccharides to encode important information about intracellular targeting of proteins, cell-cell interactions, cell differentiation and tissue development, and extracellular signals. Our discussion uses just a few examples to illustrate the diversity of structure and the range of biological activity of the glycoconjugates. In Chapter 20 we discuss the biosynthesis of polysaccharides, including peptidoglycan; and in Chapter 27, the assembly of oligosaccharide chains on glycoproteins.

Improved methods for the analysis of oligosaccharide and polysaccharide structure have revealed remarkable complexity and diversity in the oligosaccharides of glycoproteins and glycolipids. Consider the oligosaccharide chains in Figure 7–30, typical of those found in many glycoproteins. The most complex of those shown contains 14 monosaccharide residues of four different kinds, variously linked as (1→2), (1→3), (1→4), (1→6), (2→3), and (2→6), some with the α and some with the β configuration. Branched structures, not found in nucleic acids or proteins, are common in oligosaccharides. With the reasonable assumption that 20 different monosaccharide subunits are available for construction of oligosaccharides, we can calculate that many billions of different hexameric oligosaccharides are possible; this compares with 6.4×10^7 (20^6) different hexapeptides possible with the 20 common amino acids, and 4,096 (4^6) different hexanucleotides with the four nucleotide subunits. If we also allow for variations in oligosaccharides resulting from sulfation of one or more residues, the number of possible oligosaccharides increases by two orders of magnitude. In reality, only a subset of possible combinations is found, given the restrictions imposed by the biosynthetic enzymes and the availability of precursors. Nevertheless, the enormously rich structural information in glycans does not merely rival but far surpasses that of nucleic acids in the density of information contained in a molecule of modest size. Each of the oligosaccharides represented in Figures 7–24 and 7–30 presents a unique, three-dimensional face—a word in the sugar code—readable by the proteins that interact with it.

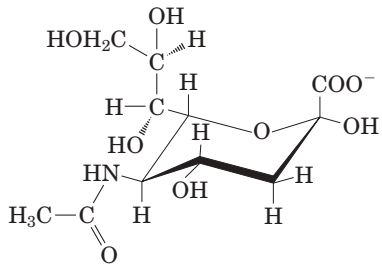
Lectins Are Proteins That Read the Sugar Code and Mediate Many Biological Processes

Lectins, found in all organisms, are proteins that bind carbohydrates with high specificity and with moderate to high affinity. Lectins serve in a wide variety of cell-cell recognition, signaling, and adhesion processes and in intracellular targeting of newly synthesized proteins. Plant lectins, abundant in seeds, probably serve as deterrents to insects and other predators. In the laboratory, purified plant lectins are useful reagents for detecting and separating glycans and glycoproteins with different oligosaccharide moieties. Here we discuss just a few examples of the roles of lectins in animal cells.

Some peptide hormones that circulate in the blood have oligosaccharide moieties that strongly influence their circulatory half-life. Luteinizing hormone and thyrotropin (polypeptide hormones produced in the pituitary) have *N*-linked oligosaccharides that end with the disaccharide GalNAc4S(β1→4)GlcNAc, which is recognized by a lectin (receptor) of hepatocytes. (GalNAc4S is *N*-acetylgalactosamine sulfated on the —OH group at C-4.) Receptor-hormone interaction mediates the uptake and destruction of luteinizing hormone and thyrotropin, reducing their concentration in the blood. Thus the blood levels of these hormones undergo a periodic rise (due to pulsatile secretion by the pituitary) and fall (due to continual destruction by hepatocytes).

The residues of Neu5Ac (a sialic acid) situated at the ends of the oligosaccharide chains of many plasma glycoproteins (Fig. 7–24) protect those proteins from uptake and degradation in the liver. For example, ceruloplasmin, a copper-containing serum glycoprotein, has several oligosaccharide chains ending in Neu5Ac. The mechanism that removes sialic acid residues from serum glycoproteins is unclear. It may be due to the activity of the enzyme neuraminidase (also called sialidase) produced by invading organisms or to a steady, slow release by extracellular enzymes. The plasma membrane of hepatocytes has lectin molecules (asialoglycoprotein receptors; "asialo-" indicating "without sialic acid") that specifically bind oligosaccharide chains with galactose residues no longer "protected" by a terminal Neu5Ac residue. Receptor-ceruloplasmin

interaction triggers endocytosis and destruction of the ceruloplasmin.



N-Acetylneuraminic acid (Neu5Ac)
(a sialic acid)

A similar mechanism is apparently responsible for removing "old" erythrocytes from the mammalian bloodstream. Newly synthesized erythrocytes have several membrane glycoproteins with oligosaccharide chains that end in Neu5Ac. When the sialic acid residues are removed by withdrawing a sample of blood from experimental animals, treating it with neuraminidase in vitro, and reintroducing it into the circulation, the treated erythrocytes disappear from the bloodstream within a few hours; erythrocytes with intact oligosaccharides (withdrawn and reintroduced without neuraminidase treatment) continue to circulate for days.

Cell surface lectins are important in the development of some human diseases—both human lectins and the lectins of infectious agents. **Selectins** are a family of plasma membrane lectins that mediate cell-cell recognition and adhesion in a wide range of cellular processes. One such process is the movement of immune cells (leukocytes) through the capillary wall, from blood to tissues, at sites of infection or inflammation **(Fig. 7–32)**. At an infection site, P-selectin on the surface of capillary endothelial cells interacts with a specific oligosaccharide of the surface glycoproteins of circulating leukocytes. This interaction slows the leukocytes as they roll along the endothelial lining of the capillaries. A second interaction, between integrin molecules (p. 470) in the leukocyte plasma membrane and an adhesion protein on the endothelial cell surface, now stops the leukocyte and allows it to move through the capillary wall into the infected tissues to initiate the immune attack. Two other selectins participate in this "lymphocyte homing": E-selectin on the endothelial cell and L-selectin on the leukocyte bind their cognate oligosaccharides on the leukocyte and endothelial cell, respectively.

Human selectins mediate the inflammatory responses in rheumatoid arthritis, asthma, psoriasis, multiple sclerosis, and the rejection of transplanted organs, and thus there is great interest in developing drugs that inhibit selectin-mediated cell adhesion. Many carcinomas express an antigen normally present only in fetal cells (sialyl Lewis x, or sialyl Le^x) that, when shed into the circulation, facilitates tumor cell survival and metastasis. Carbohydrate derivatives that mimic the sialyl Le^x portion of sialoglycoproteins or that alter the biosynthesis of the oligosaccharide might prove effective as selectin-specific drugs for treating chronic inflammation or metastatic disease.

Several animal viruses, including the influenza virus, attach to their host cells through interactions with oligosaccharides displayed on the host cell surface. The

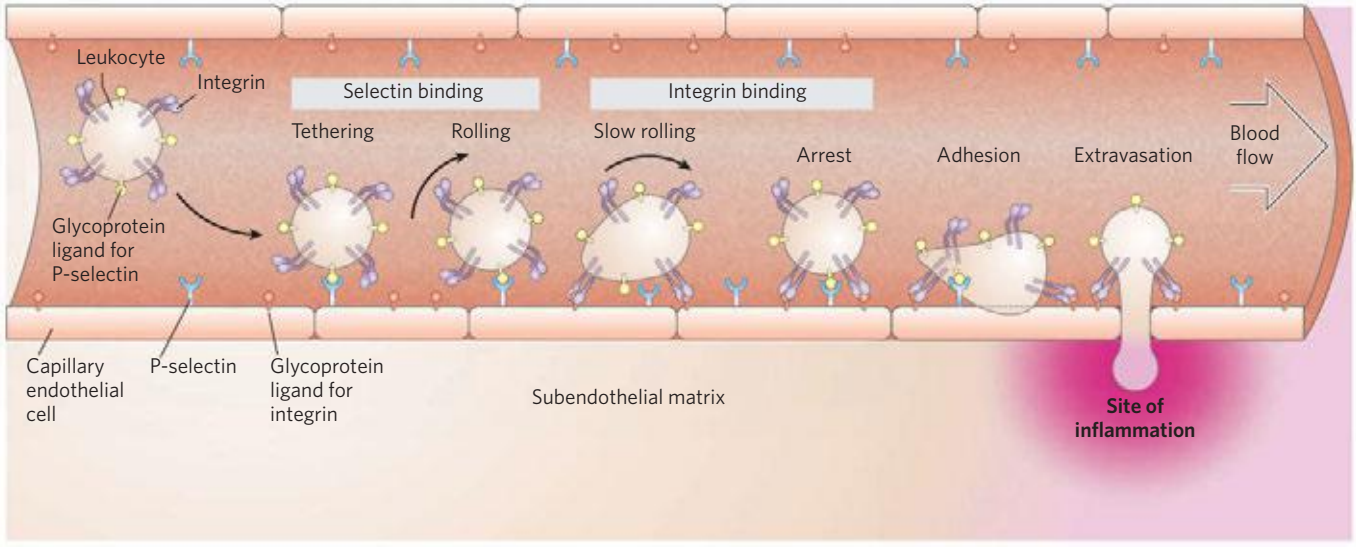
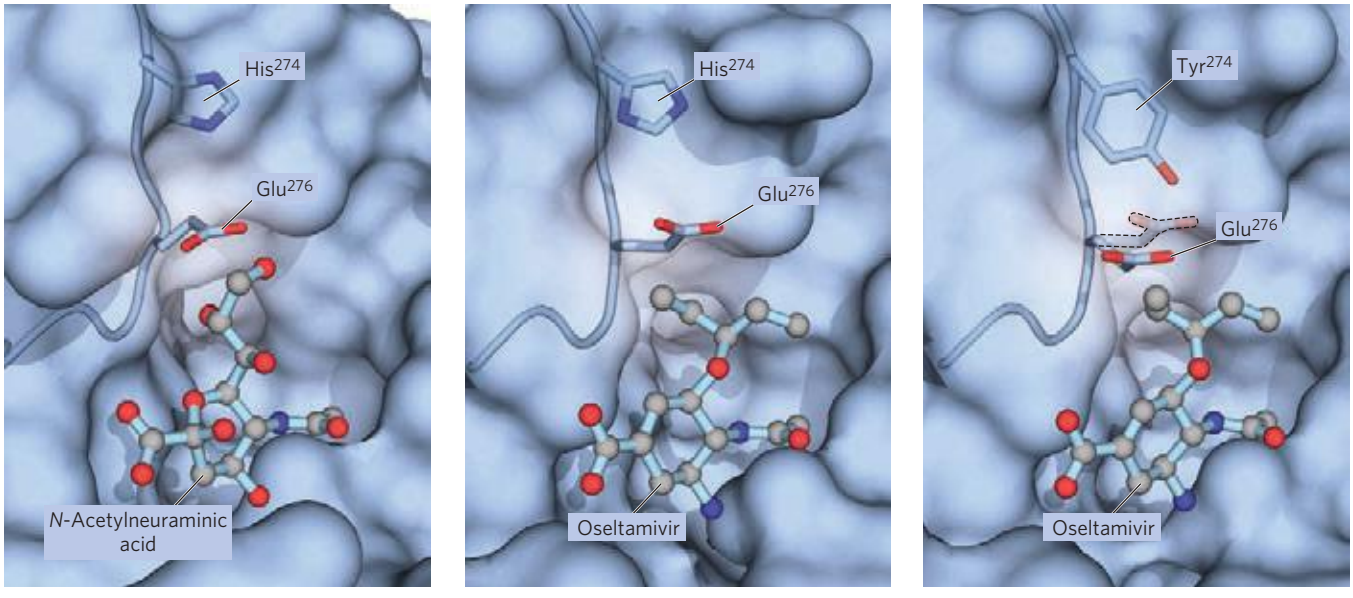


FIGURE 7–32 Role of lectin-ligand interactions in leukocyte movement to the site of an infection or injury. A leukocyte circulating through a capillary is slowed by transient interactions between P-selectin molecules in the plasma membrane of the capillary endothelial cells and glycoprotein ligands for P-selectin on the leukocyte surface. As it interacts with successive P-selectin molecules, the leukocyte rolls along the capillary surface. Near a site of inflammation, stronger interactions between integrin in the leukocyte surface and its ligand in the capillary surface lead to tight adhesion. The leukocyte stops rolling and, under the influence of signals sent out from the site of inflammation, begins extravasation—escape through the capillary wall—as it moves toward the site of inflammation.

(a)



(b) (c) (d)

FIGURE 7–33 Binding site on influenza neuraminidase for *N*-acetylneuraminic acid and an antiviral drug, oseltamivir. (a) The normal binding ligand for this enzyme is a sialic acid, *N*-acetylneuraminic acid. The drugs oseltamivir and zanamivir occupy the same site on the enzyme, competitively inhibiting it and blocking viral release from the host cell. **(b)** The normal interaction with *N*-acetylneuraminic acid in the binding site (PDB ID 2BAT). **(c)** Oseltamivir can fit into this site by pushing a nearby Glu residue out of the way (PDB ID 2HU4). **(d)** A mutation in the influenza virus's gene for neuraminidase replaces a His near this Glu residue with the larger side chain of a Tyr (PDB ID 3CL0). Now, oseltamivir is not as effective at pushing the Glu out of its way, and binds much less well to the binding site, making the mutant virus effectively resistant to oseltamivir.

lectin of the influenza virus, known as the HA (hemagglutinin) protein, is essential for viral entry and infection. After the virus has entered a host cell and has been replicated, the newly synthesized viral particles bud out of the cell, wrapped in a portion of its plasma membrane. A viral sialidase (neuraminidase) trims the terminal sialic acid residue from the host cell's oligosaccharides, releasing the viral particles from their interaction with the cell and preventing their aggregation with one another. Another round of infection can now begin. The antiviral drugs oseltamivir (Tamiflu) and zanamivir (Relenza) are used clinically in the treatment of influenza. These drugs are sugar analogs; they inhibit the viral sialidase by competing with the host cell's oligosaccharides for binding **(Fig. 7–33)**. This prevents the release of viruses from the infected cell and also causes viral particles to aggregate, both of which block another cycle of infection.

Some microbial pathogens have lectins that mediate bacterial adhesion to host cells or the entry of toxin into cells. For example, *Helicobacter pylori* has a surface lectin that adheres to oligosaccharides on the surface of epithelial cells that line the inner surface of the stomach **(Fig. 7–34)**. Among the binding sites

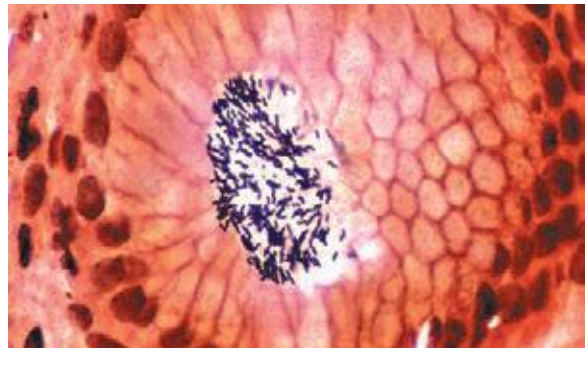


FIGURE 7–34 An ulcer in the making. *Helicobacter pylori* cells adhering to the gastric surface. This bacterium causes ulcers through interactions between a bacterial surface lectin and the Le^b oligosaccharide (a blood group antigen) of the epithelial cells lining the inside surface of the stomach.

recognized by the *H. pylori* lectin is the oligosaccharide Lewis b (Le^b), which is present in the glycoproteins and glycolipids that define the type O blood group determinant (see Fig. 10–15). This observation helps to explain the severalfold greater incidence of gastric ulcers in people of blood type O than in those of type A or B; *H. pylori* attacks their epithelial cells more effectively. Chemically synthesized analogs of the Le^b oligosaccharide may prove useful in treating this type of ulcer. Administered orally, they could prevent bacterial adhesion (and thus infection) by competing with the gastric glycoproteins for binding to the bacterial lectin.

Some of the most devastating of the human parasitic diseases, widespread in much of the developing world, are caused by eukaryotic microorganisms that display unusual surface oligosaccharides, which in some cases are known to be protective for the parasites. These organisms include the trypanosomes, responsible for African sleeping sickness and Chagas disease (see Box 6–3); *Plasmodium falciparum*, the malaria parasite; and *Entamoeba histolytica*, the causative agent of amoebic dysentery. The prospect of finding drugs that interfere with the synthesis of these unusual oligosaccharide chains, and therefore with the replication of the parasites, has inspired much recent work on the biosynthetic pathways of these oligosaccharides. ■

Lectins also act intracellularly, in sorting proteins for transportation to specific cellular compartments (see Chapter 27). For example, an oligosaccharide containing mannose 6-phosphate, recognized by a lectin, marks newly synthesized proteins in the Golgi complex for transfer to the lysosome (see Fig. 27–39).

Lectin-Carbohydrate Interactions Are Highly Specific and Often Multivalent

The high density of information in the structure of oligosaccharides provides a sugar code with an essentially unlimited number of unique "words" small enough to be read by a single protein. In their carbohydrate-binding sites, lectins have a subtle molecular complementarity that allows interaction only with their correct carbohydrate cognates. The result is an extraordinarily high specificity in these interactions. The affinity between an oligosaccharide and an individual carbohydrate-binding domain (CBD) of a lectin is sometimes modest (micromolar to millimolar K_d values), but the effective affinity is in many cases greatly increased by lectin multivalency, in which a single lectin molecule has multiple CBDs. In a cluster of oligosaccharides—as is commonly found on a membrane surface, for example—each oligosaccharide can engage one of the lectin's CBDs, strengthening the interaction. When cells express multiple lectin receptors, the avidity of the interaction can be very high, enabling highly cooperative events such as cell attachment and rolling (Fig. 7–32).

X-ray crystallographic studies of the structure of the mannose 6-phosphate receptor/lectin reveal details of its interaction with mannose 6-phosphate that explain the specificity of the binding and the role for a divalent cation in the lectin-sugar interaction **(Fig. 7–35a)**. His^{105} is hydrogen-bonded to one of the oxygen atoms of

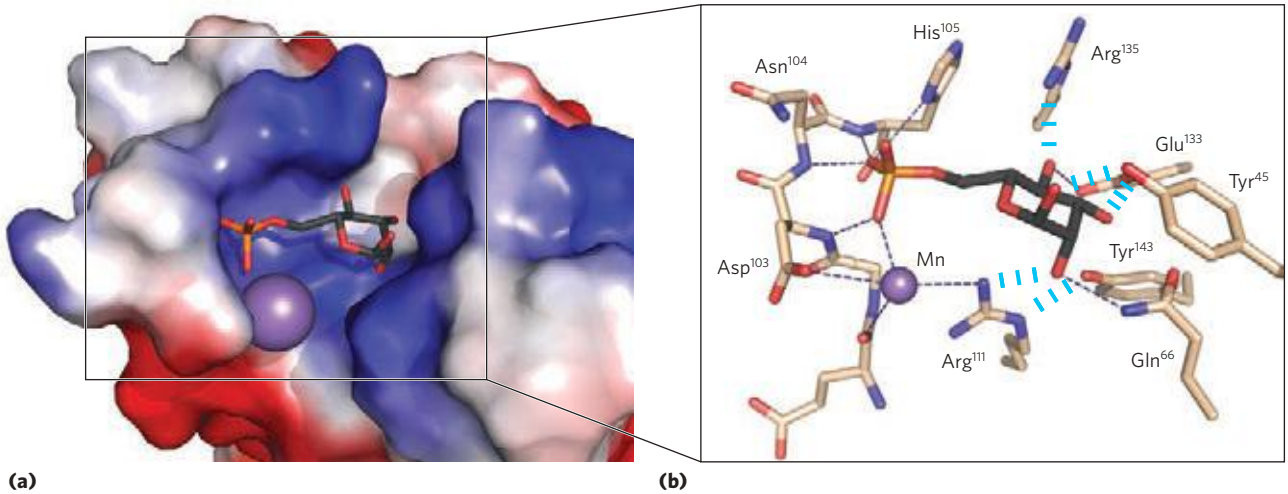


FIGURE 7–35 Details of a lectin-carbohydrate interaction. Structure of the bovine mannose 6-phosphate receptor complexed with mannose 6-phosphate (PDB ID 1M6P). The protein is represented as a surface contour image, showing the surface as predominantly negatively charged (red) or positively charged (blue). Mannose 6-phosphate is shown as a stick structure; a manganese ion is shown as a violet sphere. **(b)** An enlarged view of the binding site. Mannose 6-phosphate is hydrogen-bonded to Arg^{111} and coordinated with the manganese ion (shown smaller than its van der Waals radius for clarity). Each hydroxyl group of mannose is hydrogen-bonded to the protein. The His^{105} hydrogen-bonded to a phosphate oxygen of mannose 6-phosphate may be the residue that, when protonated at low pH, causes the receptor to release mannose 6-phosphate into the lysosome.

the phosphate (Fig. 7–35b). When the protein tagged with mannose 6-phosphate reaches the lysosome (which has a lower internal pH than the Golgi complex), the receptor loses its affinity for mannose 6-phosphate. Protonation of His^{105} may be responsible for this change in binding.

In addition to such very specific interactions, there are more general interactions that contribute to the binding of many carbohydrates to their lectins. For example, many sugars have a more polar and a less polar side **(Fig. 7–36)**; the more polar side hydrogen-bonds with the lectin, while the less polar undergoes hydrophobic interactions with nonpolar amino acid residues. The sum of all these interactions produces high-affinity binding and high specificity of lectins for their carbohydrates. This represents a kind of information transfer that is clearly central in many processes within and between cells. **Figure 7–37** summarizes some of the biological interactions mediated by the sugar code.

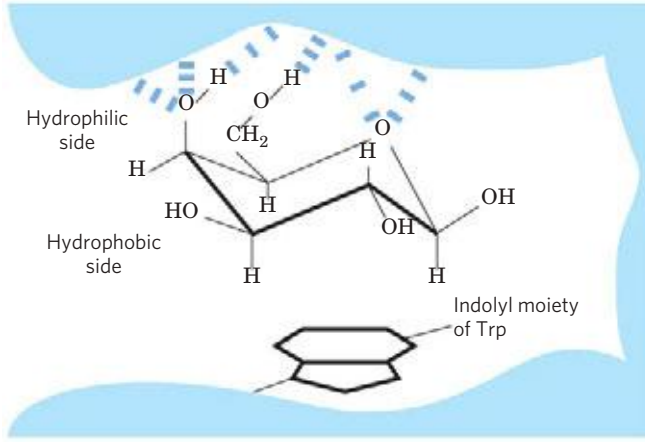


FIGURE 7–36 Hydrophobic interactions of sugar residues. Sugar units such as galactose have a more polar side (the top of the chair as shown here, with the ring oxygen and several hydroxyls) that is available to hydrogen-bond with the lectin, and a less polar side that can have hydrophobic interactions with nonpolar side chains in the protein, such as the indole ring of Trp residues.

SUMMARY 7.4 Carbohydrates as Informational Molecules: The Sugar Code

- Monosaccharides can be assembled into an almost limitless variety of oligosaccharides, which differ in the stereochemistry and position of glycosidic bonds, the type and orientation of substituent groups, and the number and type of branches. Glycans are far more information-dense than nucleic acids or proteins.
- Lectins, proteins with highly specific carbohydrate-binding domains, are commonly found on the outer surface of cells, where they initiate interaction with other cells. In vertebrates, oligosaccharide tags "read" by lectins govern the rate of degradation of certain peptide hormones, circulating proteins, and blood cells.

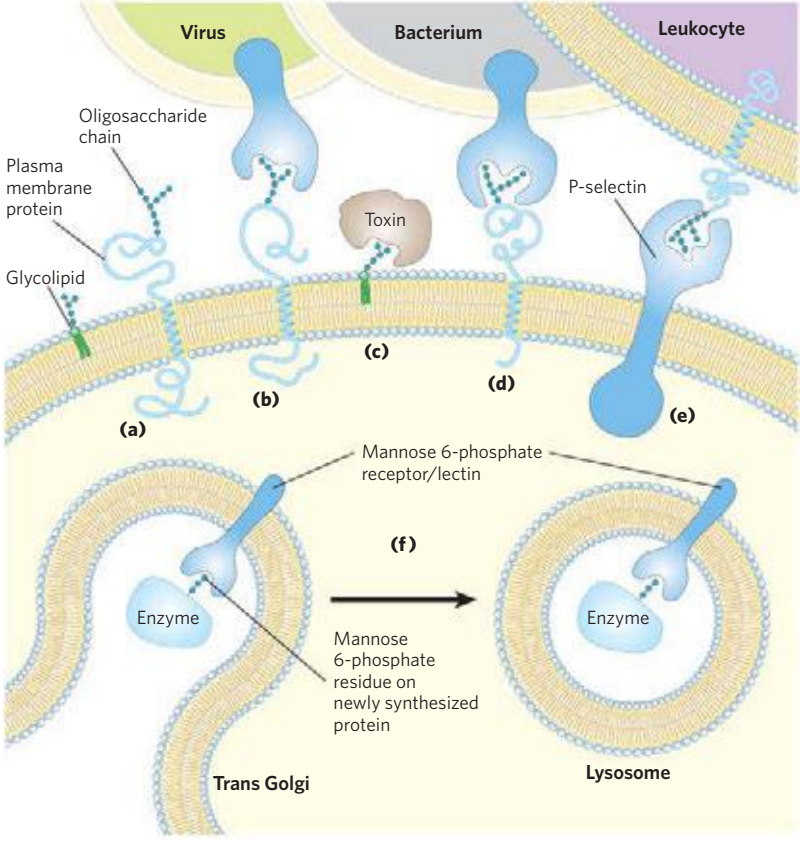


FIGURE 7–37 Role of oligosaccharides in recognition events at the cell surface and in the endomembrane system. (a) Oligosaccharides with unique structures (represented as strings of hexagons) are components of a variety of glycoproteins or glycolipids on the outer surface of plasma membranes. Their oligosaccharide moieties are bound by extracellular lectins with high specificity and affinity. **(b)** Viruses that infect animal cells, such as the influenza virus, bind to cell surface glycoproteins as the first step in infection. **(c)** Bacterial toxins, such as the cholera and pertussis toxins, bind to a surface glycolipid before entering a cell. **(d)** Some bacteria, such as *H. pylori*, adhere to and then colonize or infect animal cells. **(e)** Selectins (lectins) in the plasma membrane of certain cells mediate cell-cell interactions, such as those of leukocytes with the endothelial cells of the capillary wall at an infection site. **(f)** The mannose 6-phosphate receptor/lectin of the trans Golgi complex binds to the oligosaccharide of lysosomal enzymes, targeting them for transfer into the lysosome.

- Bacterial and viral pathogens and some eukaryotic parasites adhere to their animal cell targets by the binding of lectins in the pathogens to oligosaccharides on the target cell surface.
- X-ray crystallography of lectin-sugar complexes shows the detailed complementarity between the two molecules, which accounts for the strength and specificity of lectin interactions with carbohydrates.

7.5 Working with Carbohydrates

The growing appreciation of the importance of oligosaccharide structure in biological signaling and recognition has been the driving force behind the development of methods for analyzing the structure and stereochemistry of complex oligosaccharides. Oligosaccharide analysis is complicated by the fact that, unlike nucleic acids and proteins, oligosaccharides can be branched and are joined by a variety of linkages. The high charge density of many oligosaccharides and polysaccharides, and the relative lability of the sulfate esters in glycosaminoglycans, present further difficulties.

For simple, linear polymers such as amylose, the positions of the glycosidic bonds are determined by the classical method of exhaustive methylation: treating the intact polysaccharide with methyl iodide in a strongly basic medium to convert all free hydroxyls to acid-stable methyl ethers, then hydrolyzing the methylated polysaccharide in acid. The only free hydroxyls present in the monosaccharide derivatives so produced are those that were involved in glycosidic bonds. To determine the sequence of monosaccharide residues, including any branches that are present, exoglycosidases of known specificity are used to remove residues one at a time from the nonreducing end(s). The known specificity of these exoglycosidases often allows deduction of the position and stereochemistry of the linkages.

For analysis of the oligosaccharide moieties of glycoproteins and glycolipids, the oligosaccharides are released by purified enzymes—glycosidases that specifically cleave *O*- or *N*-linked oligosaccharides or lipases that remove lipid head groups. Alternatively, *O*-linked glycans can be released from glycoproteins by treatment with hydrazine.

The resulting mixtures of carbohydrates are resolved into their individual components by a variety of methods **(Fig. 7–38)**, including the same techniques used in protein and amino acid separation: fractional precipitation by solvents, and ion-exchange and size-exclusion chromatography (see Fig. 3–17). Highly purified lectins, attached covalently to an insoluble support, are commonly used in affinity chromatography of carbohydrates (see Fig. 3–17c).

Hydrolysis of oligosaccharides and polysaccharides in strong acid yields a mixture of monosaccharides, which may be identified and quantified by chromatographic techniques to yield the overall composition of the polymer.

Oligosaccharide analysis relies increasingly on mass spectrometry and high-resolution NMR spectroscopy. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS) and tandem mass spectrometry (MS/MS), both described in Chapter 3, are readily applicable to polar compounds such as oligosaccharides. MALDI MS is a very sensitive method for determining the mass of a molecular ion (in this case, the entire oligosaccharide chain; **Fig. 7–39**). MS/MS reveals the mass of the molecular ion and many of its fragments, which are usually the result of breakage of the glycosidic bonds. NMR analysis alone (see Box 4–5), especially for oligosaccharides of moderate size, can yield much information about sequence, linkage position, and anomeric carbon configuration. For example, the structure of the heparin segment shown as a space-filling model in Figure 7–22 was obtained entirely by NMR spectroscopy. Automated procedures and commercial instruments are used for the routine determination of oligosaccharide structure, but the sequencing of branched oligosaccharides joined by more than one type of bond remains a far more formidable task than determining the linear sequences of proteins and nucleic acids.

Another important tool in working with carbohydrates is chemical synthesis, which has proved to be a powerful approach to understanding the biological functions of glycosaminoglycans and oligosaccharides. The chemistry involved in such syntheses is difficult, but carbohydrate chemists can now synthesize short segments of almost any glycosaminoglycan, with correct stereochemistry, chain length, and sulfation pattern, and oligosaccharides significantly more complex than those shown in Figure 7–30. Solid-phase oligosaccharide synthesis is based on the same principles (and has the same advantages) as peptide synthesis (see Fig. 3–32), but requires a set of tools unique to carbohydrate chemistry: blocking groups and activating groups that allow the synthesis of glycosidic linkages with the correct hydroxyl group. Synthetic approaches of this type currently represent an area of great interest, because it is difficult to purify defined oligosaccharides in adequate quantities from natural sources.

To identify proteins with specific affinity for particular oligosaccharides, **oligosaccharide microarrays** are used. The principle is the same as for DNA microarrays (Figs 9–22 and 9–23), but the technical problems are more challenging. Pure oligosaccharides are attached to a glass slide in microdroplets, and the slide is exposed to a potential lectin (glycan-binding protein) that has been tagged with a fluorescent molecule **(Fig. 7–40)**. After all the unadsorbed protein is washed away, observation of the microarrays with a fluorescence microscope identifies the oligosaccharides that the lectin recognizes, and quantification of the fluorescence gives a rough measure of lectin-oligosaccharide affinity.

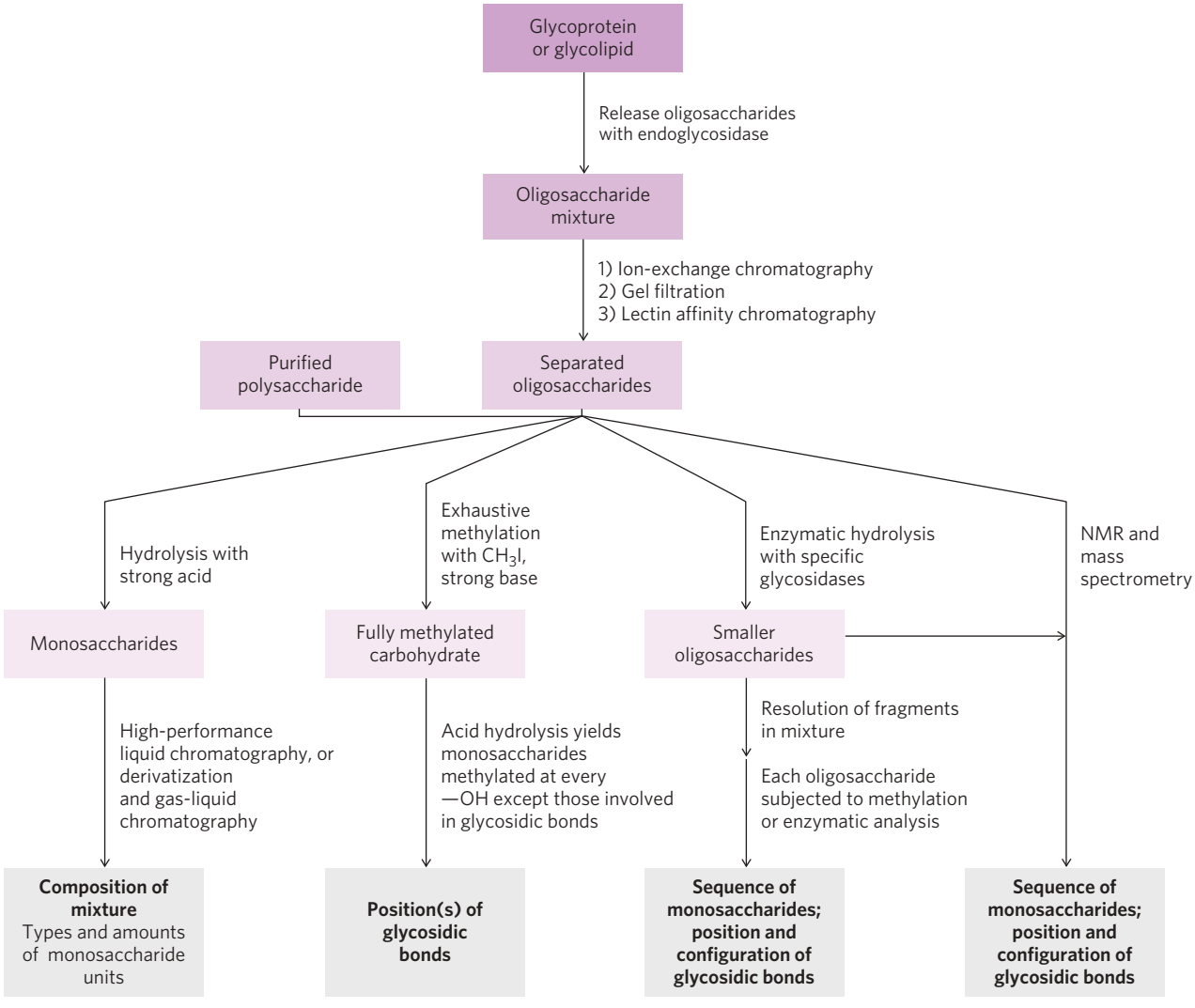


FIGURE 7–38 Methods of carbohydrate analysis. A carbohydrate purified in the first stage of the analysis often requires all four analytical routes for its complete characterization.

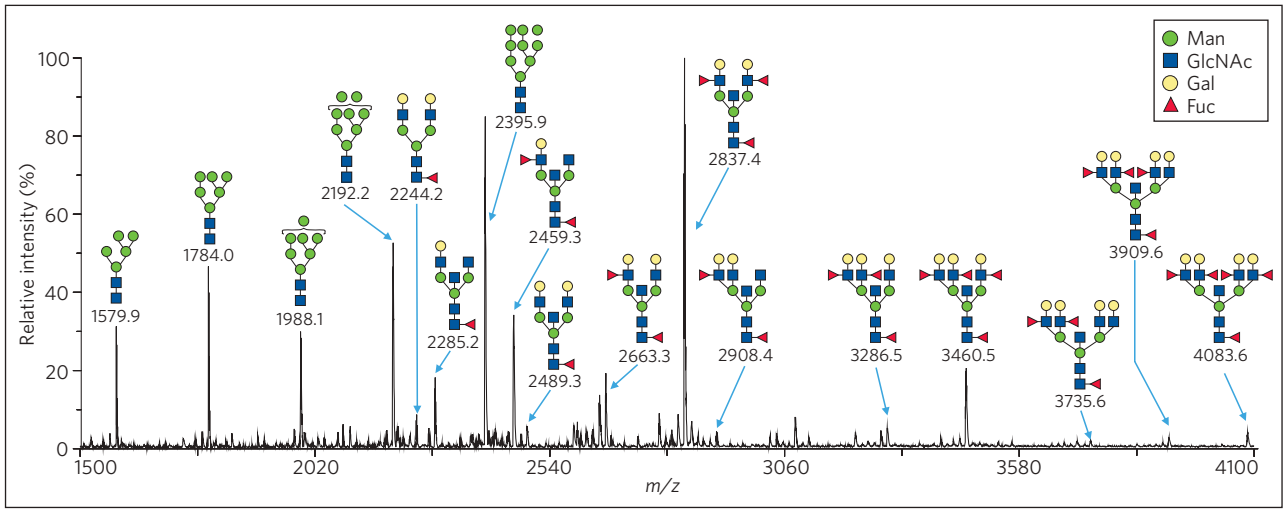


FIGURE 7–39 Separation and quantification of the oligosaccharides in a group of glycoproteins. In this experiment, the mixture of proteins extracted from kidney tissue was treated to release oligosaccharides from glycoproteins, and the oligosaccharides were analyzed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS). Each distinct oligosaccharide produces a peak at its molecular mass, and the area under the curve reflects the quantity of that oligosaccharide. The most prominent oligosaccharide here (mass 2837.4 u) is composed of 13 sugar residues; other oligosaccharides, containing as few as 7 and as many as 19 residues, were also resolved by this method.

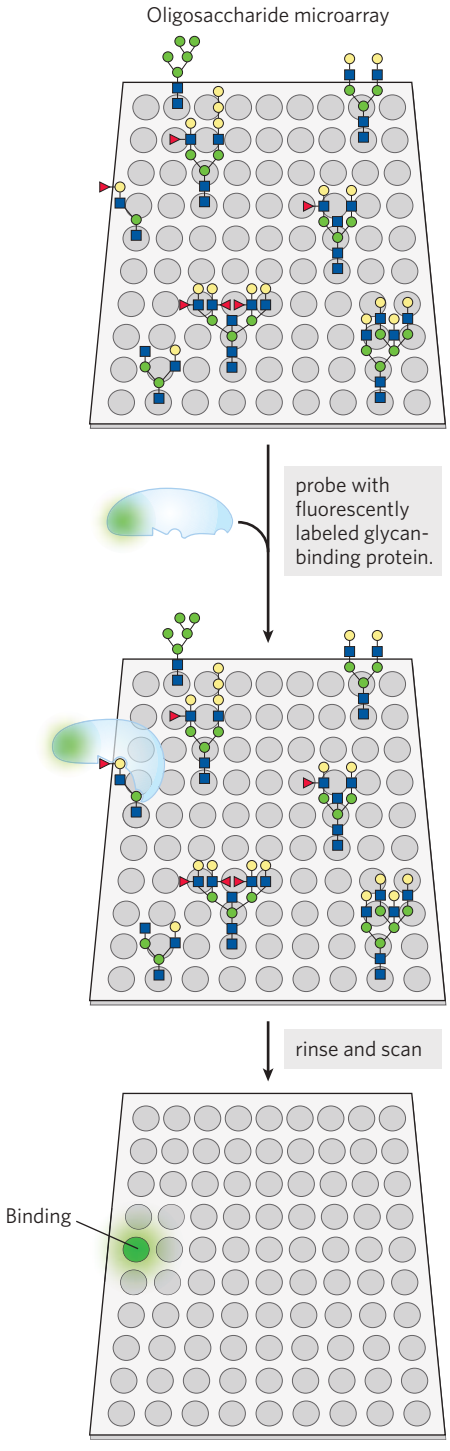


FIGURE 7–40 Oligosaccharide microarrays to determine the specificity and affinity of carbohydrate binding by lectins. Solutions of pure samples of oligosaccharides, synthesized or isolated from nature, are placed in microscopic droplets on a glass slide and attached to the glass through an inert spacer. Each spot represents a different oligosaccharide. The protein sample to be tested for its affinity for oligosaccharides is first conjugated with a fluorescent marker, then the sample is poured over the slide and allowed to equilibrate, and any nonadsorbed protein is washed away. Observation of the microarray with a fluorescence microscope shows which spots have adsorbed protein (they glow green), and assessment of the fluorescence intensity gives a rough measure of protein-oligosaccharide binding affinity.

SUMMARY 7.5 Working with Carbohydrates

- Establishing the complete structure of oligosaccharides and polysaccharides requires determination of linear sequence, branching positions, the configuration of each monosaccharide unit, and the positions of the glycosidic linkages—a more complex problem than protein and nucleic acid analysis.
- The structures of oligosaccharides and polysaccharides are usually determined by a combination of methods: specific enzymatic hydrolysis to determine stereochemistry at the glycosidic bond and to produce smaller fragments for further analysis; methylation to locate glycosidic bonds; and stepwise degradation to determine sequence and configuration of anomeric carbons.
- Mass spectrometry and high-resolution NMR spectroscopy, applicable to small samples of carbohydrate, yield essential information about sequence, configuration at anomeric and other carbons, and positions of glycosidic bonds.
- Solid-phase synthetic methods yield defined oligosaccharides that are of great value in exploring lectin-oligosaccharide interactions and may prove clinically useful.
- Microarrays of pure oligosaccharides are useful in determining the specificity and affinity of lectin binding to specific oligosaccharides.

Key Terms

Terms in bold are defined in the glossary.

glycoconjugate 243
carbohydrate 243
monosaccharide 243
oligosaccharide 243
disaccharide 243
polysaccharide 243
aldose 244
ketose 244
Fischer projection formulas 244
epimers 245
hemiacetal 245
hemiketal 245
anomers 246
anomeric carbon 246
pyranose 246
furanose 247
Haworth perspective formulas 247
mutarotation 248
reducing sugar 250
hemoglobin glycation 251
***O*-glycosidic bonds** 252
reducing end 252
glycan 254
starch 255
glycogen 255
cellulose 256
extracellular matrix (ECM) 260
glycosaminoglycan 260
hyaluronan 260
chondroitin sulfate 261
heparan sulfate 261
proteoglycan 263
glycoprotein 263
glycosphingolipid 264
syndecan 264
glypican 264
glycomics 267
lectin 269
selectins 270
oligosaccharide microarrays 274

Further Reading

General

Assadi-Porter, F.M., Maillet, E.L., Radek, J.T., Quijaada, J., Markley, J.L., & Max, M. (2010) Key amino acid residues involved in multi-point binding interactions between brazzein, a sweet protein, and the T1R2–T1R3 human sweet receptor. *J. Mol. Biol.* **398**, 584–599.

Hayes, J.E. (2007) Transdisciplinary perspectives on sweetness. *Chemosens. Percept*. **1**, 48–57.

Description of the theory of sweetness discussed in Box 7–2.

Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., & Etzler, M.E. (eds). (2009) *Essentials of Glycobiology*, 2nd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Structure, biosynthesis, metabolism, and function of glycosaminoglycans, proteoglycans, glycoproteins, and glycolipids, all presented at an intermediate level and very well illustrated. The book is available free online (www.ncbi.nlm.nih.gov/books/NBK1908).

Watanabe, H. & Tokuda, G. (2010) Cellulolytic systems in insects. *Annu. Rev. Entomol*. **55**, 609–632.

Glycosaminoglycans and Proteoglycans

Bishop, J.R., Schuksz, M., & Esko, J.D. (2007) Heparan sulfate proteoglycans fine-tune mammalian physiology. *Nature* **446**, 1030–1037.

Couchman, J.R. (2010) Transmembrane signaling proteoglycans. *Annu. Rev. Cell Dev. Biol*. **26**, 89–114.

Advanced review of the role of proteoglycans in signal transduction in vertebrates.

Fears, C.Y. & Woods, A. (2006) The role of syndecans in disease and wound healing. *Matrix Biol.* **25**, 443–456.

Intermediate-level review.

Frantz, C., Stewart, K.M., & Weaver, V.M. (2010) The extracellular matrix at a glance. *J. Cell Sci.* **123**, 4195–4200.

Poster-style summary of roles of glycans and glycoconjugates in the extracellular matrix.

Kirkpatrick, C.A. & Selleck, S.B. (2007) Heparan sulfate proteoglycans at a glance. *J. Cell Sci*. **120**, 1829–1832.

Poster-style summary of much useful information about proteoglycans.

Manon-Jensen, T., Itoh, Y., & Couchman, J.R. (2010) Proteoglycans in health and disease: the multiple roles of syndecan shedding. *FEBS J.* **277**, 3876–3889.

Roseman, S. (2001) Reflections on glycobiology. *J. Biol. Chem.* **276**, 41,527–41,542.

A masterful review of the history of carbohydrate and glycosaminoglycan studies, by one of the major contributors to this field.

Glycoproteins

Boraston, A. & Mulloy, B. (2010) Structural glycobiology: biosynthesis, recognition events, and new methods. *Curr. Opin. Struct. Biol.* **20**, 533–535.

Editorial introduction to a series of excellent reviews on these subjects published in this issue.

Luac, G. & Zoldos, V. (2010) Protein glycosylation—an evolutionary crossroad between genes and environment. *Mol. Biol. Syst.* **6**, 2372–2379.

A detailed discussion of the factors that determine whether and where a protein will be glycosylated.

Molinaro, M. (2007) *N*-glycan structure dictates extension of protein folding or onset of disposal. *Nat. Chem. Biol.* **3**, 313–320.

Intermediate-level review of the importance of protein glycosylation in the Golgi complex.

Sharon, N. & Gallagher, J. (2009) *Curr. Opin. Struct. Biol.* **19**, 495–497.

Editorial introduction to a series of good reviews on glycoproteins and glycolipids in this issue of the journal.

Weerapana, E. & Imperiali, B. (2006) Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16**, 91R–101R.

Intermediate-level review of the biosynthetic process of protein glycosylation.

Glycobiology and the Sugar Code

Boraston, A.B., Bolam, D.N., Gilbert, H.J., & Davies, G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem*. J. **382**, 769–781.

Excellent review of the structural basis for the specificity of sugar-binding proteins.

Gabius, H.-J., Andre, S., Jimenez-Barbero, J., Romero, A., & Solis, D. (2011) From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* **36**, 298–313.

Intermediate-level review of structural basis for lectin-sugar recognition.

Ghosh, P., Dahms, N.M., & Kornfeld, S. (2003) Mannose 6-phosphate receptors: new twists in the tale. *Nat. Rev. Mol. Cell Biol.* **4**, 202–212.

Hebert, D.N., Garman, S.C., & Molinari, M. (2005) The glycan code of the endoplasmic reticulum: asparagine-linked carbohydrates as protein maturation and quality control tags. *Trends Cell Biol.* **15**, 364–370.

Intermediate-level review.

Helenius, A. & Aebi, M. (2004) Roles of *N*-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049.

Lütteke, T., Bohne-Lang, A., Loss, A., Goetz, T., Frank, M., & von der Lieth, C.-W. (2006) Glycosciences.de: an internet portal to support glycomics and glycobiology research. *Glycobiology* **16**, 71R–81R.

McEver, R.P. & Zhu, C. (2010) Rolling cell adhesion. *Annu. Rev. Cell Dev. Biol*. **26,** 363–396.

Taylor, M.E. & Drickamer, K. (2006) *Introduction to Glycobiology*, 2nd edn, Oxford University Press, Oxford.

Working with Carbohydrates

Fukuda, M. (ed.). (2006) *Functional Glycomics*, Methods in Enzymology, Vol. 417, Academic Press, Inc., New York.

Fukuda, M. (ed.). (2006) *Glycobiology*, Methods in Enzymology, Vol. 415, Academic Press, Inc., New York.

Fukuda, M. (ed.). (2006) *Glycomics*, Methods in Enzymology, Vol. 416, Academic Press, Inc., New York.

Jay, A. (1996) The methylation reaction in carbohydrate analysis. *J. Carbohydr. Chem.* **15**, 897–923.

Paulson, J.C., Blixt, O., & Collins, B.E. (2006) Sweet spots in functional glycomics. *Nat. Chem. Biol.* **2**, 238–248.

Intermediate-level review of newly developed tools in glycobiology.

Zaia, J. (2008) Mass spectrometry and the emerging field of glycomics. *Chem. Biol.* **15**, 881–892.

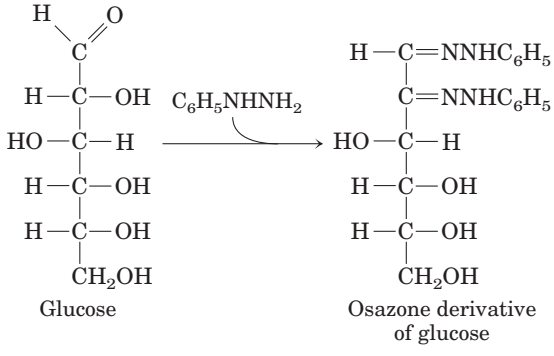
Excellent introduction to the use of mass spectrometry in studies of glycan structure and function.

Problems

1. Sugar Alcohols In the monosaccharide derivatives known as sugar alcohols, the carbonyl oxygen is reduced to a hydroxyl group. For example, D-glyceraldehyde can be reduced to glycerol. However, this sugar alcohol is no longer designated D or L. Why?

2. Recognizing Epimers Using Figure 7–3, identify the epimers of (a) D-allose, (b) D-gulose, and (c) D-ribose at C-2, C-3, and C-4.

3. Melting Points of Monosaccharide Osazone Derivatives Many carbohydrates react with phenylhydrazine ($C_6H_5NHNH_2$) to form bright yellow crystalline derivatives known as osazones:



The melting temperatures of these derivatives are easily determined and are characteristic for each osazone. This information was used to help identify monosaccharides before the development of HPLC or gas-liquid chromatography. Listed below are the melting points (MPs) of some aldose-osazone derivatives.

Monosaccharide	MP of anhydrous monosaccharide (°C)	MP of osazone derivative (°C)
Glucose	146	205
Mannose	132	205
Galactose	165–168	201
Talose	128–130	201

As the table shows, certain pairs of derivatives have the same melting points, although the underivatized monosaccharides do not. Why do glucose and mannose, and similarly galactose and talose, form osazone derivatives with the same melting points?

4. Configuration and Conformation Which bond(s) in α-D-glucose must be broken to change its configuration to β-D-glucose? Which bond(s) to convert D-glucose to D-mannose? Which bond(s) to convert one "chair" form of D-glucose to the other?

5. Deoxysugars Is D-2-deoxygalactose the same chemical as D-2-deoxyglucose? Explain.

6. Sugar Structures Describe the common structural features and the differences for each pair: (a) cellulose and glycogen; (b) D-glucose and D-fructose; (c) maltose and sucrose.

7. Reducing Sugars Draw the structural formula for α-D-glucosyl-(1→6)-D-mannosamine and circle the part of this structure that makes the compound a reducing sugar.

8. Hemiacetal and Glycosidic Linkages Explain the difference between a hemiacetal and a glycoside.

9. A Taste of Honey The fructose in honey is mainly in the β-D-pyranose form. This is one of the sweetest carbohydrates known, about twice as sweet as glucose; the β-D-furanose form of fructose is much less sweet. The sweetness of honey gradually decreases at a high temperature. Also, high-fructose corn syrup (a commercial product in which much of the glucose in corn syrup is converted to fructose) is used for sweetening *cold* but not *hot* drinks. What chemical property of fructose could account for both these observations?

10. Glucose Oxidase in Determination of Blood Glucose The enzyme glucose oxidase isolated from the mold *Penicillium notatum* catalyzes the oxidation of β-D-glucose to D-glucono-δ-lactone. This enzyme is highly specific for the β anomer of glucose and does not affect the α anomer. In spite of this specificity, the reaction catalyzed by glucose oxidase is commonly used in a clinical assay for total blood glucose—that is, for solutions consisting of a mixture of β- and α-D-glucose. What are the circumstances required to make this possible? Aside from allowing the detection of smaller quantities of glucose, what advantage does glucose oxidase offer over Fehling's reagent for measuring blood glucose?

11. Invertase "Inverts" Sucrose The hydrolysis of sucrose (specific rotation +66.5°) yields an equimolar mixture of D-glucose (specific rotation +52.5°) and D-fructose (specific rotation −92°). (See Problem 4 for details of specific rotation.)

(a) Suggest a convenient way to determine the rate of hydrolysis of sucrose by an enzyme preparation extracted from the lining of the small intestine.

(b) Explain why, in the food industry, an equimolar mixture of D-glucose and D-fructose formed by hydrolysis of sucrose is called invert sugar.

(c) The enzyme invertase (now commonly called sucrase) is allowed to act on a 10% (0.1 g/mL) solution of sucrose until hydrolysis is complete. What will be the observed optical rotation of the solution in a 10 cm cell? (Ignore a possible small contribution from the enzyme.)

12. Manufacture of Liquid-Filled Chocolates The manufacture of chocolates containing a liquid center is an interesting application of enzyme engineering. The flavored liquid center consists largely of an aqueous solution of sugars rich in fructose to provide sweetness. The technical dilemma is the following: the chocolate coating must be prepared by pouring hot melted chocolate over a solid (or almost solid) core, yet the final product must have a liquid, fructose-rich center. Suggest a way to solve this problem. (Hint: Sucrose is much less soluble than a mixture of glucose and fructose.)

13. Anomers of Sucrose? Lactose exists in two anomeric forms, but no anomeric forms of sucrose have been reported. Why?

14. Gentiobiose Gentiobiose (D-Glc(β1→6)D-Glc) is a disaccharide found in some plant glycosides. Draw the structure of gentiobiose based on its abbreviated name. Is it a reducing sugar? Does it undergo mutarotation?

15. Identifying Reducing Sugars Is *N*-acetyl-β-D-glucosamine (Fig. 7–9) a reducing sugar? What about D-gluconate? Is the disaccharide GlcN(α1↔1α)Glc a reducing sugar?

16. Cellulose Digestion Cellulose could provide a widely available and cheap form of glucose, but humans cannot digest it. Why not? If you were offered a procedure that allowed you to acquire this ability, would you accept? Why or why not?

17. Physical Properties of Cellulose and Glycogen The almost pure cellulose obtained from the seed threads of *Gossypium* (cotton) is tough, fibrous, and completely insoluble in water. In contrast, glycogen obtained from muscle or liver disperses readily in hot water to make a turbid solution. Despite their markedly different physical properties, both substances are (1→4)-linked D-glucose polymers of comparable molecular weight. What structural features of these two polysaccharides underlie their different physical properties? Explain the biological advantages of their respective properties.

18. Dimensions of a Polysaccharide Compare the dimensions of a molecule of cellulose and a molecule of amylose, each of M_r 200,000.

19. Growth Rate of Bamboo The stems of bamboo, a tropical grass, can grow at the phenomenal rate of 0.3 m/day under optimal conditions. Given that the stems are composed almost entirely of cellulose fibers oriented in the direction of growth, calculate the number of sugar residues per second that must be added enzymatically to growing cellulose chains to account for the growth rate. Each D-glucose unit contributes ~0.5 nm to the length of a cellulose molecule.

20. Glycogen as Energy Storage: How Long Can a Game Bird Fly? Since ancient times it has been observed that certain game birds, such as grouse, quail, and pheasants, are easily fatigued. The Greek historian Xenophon wrote, "The bustards . . . can be caught if one is quick in starting them up, for they will fly only a short distance, like partridges, and soon tire; and their flesh is delicious." The flight muscles of game birds rely almost entirely on the use of glucose 1-phosphate for energy, in the form of ATP (Chapter 14). The glucose 1-phosphate is formed by the breakdown of stored muscle glycogen, catalyzed by the enzyme glycogen phosphorylase. The rate of ATP production is limited by the rate at which glycogen can be broken down. During a "panic flight," the game bird's rate of glycogen breakdown is quite high, approximately 120 μmol/min of glucose 1-phosphate produced per gram of fresh tissue. Given that the flight muscles usually contain about 0.35% glycogen by weight, calculate how long a game bird can fly. (Assume the average molecular weight of a glucose residue in glycogen is 162 g/mol.)

21. Relative Stability of Two Conformers Explain why the two structures shown in Figure 7–18b are so different in energy (stability). Hint: See Figure 1–22.

22. Volume of Chondroitin Sulfate in Solution One critical function of chondroitin sulfate is to act as a lubricant in skeletal joints by creating a gel-like medium that is resilient to friction and shock. This function seems to be related to a distinctive property of chondroitin sulfate: the volume occupied by the molecule is much greater in solution than in the dehydrated solid. Why is the volume so much larger in solution?

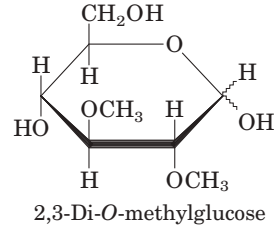
23. Heparin Interactions Heparin, a highly negatively charged glycosaminoglycan, is used clinically as an anticoagulant. It acts by binding several plasma proteins, including antithrombin III, an inhibitor of blood clotting. The 1:1 binding of heparin to antithrombin III seems to cause a conformational change in the protein that greatly increases its ability to inhibit clotting. What amino acid residues of antithrombin III are likely to interact with heparin?

24. Permutations of a Trisaccharide Think about how one might estimate the number of possible trisaccharides composed of *N*-acetylglucosamine 4-sulfate (GlcNAc4S) and glucuronic acid (GlcA), and draw 10 of them.

25. Effect of Sialic Acid on SDS Polyacrylamide Gel Electrophoresis Suppose you have four forms of a protein, all with identical amino acid sequence but containing zero, one, two, or three oligosaccharide chains, each ending in a single sialic acid residue. Draw the gel pattern you would expect when a mixture of these four glycoproteins is subjected to SDS polyacrylamide gel electrophoresis (see Fig. 3–18) and stained for protein. Identify any bands in your drawing.

26. Information Content of Oligosaccharides The carbohydrate portion of some glycoproteins may serve as a cellular recognition site. To perform this function, the oligosaccharide moiety must have the potential to exist in a large variety of forms. Which can produce a greater variety of structures: oligopeptides composed of five different amino acid residues, or oligosaccharides composed of five different monosaccharide residues? Explain.

27. Determination of the Extent of Branching in Amylopectin The amount of branching (number of (α1→6) glycosidic bonds) in amylopectin can be determined by the following procedure. A sample of amylopectin is exhaustively methylated—treated with a methylating agent (methyl iodide) that replaces the hydrogen of every sugar hydroxyl with a methyl group, converting —OH to —OCH_3. All the glycosidic bonds in the treated sample are then hydrolyzed in aqueous acid, and the amount of 2,3-di-*O*-methylglucose so formed is determined.



2,3-Di-*O*-methylglucose

(a) Explain the basis of this procedure for determining the number of (α1→6) branch points in amylopectin. What happens to the unbranched glucose residues in amylopectin during the methylation and hydrolysis procedure?

(b) A 258 mg sample of amylopectin treated as described above yielded 12.4 mg of 2,3-di-*O*-methylglucose. Determine

what percentage of the glucose residues in the amylopectin contained an (α1→6) branch. (Assume that the average molecular weight of a glucose residue in amylopectin is 162 g/mol.)

28. Structural Analysis of a Polysaccharide A polysaccharide of unknown structure was isolated, subjected to exhaustive methylation, and hydrolyzed. Analysis of the products revealed three methylated sugars: 2,3,4-tri-*O*-methyl-D-glucose, 2,4-di-*O*-methyl-D-glucose, and 2,3,4,6-tetra-*O*-methyl-D-glucose, in the ratio 20:1:1. What is the structure of the polysaccharide?

Data Analysis Problem

29. Determining the Structure of ABO Blood Group Antigens The human ABO blood group system was first discovered in 1901, and in 1924 this trait was shown to be inherited at a single gene locus with three alleles. In 1960, W. T. J. Morgan published a paper summarizing what was known at that time about the structure of the ABO antigen molecules. When the paper was published, the complete structures of the A, B, and O antigens were not yet known; this paper is an example of what scientific knowledge looks like "in the making."

In any attempt to determine the structure of an unknown biological compound, researchers must deal with two fundamental problems: (1) If you don't know what *it* is, how do you know if *it* is pure? (2) If you don't know what *it* is, how do you know that your extraction and purification conditions have not changed *its* structure? Morgan addressed problem 1 through several methods. One method is described in his paper (p. 312) as observing "constant analytical values after fractional solubility tests." In this case, "analytical values" are measurements of chemical composition, melting point, and so forth.

(a) Based on your understanding of chemical techniques, what could Morgan mean by "fractional solubility tests"?

(b) Why would the analytical values obtained from fractional solubility tests of a *pure* substance be constant, and those of an *impure* substance not be constant?

Morgan addressed problem 2 by using an assay to measure the immunological activity of the substance present in different samples.

(c) Why was it important for Morgan's studies, and especially for addressing problem 2, that this activity assay be quantitative (measuring a level of activity) rather than simply qualitative (measuring only the presence or absence of a substance)?

The structure of the blood group antigens is shown in Figure 10–15. In his paper (p. 314), Morgan listed several properties of the three antigens, A, B, and O, that were known at that time:

1. Type B antigen has a higher content of galactose than A or O.
2. Type A antigen contains more total amino sugars than B or O.
3. The glucosamine/galactosamine ratio for the A antigen is roughly 1.2; for B, it is roughly 2.5.

(d) Which of these findings is (are) consistent with the known structures of the blood group antigens?

(e) How do you explain the discrepancies between Morgan's data and the known structures?

In later work, Morgan and his colleagues used a clever technique to obtain structural information about the blood group antigens. Enzymes had been found that would specifically degrade the antigens. However, these were available only as crude enzyme preparations, perhaps containing more than one enzyme of unknown specificity. Degradation of the blood type antigens by these crude enzymes could be inhibited by the addition of particular sugar molecules to the reaction. Only sugars found in the blood type antigens would cause this inhibition. One enzyme preparation, isolated from the protozoan *Trichomonas foetus*, would degrade all three antigens and was inhibited by the addition of particular sugars. The results of these studies are summarized in the table below, showing the percentage of substrate remaining unchanged when the *T. foetus* enzyme acted on the blood group antigens in the presence of sugars.

	Unchanged substrate (%)		
Sugar added	**A antigen**	**B antigen**	**O antigen**
Control—no sugar	3	1	1
L-Fucose	3	1	100
D-Fucose	3	1	1
L-Galactose	3	1	3
D-Galactose	6	100	1
N-Acetylglucosamine	3	1	1
N-Acetylgalactosamine	100	6	1

For the O antigen, a comparison of the control and L-fucose results shows that L-fucose inhibits the degradation of the antigen. This is an example of product inhibition, in which an excess of reaction product shifts the equilibrium of the reaction, preventing further breakdown of substrate.

(f) Although the O antigen contains galactose, *N*-acetylglucosamine, and *N*-acetylgalactosamine, none of these sugars inhibited the degradation of this antigen. Based on these data, is the enzyme preparation from *T. foetus* an endo- or exoglycosidase? (Endoglycosidases cut bonds between interior residues; exoglycosidases remove one residue at a time from the end of a polymer.) Explain your reasoning.

(g) Fucose is also present in the A and B antigens. Based on the structure of these antigens, why does fucose fail to prevent their degradation by the *T. foetus* enzyme? What structure would be produced?

(h) Which of the results in (f) and (g) are consistent with the structures shown in Figure 10–15? Explain your reasoning.

Reference

Morgan, W.T.J. (1960) The Croonian Lecture: a contribution to human biochemical genetics; the chemical basis of blood-group specificity. *Proc. R. Soc. Lond. B Biol. Sci.* **151**, 308–347.

8

Nucleotides and Nucleic Acids

Nucleotides have a variety of roles in cellular metabolism. They are the energy currency in metabolic transactions, the essential chemical links in the response of cells to hormones and other extracellular stimuli, and the structural components of an array of enzyme cofactors and metabolic intermediates. And, last but certainly not least, they are the constituents of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), the molecular repositories of genetic information. The structure of every protein, and ultimately of every biomolecule and cellular component, is a product of information programmed into the nucleotide sequence of a cell's nucleic acids. The ability to store and transmit genetic information from one generation to the next is a fundamental condition for life.

This chapter provides an overview of the chemical nature of the nucleotides and nucleic acids found in most cells; a more detailed examination of the function of nucleic acids is the focus of Part III of this text.

8.1 Some Basics

Nucleotides, Building Blocks of Nucleic Acids The amino acid sequence of every protein in a cell, and the nucleotide sequence of every RNA, is specified by a nucleotide sequence in the cell's DNA. A segment of a DNA molecule that contains the information required for the synthesis of a functional biological product, whether protein or RNA, is referred to as a **gene**. A cell typically has many thousands of genes, and DNA molecules, not surprisingly, tend to be very large. The storage and transmission of biological information are the only known functions of DNA.

RNAs have a broader range of functions, and several classes are found in cells. **Ribosomal RNAs (rRNAs)** are components of ribosomes, the complexes that carry out the synthesis of proteins. **Messenger RNAs (mRNAs)** are intermediaries, carrying genetic information from one or a few genes to a ribosome, where the corresponding proteins can be synthesized. **Transfer RNAs (tRNAs)** are adapter molecules that faithfully translate the information in mRNA into a specific sequence of amino acids. In addition to these major classes there is a wide variety of RNAs with special functions, described in depth in Part III.

Nucleotides and Nucleic Acids Have Characteristic Bases and Pentoses

Nucleotides have three characteristic components: (1) a nitrogenous (nitrogen-containing) base, (2) a pentose, and (3) one or more phosphates **(Fig. 8–1)**. The molecule without a phosphate group is called a **nucleoside**. The nitrogenous bases are derivatives of two parent

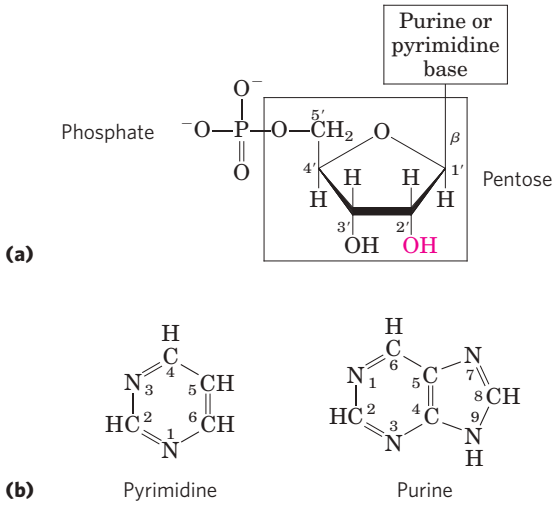


FIGURE 8–1 Structure of nucleotides. (a) General structure showing the numbering convention for the pentose ring. This is a ribonucleotide. In deoxyribonucleotides the —OH group on the 2′ carbon (in red) is replaced with H. **(b)** The parent compounds of the pyrimidine and purine bases of nucleotides and nucleic acids, showing the numbering conventions.

compounds, **pyrimidine** and **purine**. The bases and pentoses of the common nucleotides are heterocyclic compounds.

KEY CONVENTION: The carbon and nitrogen atoms in the parent structures are conventionally numbered to facilitate the naming and identification of the many derivative compounds. The convention for the pentose ring follows rules outlined in Chapter 7, but in the pentoses of nucleotides and nucleosides the carbon numbers are given a prime (′) designation to distinguish them from the numbered atoms of the nitrogenous bases. ■

The base of a nucleotide is joined covalently (at N-1 of pyrimidines and N-9 of purines) in an *N*-β-glycosyl bond to the 1′ carbon of the pentose, and the phosphate is esterified to the 5′ carbon. The *N*-β-glycosyl bond is formed by removal of the elements of water (a hydroxyl group from the pentose and hydrogen from the base), as in *O*-glycosidic bond formation (see Fig. 7–30).

Both DNA and RNA contain two major purine bases, **adenine** (A) and **guanine** (G), and two major pyrimidines. In both DNA and RNA one of the pyrimidines is **cytosine** (C), but the second common pyrimidine is not the same in both: it is **thymine** (T) in DNA and **uracil** (U) in RNA. Only occasionally does thymine occur in RNA or uracil in DNA. The structures of the five major bases are shown in **Figure 8–2**, and the nomenclature of their corresponding nucleotides and nucleosides is summarized in Table 8–1.

Nucleic acids have two kinds of pentoses. The recurring deoxyribonucleotide units of DNA contain 2′-deoxy-D-ribose, and the ribonucleotide units of RNA contain D-ribose. In nucleotides, both types of pentoses are in their β-furanose (closed five-membered ring) form. As **Figure 8–3** shows, the pentose ring is not planar but occurs in one of a variety of conformations generally described as "puckered."

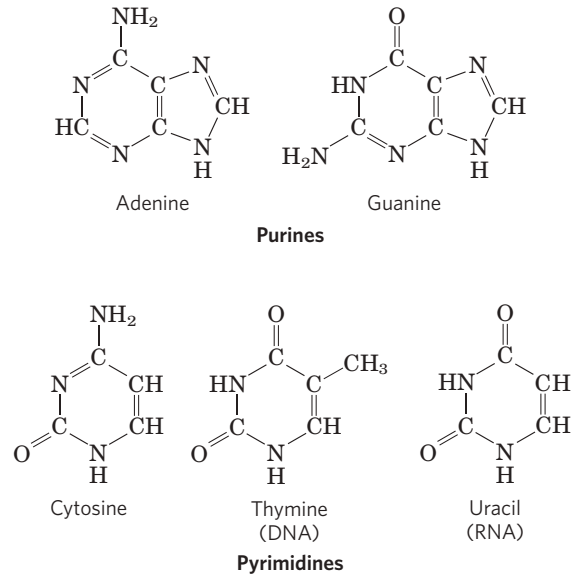


FIGURE 8–2 Major purine and pyrimidine bases of nucleic acids. Some of the common names of these bases reflect the circumstances of their discovery. Guanine, for example, was first isolated from guano (bird manure), and thymine was first isolated from thymus tissue.

KEY CONVENTION: Although DNA and RNA seem to have two distinctions—different pentoses and the presence of uracil in RNA and thymine in DNA—it is the pentoses that define the identity of a nucleic acid. If the nucleic acid contains 2′-deoxy-D-ribose, it is DNA by definition even though it may contain uracil. Similarly, if the nucleic acid contains D-ribose, it is RNA regardless of its base composition. ■

TABLE 8–1 Nucleotide and Nucleic Acid Nomenclature

Base	Nucleoside	Nucleotide	Nucleic acid
Purines			
Adenine	Adenosine	Adenylate	RNA
	Deoxyadenosine	Deoxyadenylate	DNA
Guanine	Guanosine	Guanylate	RNA
	Deoxyguanosine	Deoxyguanylate	DNA
Pyrimidines			
Cytosine	Cytidine	Cytidylate	RNA
	Deoxycytidine	Deoxycytidylate	DNA
Thymine	Thymidine or deoxythymidine	Thymidylate or deoxythymidylate	DNA
Uracil	Uridine	Uridylate	RNA

Note: "Nucleoside" and "nucleotide" are generic terms that include both ribo- and deoxyribo- forms. Also, ribonucleosides and ribonucleotides are here designated simply as nucleosides and nucleotides (e.g., riboadenosine as adenosine), and deoxyribonucleosides and deoxyribonucleotides as deoxynucleosides and deoxynucleotides (e.g., deoxyriboadenosine as deoxyadenosine). Both forms of naming are acceptable, but the shortened names are more commonly used. Thymine is an exception; "ribothymidine" is used to describe its unusual occurrence in RNA.

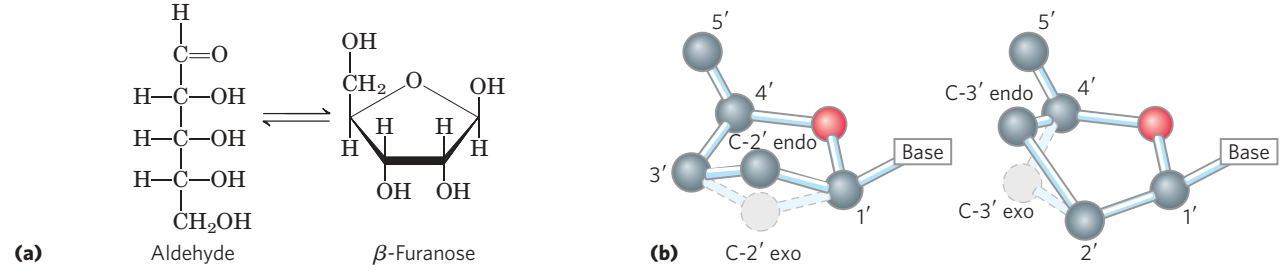
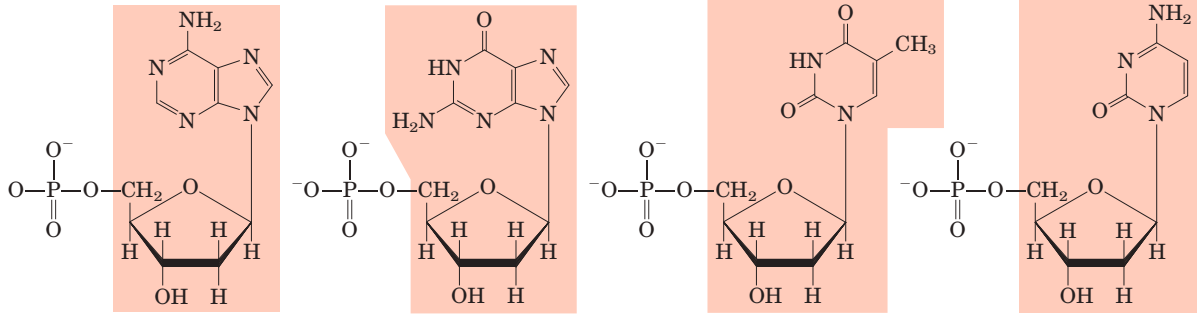


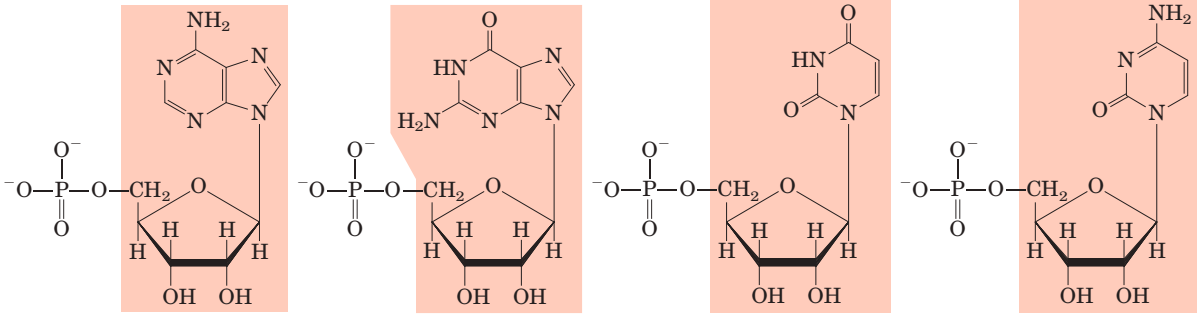
FIGURE 8–3 Conformations of ribose. (a) In solution, the straight-chain (aldehyde) and ring (β-furanose) forms of free ribose are in equilibrium. RNA contains only the ring form, β-D-ribofuranose. Deoxyribose undergoes a similar interconversion in solution, but in DNA exists solely as β-2′-deoxy-D-ribofuranose. **(b)** Ribofuranose rings in nucleotides can exist in four different puckered conformations. In all cases, four of the five atoms are nearly in a single plane. The fifth atom (C-2′ or C-3′) is on either the same (endo) or the opposite (exo) side of the plane relative to the C-5′ atom.

Figure 8–4 gives the structures and names of the four major **deoxyribonucleotides** (deoxyribonucleoside 5′-monophosphates), the structural units of DNAs, and the four major **ribonucleotides** (ribonucleoside 5′-monophosphates), the structural units of RNAs.



Nucleotide:	Deoxyadenylate (deoxyadenosine 5′-monophosphate)	Deoxyguanylate (deoxyguanosine 5′-monophosphate)	Deoxythymidylate (deoxythymidine 5′-monophosphate)	Deoxycytidylate (deoxycytidine 5′-monophosphate)
Symbols:	A, dA, dAMP	G, dG, dGMP	T, dT, dTMP	C, dC, dCMP
Nucleoside:	Deoxyadenosine	Deoxyguanosine	Deoxythymidine	Deoxycytidine

(a) Deoxyribonucleotides



Nucleotide:	Adenylate (adenosine 5′-monophosphate)	Guanylate (guanosine 5′-monophosphate)	Uridylate (uridine 5′-monophosphate)	Cytidylate (cytidine 5′-monophosphate)
Symbols:	A, AMP	G, GMP	U, UMP	C, CMP
Nucleoside:	Adenosine	Guanosine	Uridine	Cytidine

(b) Ribonucleotides

FIGURE 8–4 Deoxyribonucleotides and ribonucleotides of nucleic acids. All nucleotides are shown in their free form at pH 7.0. The nucleotide units of DNA **(a)** are usually symbolized as A, G, T, and C, sometimes as dA, dG, dT, and dC; those of RNA **(b)** as A, G, U, and C. In their free form the deoxyribonucleotides are commonly abbreviated dAMP, dGMP, dTMP, and dCMP; the ribonucleotides, AMP, GMP, UMP, and CMP. For each nucleotide in the figure, the more common name is followed by the complete name in parentheses. All abbreviations assume that the phosphate group is at the 5′ position. The nucleoside portion of each molecule is shaded in light red. In this and the following illustrations, the ring carbons are not shown.

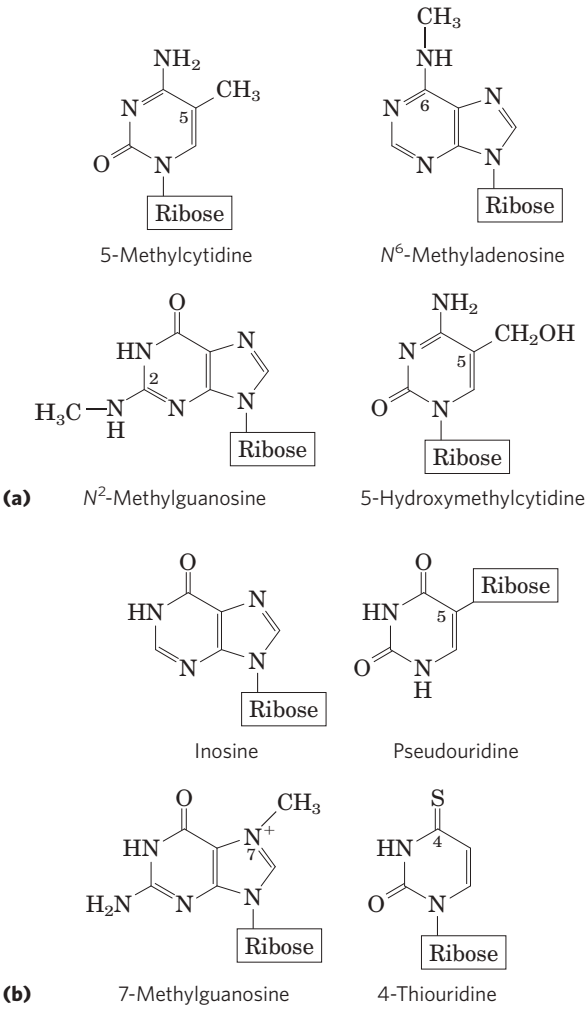


FIGURE 8–5 Some minor purine and pyrimidine bases, shown as the nucleosides. (a) Minor bases of DNA. 5-Methylcytidine occurs in the DNA of animals and higher plants, N^6-methyladenosine in bacterial DNA, and 5-hydroxymethylcytidine in the DNA of animals and of bacteria infected with certain bacteriophages. **(b)** Some minor bases of tRNAs. Inosine contains the base hypoxanthine. Note that pseudouridine, like uridine, contains uracil; they are distinct in the point of attachment to the ribose—in uridine, uracil is attached through N-1, the usual attachment point for pyrimidines; in pseudouridine, through C-5.

Although nucleotides bearing the major purines and pyrimidines are most common, both DNA and RNA also contain some minor bases **(Fig. 8–5)**. In DNA the most common of these are methylated forms of the major bases; in some viral DNAs, certain bases may be hydroxymethylated or glucosylated. Altered or unusual bases in DNA molecules often have roles in regulating or protecting the genetic information. Minor bases of many types are also found in RNAs, especially in tRNAs (see Fig. 8–25 and Fig. 26–22).

KEY CONVENTION: The nomenclature for the minor bases can be confusing. Like the major bases, many have common names—hypoxanthine, for example, shown as its nucleoside inosine in Figure 8–5. When an atom in the purine or pyrimidine ring is substituted, the usual convention (used here) is simply to indicate the ring position of the substituent by its number—for example, 5-methylcytosine, 7-methylguanine, and 5-hydroxymethylcytosine (shown as the nucleosides in Fig. 8–5). The element to which the substituent is attached (N, C, O) is not identified. The convention changes when the substituted atom is exocyclic (not within the ring structure), in which case the type of atom is identified and the ring position to which it is attached is denoted with a superscript. The amino nitrogen attached to C-6 of adenine is N^6; similarly, the carbonyl oxygen and amino nitrogen at C-6 and C-2 of guanine are O^6 and N^2, respectively. Examples of this nomenclature are N^6-methyladenosine and N^2-methylguanosine (Fig. 8–5). ■

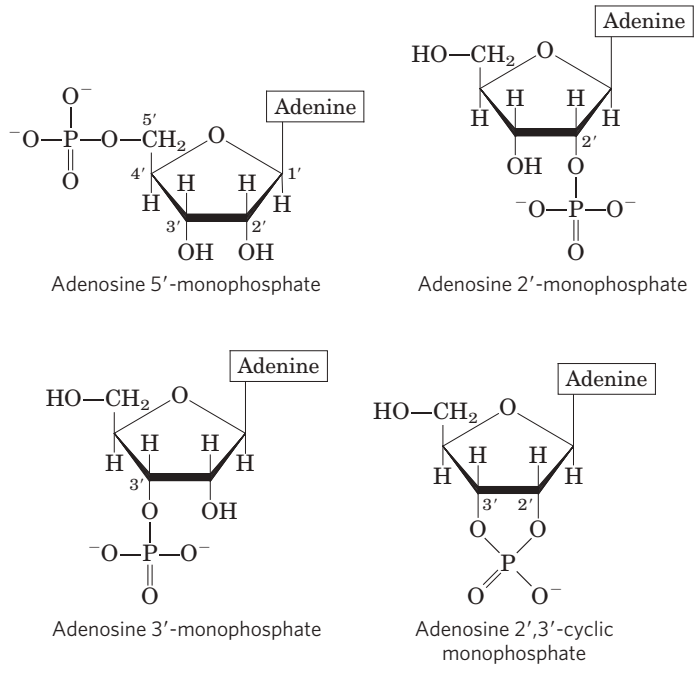


FIGURE 8–6 Some adenosine monophosphates. Adenosine 2′-monophosphate, 3′-monophosphate, and 2′,3′-cyclic monophosphate are formed by enzymatic and alkaline hydrolysis of RNA.

Cells also contain nucleotides with phosphate groups in positions other than on the 5′ carbon **(Fig. 8–6)**. **Ribonucleoside 2′,3′-cyclic monophosphates** are isolatable intermediates, and **ribonucleoside 3′-monophosphates** are end products of the hydrolysis of RNA by certain ribonucleases. Other variations are adenosine 3′,5′-cyclic monophosphate (cAMP) and guanosine 3′,5′-cyclic monophosphate (cGMP), considered at the end of this chapter.

Phosphodiester Bonds Link Successive Nucleotides in Nucleic Acids

The successive nucleotides of both DNA and RNA are covalently linked through phosphate-group "bridges," in which the 5′-phosphate group of one nucleotide unit is joined to the 3′-hydroxyl group of the next nucleotide,

creating a **phosphodiester linkage** (Fig. 8–7). Thus the covalent backbones of nucleic acids consist of alternating phosphate and pentose residues, and the nitrogenous bases may be regarded as side groups joined to the backbone at regular intervals. The backbones of both DNA and RNA are hydrophilic. The hydroxyl groups of the sugar residues form hydrogen bonds with water. The phosphate groups, with a pK_a near 0, are completely ionized and negatively charged at pH 7, and the negative charges are generally neutralized by ionic interactions with positive charges on proteins, metal ions, and polyamines.

KEY CONVENTION: All the phosphodiester linkages in DNA and RNA have the same orientation along the chain (Fig. 8–7), giving each linear nucleic acid strand a specific polarity and distinct 5′ and 3′ ends. By definition, the **5′ end** lacks a nucleotide at the 5′ position and the **3′ end** lacks a nucleotide at the 3′ position. Other groups (most often one or more phosphates) may be present on one or both ends. The 5′ to 3′ orientation of a strand of nucleic acid refers to the *ends* of the strand, not the orientation of the individual phosphodiester bonds linking its constituent nucleotides. ■

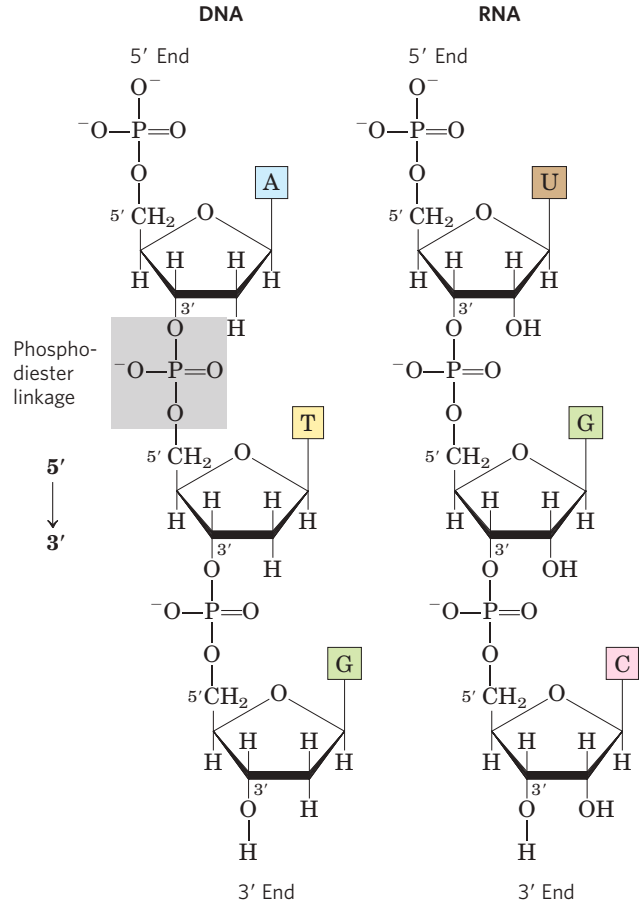


FIGURE 8–7 Phosphodiester linkages in the covalent backbone of DNA and RNA. The phosphodiester bonds (one of which is shaded in the DNA) link successive nucleotide units. The backbone of alternating pentose and phosphate groups in both types of nucleic acid is highly polar. The 5′ and 3′ ends of the macromolecule may be free or may have an attached phosphoryl group.

The covalent backbone of DNA and RNA is subject to slow, nonenzymatic hydrolysis of the phosphodiester bonds. In the test tube, RNA is hydrolyzed rapidly under alkaline conditions, but DNA is not; the 2′-hydroxyl groups in RNA (absent in DNA) are directly involved in the process. Cyclic 2′,3′-monophosphate nucleotides are the first products of the action of alkali on RNA and are rapidly hydrolyzed further to yield a mixture of 2′- and 3′-nucleoside monophosphates (Fig. 8–8).

The nucleotide sequences of nucleic acids can be represented schematically, as illustrated below by a segment of DNA with five nucleotide units. The phosphate

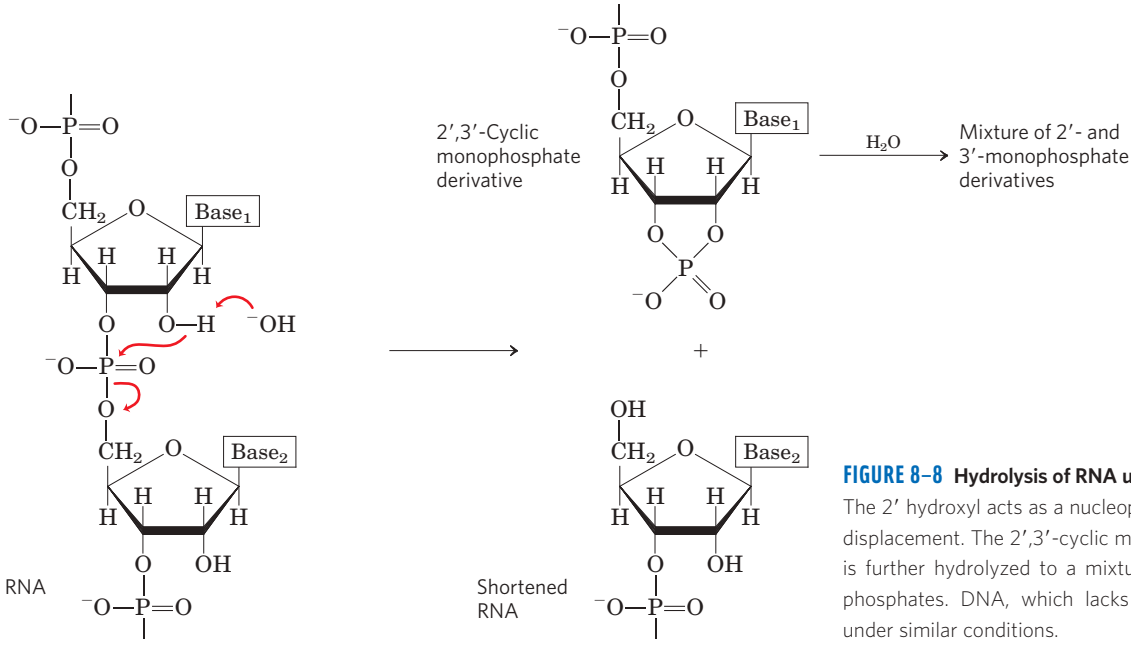
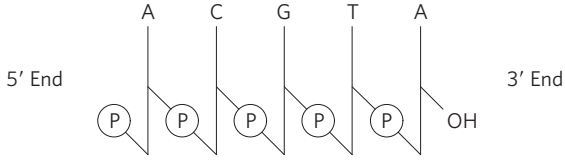


FIGURE 8–8 Hydrolysis of RNA under alkaline conditions. The 2′ hydroxyl acts as a nucleophile in an intramolecular displacement. The 2′,3′-cyclic monophosphate derivative is further hydrolyzed to a mixture of 2′- and 3′-monophosphates. DNA, which lacks 2′ hydroxyls, is stable under similar conditions.

groups are symbolized by Ⓟ, and each deoxyribose is symbolized by a vertical line, from C-1′ at the top to C-5′ at the bottom (but keep in mind that the sugar is always in its closed-ring β-furanose form in nucleic acids). The connecting lines between nucleotides (which pass through Ⓟ) are drawn diagonally from the middle (C-3′) of the deoxyribose of one nucleotide to the bottom (C-5′) of the next.



Some simpler representations of this pentadeoxyribonucleotide are pA-C-G-T-A_{OH}, pApCpGpTpA, and pACGTA.

KEY CONVENTION: The sequence of a single strand of nucleic acid is always written with the 5′ end at the left and the 3′ end at the right—that is, in the 5′→3′ direction. ■

A short nucleic acid is referred to as an **oligonucleotide**. The definition of "short" is somewhat arbitrary, but polymers containing 50 or fewer nucleotides are generally called oligonucleotides. A longer nucleic acid is called a **polynucleotide**.

The Properties of Nucleotide Bases Affect the Three-Dimensional Structure of Nucleic Acids

Free pyrimidines and purines are weakly basic compounds and thus are called bases. The purines and pyrimidines common in DNA and RNA are aromatic molecules (Fig. 8–2), a property with important consequences for the structure, electron distribution, and light absorption of nucleic acids. Electron delocalization among atoms in the ring gives most of the bonds partial double-bond character. One result is that pyrimidines are planar molecules and purines are very nearly planar, with a slight pucker. Free pyrimidine and purine bases

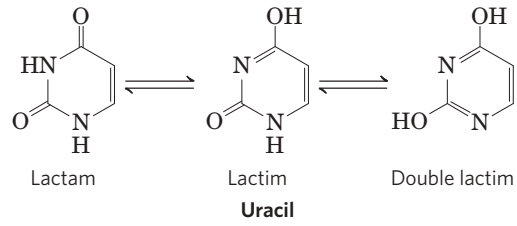


FIGURE 8–9 Tautomeric forms of uracil. The lactam form predominates at pH 7.0; the other forms become more prominent as pH decreases. The other free pyrimidines and the free purines also have tautomeric forms, but they are more rarely encountered.

may exist in two or more tautomeric forms depending on the pH. Uracil, for example, occurs in lactam, lactim, and double lactim forms **(Fig. 8–9)**. The structures shown in Figure 8–2 are the tautomers that predominate at pH 7.0. All nucleotide bases absorb UV light, and nucleic acids are characterized by a strong absorption at wavelengths near 260 nm **(Fig. 8–10)**.

The purine and pyrimidine bases are hydrophobic and relatively insoluble in water at the near-neutral pH of the cell. At acidic or alkaline pH the bases become charged and their solubility in water increases. Hydrophobic stacking interactions in which two or more bases are positioned with the planes of their rings parallel (like a stack of coins) are one of two important modes of interaction between bases in nucleic acids. The stacking also involves a combination of van der Waals and dipole-dipole interactions between the bases. Base stacking helps to minimize contact of the bases with water, and base-stacking interactions are very important in stabilizing the three-dimensional structure of nucleic acids, as described later.

The functional groups of pyrimidines and purines are ring nitrogens, carbonyl groups, and exocyclic amino groups. Hydrogen bonds involving the amino and carbonyl groups are the most important mode of interaction between two (and occasionally three or four) complementary strands of nucleic acid. The most common

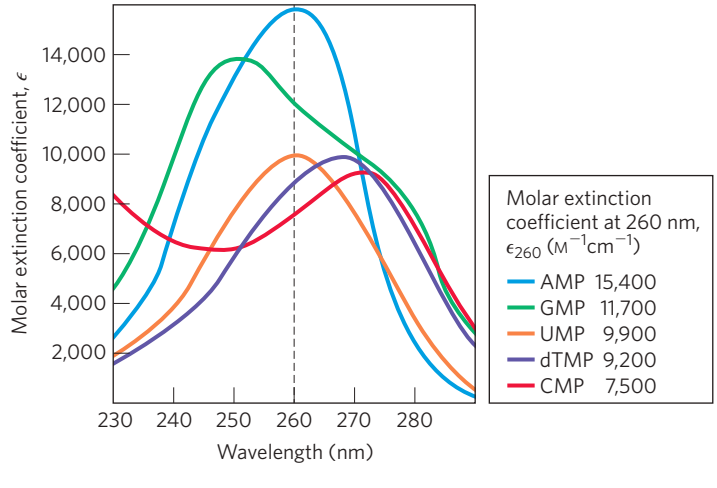


FIGURE 8–10 Absorption spectra of the common nucleotides. The spectra are shown as the variation in molar extinction coefficient with wavelength. The molar extinction coefficients at 260 nm and pH 7.0 (ε_{260}) are listed in the table. The spectra of corresponding ribonucleotides and deoxyribonucleotides, as well as the nucleosides, are essentially identical. For mixtures of nucleotides, a wavelength of 260 nm (dashed vertical line) is used for absorption measurements.

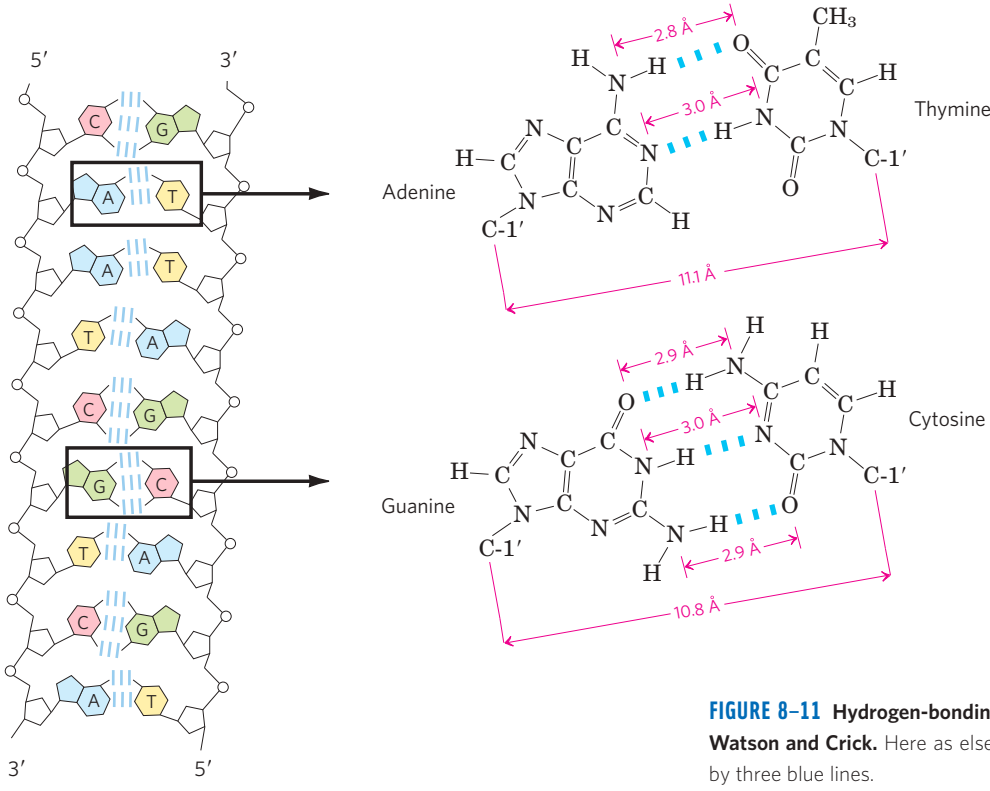


FIGURE 8–11 Hydrogen-bonding patterns in the base pairs defined by Watson and Crick. Here as elsewhere, hydrogen bonds are represented by three blue lines.

hydrogen-bonding patterns are those defined by James D. Watson and Francis Crick in 1953, in which A bonds specifically to T (or U) and G bonds to C **(Fig. 8–11)**. These two types of **base pairs** predominate in double-stranded DNA and RNA, and the tautomers shown in Figure 8–2 are responsible for these patterns. It is this specific pairing of bases that permits the duplication of genetic information, as we shall discuss later in this chapter.

James D. Watson

Francis Crick, 1916–2004

SUMMARY 8.1 Some Basics

- A nucleotide consists of a nitrogenous base (purine or pyrimidine), a pentose sugar, and one or more phosphate groups. Nucleic acids are polymers of nucleotides, joined together by phosphodiester linkages between the 5′-hydroxyl group of one pentose and the 3′-hydroxyl group of the next.
- There are two types of nucleic acid: RNA and DNA. The nucleotides in RNA contain ribose, and the common pyrimidine bases are uracil and cytosine. In DNA, the nucleotides contain 2′-deoxyribose, and the common pyrimidine bases are thymine and cytosine. The primary purines are adenine and guanine in both RNA and DNA.

8.2 Nucleic Acid Structure

The discovery of the structure of DNA by Watson and Crick in 1953 gave rise to entirely new disciplines and influenced the course of many established ones. In this section we focus on DNA structure, some of the events that led to its discovery, and more recent refinements in our understanding of DNA. RNA structure is also introduced.

As in the case of protein structure (Chapter 4), it is sometimes useful to describe nucleic acid structure in terms of hierarchical levels of complexity (primary, secondary, tertiary). The primary structure of a nucleic acid is its covalent structure and nucleotide sequence. Any regular, stable structure taken up by some or all of the nucleotides in a nucleic acid can be referred to as secondary structure. All structures considered in the remainder of this chapter fall under the heading of secondary structure. The complex folding of large chromosomes within eukaryotic chromatin and bacterial nucleoids, or the elaborate folding of large tRNA

or rRNA molecules, is generally considered tertiary structure. DNA tertiary structure is discussed in Chapter 24, and RNA tertiary structure is considered in Chapter 26.

DNA Is a Double Helix That Stores Genetic Information

DNA was first isolated and characterized by Friedrich Miescher in 1868. He called the phosphorus-containing substance "nuclein." Not until the 1940s, with the work of Oswald T. Avery, Colin MacLeod, and Maclyn McCarty, was there any compelling evidence that DNA was the genetic material. Avery and his colleagues found that DNA extracted from a virulent (disease-causing) strain of the bacterium *Streptococcus pneumoniae* and injected into a nonvirulent strain of the same bacterium transformed the nonvirulent strain into a virulent strain. They concluded that the DNA from the virulent strain carried the genetic information for virulence. Then in 1952, experiments by Alfred D. Hershey and Martha Chase, in which they studied the infection of bacterial cells by a virus (bacteriophage) with radioactively labeled DNA or protein, removed any remaining doubt that DNA, not protein, carried the genetic information.

Another important clue to the structure of DNA came from the work of Erwin Chargaff and his colleagues in the late 1940s. They found that the four nucleotide bases of DNA occur in different ratios in the DNAs of different organisms and that the amounts of certain bases are closely related. These data, collected from DNAs of a great many different species, led Chargaff to the following conclusions:

1. The base composition of DNA generally varies from one species to another.
2. DNA specimens isolated from different tissues of the same species have the same base composition.
3. The base composition of DNA in a given species does not change with an organism's age, nutritional state, or changing environment.
4. In all cellular DNAs, regardless of the species, the number of adenosine residues is equal to the number of thymidine residues (that is, A = T), and the number of guanosine residues is equal to the number of cytidine residues (G = C). From these relationships it follows that the sum of the purine residues equals the sum of the pyrimidine residues; that is, A + G = T + C.

These quantitative relationships, sometimes called "Chargaff's rules," were confirmed by many subsequent researchers. They were a key to establishing the three-dimensional structure of DNA and yielded clues to how genetic information is encoded in DNA and passed from one generation to the next.

To shed more light on the structure of DNA, Rosalind Franklin and Maurice Wilkins used the powerful method of x-ray diffraction (see Box 4–5) to analyze DNA fibers. They showed in the early 1950s that DNA produces a characteristic x-ray diffraction pattern **(Fig. 8–12)**. From this pattern it was deduced that DNA molecules are helical with two periodicities along their long axis, a primary one of 3.4 Å and a secondary one of 34 Å. The problem then was to formulate a three-dimensional model of the DNA molecule that could account not only for the x-ray diffraction data but also for the specific A = T and G = C base equivalences discovered by Chargaff and for the other chemical properties of DNA.

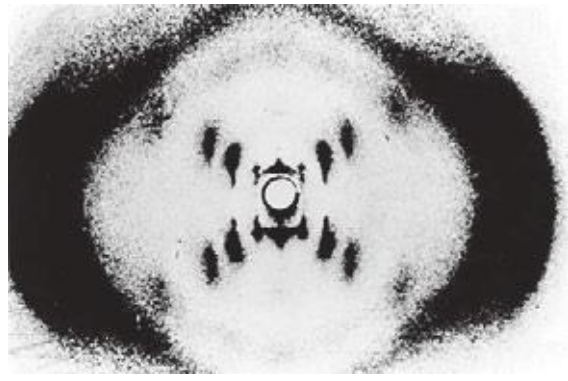


FIGURE 8–12 X-ray diffraction pattern of DNA fibers. The spots forming a cross in the center denote a helical structure. The heavy bands at the left and right arise from the recurring bases.

Rosalind Franklin, 1920–1958

Maurice Wilkins, 1916–2004

James Watson and Francis Crick relied on this accumulated information about DNA to set about deducing its structure. In 1953 they postulated a three-dimensional model of DNA structure that accounted for all the available data. It consists of two helical DNA chains wound around the same axis to form a right-handed double helix (see Box 4–1 for an explanation of the right- or left-handed sense of a helical structure). The hydrophilic backbones of alternating deoxyribose and phosphate groups are on the outside of the double helix, facing the surrounding water. The furanose ring of each deoxyribose is in the C-2′ endo conformation. The purine and pyrimidine bases of both strands are stacked inside the double helix, with their hydrophobic and nearly planar

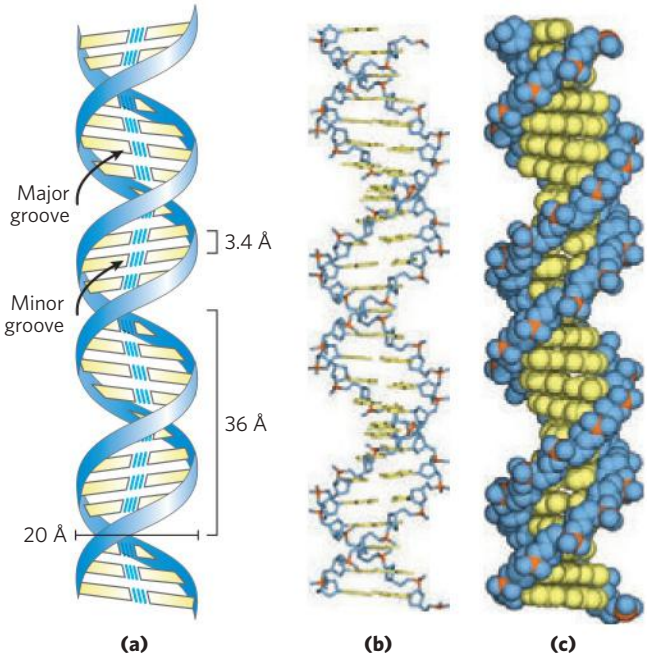


FIGURE 8–13 Watson-Crick model for the structure of DNA. The original model proposed by Watson and Crick had 10 base pairs, or 34 Å (3.4 nm), per turn of the helix; subsequent measurements revealed 10.5 base pairs, or 36 Å (3.6 nm), per turn. **(a)** Schematic representation, showing dimensions of the helix. **(b)** Stick representation showing the backbone and stacking of the bases. **(c)** Space-filling model.

ring structures very close together and perpendicular to the long axis. The offset pairing of the two strands creates a **major groove** and **minor groove** on the surface of the duplex **(Fig. 8–13)**. Each nucleotide base of one strand is paired in the same plane with a base of the other strand. Watson and Crick found that the hydrogen-bonded base pairs illustrated in Figure 8–11, G with C and A with T, are those that fit best within the structure, providing a rationale for Chargaff's rule that in any DNA, G = C and A = T. It is important to note that three hydrogen bonds can form between G and C, symbolized G≡C, but only two can form between A and T, symbolized A═T. This is one reason for the finding that separation of paired DNA strands is more difficult the higher the ratio of G≡C to A═T base pairs. Other pairings of bases tend (to varying degrees) to destabilize the double-helical structure.

When Watson and Crick constructed their model, they had to decide at the outset whether the strands of DNA should be **parallel** or **antiparallel**—whether their 3′,5′-phosphodiester bonds should run in the same or opposite directions. An antiparallel orientation produced the most convincing model, and later work with DNA polymerases (Chapter 25) provided experimental evidence that the strands are indeed antiparallel, a finding ultimately confirmed by x-ray analysis.

To account for the periodicities observed in the x-ray diffraction patterns of DNA fibers, Watson and Crick manipulated molecular models to arrive at a structure in which the vertically stacked bases inside the double helix would be 3.4 Å apart; the secondary repeat distance of about 34 Å was accounted for by the presence of 10 base pairs in each complete turn of the double helix. In aqueous solution the structure differs slightly from that in fibers, having 10.5 base pairs per helical turn (Fig. 8–13).

As **Figure 8–14** shows, the two antiparallel polynucleotide chains of double-helical DNA are not identical in either base sequence or composition. Instead they are **complementary** to each other. Wherever adenine occurs in one chain, thymine is found in the other; similarly, wherever guanine occurs in one chain, cytosine is found in the other.

The DNA double helix, or duplex, is held together by two forces, as described earlier: hydrogen bonding between complementary base pairs (Fig. 8–11) and base-stacking interactions. The complementarity between the DNA strands is attributable to the hydrogen bonding between base pairs. The base-stacking interactions, which are largely nonspecific with respect to the identity of the stacked bases, make the major contribution to the stability of the double helix.

The important features of the double-helical model of DNA structure are supported by much chemical and biological evidence. Moreover, the model immediately suggested a mechanism for the transmission of genetic information. The essential feature of the model is the complementarity of the two DNA strands. As Watson and Crick were able to see, well before confirmatory data became available, this structure could logically be replicated by (1) separating the two strands and (2) synthesizing a complementary strand for each. Because nucleotides in each new strand are joined in a sequence specified by the base-pairing rules stated above, each preexisting

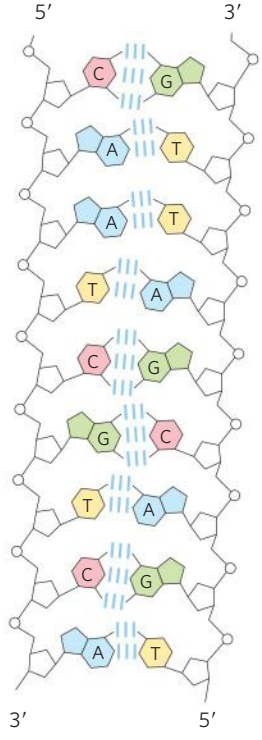


FIGURE 8–14 Complementarity of strands in the DNA double helix. The complementary antiparallel strands of DNA follow the pairing rules proposed by Watson and Crick. The base-paired antiparallel strands differ in base composition: the left strand has the composition $A_3T_2G_1C_3$; the right, $A_2T_3G_3C_1$. They also differ in sequence when each chain is read in the 5′→3′ direction. Note the base equivalences: A = T and G = C in the duplex.

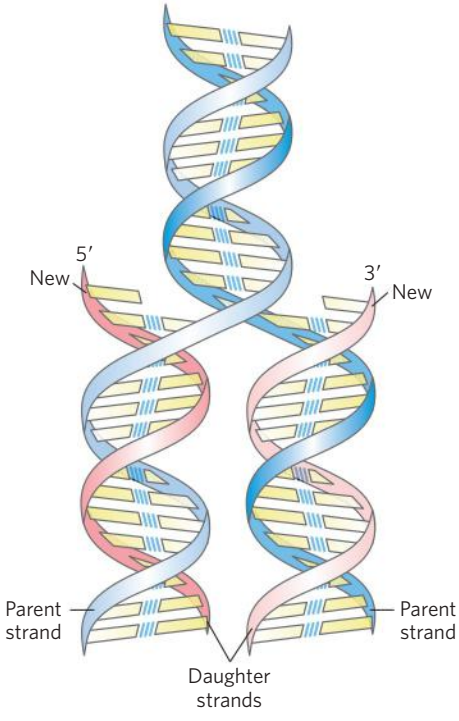


FIGURE 8–15 Replication of DNA as suggested by Watson and Crick. The preexisting or "parent" strands become separated, and each is the template for biosynthesis of a complementary "daughter" strand (in pink).

strand functions as a template to guide the synthesis of one complementary strand **(Fig. 8–15)**. These expectations were experimentally confirmed, inaugurating a revolution in our understanding of biological inheritance.

WORKED EXAMPLE 8–1 Base Pairing in DNA

In samples of DNA isolated from two unidentified species of bacteria, X and Y, adenine makes up 32% and 17%, respectively, of the total bases. What relative proportions of adenine, guanine, thymine, and cytosine would you expect to find in the two DNA samples? What assumptions have you made? One of these species was isolated from a hot spring (64 °C). Which species is most likely the thermophilic bacterium, and why?

Solution: For any double-helical DNA, A = T and G = C. The DNA from species X has 32% A and therefore must contain 32% T. This accounts for 64% of the bases and leaves 36% as G≡C pairs: 18% G and 18% C. The sample from species Y, with 17% A, must contain 17% T, accounting for 34% of the base pairs. The remaining 66% of the bases are thus equally distributed as 33% G and 33% C. This calculation is based on the assumption that both DNA molecules are double-stranded.

The higher the G + C content of a DNA molecule, the higher the melting temperature. Species Y, having the DNA with the higher G + C content (66%), most likely is the thermophilic bacterium; its DNA has a higher melting temperature and thus is more stable at the temperature of the hot spring.

DNA Can Occur in Different Three-Dimensional Forms

DNA is a remarkably flexible molecule. Considerable rotation is possible around several types of bonds in the sugar–phosphate (phosphodeoxyribose) backbone, and thermal fluctuation can produce bending, stretching, and unpairing (melting) of the strands. Many significant deviations from the Watson-Crick DNA structure are found in cellular DNA, some or all of which may be important in DNA metabolism. These structural variations generally do not affect the key properties of DNA defined by Watson and Crick: strand complementarity, antiparallel strands, and the requirement for A═T and G≡C base pairs.

Structural variation in DNA reflects three things: the different possible conformations of the deoxyribose, rotation about the contiguous bonds that make up the phosphodeoxyribose backbone **(Fig. 8–16a)**, and free rotation about the C-1′–*N*-glycosyl bond (Fig. 8–16b). Because of steric constraints, purines in purine nucleotides are restricted to two stable conformations with respect to deoxyribose, called syn and anti (Fig. 8–16b). Pyrimidines are generally restricted to the anti conformation because of steric interference between the sugar and the carbonyl oxygen at C-2 of the pyrimidine.

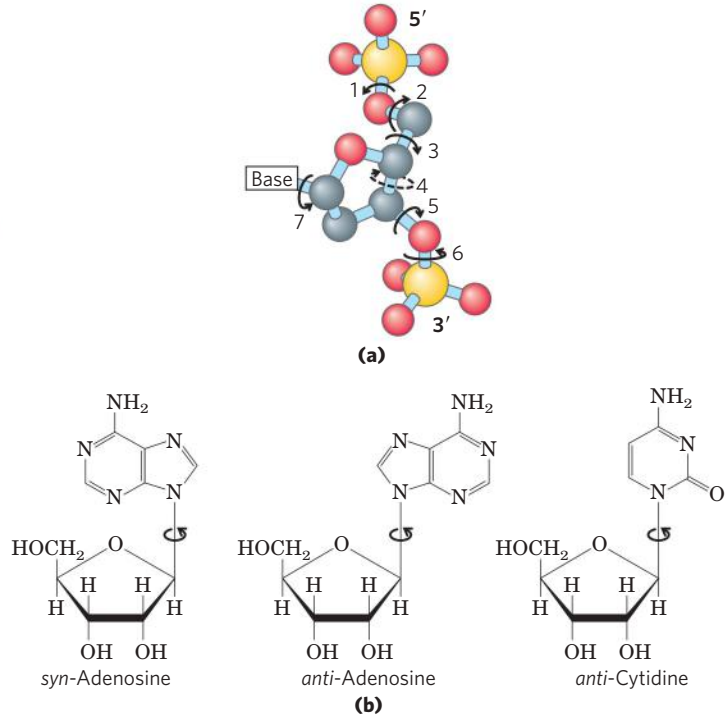


FIGURE 8–16 Structural variation in DNA. (a) The conformation of a nucleotide in DNA is affected by rotation about seven different bonds. Six of the bonds rotate freely. The limited rotation about bond 4 gives rise to ring pucker. This conformation is endo or exo, depending on whether the atom is displaced to the same side of the plane as C-5′ or to the opposite side (see Fig. 8–3b). **(b)** For purine bases in nucleotides, only two conformations with respect to the attached ribose units are sterically permitted, anti or syn. Pyrimidines occur in the anti conformation.

The Watson-Crick structure is also referred to as **B-form DNA**, or B-DNA. The B form is the most stable structure for a random-sequence DNA molecule under physiological conditions and is therefore the standard point of reference in any study of the properties of DNA. Two structural variants that have been well characterized in crystal structures are the **A** and **Z forms**. These three DNA conformations are shown in **Figure 8–17**, with a summary of their properties. The A form is favored in many solutions that are relatively devoid of water. The DNA is still arranged in a right-handed double helix, but the helix is wider and the number of base pairs per helical turn is 11, rather than 10.5 as in B-DNA. The plane of the base pairs in A-DNA is tilted about 20° relative to B-DNA base pairs, thus the base pairs in A-DNA are not perfectly perpendicular to the helix axis. These structural changes deepen the major groove while making the minor groove shallower. The reagents used to promote crystallization of DNA tend to dehydrate it, and thus most short DNA molecules tend to crystallize in the A form.

Z-form DNA is a more radical departure from the B structure; the most obvious distinction is the left-handed helical rotation. There are 12 base pairs per helical turn, and the structure appears more slender and elongated. The DNA backbone takes on a zigzag appearance. Certain nucleotide sequences fold into left-handed Z helices much more readily than others. Prominent examples are sequences in which pyrimidines alternate with purines, especially alternating C and G or 5-methyl-C and G residues. To form the left-handed helix in Z-DNA, the purine residues flip to the syn conformation, alternating with pyrimidines in the anti conformation. The major groove is barely apparent in Z-DNA, and the minor groove is narrow and deep.

Whether A-DNA occurs in cells is uncertain, but there is evidence for some short stretches (tracts) of Z-DNA in both bacteria and eukaryotes. These Z-DNA tracts may play a role (as yet undefined) in regulating the expression of some genes or in genetic recombination.

Certain DNA Sequences Adopt Unusual Structures

Other sequence-dependent structural variations found in larger chromosomes may affect the function and metabolism of the DNA segments in their immediate vicinity. For example, bends occur in the DNA helix wherever four or more adenosine residues appear sequentially in one strand. Six adenosines in a row produce a bend of about 18°. The bending observed with this and other sequences may be important in the binding of some proteins to DNA.

A rather common type of DNA sequence is a **palindrome**. A palindrome is a word, phrase, or sentence that is spelled identically read either forward or backward; two examples are ROTATOR and NURSES RUN. The term is applied to regions of DNA with **inverted repeats** of base sequence having twofold symmetry

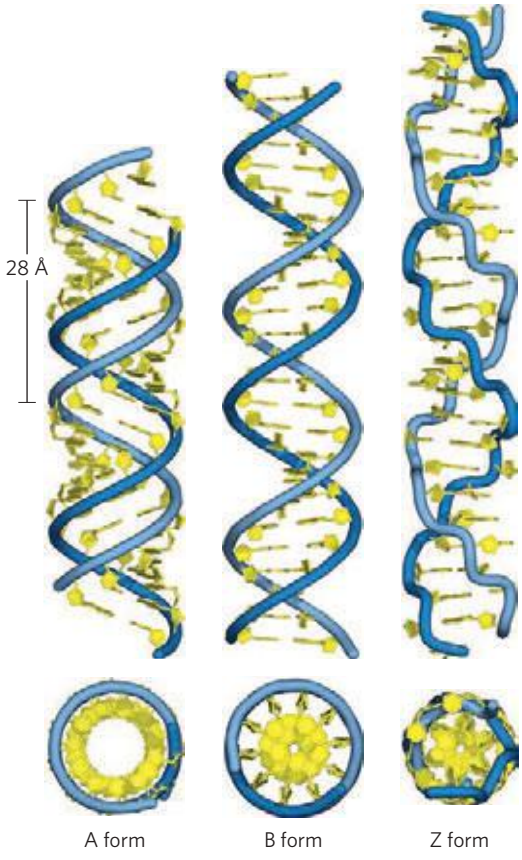


FIGURE 8–17 Comparison of A, B, and Z forms of DNA. Each structure shown here has 36 base pairs. The riboses and bases are shown in yellow. The phosphodiester backbone is represented as a blue rope. Blue is the color used to represent DNA strands in later chapters. The table summarizes some properties of the three forms of DNA.

	A form	B form	Z form
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11	10.5	12
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to the helix axis	20°	6°	7°
Sugar pucker conformation	C-3′ endo	C-2′ endo	C-2′ endo for pyrimidines; C-3′ endo for purines
Glycosyl bond conformation	Anti	Anti	Anti for pyrimidines; syn for purines

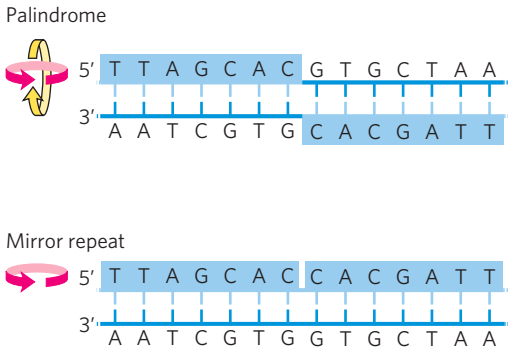


FIGURE 8–18 Palindromes and mirror repeats. Palindromes are sequences of double-stranded nucleic acids with twofold symmetry. In order to superimpose one repeat (shaded sequence) on the other, it must be rotated 180° about the horizontal axis then 180° about the vertical axis, as shown by the colored arrows. A mirror repeat, on the other hand, has a symmetric sequence within each strand. Superimposing one repeat on the other requires only a single 180° rotation about the vertical axis.

over two strands of DNA **(Fig. 8–18)**. Such sequences are self-complementary within each strand and therefore have the potential to form **hairpin** or **cruciform** (cross-shaped) structures **(Fig. 8–19)**. When the inverted repeat occurs within each individual strand of the DNA, the sequence is called a **mirror repeat**. Mirror repeats do not have complementary sequences within the same strand and cannot form hairpin or cruciform structures. Sequences of these types are found in virtually every large DNA molecule and can encompass a few base pairs or thousands. The extent to which palindromes occur as cruciforms in cells is not known, although some cruciform structures have been demonstrated in vivo in *Escherichia coli*. Self-complementary sequences cause isolated single strands of DNA (or RNA) in solution to fold into complex structures containing multiple hairpins.

Several unusual DNA structures involve three or even four DNA strands. Nucleotides participating in a Watson-Crick base pair (Fig. 8–11) can form additional hydrogen bonds, particularly with functional groups arrayed in the major groove. For example, a cytidine residue (if protonated) can pair with the guanosine residue of a G≡C nucleotide pair **(Fig. 8–20)**; a thymidine can pair with the adenosine of an A═T pair. The N-7, O^6, and N^6 of purines, the atoms that participate in the hydrogen bonding of triplex DNA, are often referred to as **Hoogsteen positions**, and the non-Watson-Crick pairing is called **Hoogsteen pairing**, after Karst Hoogsteen, who in 1963 first recognized the potential for these unusual pairings. Hoogsteen pairing allows the formation of **triplex DNAs**. The triplexes shown in Figure 8–20 (a, b) are most stable at low pH because the C≡G·C^+ triplet requires a protonated cytosine. In the triplex, the pK_a of this cytosine is >7.5, altered from its normal value of 4.2. The triplexes also form most readily within long sequences containing only pyrimidines or only purines in a given strand. Some triplex DNAs contain two pyrimidine

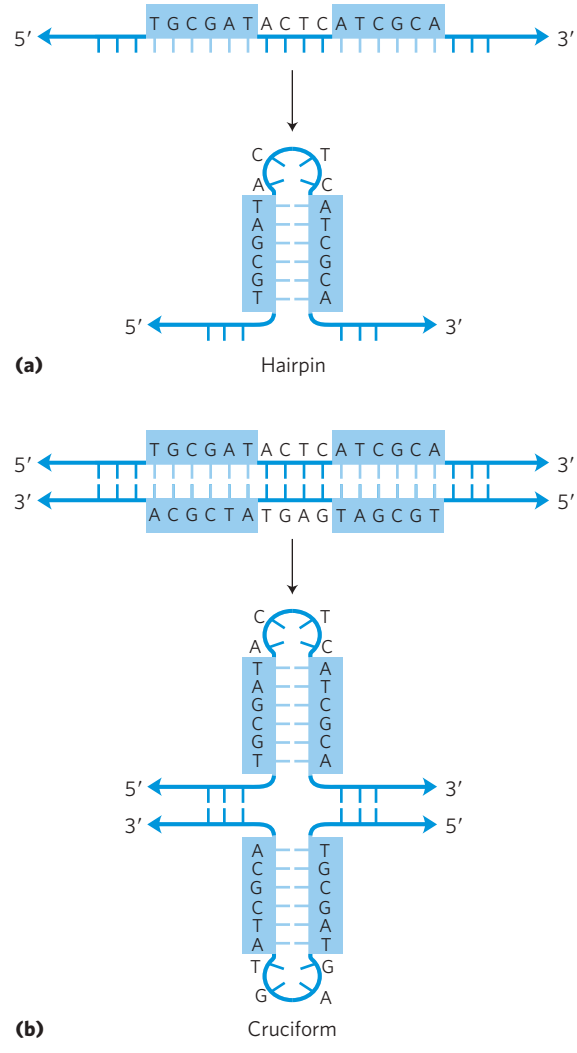
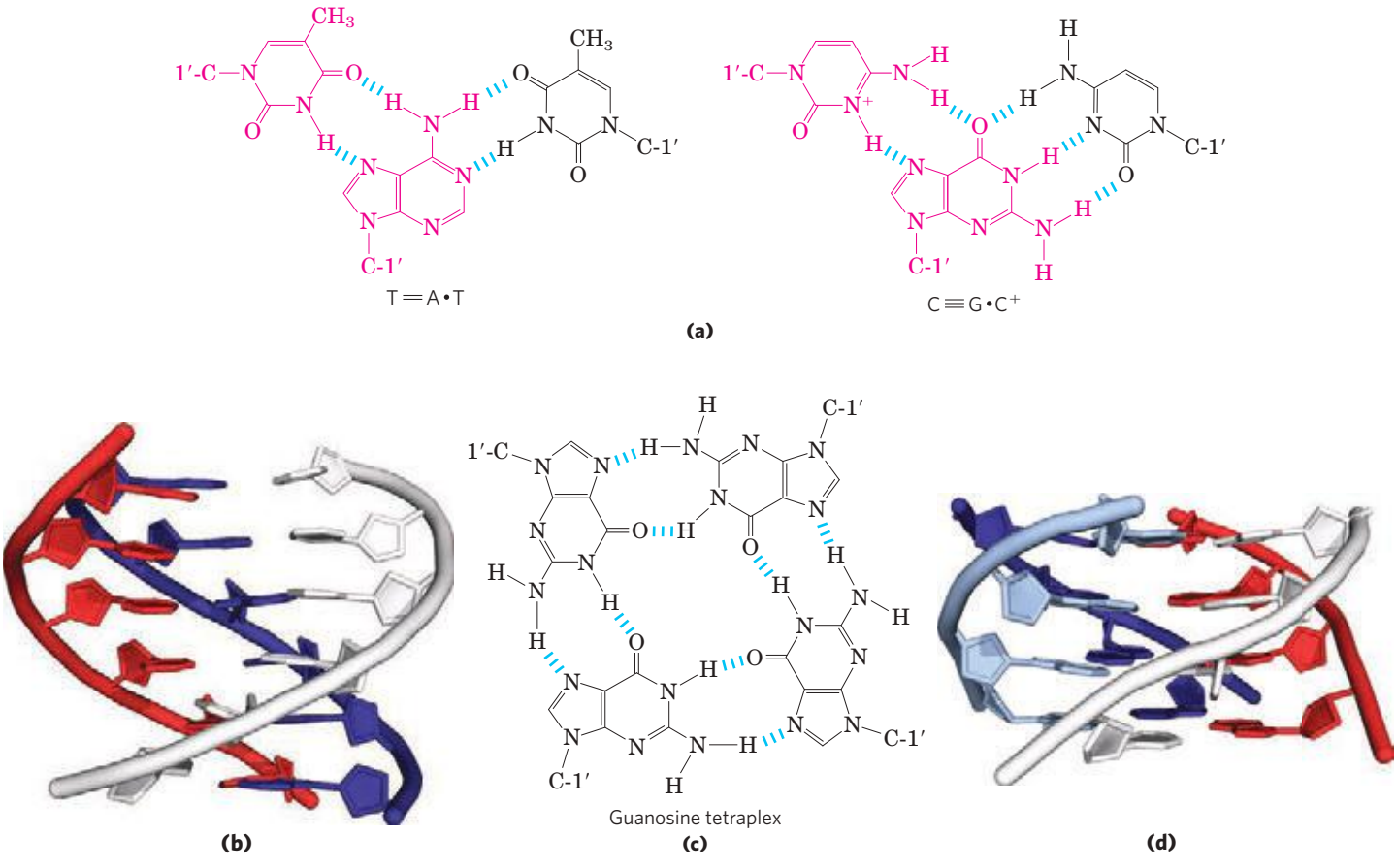


FIGURE 8–19 Hairpins and cruciforms. Palindromic DNA (or RNA) sequences can form alternative structures with intrastrand base pairing. **(a)** When only a single DNA (or RNA) strand is involved, the structure is called a hairpin. **(b)** When both strands of a duplex DNA are involved, it is called a cruciform. Blue shading highlights asymmetric sequences that can pair with the complementary sequence either in the same strand or in the complementary strand.

strands and one purine strand; others contain two purine strands and one pyrimidine strand.

Four DNA strands can also pair to form a tetraplex (quadruplex), but this occurs readily only for DNA sequences with a very high proportion of guanosine residues (Fig. 8–20c, d). The guanosine tetraplex, or **G tetraplex**, is quite stable over a wide range of conditions. The orientation of strands in the tetraplex can vary as shown in Figure 8–20e.

In the DNA of living cells, sites recognized by many sequence-specific DNA-binding proteins (Chapter 28) are arranged as palindromes, and polypyrimidine or polypurine sequences that can form triple helices are found within regions involved in the regulation of expression of some eukaryotic genes. In principle, synthetic DNA strands designed to pair with these sequences to form triplex DNA could disrupt gene



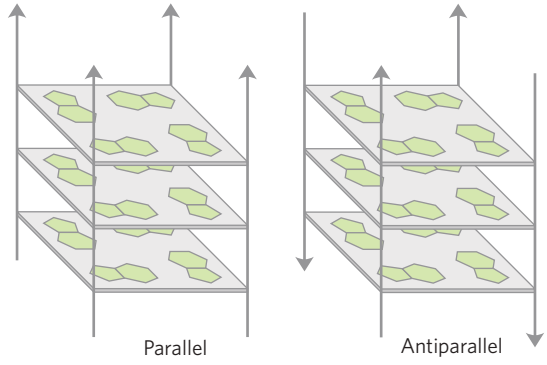


FIGURE 8–20 DNA structures containing three or four DNA strands. **(a)** Base-pairing patterns in one well-characterized form of triplex DNA. The Hoogsteen pair in each case is shown in red. **(b)** Triple-helical DNA containing two pyrimidine strands (red and white; sequence TTCCT) and one purine strand (blue; sequence AAGGAA) (derived from PDB ID 1BCE). The blue and white strands are antiparallel and paired by normal Watson-Crick base-pairing patterns. The third (all-pyrimidine) strand (red) is parallel to the purine strand and paired through non-Watson-Crick hydrogen bonds. The triplex is viewed from the side, with six triplets shown. **(c)** Base-pairing pattern in the guanosine tetraplex structure. **(d)** Four successive tetraplets from a G tetraplex structure (PDB ID 244D). **(e)** Possible variants in the orientation of strands in a G tetraplex.

expression. This approach to controlling cellular metabolism is of commercial interest for its potential application in medicine and agriculture.

Messenger RNAs Code for Polypeptide Chains

We now turn our attention to the expression of the genetic information that DNA contains. RNA, the second major form of nucleic acid in cells, has many functions. In gene expression, RNA acts as an intermediary by using the information encoded in DNA to specify the amino acid sequence of a functional protein.

Given that the DNA of eukaryotes is largely confined to the nucleus whereas protein synthesis occurs on ribosomes in the cytoplasm, some molecule other than DNA must carry the genetic message from the nucleus to the cytoplasm. As early as the 1950s, RNA was considered the logical candidate: RNA is found in both the nucleus and the cytoplasm, and an increase in protein synthesis is accompanied by an increase in the amount of cytoplasmic RNA and an increase in its rate of turnover. These and other observations led several researchers to suggest that RNA carries genetic information from DNA to the protein biosynthetic machinery of the ribosome. In 1961 François Jacob and Jacques Monod presented a unified (and essentially correct) picture of many aspects of this process. They proposed

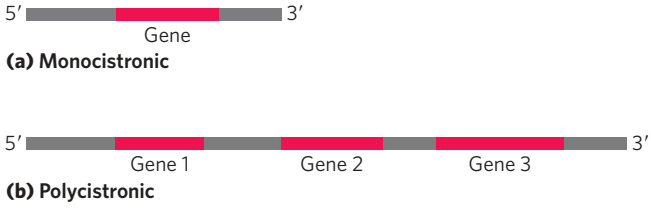


FIGURE 8–21 Bacterial mRNA. Schematic diagrams show **(a)** monocistronic and **(b)** polycistronic mRNAs of bacteria. Red segments represent RNA coding for a gene product; gray segments represent noncoding RNA. In the polycistronic transcript, noncoding RNA separates the three genes.

the name "messenger RNA" (mRNA) for that portion of the total cellular RNA carrying the genetic information from DNA to the ribosomes, where the messengers provide the templates that specify amino acid sequences in polypeptide chains. Although mRNAs from different genes can vary greatly in length, the mRNAs from a particular gene generally have a defined size. The process of forming mRNA on a DNA template is known as **transcription**.

In bacteria and archaea, a single mRNA molecule may code for one or several polypeptide chains. If it carries the code for only one polypeptide, the mRNA is **monocistronic**; if it codes for two or more different polypeptides, the mRNA is **polycistronic**. In eukaryotes, most mRNAs are monocistronic. (For the purposes of this discussion, "cistron" refers to a gene. The term itself has historical roots in the science of genetics, and its formal genetic definition is beyond the scope of this text.) The minimum length of an mRNA is set by the length of the polypeptide chain for which it codes. For example, a polypeptide chain of 100 amino acid residues requires an RNA coding sequence of at least 300 nucleotides, because each amino acid is coded by a nucleotide triplet (this and other details of protein synthesis are discussed in Chapter 27). However, mRNAs transcribed from DNA are always somewhat longer than the length needed simply to code for a polypeptide sequence (or sequences). The additional, noncoding RNA includes sequences that regulate protein synthesis. **Figure 8–21** summarizes the general structure of bacterial mRNAs.

Many RNAs Have More Complex Three-Dimensional Structures

Messenger RNA is only one of several classes of cellular RNA. Transfer RNAs are adapter molecules in protein synthesis; covalently linked to an amino acid at one end, they pair with the mRNA in such a way that amino acids are joined to a growing polypeptide in the correct sequence. Ribosomal RNAs are components of ribosomes. There is also a wide variety of special-function RNAs, including some (called ribozymes) that have enzymatic activity. All the RNAs are considered in detail in Chapter 26. The diverse and often complex functions of these RNAs reflect a diversity of structure much richer than that observed in DNA molecules.

The product of transcription of DNA is always single-stranded RNA. The single strand tends to assume a right-handed helical conformation dominated by base-stacking interactions **(Fig. 8–22)**, which are stronger between two purines than between a purine and pyrimidine or between two pyrimidines. The purine-purine interaction is so strong that a pyrimidine separating two purines is often displaced from the stacking pattern so that the purines can interact. Any self-complementary sequences in the molecule produce more complex structures. RNA can base-pair with complementary regions of either RNA or DNA. Base pairing matches the pattern for DNA: G pairs with C and A pairs with U (or with the occasional T residue in some RNAs). One difference is that base pairing between G and U residues—unusual in DNA—is allowed in RNA (see Fig. 8–24) when complementary sequences in two single strands of RNA pair with each other. The paired strands in RNA or RNA-DNA duplexes are antiparallel, as in DNA.

When two strands of RNA with perfectly complementary sequences are paired, the predominant double-stranded structure is an A-form right-handed double helix. However, strands of RNA that are perfectly paired over long regions of sequence are uncommon. The three-dimensional structures of many RNAs, like those of proteins, are complex and unique. Weak

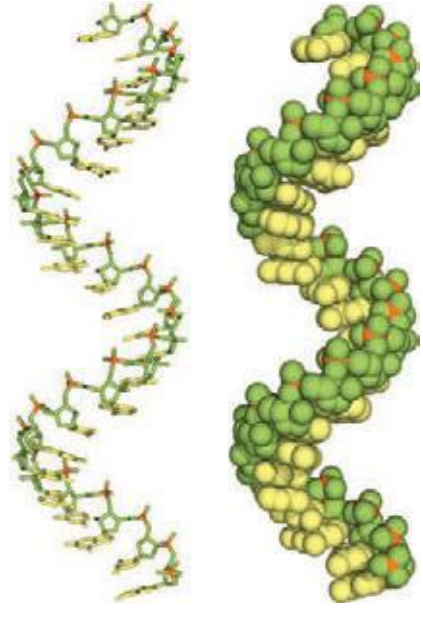


FIGURE 8–22 Typical right-handed stacking pattern of single-stranded RNA. The bases are shown in yellow, the phosphorus atoms in orange, and the riboses and phosphate oxygens in green. Green is used to represent RNA strands in succeeding chapters, just as blue is used for DNA.

interactions, especially base-stacking interactions, help stabilize RNA structures, just as they do in DNA. Z-form helices have been made in the laboratory (under very high-salt or high-temperature conditions). The B form of RNA has not been observed. Breaks in the regular A-form helix caused by mismatched or unmatched bases in one or both strands are common and result in bulges or internal loops **(Fig. 8–23)**. Hairpin loops form between nearby self-complementary (palindromic) sequences. The potential for base-paired helical segments in many RNAs is extensive **(Fig. 8–24)**, and the resulting hairpins are the most common type of secondary structure in RNA. Specific short base sequences (such as UUCG) are often found at the ends of RNA hairpins and are known to form particularly tight and stable loops. Such sequences may act as starting points for the folding of an RNA molecule into its precise three-dimensional structure. Other contributions are made by hydrogen bonds that are not part of standard Watson-Crick base pairs. For example, the 2′-hydroxyl group of ribose can hydrogen-bond with other groups. Some of these properties are evident in the structure of the phenylalanine transfer RNA of yeast—the tRNA responsible for inserting Phe residues into polypeptides—and in two RNA enzymes, or ribozymes, whose functions, like those of

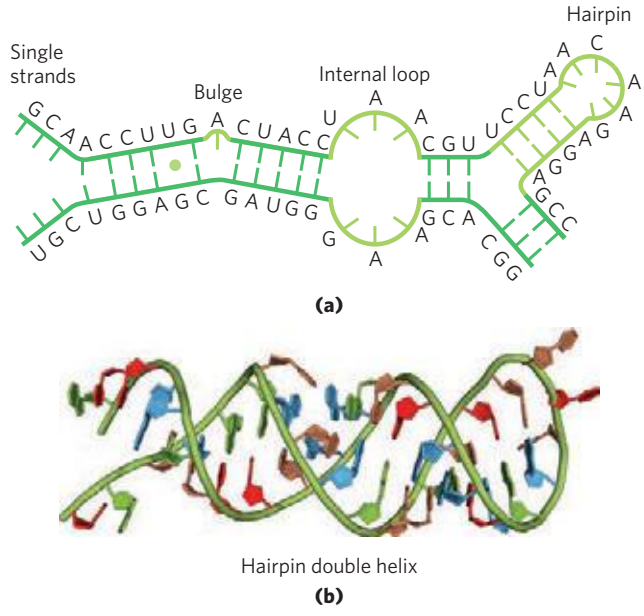


FIGURE 8–23 Secondary structure of RNAs. (a) Bulge, internal loop, and hairpin loop. **(b)** The paired regions generally have an A-form right-handed helix, as shown for a hairpin (derived from PDB ID 1GID).

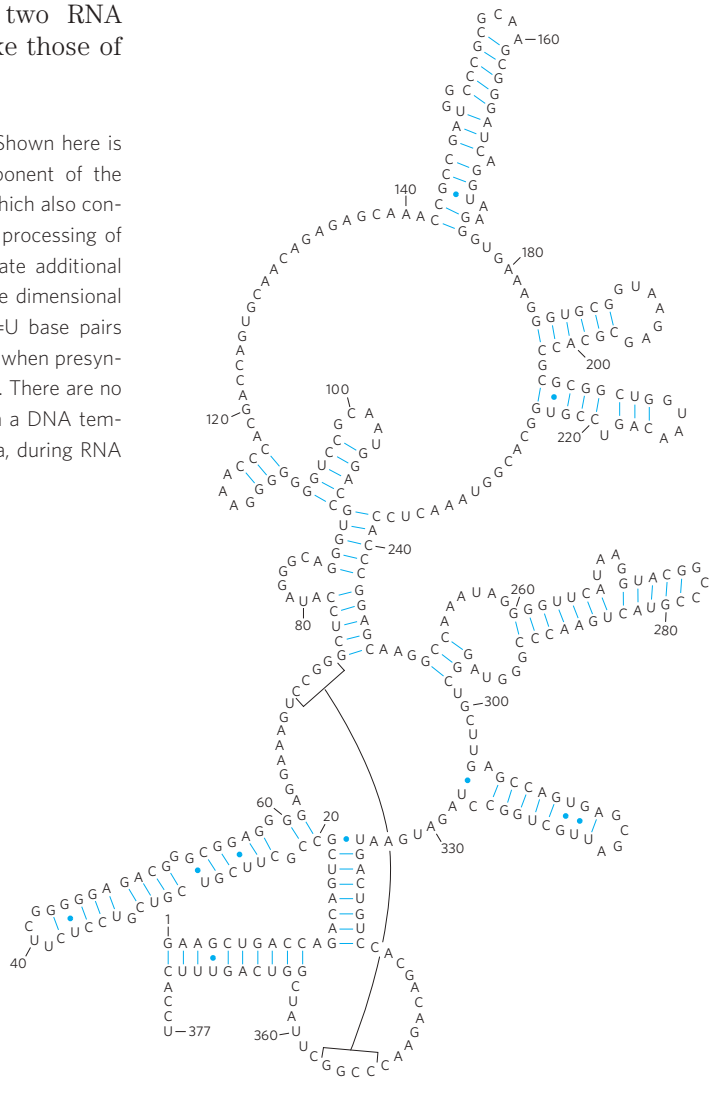
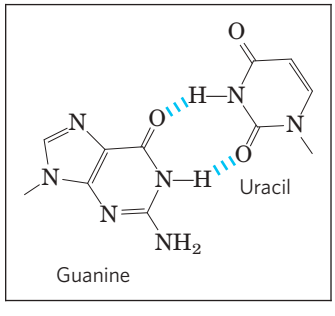


FIGURE 8–24 Base-paired helical structures in an RNA. Shown here is the possible secondary structure of the M1 RNA component of the enzyme RNase P of *E. coli*, with many hairpins. RNase P, which also contains a protein component (not shown), functions in the processing of transfer RNAs (see Fig. 26–26). The two brackets indicate additional complementary sequences that may be paired in the three dimensional structure. The blue dots indicate non-Watson-Crick G═U base pairs (boxed inset). Note that G═U base pairs are allowed only when presynthesized strands of RNA fold up or anneal with each other. There are no RNA polymerases (the enzymes that synthesize RNAs on a DNA template) that insert a U opposite a template G, or vice versa, during RNA synthesis.



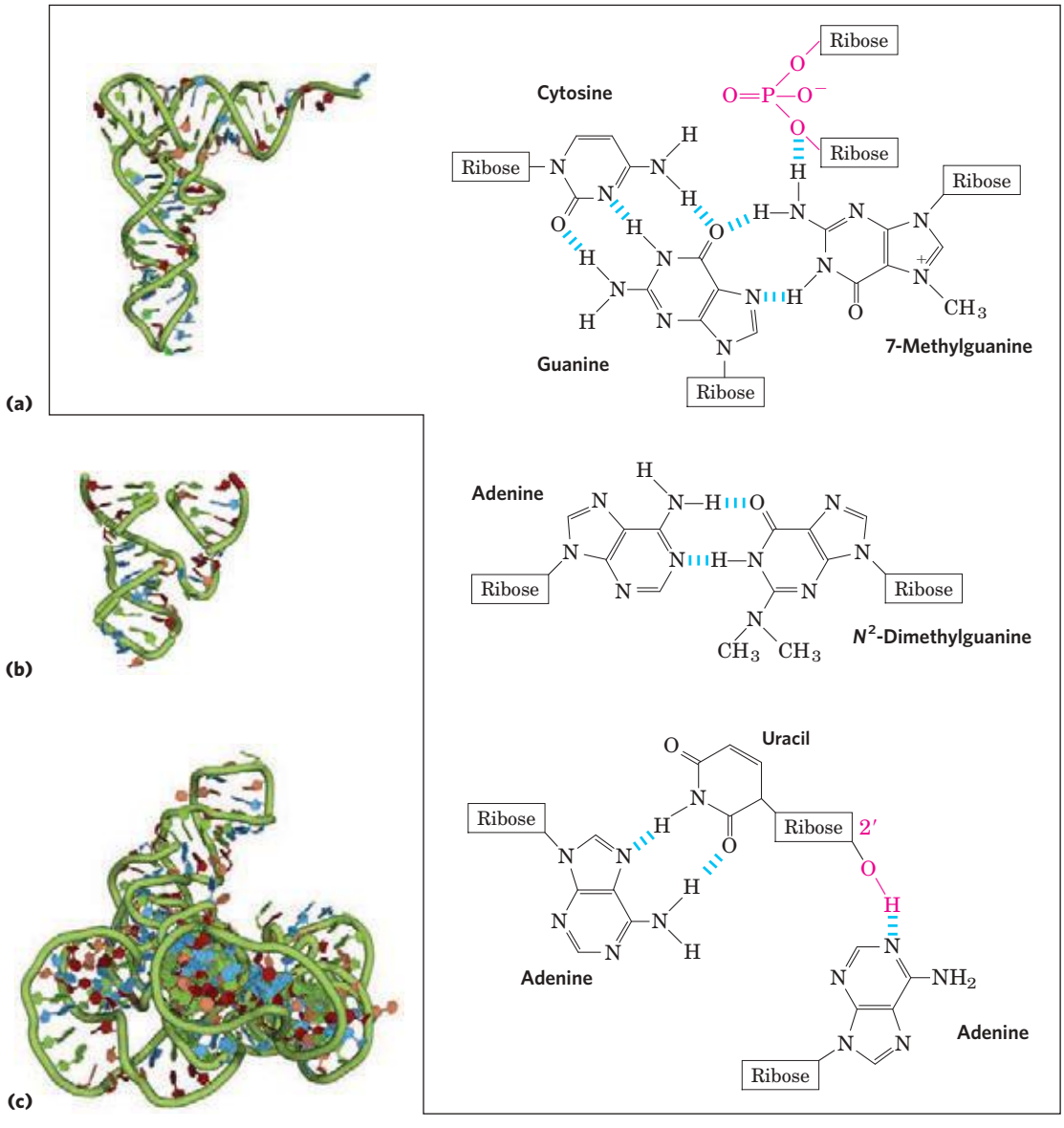


FIGURE 8–25 Three-dimensional structure in RNA. (a) Three-dimensional structure of phenylalanine tRNA of yeast (PDB ID 1TRA). Some unusual base-pairing patterns found in this tRNA are shown. Note also the involvement of the oxygen of a ribose phosphodiester bond in one hydrogen-bonding arrangement, and a ribose 2′-hydroxyl group in another (both in red). **(b)** A hammerhead ribozyme (so named because the secondary structure at the active site looks like the head of a hammer), derived from certain plant viruses (derived from PDB ID 1MME). Ribozymes, or RNA enzymes, catalyze a variety of reactions, primarily in RNA metabolism and protein synthesis. The complex three-dimensional structures of these RNAs reflect the complexity inherent in catalysis, as described for protein enzymes in Chapter 6. **(c)** A segment of mRNA known as an intron, from the ciliated protozoan *Tetrahymena thermophila* (derived from PDB ID 1GRZ). This intron (a ribozyme) catalyzes its own excision from between exons in an mRNA strand (discussed in Chapter 26).

protein enzymes, depend on their three-dimensional structures **(Fig. 8–25)**.

The analysis of RNA structure and the relationship between its structure and its function is an emerging field of inquiry that has many of the same complexities as the analysis of protein structure. The importance of understanding RNA structure grows as we become increasingly aware of the large number of functional roles for RNA molecules.

SUMMARY 8.2 Nucleic Acid Structure

- Many lines of evidence show that DNA bears genetic information. Some of the earliest evidence came from the Avery-MacLeod-McCarty experiment, which showed that DNA isolated from one bacterial strain can enter and transform the cells of another strain, endowing it with some of the inheritable characteristics of the donor. The

Hershey-Chase experiment showed that the DNA of a bacterial virus, but not its protein coat, carries the genetic message for replication of the virus in a host cell.

- Putting together the available data, Watson and Crick postulated that native DNA consists of two antiparallel chains in a right-handed double-helical arrangement. Complementary base pairs, A═T and G≡C, are formed by hydrogen bonding within the helix. The base pairs are stacked perpendicular to the long axis of the double helix, 3.4 Å apart, with 10.5 base pairs per turn.
- DNA can exist in several structural forms. Two variations of the Watson-Crick form, or B-DNA, are A- and Z-DNA. Some sequence-dependent structural variations cause bends in the DNA molecule. DNA strands with appropriate sequences can form hairpin or cruciform structures or triplex or tetraplex DNA.
- Messenger RNA transfers genetic information from DNA to ribosomes for protein synthesis. Transfer RNA and ribosomal RNA are also involved in protein synthesis. RNA can be structurally complex; single RNA strands can fold into hairpins, double-stranded regions, or complex loops.

8.3 Nucleic Acid Chemistry

The role of DNA as a repository of genetic information depends in part on its inherent stability. The chemical transformations that do occur are generally very slow in the absence of an enzyme catalyst. The long-term storage of information without alteration is so important to a cell, however, that even very slow reactions that alter DNA structure can be physiologically significant. Processes such as carcinogenesis and aging may be intimately linked to slowly accumulating, irreversible alterations of DNA. Other, nondestructive alterations also occur and are essential to function, such as the strand separation that must precede DNA replication or transcription. In addition to providing insights into physiological processes, our understanding of nucleic acid chemistry has given us a powerful array of technologies that have applications in molecular biology, medicine, and forensic science. We now examine the chemical properties of DNA and some of these technologies.

Double-Helical DNA and RNA Can Be Denatured

Solutions of carefully isolated, native DNA are highly viscous at pH 7.0 and room temperature (25 °C). When such a solution is subjected to extremes of pH or to temperatures above 80 °C, its viscosity decreases sharply, indicating that the DNA has undergone a physical change. Just as heat and extremes of pH denature globular proteins, they also cause denaturation, or melting, of double-helical DNA. Disruption of the hydrogen bonds between paired bases and of base stacking causes unwinding of the double helix to form two single strands, completely separate from each other along the entire length or part of the length (partial denaturation) of the molecule. No covalent bonds in the DNA are broken **(Fig. 8–26)**.

Renaturation of a DNA molecule is a rapid one-step process, as long as a double-helical segment of a dozen or more residues still unites the two strands. When the temperature or pH is returned to the range in which most organisms live, the unwound segments of the two strands spontaneously rewind, or **anneal**, to yield the intact duplex (Fig. 8–26). However, if the two strands are completely separated, renaturation occurs in two steps. In the first, relatively slow step, the two strands "find" each other by random collisions and form a short segment of complementary double helix. The second step is much faster: the remaining unpaired bases successively come into register as base pairs, and the two strands "zipper" themselves together to form the double helix.

The close interaction between stacked bases in a nucleic acid has the effect of decreasing its absorption of UV light relative to that of a solution with the same concentration of free nucleotides, and the absorption is decreased further when two complementary nucleic acid strands are paired. This is called the hypochromic

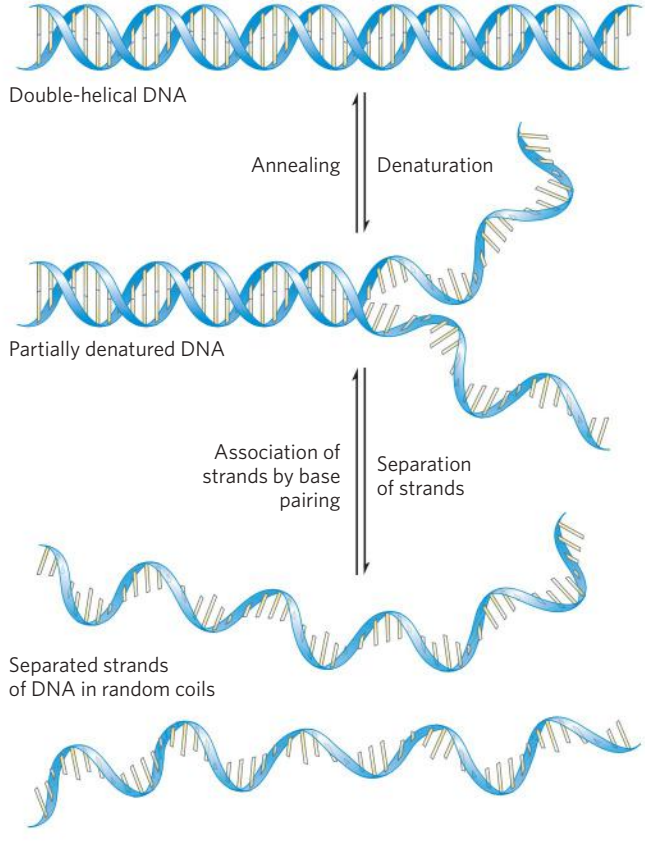


FIGURE 8–26 **Reversible denaturation and annealing (renaturation) of DNA.**

effect. Denaturation of a double-stranded nucleic acid produces the opposite result: an increase in absorption called the hyperchromic effect. The transition from double-stranded DNA to the single-stranded, denatured form can thus be detected by monitoring UV absorption at 260 nm.

Viral or bacterial DNA molecules in solution denature when they are heated slowly **(Fig. 8–27)**. Each species of DNA has a characteristic denaturation temperature, or melting point (t_m; formally, the temperature at which half the DNA is present as separated single strands): the higher its content of G≡C base pairs, the higher the melting point of the DNA. This is because G≡C base pairs, with three hydrogen bonds, require more heat energy to dissociate than A═T base pairs. Thus the melting point of a DNA molecule, determined under fixed conditions of pH and ionic strength, can yield an estimate of its base composition. If denaturation conditions are carefully controlled, regions that are rich in A═T base pairs will specifically denature while most of the DNA remains double-stranded. Such

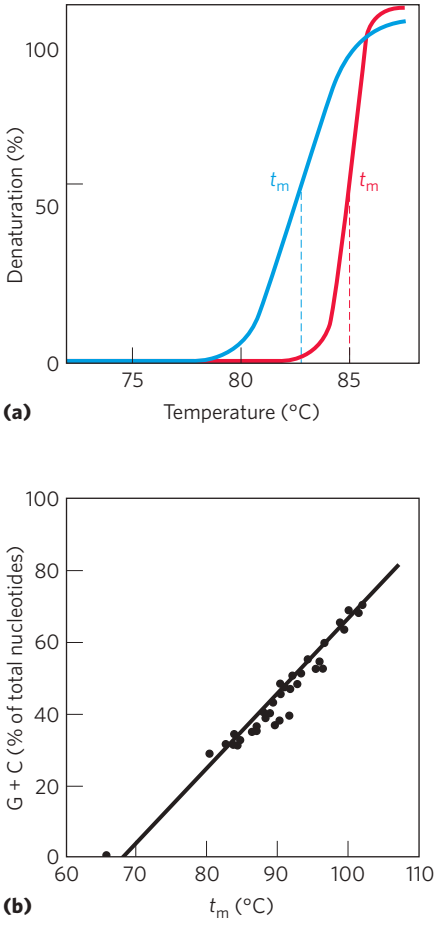


FIGURE 8–27 Heat denaturation of DNA. (a) The denaturation, or melting, curves of two DNA specimens. The temperature at the midpoint of the transition (t_m) is the melting point; it depends on pH and ionic strength and on the size and base composition of the DNA. **(b)** Relationship between t_m and the G+C content of a DNA.

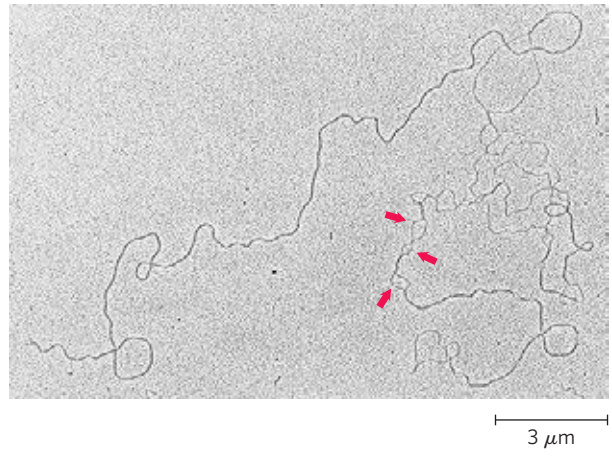


FIGURE 8–28 Partially denatured DNA. This DNA was partially denatured, then fixed to prevent renaturation during sample preparation. The shadowing method used to visualize the DNA in this electron micrograph increases its diameter approximately fivefold and obliterates most details of the helix. However, length measurements can be obtained, and single-stranded regions are readily distinguishable from double-stranded regions. The arrows point to some single-stranded bubbles where denaturation has occurred. The regions that denature are highly reproducible and are rich in A═T base pairs.

denatured regions (called bubbles) can be visualized with electron microscopy **(Fig. 8–28)**. Note that in the strand separation of DNA that occurs in vivo during processes such as DNA replication and transcription, the sites where these processes are initiated are often rich in A═T base pairs, as we shall see.

Duplexes of two RNA strands or one RNA strand and one DNA strand (RNA-DNA hybrids) can also be denatured. Notably, RNA duplexes are more stable to heat denaturation than DNA duplexes. At neutral pH, denaturation of a double-helical RNA often requires temperatures 20 °C or more higher than those required for denaturation of a DNA molecule with a comparable sequence, assuming the strands in each molecule are perfectly complementary. The stability of an RNA-DNA hybrid is generally intermediate between that of RNA and DNA duplexes. The physical basis for these differences in thermal stability is not known.

Nucleic Acids from Different Species Can Form Hybrids

The ability of two complementary DNA strands to pair with one another can be used to detect similar DNA sequences in two different species or within the genome of a single species. If duplex DNAs isolated from human cells and from mouse cells are completely denatured by heating, then mixed and kept at about 25 °C below their t_m for many hours, much of the DNA will anneal. The rate of DNA annealing is affected by temperature, the length and concentration of the DNA fragments being annealed, the concentration of salts in the reaction

mixture, and properties of the sequence itself (e.g., complexity and G≡C content). Temperature is especially important. If the temperature is too low, short sequences with coincidental similarity from distant, heterologous parts of the DNA molecules will anneal unproductively and interfere with the more general alignment of complementary DNA strands. Temperatures that are too high will favor denaturation. Most of the reannealing occurs between complementary mouse DNA strands to form mouse duplex DNA; similarly, most human DNA strands anneal with complementary human DNA strands. However, some strands of the mouse DNA will associate with human DNA strands to yield **hybrid duplexes**, in which segments of a mouse DNA strand form base-paired regions with segments of a human DNA strand **(Fig. 8–29)**. This reflects a common evolutionary heritage; different organisms generally have many proteins and RNAs with similar functions and, often, similar structures. In many cases, the DNAs encoding these proteins and RNAs have similar sequences. The closer the evolutionary relationship between two species, the more extensively their DNAs will hybridize. For example, human DNA hybridizes much more extensively with mouse DNA than with DNA from yeast.

The hybridization of DNA strands from different sources forms the basis for a powerful set of techniques essential to the practice of modern molecular genetics. A specific DNA sequence or gene can be detected in the presence of many other sequences if one already has an appropriate complementary DNA strand (usually labeled in some way) to hybridize with it (Chapter 9). The complementary DNA can be from a different species or from the same species, or it can be synthesized chemically in the laboratory using techniques described later in this chapter. Hybridization techniques can be varied to detect a specific RNA rather than DNA. The isolation and identification of specific genes and RNAs rely on these hybridization techniques. Applications of this technology make possible the identification of an individual on the basis of a single hair left at the scene of a crime or the prediction of the onset of a disease decades before symptoms appear (see Box 9–1).

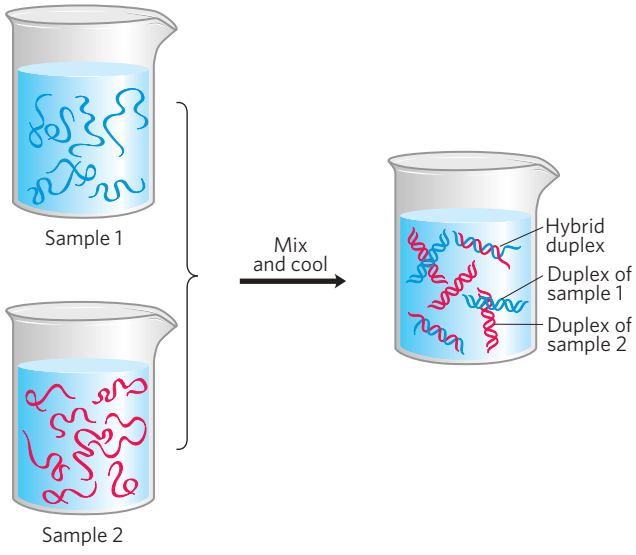


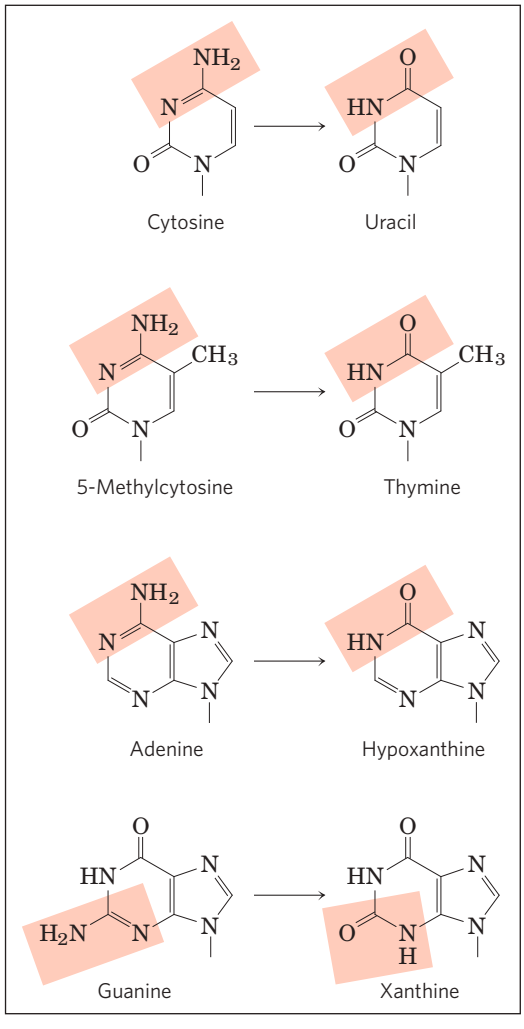
FIGURE 8–29 DNA hybridization. Two DNA samples to be compared are completely denatured by heating. When the two solutions are mixed and slowly cooled, DNA strands of each sample associate with their normal complementary partner and anneal to form duplexes. If the two DNAs have significant sequence similarity, they also tend to form partial duplexes or hybrids with each other: the greater the sequence similarity between the two DNAs, the greater the number of hybrids formed. Hybrid formation can be measured in several ways. One of the DNAs is usually labeled with a radioactive isotope to simplify their detection and measurement.

Nucleotides and Nucleic Acids Undergo Nonenzymatic Transformations

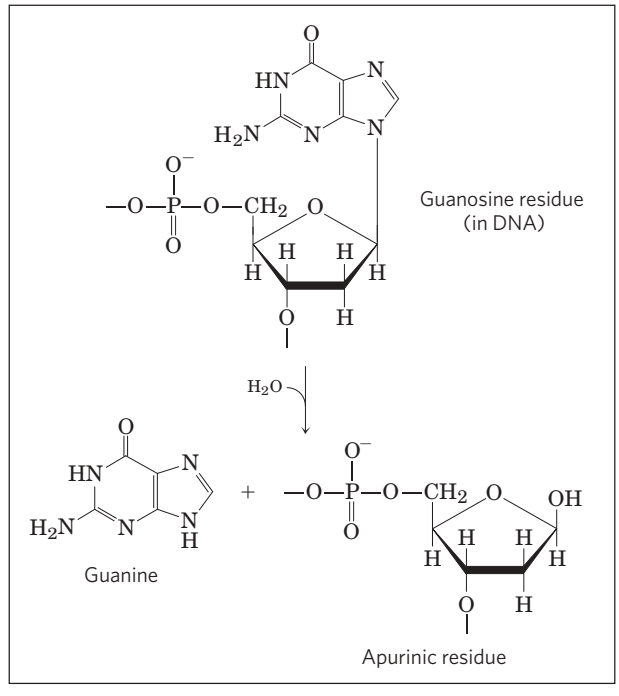
Purines and pyrimidines, along with the nucleotides of which they are a part, undergo spontaneous alterations in their covalent structure. The rate of these reactions is generally *very slow*, but they are physiologically significant because of the cell's very low tolerance for alterations in its genetic information. Alterations in DNA structure that produce permanent changes in the genetic information encoded therein are called **mutations**, and much evidence suggests an intimate link between the accumulation of mutations in an individual organism and the process of aging and carcinogenesis.

Several nucleotide bases undergo spontaneous loss of their exocyclic amino groups (deamination) **(Fig. 8–30a)**. For example, under typical cellular conditions, deamination of cytosine (in DNA) to uracil occurs in about one of every 10^7 cytidine residues in 24 hours. This corresponds to about 100 spontaneous events per day, on average, in a mammalian cell. Deamination of adenine and guanine occurs at about 1/100th this rate.

The slow cytosine deamination reaction seems innocuous enough, but is almost certainly the reason why DNA contains thymine rather than uracil. The product of cytosine deamination (uracil) is readily recognized as foreign in DNA and is removed by a repair system (Chapter 25). If DNA normally contained uracil, recognition of uracils resulting from cytosine deamination would be more difficult, and unrepaired uracils would lead to permanent sequence changes as they were paired with adenines during replication. Cytosine deamination would gradually lead to a decrease in G≡C base pairs and an increase in A═U base pairs in the DNA of all cells. Over the millennia, cytosine deamination could eliminate G≡C base pairs and the genetic code that depends on them. Establishing thymine as one of the four bases in DNA may well



(a) Deamination



(b) Depurination

FIGURE 8–30 **Some well-characterized nonenzymatic reactions of nucleotides.** **(a)** Deamination reactions. Only the base is shown. **(b)** Depurination, in which a purine is lost by hydrolysis of the *N*-β-glycosyl bond. Loss of pyrimidines via a similar reaction occurs, but much more slowly. The resulting lesion, in which the deoxyribose is present but the base is not, is called an abasic site or an AP site (apurinic site or, rarely, apyrimidinic site). The deoxyribose remaining after depurination is readily converted from the β-furanose to the aldehyde form (see Fig. 8–3), further destabilizing the DNA at this position. More nonenzymatic reactions are illustrated in Figures 8–31 and 8–32.

have been one of the crucial turning points in evolution, making the long-term storage of genetic information possible.

Another important reaction in deoxyribonucleotides is the hydrolysis of the *N*-β-glycosyl bond between the base and the pentose, to create a DNA lesion called an AP (apurinic, apyrimidinic) site or abasic site (Fig. 8–30b). This occurs at a higher rate for purines than for pyrimidines. As many as one in 10^5 purines (10,000 per mammalian cell) are lost from DNA every 24 hours under typical cellular conditions. Depurination of ribonucleotides and RNA is much slower and generally is not considered physiologically significant. In the test tube, loss of purines can be accelerated by dilute acid. Incubation of DNA at pH 3 causes selective removal of the purine bases, resulting in a derivative called apurinic acid.

Other reactions are promoted by radiation. UV light induces the condensation of two ethylene groups to form a cyclobutane ring. In the cell, the same reaction between adjacent pyrimidine bases in nucleic acids forms cyclobutane pyrimidine dimers. This happens most frequently between adjacent thymidine residues on the same DNA strand **(Fig. 8–31)**. A second type of pyrimidine dimer, called a 6-4 photoproduct, is also formed during UV irradiation. Ionizing radiation (x rays and gamma rays) can cause ring opening and fragmentation of bases as well as breaks in the covalent backbone of nucleic acids.

Virtually all forms of life are exposed to energy-rich radiation capable of causing chemical changes in DNA. Near-UV radiation (with wavelengths of 200 to 400 nm), which makes up a significant portion of the solar spectrum, is known to cause pyrimidine dimer formation and other chemical changes in the DNA of bacteria and of human skin cells. We are subject to a constant field of ionizing radiation in the form of cosmic rays, which can penetrate deep into the earth, as well as radiation emitted from radioactive elements, such as radium, plutonium, uranium, radon, ^{14}C, and ^{3}H. X rays used in medical and dental examinations and in radiation therapy of cancer and other diseases are another form of ionizing radiation. It is estimated that UV and ionizing radiations are responsible for about 10% of all DNA damage caused by environmental agents.

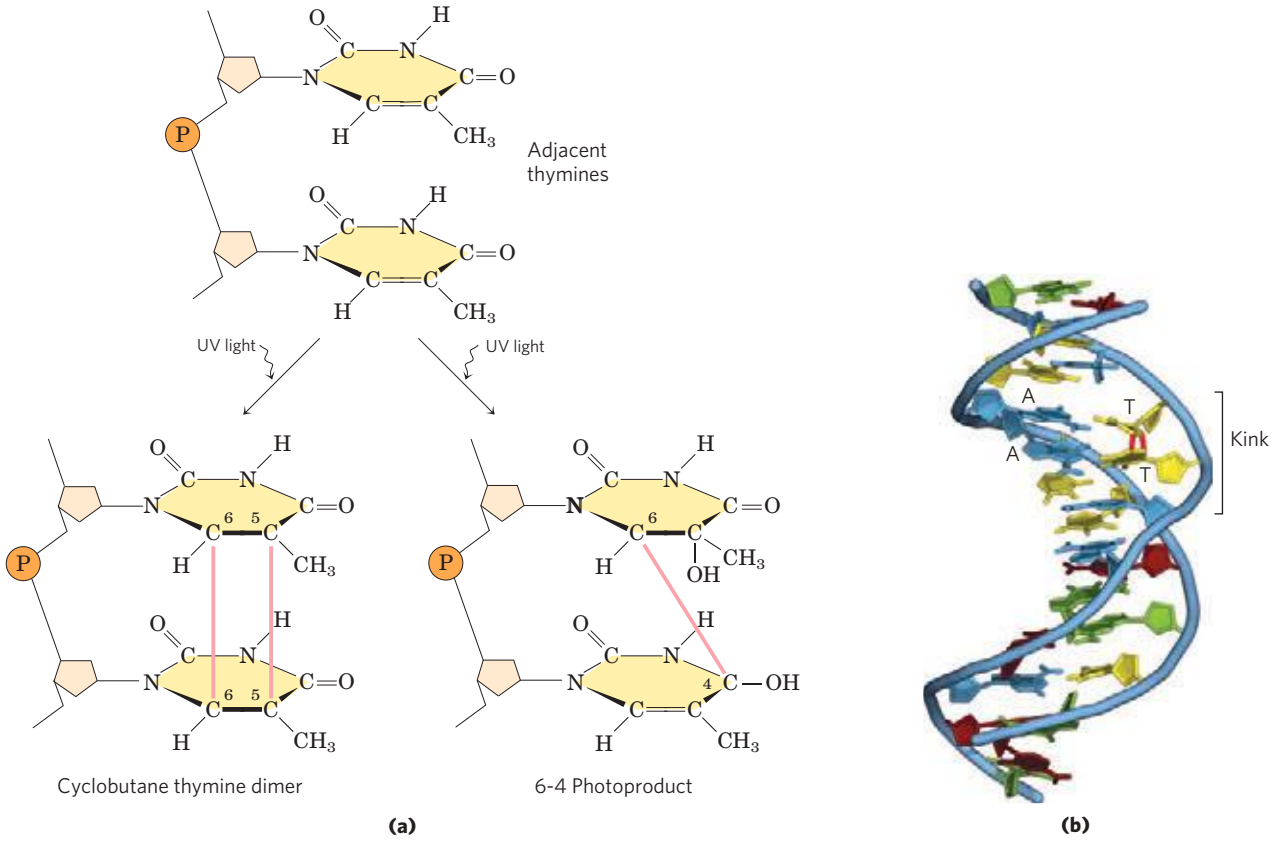
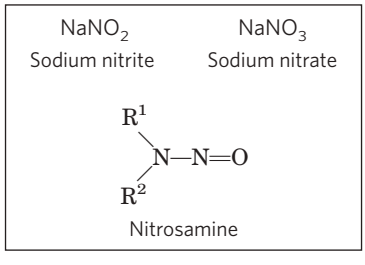


FIGURE 8–31 **Formation of pyrimidine dimers induced by UV light.** **(a)** One type of reaction (on the left) results in the formation of a cyclobutyl ring involving C-5 and C-6 of adjacent pyrimidine residues. An alternative reaction (on the right) results in a 6-4 photoproduct, with a linkage between C-6 of one pyrimidine and C-4 of its neighbor. **(b)** Formation of a cyclobutane pyrimidine dimer introduces a bend or kink into the DNA (PDB ID 1TTD).

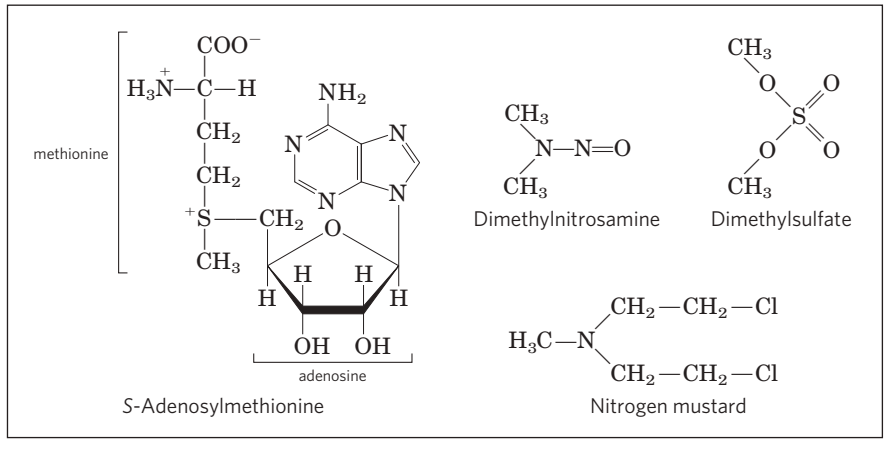
DNA also may be damaged by reactive chemicals introduced into the environment as products of industrial activity. Such products may not be injurious per se but may be metabolized by cells into forms that are. There are two prominent classes of such agents **(Fig. 8–32)**: (1) deaminating agents, particularly nitrous acid (HNO_2) or compounds that can be metabolized to nitrous acid or nitrites, and (2) alkylating agents.

Nitrous acid, formed from organic precursors such as nitrosamines and from nitrite and nitrate salts, is a potent accelerator of the deamination of bases. Bisulfite has similar effects. Both agents are used as preservatives in processed foods to prevent the growth of toxic bacteria. They do not seem to increase cancer risks significantly when used in this way, perhaps because they are used in small amounts and make only a minor contribution to the overall levels of DNA damage. (The potential health risk from food spoilage if these preservatives were not used is much greater.)

FIGURE 8–32 **Chemical agents that cause DNA damage.** **(a)** Precursors of nitrous acid, which promotes deamination reactions. **(b)** Alkylating agents. Only *S*-adenosylmethionine acts enzymatically.

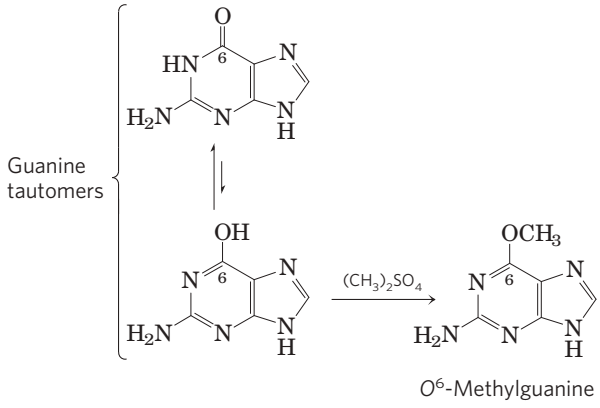


(a) Nitrous acid precursors



(b) Alkylating agents

Alkylating agents can alter certain bases of DNA. For example, the highly reactive chemical dimethylsulfate (Fig. 8–32b) can methylate a guanine to yield O^6-methylguanine, which cannot base-pair with cytosine.



Many similar reactions are brought about by alkylating agents normally present in cells, such as *S*-adenosyl methionine.

The most important source of mutagenic alterations in DNA is oxidative damage. Excited-oxygen species such as hydrogen peroxide, hydroxyl radicals, and superoxide radicals arise during irradiation or as a byproduct of aerobic metabolism. Of these species, the hydroxyl radicals are responsible for most oxidative DNA damage. Cells have an elaborate defense system to destroy reactive oxygen species, including enzymes such as catalase and superoxide dismutase that convert reactive oxygen species to harmless products. A fraction of these oxidants inevitably escape cellular defenses, however, and damage to DNA occurs through any of a large, complex group of reactions ranging from oxidation of deoxyribose and base moieties to strand breaks. Accurate estimates for the extent of this damage are not yet available, but every day the DNA of each human cell is subjected to thousands of damaging oxidative reactions.

This is merely a sampling of the best-understood reactions that damage DNA. Many carcinogenic compounds in food, water, or air exert their cancer-causing effects by modifying bases in DNA. Nevertheless, the integrity of DNA as a polymer is better maintained than that of either RNA or protein, because DNA is the only macromolecule that has the benefit of extensive biochemical repair systems. These repair processes (described in Chapter 25) greatly lessen the impact of damage to DNA. ■

Some Bases of DNA Are Methylated

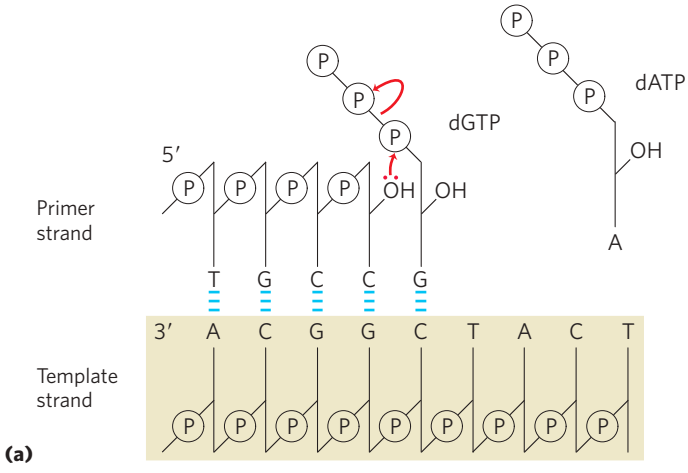
Certain nucleotide bases in DNA molecules are enzymatically methylated. Adenine and cytosine are methylated more often than guanine and thymine. Methylation is generally confined to certain sequences or regions of a DNA molecule. In some cases the function of methylation is well understood; in others the function remains unclear. All known DNA methylases use *S*-adenosylmethionine as a methyl group donor (Fig. 8–32b). *E. coli* has two prominent methylation systems. One serves as part of a defense mechanism that helps the cell to distinguish its DNA from foreign DNA by marking its own DNA with methyl groups and destroying (foreign) DNA without the methyl groups (this is known as a restriction-modification system; see p. 314). The other system methylates adenosine residues within the sequence (5′)GATC(3′) to N^6-methyladenosine (Fig. 8–5a). This is mediated by the Dam (*D*NA *a*denine *m*ethylation) methylase, a component of a system that repairs mismatched base pairs formed occasionally during DNA replication (see Fig. 25–21).

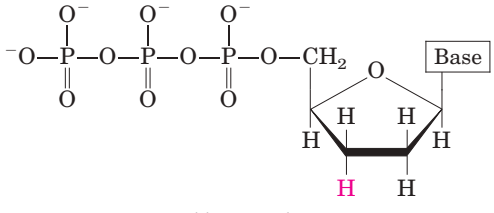
In eukaryotic cells, about 5% of cytidine residues in DNA are methylated to 5-methylcytidine (Fig. 8–5a). Methylation is most common at CpG sequences, producing methyl-CpG symmetrically on both strands of the DNA. The extent of methylation of CpG sequences varies by molecular region in large eukaryotic DNA molecules.

The Sequences of Long DNA Strands Can Be Determined

In its capacity as a repository of information, a DNA molecule's most important property is its nucleotide sequence. Until the late 1970s, determining the sequence of a nucleic acid containing even five or ten nucleotides was very laborious. The development of two new techniques in 1977, one by Alan Maxam and Walter Gilbert and the other by Frederick Sanger, made possible the sequencing of larger DNA molecules with an ease unimagined just a few years before. The techniques depend on an improved understanding of nucleotide chemistry and DNA metabolism, and on electrophoretic methods for separating DNA strands differing in size by only one nucleotide. Electrophoresis of DNA is similar to that of proteins (see Fig. 3–18). Polyacrylamide is often used as the gel matrix in work with short DNA molecules (up to a few hundred nucleotides); agarose is generally used for longer pieces of DNA.

In both Sanger and Maxam-Gilbert sequencing, the general principle is to reduce the DNA to four sets of labeled fragments. The reaction producing each set is base-specific, so the lengths of the fragments correspond to positions in the DNA sequence where a certain base occurs. For example, for an oligonucleotide with the sequence pAATCGACT, labeled at the 5′ end (the left end), a reaction that breaks the DNA after each C residue will generate two labeled fragments: a four-nucleotide and a seven-nucleotide fragment; a reaction that breaks the DNA after each G will produce only one labeled, five-nucleotide fragment. Because the fragments are radioactively labeled at their 5′ ends, only the fragment to the 5′ side of the break is visualized. The fragment sizes correspond to the relative positions of C and G residues in the sequence. When the sets of fragments corresponding to each of the four bases are electrophoretically separated side by side, they produce a ladder of bands from which the sequence can be read directly **(Fig. 8–33)**. We illustrate only the Sanger





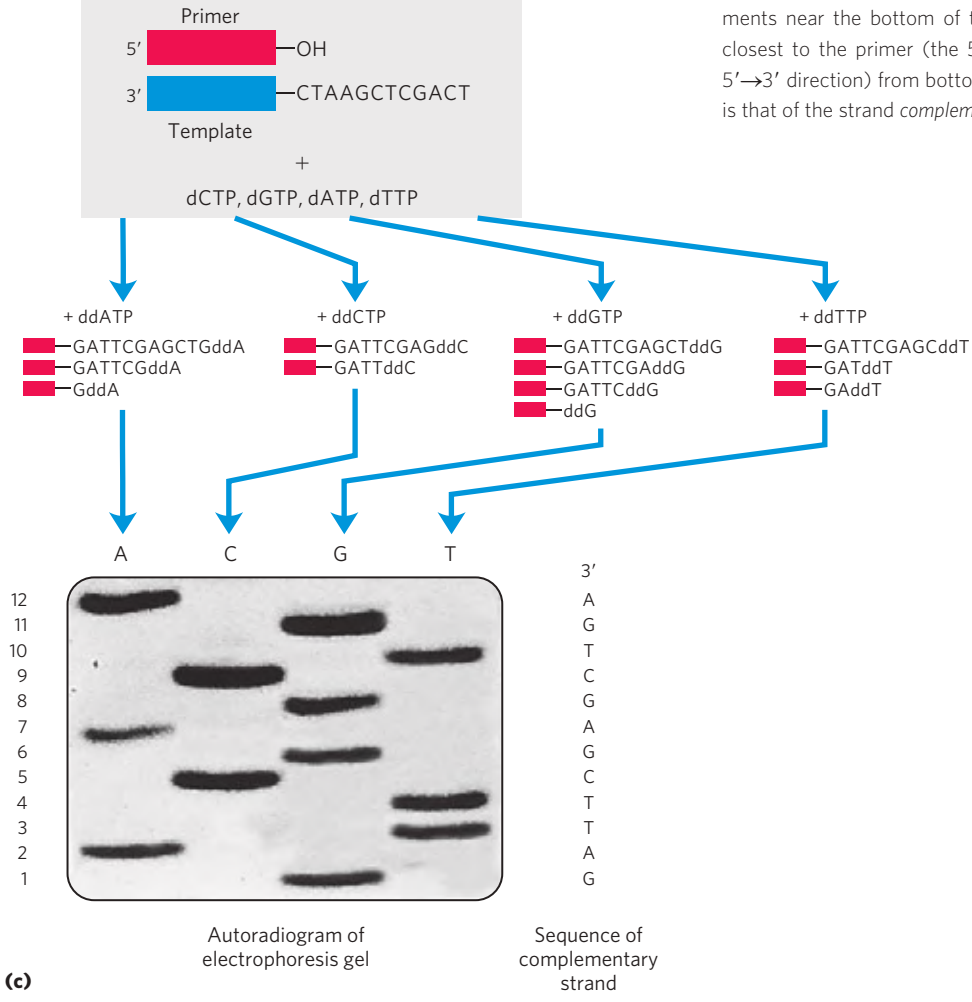


FIGURE 8–33 DNA sequencing by the Sanger method. This method makes use of the mechanism of DNA synthesis by DNA polymerases (Chapter 25). **(a)** DNA polymerases require both a primer (a short oligonucleotide strand), to which nucleotides are added, and a template strand to guide selection of each new nucleotide. In cells, the 3′-hydroxyl group of the primer reacts with an incoming deoxynucleoside triphosphate (dNTP) to form a new phosphodiester bond. **(b)** The Sanger sequencing procedure uses dideoxynucleoside triphosphate (ddNTP) analogs to interrupt DNA synthesis. (The Sanger method is also known as the dideoxy method.) When a ddNTP is inserted in place of a dNTP, strand elongation is halted after the analog is added, because it lacks the 3′-hydroxyl group needed for the next step. **(c)** The DNA to be sequenced is used as the template strand, and a short primer, radioactively or fluorescently labeled, is annealed to it. By addition of small amounts of a single ddNTP, for example ddCTP, to an otherwise normal reaction system, the synthesized strands will be prematurely terminated at some locations where dC normally occurs. Given the excess of dCTP over ddCTP, the chance that the analog will be incorporated whenever a dC is to be added is small. However, ddCTP is present in sufficient amounts to ensure that each new strand has a high probability of acquiring at least one ddC at some point during synthesis. The result is a solution containing a mixture of labeled fragments, each ending with a C residue. Each C residue in the sequence generates a set of fragments of a particular length, such that the different-sized fragments, separated by electrophoresis, reveal the location of C residues. This procedure is repeated separately for each of the four ddNTPs, and the sequence can be read directly from an autoradiogram of the gel. Because shorter DNA fragments migrate faster, the fragments near the bottom of the gel represent the nucleotide positions closest to the primer (the 5′ end), and the sequence is read (in the 5′→3′ direction) from bottom to top. Note that the sequence obtained is that of the strand *complementary* to the strand being analyzed.

method, because it has proved to be technically easier and is in more widespread use. It requires the enzymatic synthesis of a DNA strand complementary to the strand under analysis, using a radioactively labeled "primer" and dideoxynucleotides.

Since these first practical DNA-sequencing methods appeared, methodology has improved rapidly. Much of the advance has been fueled by the Human Genome Project, described in Chapter 9. A variation of Sanger's sequencing method, in which the dideoxynucleotides used for each reaction are labeled with a differently colored fluorescent tag **(Fig. 8–34)**, was used in early efforts to automate large DNA-sequencing efforts. With this technology, researchers can sequence DNA molecules containing thousands of nucleotides in a few hours. This approach was heavily used in the initial efforts to sequence entire organism genomes, and is still used for routine sequencing of genes or DNA segments. However, modern genomic sequencing now makes use of vastly more efficient methods, sometimes referred to as next-generation or **next-gen sequencing**. These are described in Chapter 9. **Dideoxy Sequencing of DNA**

The Chemical Synthesis of DNA Has Been Automated

An important practical advance in nucleic acid chemistry was the rapid and accurate synthesis of short oligonucleotides of known sequence. The methods were pioneered by H. Gobind Khorana and his colleagues in the 1970s. Refinements by Robert Letsinger and Marvin Caruthers led to the chemistry now in widest use, called the phosphoramidite method **(Fig. 8–35)**. The synthesis is carried out with the growing strand attached to a solid support, using principles similar to those used by Merrifield for peptide synthesis (see Fig. 3–32), and is readily automated. The efficiency of each addition step is very high, allowing the routine synthesis of polymers containing 70 or 80 nucleotides and, in some laboratories, much longer strands. The availability of relatively inexpensive DNA polymers with predesigned sequences is having a powerful impact on all areas of biochemistry (Chapter 9).

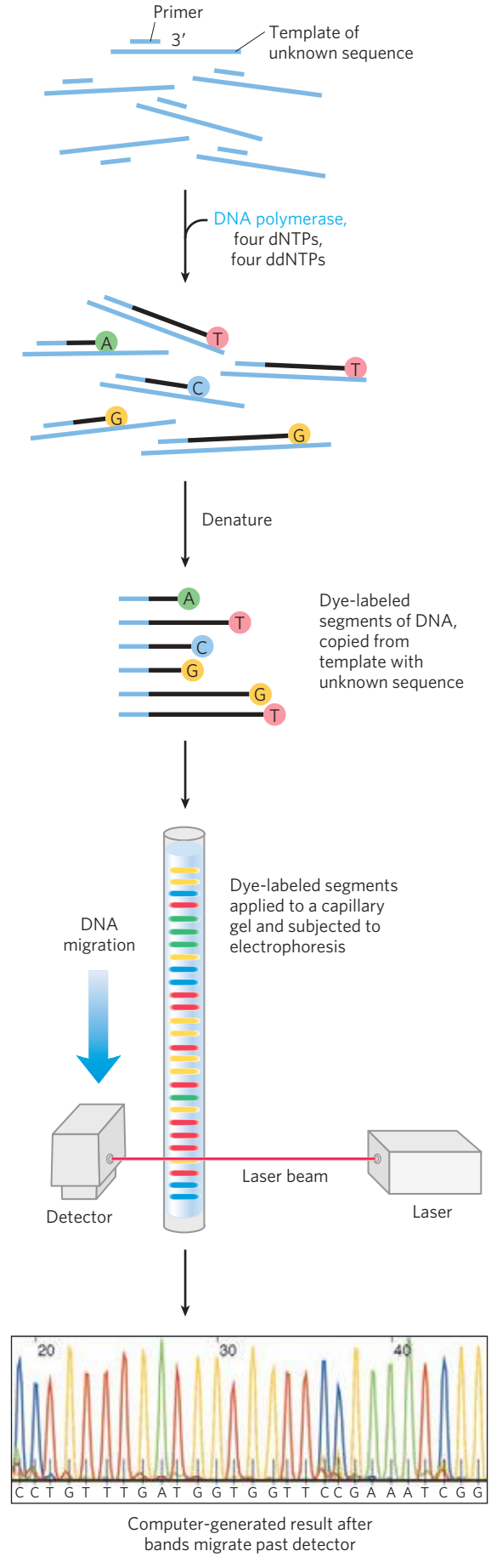


FIGURE 8–34 Strategy for automating DNA-sequencing reactions. Each dideoxynucleotide used in the Sanger method can be linked to a fluorescent molecule that gives all the fragments terminating in that nucleotide a particular color. All four labeled ddNTPs are added to a single tube. The resulting colored DNA fragments are then separated by size in a single electrophoretic gel contained in a capillary tube (a refinement of gel electrophoresis that allows for faster separations). All fragments of a given length migrate through the capillary gel in a single peak, and the color associated with each peak is detected using a laser beam. The DNA sequence is read by determining the sequence of colors in the peaks as they pass the detector. This information is fed directly to a computer, which determines the sequence.

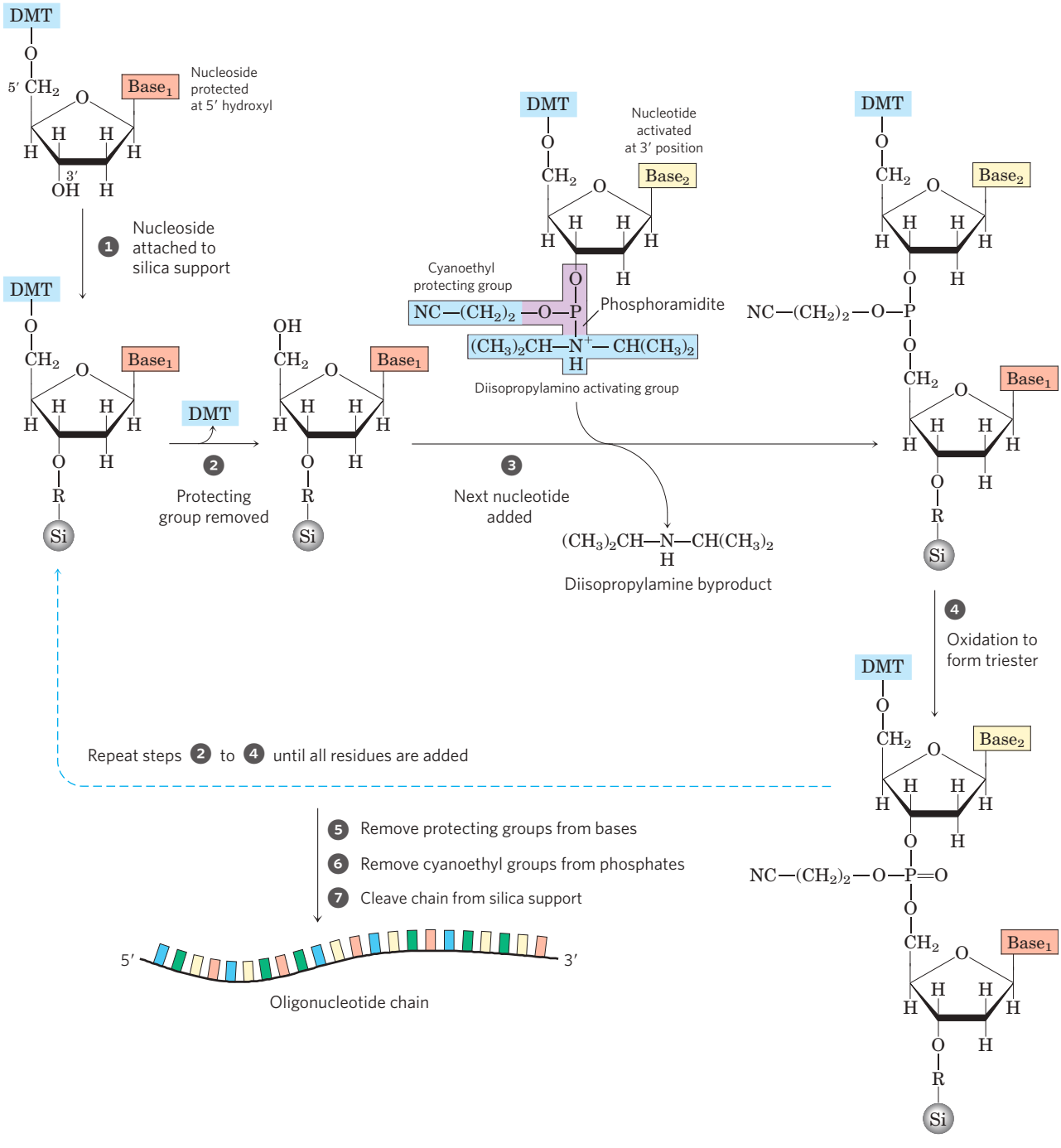


FIGURE 8–35 Chemical synthesis of DNA by the phosphoramidite method. Automated DNA synthesis is conceptually similar to the synthesis of polypeptides on a solid support. The oligonucleotide is built up on the solid support (silica), one nucleotide at a time, in a repeated series of chemical reactions with suitably protected nucleotide precursors. ❶ The first nucleoside (which will be the 3′ end) is attached to the silica support at the 3′ hydroxyl (through a linking group, R) and is protected at the 5′ hydroxyl with an acid-labile dimethoxytrityl group (DMT). The reactive groups on all bases are also chemically protected. ❷ The protecting DMT group is removed by washing the column with acid (the DMT group is colored, so this reaction can be followed spectrophotometrically). ❸ The next nucleotide has a reactive phosphoramidite at its 3′ position: a trivalent phosphite (as opposed to the more oxidized pentavalent phosphate normally present in nucleic acids) with one linked oxygen replaced by an amino group or substituted amine. In the common variant shown, one of the phosphoramidite oxygens is bonded to the deoxyribose, the other is protected by a cyanoethyl group, and the third position is occupied by a readily displaced diisopropylamino group. Reaction with the immobilized nucleotide forms a 5′,3′ linkage, and the diisopropylamino group is eliminated. In step ❹, the phosphite linkage is oxidized with iodine to produce a phosphotriester linkage. Reactions ❷ through ❹ are repeated until all nucleotides are added. At each step, excess nucleotide is removed before addition of the next nucleotide. In steps ❺ and ❻ the remaining protecting groups on the bases and the phosphates are removed, and in ❼ the oligonucleotide is separated from the solid support and purified. The chemical synthesis of RNA is somewhat more complicated because of the need to protect the 2′ hydroxyl of ribose without adversely affecting the reactivity of the 3′ hydroxyl.

SUMMARY 8.3 Nucleic Acid Chemistry

- Native DNA undergoes reversible unwinding and separation of strands (melting) on heating or at extremes of pH. DNAs rich in G≡C pairs have higher melting points than DNAs rich in A═T pairs.
- Denatured single-stranded DNAs from two species can form a hybrid duplex, the degree of hybridization depending on the extent of sequence similarity. Hybridization is the basis for important techniques used to study and isolate specific genes and RNAs.
- DNA is a relatively stable polymer. Spontaneous reactions such as deamination of certain bases, hydrolysis of base-sugar *N*-glycosyl bonds, radiation-induced formation of pyrimidine dimers, and oxidative damage occur at very low rates, yet are important because of a cell's very low tolerance for changes in genetic material.
- DNA sequences can be determined with a range of modern methods.
- Oligonucleotides of known sequence can be synthesized rapidly and accurately.

8.4 Other Functions of Nucleotides

In addition to their roles as the subunits of nucleic acids, nucleotides have a variety of other functions in every cell: as energy carriers, components of enzyme cofactors, and chemical messengers.

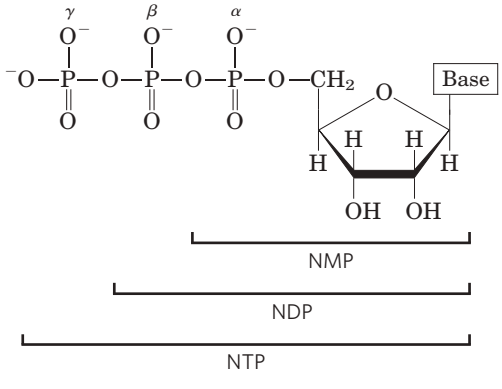
Nucleotides Carry Chemical Energy in Cells

The phosphate group covalently linked at the 5′ hydroxyl of a ribonucleotide may have one or two additional phosphates attached. The resulting molecules are referred to as nucleoside mono-, di-, and triphosphates **(Fig. 8–36)**. Starting from the ribose, the three phosphates are generally labeled α, β, and γ. Hydrolysis of nucleoside triphosphates provides the chemical energy to drive many cellular reactions. Adenosine 5′-triphosphate, ATP, is by far the most widely used for this purpose, but UTP, GTP, and CTP are also used in some reactions. Nucleoside triphosphates also serve as the activated precursors of DNA and RNA synthesis, as described in Chapters 25 and 26.

The energy released by hydrolysis of ATP and the other nucleoside triphosphates is accounted for by the structure of the triphosphate group. The bond between the ribose and the α phosphate is an ester linkage. The α,β and β,γ linkages are phosphoanhydrides **(Fig. 8–37)**. Hydrolysis of the ester linkage yields about 14 kJ/mol under standard conditions, whereas hydrolysis of each anhydride bond yields about 30 kJ/mol. ATP hydrolysis often plays an important thermodynamic role in biosynthesis. When coupled to a reaction with a positive free-energy change, ATP hydrolysis shifts the equilibrium of the overall process to favor product formation (recall the relationship between equilibrium constant and free-energy change described by Eqn 6–3 on p. 194).

Adenine Nucleotides Are Components of Many Enzyme Cofactors

A variety of enzyme cofactors serving a wide range of chemical functions include adenosine as part of their structure **(Fig. 8–38)**. They are unrelated structurally except for the presence of adenosine. In none of these cofactors does the adenosine portion participate directly in the primary function, but removal of adenosine generally results in a drastic reduction of cofactor activities. For example, removal of the adenine nucleotide (3′-phosphoadenosine diphosphate) from acetoacetyl-CoA, the coenzyme A derivative of acetoacetate, reduces its reactivity as a substrate for β-ketoacyl-CoA transferase (an enzyme of lipid metabolism) by a factor of 10^6. Although this requirement for adenosine has not been investigated in detail, it must involve the binding energy between



Abbreviations of ribonucleoside 5′-phosphates			
Base	Mono-	Di-	Tri-
Adenine	AMP	ADP	ATP
Guanine	GMP	GDP	GTP
Cytosine	CMP	CDP	CTP
Uracil	UMP	UDP	UTP

Abbreviations of deoxyribonucleoside 5′-phosphates			
Base	Mono-	Di-	Tri-
Adenine	dAMP	dADP	dATP
Guanine	dGMP	dGDP	dGTP
Cytosine	dCMP	dCDP	dCTP
Thymine	dTMP	dTDP	dTTP

FIGURE 8–36 Nucleoside phosphates. General structure of the nucleoside 5′-mono-, di-, and triphosphates (NMPs, NDPs, and NTPs) and their standard abbreviations. In the deoxyribonucleoside phosphates (dNMPs, dNDPs, and dNTPs), the pentose is 2′-deoxy-D-ribose.

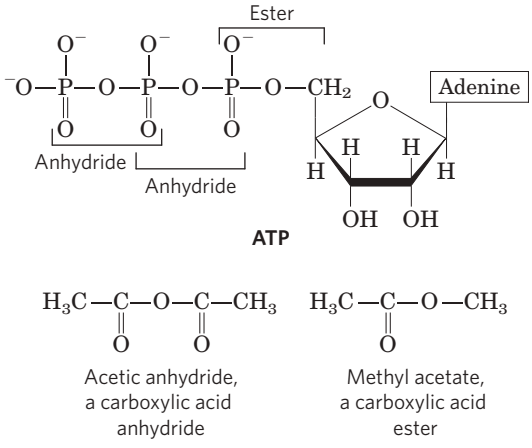


FIGURE 8–37 The phosphate ester and phosphoanhydride bonds of ATP. Hydrolysis of an anhydride bond yields more energy than hydrolysis of the ester. A carboxylic acid anhydride and carboxylic acid ester are shown for comparison.

enzyme and substrate (or cofactor) that is used both in catalysis and in stabilizing the initial enzyme-substrate complex (Chapter 6). In the case of β-ketoacyl-CoA transferase, the nucleotide moiety of coenzyme A seems to be a binding "handle" that helps to pull the substrate (acetoacetyl-CoA) into the active site. Similar roles may be found for the nucleoside portion of other nucleotide cofactors.

Why is adenosine, rather than some other large molecule, used in these structures? The answer here may involve a form of evolutionary economy. Adenosine is certainly not unique in the amount of potential binding energy it can contribute. The importance of adenosine probably lies not so much in some special chemical characteristic as in the evolutionary advantage of using one compound for multiple roles. Once ATP became the universal source of chemical energy, systems developed

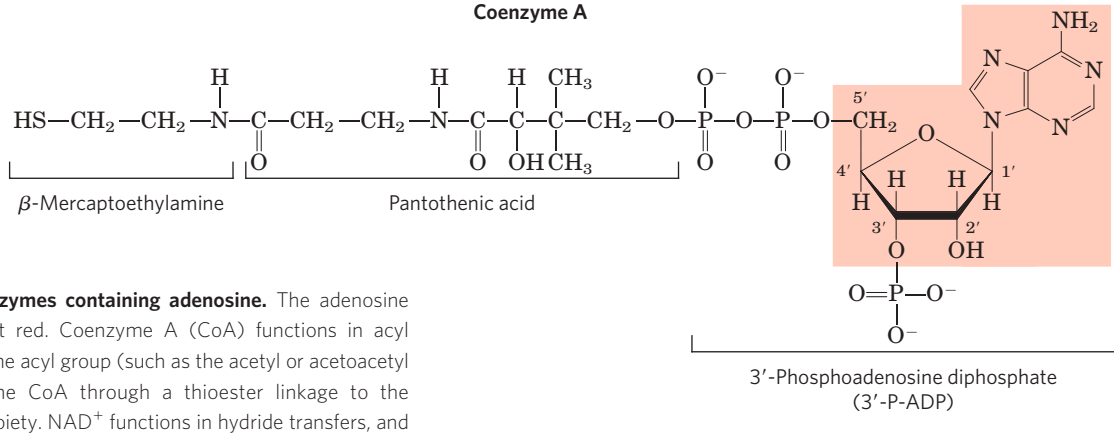
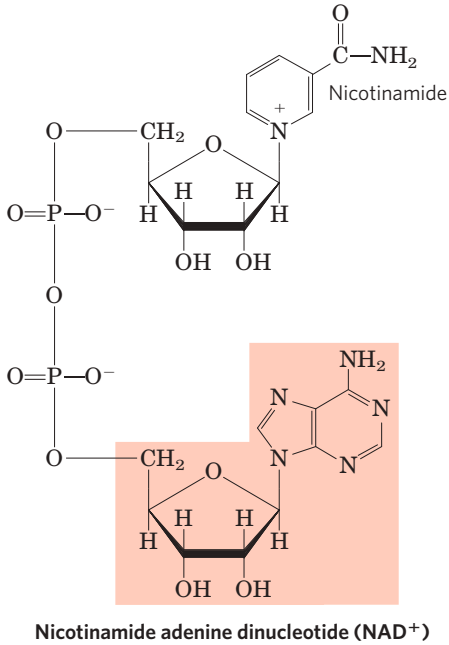
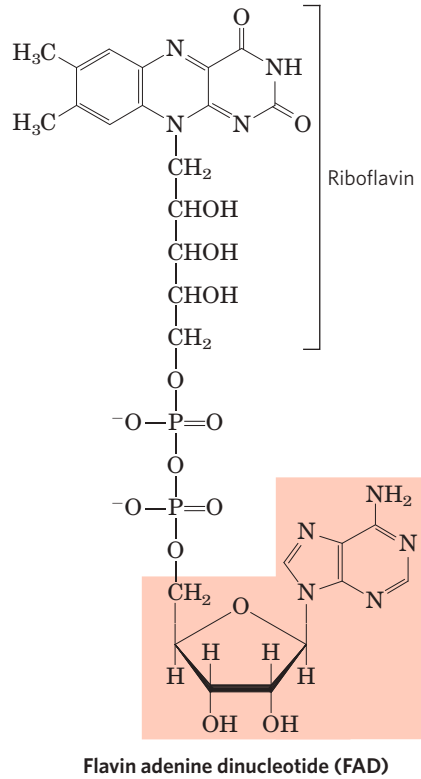


FIGURE 8–38 Some coenzymes containing adenosine. The adenosine portion is shaded in light red. Coenzyme A (CoA) functions in acyl group transfer reactions; the acyl group (such as the acetyl or acetoacetyl group) is attached to the CoA through a thioester linkage to the β-mercaptoethylamine moiety. NAD^+ functions in hydride transfers, and FAD, the active form of vitamin B_2 (riboflavin), in electron transfers. Another coenzyme incorporating adenosine is 5′-deoxyadenosylcobalamin, the active form of vitamin B_{12} (see Box 17–2), which participates in intramolecular group transfers between adjacent carbons.



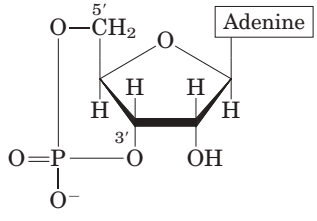


to synthesize ATP in greater abundance than the other nucleotides; because it is abundant, it becomes the logical choice for incorporation into a wide variety of structures. The economy extends to protein structure. A single protein domain that binds adenosine can be used in different enzymes. Such a domain, called a **nucleotide-binding fold**, is found in many enzymes that bind ATP and nucleotide cofactors.

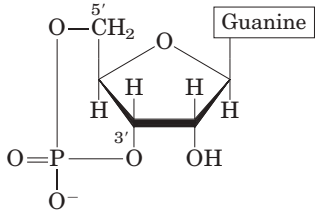
Some Nucleotides Are Regulatory Molecules

Cells respond to their environment by taking cues from hormones or other external chemical signals. The interaction of these extracellular chemical signals ("first messengers") with receptors on the cell surface often leads to the production of **second messengers** inside the cell, which in turn leads to adaptive changes in the cell interior (Chapter 12). Often, the second messenger is a nucleotide **(Fig. 8–39)**. One of the most common is **adenosine 3′,5′-cyclic monophosphate (cyclic AMP**, or **cAMP)**, formed from ATP in a reaction catalyzed by adenylyl cyclase, an enzyme associated with the inner face of the plasma membrane. Cyclic AMP serves regulatory functions in virtually every cell outside the plant kingdom. Guanosine 3′,5′-cyclic monophosphate (cGMP) occurs in many cells and also has regulatory functions.

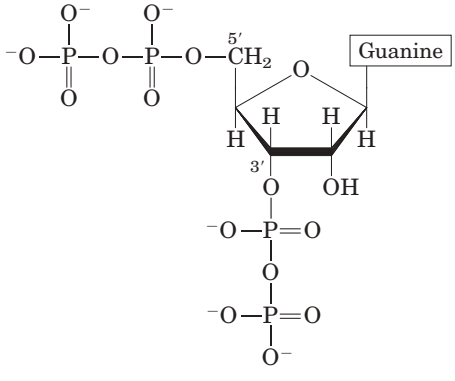
Another regulatory nucleotide, ppGpp (Fig. 8–39), is produced in bacteria in response to a slowdown in protein synthesis during amino acid starvation. This nucleotide inhibits the synthesis of the rRNA and tRNA molecules (see Fig. 28–22) needed for protein synthesis, preventing the unnecessary production of nucleic acids.



Adenosine 3′,5′-cyclic monophosphate
(cyclic AMP; cAMP)



Guanosine 3′,5′-cyclic monophosphate
(cyclic GMP; cGMP)



Guanosine 5′-diphosphate,3′-diphosphate
(guanosine tetraphosphate)
(ppGpp)

FIGURE 8–39 **Three regulatory nucleotides.**

SUMMARY 8.4 Other Functions of Nucleotides

- ATP is the central carrier of chemical energy in cells. The presence of an adenosine moiety in a variety of enzyme cofactors may be related to binding-energy requirements.
- Cyclic AMP, formed from ATP in a reaction catalyzed by adenylyl cyclase, is a common second messenger produced in response to hormones and other chemical signals.

Key Terms

Terms in bold are defined in the glossary.

gene 281
ribosomal RNA (rRNA) 281
messenger RNA (mRNA) 281
transfer RNA (tRNA) 281
nucleotide 281
nucleoside 281
pyrimidine 282
purine 282
deoxyribonucleotides 283
ribonucleotide 283
phosphodiester linkage 285
5′ end 285
3′ end 285
oligonucleotide 286
polynucleotide 286
base pair 287
major groove 289
minor groove 289
B-form DNA 291
A-form DNA 291
Z-form DNA 291
palindrome 291
hairpin 292
cruciform 292
triplex DNA 292
G tetraplex 292
transcription 294
monocistronic mRNA 294
polycistronic mRNA 294
mutation 299
second messenger 308
adenosine 3′,5′-cyclic monophosphate (cyclic AMP, cAMP) 308

Further Reading

General

Cox, M.M., Doudna, J.A., & O'Donnell, M. (2012) *Molecular Biology: Principles and Practice*, W. H. Freeman and Company, New York.

The best place to start to learn more about nucleic acid structure and function.

Friedberg, E.C., Walker, G.C., Siede, W., Wood, R.D., Schultz, R.A., & Ellenberger, T. (2006) *DNA Repair and Mutagenesis*, 2nd edn, ASM Press, Washington, DC.

A good source for more information on the chemistry of nucleotides and nucleic acids.

Historical

Judson, H.F. (1996) *The Eighth Day of Creation: Makers of the Revolution in Biology*, expanded edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Olby, R.C. (1994) *The Path to the Double Helix: The Discovery of DNA*, Dover Publications, Inc., New York.

Sayre, A. (1978) *Rosalind Franklin and DNA*, W. W. Norton & Co., Inc., New York.

Watson, J.D. (1968) *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*, Atheneum, New York. [Paperback edition, Touchstone Books, 2001.]

Nucleic Acid Structure

Frank-Kamenetskii, M.D. & Mirkin, S.M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65–95.

Holbrook, S.R. (2008) Structural principles from large RNAs. *Annu. Rev. Biophys*. **37**, 445–464.

Keniry, M.A. (2000) Quadruplex structures in nucleic acids. *Biopolymers* **56**, 123–146.

Good summary of the structural properties of quadruplexes.

Nucleic Acid Chemistry

Bonetta, L. (2006) Genome sequencing in the fast lane. *Nat. Methods* **3**, 141–147.

This paper introduces a newer generation of sequencing methods that are described in Chapter 9.

Collins, A.R. (1999) Oxidative DNA damage, antioxidants, and cancer. *Bioessays* **21**, 238–246.

Cooke, M.S., Evans, M.D., Dizdaroglu, M., & Lunt J. (2003) Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214.

Imlay, J.A. (2008) Cellular defenses against superoxide and hydrogen peroxide. *Annu. Rev. Biochem.* **77**, 755–776.

Marnett, L.J. & Plastaras, J.P. (2001) Endogenous DNA damage and mutation. *Trends Genet.* **17**, 214–221.

ATP as Energy Carrier

Jencks, W.P. (1987) Economics of enzyme catalysis. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 65–73.

A relatively short article, full of insights.

Problems

1. Nucleotide Structure Which positions in the purine ring of a purine nucleotide in DNA have the potential to form hydrogen bonds but are not involved in Watson-Crick base pairing?

2. Base Sequence of Complementary DNA Strands One strand of a double-helical DNA has the sequence (5′)GCGCAATATTTCTCAAAATATTGCGC(3′). Write the base sequence of the complementary strand. What special type of sequence is contained in this DNA segment? Does the double-stranded DNA have the potential to form any alternative structures?

3. DNA of the Human Body Calculate the weight in grams of a double-helical DNA molecule stretching from the Earth to the moon (~320,000 km). The DNA double helix weighs about 1×10^{-18} g per 1,000 nucleotide pairs; each base pair extends 3.4 Å. For an interesting comparison, your body contains about 0.5 g of DNA!

4. DNA Bending Assume that a poly(A) tract five base pairs long produces a 20° bend in a DNA strand. Calculate the total (net) bend produced in a DNA if the center base pairs (the third of five) of two successive $(dA)_5$ tracts are located (a) 10 base pairs apart; (b) 15 base pairs apart. Assume 10 base pairs per turn in the DNA double helix.

5. Distinction between DNA Structure and RNA Structure Hairpins may form at palindromic sequences in single strands of either RNA or DNA. How is the helical structure of a long and fully base-paired (except at the end) hairpin in RNA different from that of a similar hairpin in DNA?

6. Nucleotide Chemistry The cells of many eukaryotic organisms have highly specialized systems that specifically repair G–T mismatches in DNA. The mismatch is repaired to form a G≡C (not A═T) base pair. This G–T mismatch repair mechanism occurs in addition to a more general system that repairs virtually all mismatches. Suggest why cells might require a specialized system to repair G–T mismatches.

7. Denaturation of Nucleic Acids A duplex DNA oligonucleotide in which one of the strands has the sequence TAATACGACTCACTATAGGG has a melting temperature (t_m) of 59 °C. If an RNA duplex oligonucleotide of identical sequence (substituting U for T) is constructed, will its melting temperature be higher or lower?

8. Spontaneous DNA Damage Hydrolysis of the *N*-glycosyl bond between deoxyribose and a purine in DNA creates an AP site. An AP site generates a thermodynamic destabilization greater than that created by any DNA mismatched base pair. This effect is not completely understood. Examine the structure of an AP site (see Fig. 8–30b) and describe some chemical consequences of base loss.

9. Prediction of Nucleic Acid Structure from Its Sequence A part of a sequenced chromosome has the sequence (on one strand) ATTGCATCCGCGCGTGCGCGCGCG-ATCCCGTTACTTTCCG. Which part of this sequence is most likely to take up the Z conformation?

10. Nucleic Acid Structure Explain why the absorption of UV light by double-stranded DNA increases (the hyperchromic effect) when the DNA is denatured.

11. Determination of Protein Concentration in a Solution Containing Proteins and Nucleic Acids The concentration of protein or nucleic acid in a solution containing both can be estimated by using their different light absorption properties: proteins absorb most strongly at 280 nm and nucleic acids at 260 nm. Estimates of their respective concentrations in a mixture can be made by measuring the absorbance (A) of the solution at 280 and 260 nm and using the table on the next page, which gives $R_{280/260}$, the ratio of absorbances at 280 and 260 nm; the percentage of total mass that is nucleic acid; and a factor, F, that corrects the A_{280} reading and gives a more accurate protein

estimate. The protein concentration (in mg/mL) = $F \times A_{280}$ (assuming the cuvette is 1 cm wide). Calculate the protein concentration in a solution of $A_{280} = 0.69$ and $A_{260} = 0.94$.

$R_{280/260}$	Proportion of nucleic acid (%)	F
1.75	0.00	1.116
1.63	0.25	1.081
1.52	0.50	1.054
1.40	0.75	1.023
1.36	1.00	0.994
1.30	1.25	0.970
1.25	1.50	0.944
1.16	2.00	0.899
1.09	2.50	0.852
1.03	3.00	0.814
0.979	3.50	0.776
0.939	4.00	0.743
0.874	5.00	0.682
0.846	5.50	0.656
0.822	6.00	0.632
0.804	6.50	0.607
0.784	7.00	0.585
0.767	7.50	0.565
0.753	8.00	0.545
0.730	9.00	0.508
0.705	10.00	0.478
0.671	12.00	0.422
0.644	14.00	0.377
0.615	17.00	0.322
0.595	20.00	0.278

12. Solubility of the Components of DNA Draw the following structures and rate their relative solubilities in water (most soluble to least soluble): deoxyribose, guanine, phosphate. How are these solubilities consistent with the three-dimensional structure of double-stranded DNA?

13. Sanger Sequencing Logic In the Sanger (dideoxy) method for DNA sequencing, a small amount of a dideoxynucleotide triphosphate—say, ddCTP—is added to the sequencing reaction along with a larger amount of the corresponding dCTP. What result would be observed if the dCTP were omitted?

14. DNA Sequencing The following DNA fragment was sequenced by the Sanger method. The red asterisk indicates a fluorescent label.

*5′ ——— 3′-OH
3′ ——— ATTACGCAAGGACATTAGAC---5′

A sample of the DNA was reacted with DNA polymerase and each of the nucleotide mixtures (in an appropriate buffer) listed below. Dideoxynucleotides (ddNTPs) were added in relatively small amounts.

1. dATP, dTTP, dCTP, dGTP, ddTTP
2. dATP, dTTP, dCTP, dGTP, ddGTP
3. dATP, dCTP, dGTP, ddTTP
4. dATP, dTTP, dCTP, dGTP

The resulting DNA was separated by electrophoresis on an agarose gel, and the fluorescent bands on the gel were located. The band pattern resulting from nucleotide mixture 1 is shown below. Assuming that all mixtures were run on the same gel, what did the remaining lanes of the gel look like?

15. Snake Venom Phosphodiesterase An exonuclease is an enzyme that sequentially cleaves nucleotides from the end of a polynucleotide strand. Snake venom phosphodiesterase, which hydrolyzes nucleotides from the 3′ end of any oligonucleotide with a free 3′-hydroxyl group, cleaves between the 3′ hydroxyl of the ribose or deoxyribose and the phosphoryl group of the next nucleotide. It acts on single-stranded DNA or RNA and has no base specificity. This enzyme was used in sequence determination experiments before the development of modern nucleic acid sequencing techniques. What are the products of partial digestion by snake venom phosphodiesterase of an oligonucleotide with the following sequence?

(5′)GCGCCAUUGC(3′)—OH

16. Preserving DNA in Bacterial Endospores Bacterial endospores form when the environment is no longer conducive to active cell metabolism. The soil bacterium *Bacillus subtilis*, for example, begins the process of sporulation when one or more nutrients are depleted. The end product is a small, metabolically dormant structure that can survive almost indefinitely with no detectable metabolism. Spores have mechanisms to prevent accumulation of potentially lethal mutations in their DNA over periods of dormancy that can exceed 1,000 years. *B. subtilis* spores are much more resistant than are the organism's growing cells to heat, UV radiation, and oxidizing agents, all of which promote mutations.

(a) One factor that prevents potential DNA damage in spores is their greatly decreased water content. How would this affect some types of mutations?

(b) Endospores have a category of proteins called small acid-soluble proteins (SASPs) that bind to their DNA, preventing formation of cyclobutane-type dimers. What causes cyclobutane dimers, and why do bacterial endospores need mechanisms to prevent their formation?

17. Oligonucleotide Synthesis In the scheme of Figure 8–35, each new base to be added to the growing oligonucleotide is modified so that its 3′ hydroxyl is activated and the 5′ hydroxyl has a dimethoxytrityl (DMT) group attached. What is the function of the DMT group on the incoming base?

Using the Web

18. The Structure of DNA Elucidation of the three-dimensional structure of DNA helped researchers understand how this molecule conveys information that can be faithfully replicated from one generation to the next. To see the secondary structure of double-stranded DNA, go to the Protein Data Bank website (www.pdb.org). Use the PDB identifiers listed below to retrieve the structure summaries for the two forms of DNA. Open the structures using Jmol, and use the controls in the Jmol menu (accessed with a control-click or by clicking on the Jmol logo in the lower right corner of the image screen) to complete the following exercises. Refer to the Jmol help links as needed.

(a) Obtain the file for 141D, a highly conserved, repeated DNA sequence from the end of the HIV-1 (the virus that causes AIDS) genome. Display the molecule as a ball-and-stick structure and color by element. Identify the sugar–phosphate backbone for each strand of the DNA duplex. Locate and identify individual bases. Identify the 5′ end of each strand. Locate the major and minor grooves. Is this a right- or left-handed helix?

(b) Obtain the file for 145D, a DNA with the Z conformation. Display the molecule as a ball-and-stick structure and color by element. Identify the sugar–phosphate backbone for each strand of the DNA duplex. Is this a right- or left-handed helix?

(c) To fully appreciate the secondary structure of DNA, view the molecules in stereo. On the control menu, Select > All, then Style > Stereographic > Cross-eyed viewing or Wall-eyed viewing. (If you have stereographic glasses available, select the appropriate option.) You will see two images of the DNA molecule. Sit with your nose approximately 10 inches from the monitor and focus on the tip of your nose (cross-eyed) or the opposite edges of the screen (wall-eyed). In the background you should see three images of the DNA helix. Shift your focus to the middle image, which should appear three-dimensional. (Note that only one of the two authors can make this work.)

Data Analysis Problem

19. Chargaff's Studies of DNA Structure The chapter section "DNA Is a Double Helix That Stores Genetic Information" includes a summary of the main findings of Erwin Chargaff and his coworkers, listed as four conclusions ("Chargaff's rules"; p. 288). In this problem, you will examine the data Chargaff collected in support of these conclusions.

In one paper, Chargaff (1950) described his analytical methods and some early results. Briefly, he treated DNA samples with acid to remove the bases, separated the bases by paper chromatography, and measured the amount of each base with UV spectroscopy. His results are shown in the three tables below. The *molar ratio* is the ratio of the number of moles of each base in the sample to the number of moles of phosphate in the sample—this gives the fraction of the total number of bases represented by each particular base. The *recovery* is the sum of all four bases (the sum of the molar ratios); full recovery of all bases in the DNA would give a recovery of 1.0.

Molar ratios in ox DNA

	Thymus			Spleen		Liver
Base	**Prep. 1**	**Prep. 2**	**Prep. 3**	**Prep. 1**	**Prep. 2**	**Prep. 1**
Adenine	0.26	0.28	0.30	0.25	0.26	0.26
Guanine	0.21	0.24	0.22	0.20	0.21	0.20
Cytosine	0.16	0.18	0.17	0.15	0.17	
Thymine	0.25	0.24	0.25	0.24	0.24	
Recovery	*0.88*	*0.94*	*0.94*	*0.84*	*0.88*	

Molar ratios in human DNA

	Sperm		Thymus	Liver	
Base	**Prep. 1**	**Prep. 2**	**Prep. 1**	**Normal**	**Carcinoma**
Adenine	0.29	0.27	0.28	0.27	0.27
Guanine	0.18	0.17	0.19	0.19	0.18
Cytosine	0.18	0.18	0.16		0.15
Thymine	0.31	0.30	0.28		0.27
Recovery	*0.96*	*0.92*	*0.91*		*0.87*

Molar ratios in DNA of microorganisms

	Yeast		Avian tubercle bacilli
Base	**Prep. 1**	**Prep. 2**	**Prep. 1**
Adenine	0.24	0.30	0.12
Guanine	0.14	0.18	0.28
Cytosine	0.13	0.15	0.26
Thymine	0.25	0.29	0.11
Recovery	*0.76*	*0.92*	*0.77*

(a) Based on these data, Chargaff concluded that "no differences in composition have so far been found in DNA from different tissues of the same species." This corresponds to conclusion 2 in this chapter. However, a skeptic looking at the data above might say, "They certainly look different to me!" If you were Chargaff, how would you use the data to convince the skeptic to change her mind?

(b) The base composition of DNA from normal and cancerous liver cells (hepatocarcinoma) was not distinguishably different. Would you expect Chargaff's technique to be capable of detecting a difference between the DNA of normal and cancerous cells? Explain your reasoning.

As you might expect, Chargaff's data were not completely convincing. He went on to improve his techniques, as described

in his 1951 paper, in which he reported molar ratios of bases in DNA from a variety of organisms:

Source	A:G	T:C	A:T	G:C	Purine:pyrimidine
Ox	1.29	1.43	1.04	1.00	1.1
Human	1.56	1.75	1.00	1.00	1.0
Hen	1.45	1.29	1.06	0.91	0.99
Salmon	1.43	1.43	1.02	1.02	1.02
Wheat	1.22	1.18	1.00	0.97	0.99
Yeast	1.67	1.92	1.03	1.20	1.0
Haemophilus influenzae type c	1.74	1.54	1.07	0.91	1.0
E. coli K-12	1.05	0.95	1.09	0.99	1.0
Avian tubercle bacillus	0.4	0.4	1.09	1.08	1.1
Serratia marcescens	0.7	0.7	0.95	0.86	0.9
Bacillus schatz	0.7	0.6	1.12	0.89	1.0

(c) According to Chargaff, as stated in conclusion 1 in this chapter, "The base composition of DNA generally varies from one species to another." Provide an argument, based on the data presented so far, that supports this conclusion.

(d) According to conclusion 4, "In all cellular DNAs, regardless of the species . . . A + G = T + C." Provide an argument, based on the data presented so far, that supports this conclusion.

Part of Chargaff's intent was to disprove the "tetranucleotide hypothesis"; this was the idea that DNA was a monotonous tetranucleotide polymer $(AGCT)_n$ and therefore not capable of containing sequence information. Although the data presented above show that DNA cannot be simply a tetranucleotide—if so, all samples would have molar ratios of 0.25 for each base—it was still possible that the DNA from different organisms was a slightly more complex, but still monotonous, repeating sequence.

To address this issue, Chargaff took DNA from wheat germ and treated it with the enzyme deoxyribonuclease for different time intervals. At each time interval, some of the DNA was converted to small fragments; the remaining, larger fragments he called the "core." In the table below, the "19% core" corresponds to the larger fragments left behind when 81% of the DNA was degraded; the "8% core" corresponds to the larger fragments left after 92% degradation.

Base	Intact DNA	19% Core	8% Core
Adenine	0.27	0.33	0.35
Guanine	0.22	0.20	0.20
Cytosine	0.22	0.16	0.14
Thymine	0.27	0.26	0.23
Recovery	*0.98*	*0.95*	*0.92*

(e) How would you use these data to argue that wheat germ DNA is not a monotonous repeating sequence?

References

Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymic degradation. *Experientia* **6**, 201–209.

Chargaff, E. (1951) Structure and function of nucleic acids as cell constituents. *Fed. Proc.* **10**, 654–659.

9

DNA-Based Information Technologies

The complexity of the molecules and systems revealed in this book can sometimes conceal a biochemical reality—what we have learned is just a beginning. Novel proteins and lipids and carbohydrates and nucleic acids are discovered every day, and we often have no clue as to their functions. How many have yet to be encountered, and what might they do? Even well-characterized biomolecules continue to challenge researchers with countless unresolved mechanistic and functional questions. The thousands of completed genome sequences in hand have provided one look at the immensity of the task ahead. Simply put, we do not know the function of most of the DNA—often including half or more of the genes—in a typical genome. Those same genomic sequences also provide an unprecedented opportunity. There is no greater source of information about a cell or organism than that buried in its own DNA. The task of extracting that information has given rise to sophisticated DNA-based technologies that are now at the heart of almost every biochemical story. The technologies we now turn to touch every topic we explore in subsequent chapters.

As objects of study, DNA molecules present a special problem—their size. Chromosomes are far and away the largest biomolecules in any cell. How does a researcher find the information he or she seeks when it is just a small part of a chromosome that can include millions or even billions of contiguous base pairs? Solutions to these problems began to emerge in the 1970s.

Decades of advances by thousands of scientists working in genetics, biochemistry, cell biology, and physical chemistry came together in the laboratories of Paul Berg, Herbert Boyer, and Stanley Cohen to yield techniques for locating, isolating, preparing, and studying small segments of DNA derived from much larger chromosomes. Techniques for DNA cloning paved the way to the modern field of **genomics** and more broadly to many of the technologies that contribute to **systems biology**, the study of biochemistry on the scale of whole cells and organisms.

Every student and instructor, when considering the topics we present in this chapter, encounters a conflict. First, the methods we describe were made possible by advances in our understanding of DNA and RNA metabolism. Hence, one must understand some fundamental concepts of DNA replication, RNA transcription, protein synthesis, and gene regulation to appreciate how these methods work. At the same time, however, modern biochemistry relies on these same methods to such an extent that a current treatment of any aspect of the discipline becomes very difficult without a proper introduction to them. By presenting these methods early in the book, we acknowledge that they are inextricably interwoven with both the advances that gave rise to them and the newer discoveries they

Paul Berg

Herbert Boyer

Stanley N. Cohen

now make possible. The background we necessarily provide makes the discussion here not just an introduction to technology, but also a preview of many of the fundamentals of DNA and RNA biochemistry encountered in later chapters.

We begin by outlining the principles of DNA cloning, then illustrate the range of applications and the potential of many newer technologies that support and accelerate the advance of biochemistry.

9.1 Studying Genes and Their Products

A researcher has isolated a new enzyme that she knows is the key to a human disease. She hopes to isolate large amounts of the protein to crystallize it for structural analysis and to study it. She wants to alter amino acid residues at its active site to understand the reaction it catalyzes. She plans an elaborate research program to elucidate how this enzyme interacts with, and is regulated by, other proteins in the cell. All of this, and much more, becomes possible if she can obtain the gene encoding her enzyme. Unfortunately, that gene is just a few thousand base pairs within a human chromosome with a size measured in hundreds of millions of base pairs. How does she isolate the small segment that she needs and then study it? The answer lies in DNA cloning and methods developed to manipulate cloned genes.

Genes Can Be Isolated by DNA Cloning

A *clone* is an identical copy. This term originally applied to cells of a single type, isolated and allowed to reproduce to create a population of identical cells. When applied to DNA, a clone represents many identical copies of a particular gene segment. In brief, our researcher must cut the gene out of the larger chromosome, attach it to a much smaller piece of carrier DNA, and allow microorganisms to make many copies of it for her. This is the process of **DNA cloning**. The result is selective amplification of a particular gene or DNA segment, facilitating its isolation and study. Classically, the cloning of DNA from any organism entails five general procedures:

1. *Cutting target DNA at precise locations.* Sequence-specific endonucleases (restriction endonucleases) provide the necessary molecular scissors.
2. *Selecting a small carrier molecule of DNA capable of self-replication.* These DNAs are called **cloning vectors** (a vector is a delivery agent). They are typically plasmids or viral DNAs.
3. *Joining two DNA fragments covalently.* The enzyme DNA ligase links the cloning vector and the DNA to be cloned. Composite DNA molecules comprising covalently linked segments from two or more sources are called **recombinant DNAs**.
4. *Moving recombinant DNA from the test tube to a host cell* that will provide the enzymatic machinery for DNA replication.
5. *Selecting or identifying host cells that contain recombinant DNA.*

The methods used to accomplish these and related tasks are collectively referred to as **recombinant DNA technology** or, more informally, **genetic engineering**.

Much of our initial discussion will focus on DNA cloning in the bacterium *Escherichia coli*, the first organism used for recombinant DNA work and still the most common host cell. *E. coli* has many advantages: its DNA metabolism (like many other of its biochemical processes) is well understood; many naturally occurring cloning vectors associated with *E. coli*, such as plasmids and bacteriophages (bacterial viruses; also called phages), are well characterized; and techniques are available for moving DNA expeditiously from one bacterial cell to another. The principles discussed here are broadly applicable to DNA cloning in other organisms, a topic discussed more fully later in the section.

Restriction Endonucleases and DNA Ligases Yield Recombinant DNA

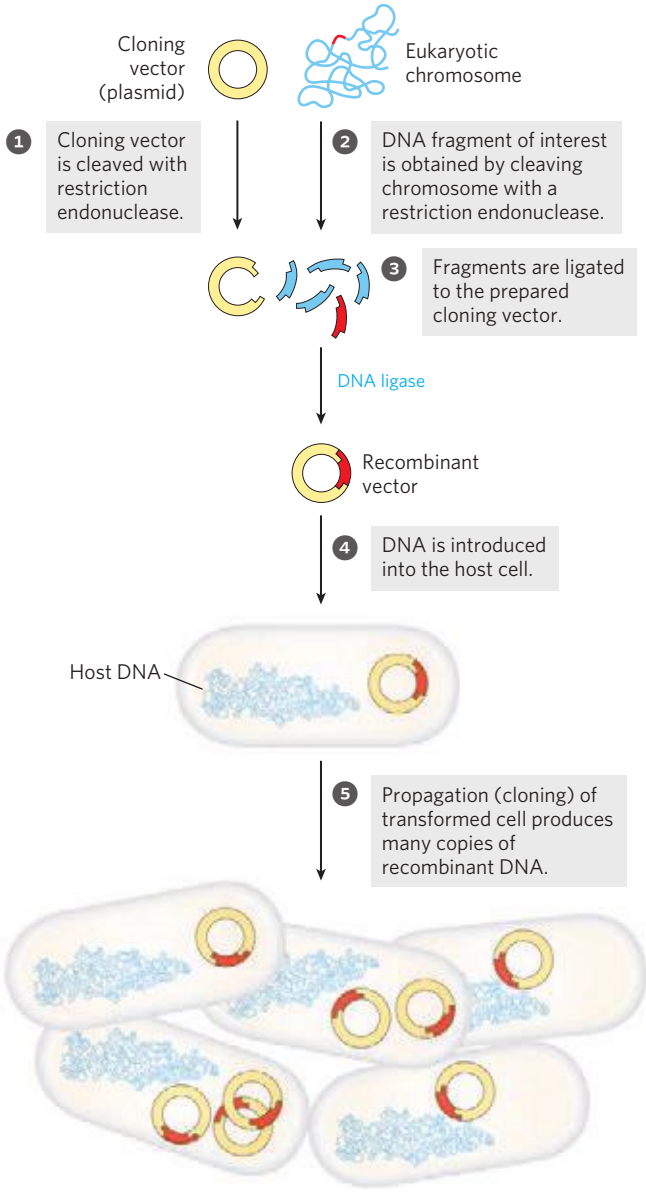
Particularly important to recombinant DNA technology is a set of enzymes (Table 9–1) made available through decades of research on nucleic acid metabolism. Two classes of enzymes lie at the heart of the classic approach to generating and propagating a recombinant DNA molecule **(Fig. 9–1)**. First, **restriction endonucleases** (also called restriction enzymes) recognize and cleave DNA at specific sequences (recognition sequences or restriction sites) to generate a set of smaller fragments. Second, the DNA fragment to be cloned is joined to a suitable cloning vector by using **DNA ligases** to link the DNA molecules together. The recombinant vector is then introduced into a host cell, which amplifies the fragment in the course of many generations of cell division.

Restriction endonucleases are found in a wide range of bacterial species. Werner Arber discovered in the early 1960s that their biological function is to recognize and cleave foreign DNA (the DNA of an infecting virus, for example); such DNA is said to be *restricted*. In the host cell's DNA, the sequence that would be recognized by its own restriction endonuclease is protected from digestion by methylation of the DNA, catalyzed by a specific DNA methylase. The restriction endonuclease and the corresponding methylase are sometimes referred to as a **restriction-modification system**.

There are three types of restriction endonucleases, designated I, II, and III. Types I and III are generally large, multisubunit complexes containing both the

TABLE 9–1 Some Enzymes Used in Recombinant DNA Technology

Enzyme(s)	Function
Type II restriction endonucleases	Cleave DNAs at specific base sequences
DNA ligase	Joins two DNA molecules or fragments
DNA polymerase I (*E. coli*)	Fills gaps in duplexes by stepwise addition of nucleotides to 3′ ends
Reverse transcriptase	Makes a DNA copy of an RNA molecule
Polynucleotide kinase	Adds a phosphate to the 5′-OH end of a polynucleotide to label it or permit ligation
Terminal transferase	Adds homopolymer tails to the 3′-OH ends of a linear duplex
Exonuclease III	Removes nucleotide residues from the 3′ ends of a DNA strand
Bacteriophage λ exonuclease	Removes nucleotides from the 5′ ends of a duplex to expose single-stranded 3′ ends
Alkaline phosphatase	Removes terminal phosphates from either the 5′ or 3′ end (or both)



endonuclease and methylase activities. Type I restriction endonucleases cleave DNA at random sites that can be more than 1,000 base pairs (bp) from the recognition sequence. Type III restriction endonucleases cleave the DNA about 25 bp from the recognition sequence. Both types move along the DNA in a reaction that requires the energy of ATP. **Type II restriction endonucleases**, first isolated by Hamilton Smith in 1970, are simpler, require no ATP, and catalyze the hydrolytic cleavage of particular phosphodiester bonds in the DNA within the recognition sequence itself. The extraordinary utility of this group of restriction endonucleases was demonstrated by Daniel Nathans, who first used them to develop novel methods for mapping and analyzing genes and genomes.

Thousands of type II restriction endonucleases have been discovered in different bacterial species, and more than 100 different DNA sequences are recognized by one or more of these enzymes. The recognition sequences are usually 4 to 6 bp long and palindromic (see Fig. 8–18). Table 9–2 lists sequences recognized by a few type II restriction endonucleases.

Some restriction endonucleases make staggered cuts on the two DNA strands, leaving two to four nucleotides of one strand unpaired at each resulting end.

FIGURE 9–1 Schematic illustration of DNA cloning. A cloning vector and eukaryotic chromosomes are separately cleaved with the same restriction endonuclease. (A single chromosome is shown here for simplicity.) The fragments to be cloned are then ligated to the cloning vector. The resulting recombinant DNA (only one recombinant vector is shown here) is introduced into a host cell, where it can be propagated (cloned). Note that this drawing is not to scale: the size of the *E. coli* chromosome relative to that of a typical cloning vector (such as a plasmid) is much greater than depicted here.

TABLE 9–2 Recognition Sequences for Some Type II Restriction Endonucleases

*Bam*HI	(5′) G G A T C C (3′) C C T A G G	*Hin*dIII	(5′) A A G C T T (3′) T T C G A A
*Cla*I	(5′) A T C G A T (3′) T A G C T A	*Not*I	(5′) G C G G C C G C (3′) C G C C G G C G
*Eco*RI	(5′) G A A T T C (3′) C T T A A G	*Pst*I	(5′) C T G C A G (3′) G A C G T C
*Eco*RV	(5′) G A T A T C (3′) C T A T A G	*Pvu*II	(5′) C A G C T G (3′) G T C G A C
*Hae*III	(5′) G G C C (3′) C C G G	*Tth*111I	(5′) G A C N N N G T C (3′) C T G N N N C A G

Note: Arrows indicate the phosphodiester bonds cleaved by each restriction endonuclease. Asterisks indicate bases that are methylated by the corresponding methylase (where known). N denotes any base. Note that the name of each enzyme consists of a three-letter abbreviation (in italics) of the bacterial species from which it is derived, sometimes followed by a strain designation and Roman numerals to distinguish different restriction endonucleases isolated from the same bacterial species. Thus *Bam*HI is the first (I) restriction endonuclease characterized from *B*acillus *amy*loliquefaciens, strain H.

These unpaired strands are referred to as **sticky ends (Fig. 9–2a)** because they can base-pair with each other or with complementary sticky ends of other DNA fragments. Other restriction endonucleases cleave both strands of DNA at the opposing phosphodiester bonds, leaving no unpaired bases on the ends, often called **blunt ends** (Fig. 9–2b).

The average size of the DNA fragments produced by cleaving genomic DNA with a restriction endonuclease depends on the frequency with which a particular restriction site occurs in the DNA molecule; this in turn depends largely on the size of the recognition sequence.

In a DNA molecule with a random sequence in which all four nucleotides were equally abundant, a 6 bp sequence recognized by a restriction endonuclease such as *Bam*HI would occur on average once every 4^6 (4,096) bp. Enzymes that recognize a 4 bp sequence would produce smaller DNA fragments from a random-sequence DNA molecule; a recognition sequence of this size would be expected to occur about once every 4^4 (256) bp. In natural DNA molecules, particular recognition sequences tend to occur less frequently than this because nucleotide sequences in DNA are not random and the four nucleotides are not equally abundant. In laboratory experiments, the average size of the fragments produced by restriction endonuclease cleavage of a large DNA can be increased by simply terminating the reaction before completion; the result is called a partial digest. Average fragment size can also be increased by using a special class of endonucleases called homing endonucleases (see Fig. 26–37). These recognize and cleave much longer DNA sequences (14 to 20 bp).

Once a DNA molecule has been cleaved into fragments, a particular fragment of known size can be partially purified by agarose or acrylamide gel electrophoresis (p. 302) or by HPLC (p. 92). For a typical mammalian genome, however, cleavage by a restriction endonuclease usually yields too many different DNA fragments to permit convenient isolation of a particular fragment. A common intermediate step in the cloning of a specific gene or DNA segment is the construction of a DNA library (described in Section 9.2).

After the target DNA fragment is isolated, DNA ligase can be used to join it to a similarly digested cloning vector—that is, a vector digested by the *same* restriction endonuclease; a fragment generated by *Eco*RI, for example, generally will not link to a fragment generated by *Bam*HI. As described in more detail in Chapter 25 (see Fig. 25–16), DNA ligase catalyzes the formation of new phosphodiester bonds in a reaction that uses ATP or a similar cofactor. The base pairing of complementary sticky ends greatly facilitates the ligation reaction (Fig. 9–2a). Blunt ends can also be ligated, albeit less efficiently. Researchers can create new DNA sequences by inserting synthetic DNA fragments (called **linkers**) between the ends that are being ligated.

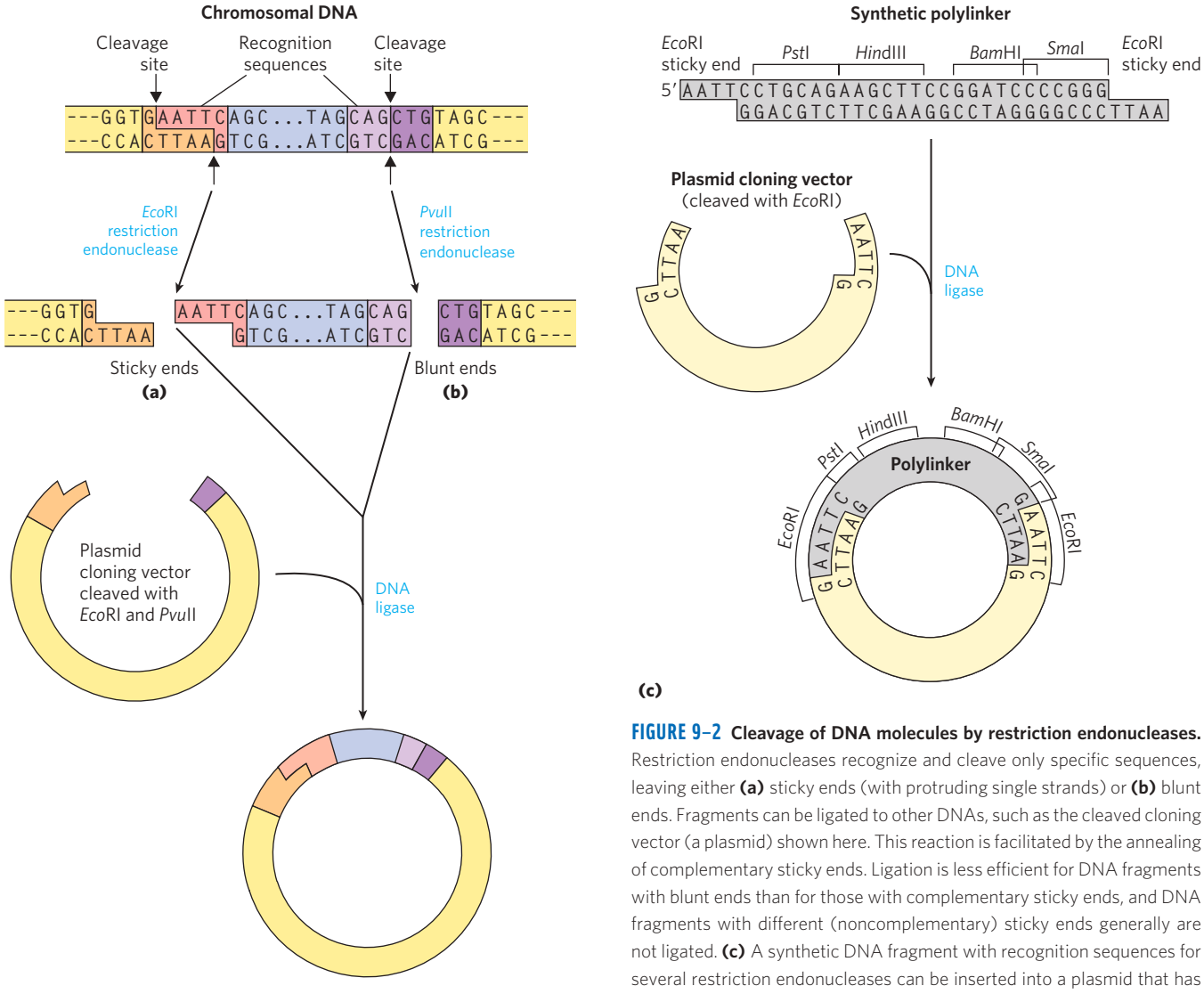


FIGURE 9–2 Cleavage of DNA molecules by restriction endonucleases. Restriction endonucleases recognize and cleave only specific sequences, leaving either **(a)** sticky ends (with protruding single strands) or **(b)** blunt ends. Fragments can be ligated to other DNAs, such as the cleaved cloning vector (a plasmid) shown here. This reaction is facilitated by the annealing of complementary sticky ends. Ligation is less efficient for DNA fragments with blunt ends than for those with complementary sticky ends, and DNA fragments with different (noncomplementary) sticky ends generally are not ligated. **(c)** A synthetic DNA fragment with recognition sequences for several restriction endonucleases can be inserted into a plasmid that has been cleaved by a restriction endonuclease. The insert is called a linker; an insert with multiple restriction sites is called a polylinker. **Restriction Endonucleases**

Inserted DNA fragments with multiple recognition sequences for restriction endonucleases (often useful later as points for inserting additional DNA by cleavage and ligation) are called **polylinkers** (Fig. 9–2c).

The effectiveness of sticky ends in selectively joining two DNA fragments was apparent in the earliest recombinant DNA experiments. Before restriction endonucleases were widely available, some workers found they could generate sticky ends by the combined action of the bacteriophage λ exonuclease and terminal transferase (Table 9–1). The fragments to be joined were given complementary homopolymeric tails. Peter Lobban and Dale Kaiser used this method in 1971 in the first experiments to join naturally occurring DNA fragments. Similar methods were used soon after in the laboratory of Paul Berg to join DNA segments from simian virus 40 (SV40) to DNA derived from bacteriophage λ, thereby creating the first recombinant DNA molecule with DNA segments from different species.

Cloning Vectors Allow Amplification of Inserted DNA Segments

The principles that govern the delivery of recombinant DNA in clonable form to a host cell, and its subsequent amplification in the host, are well illustrated by considering three popular cloning vectors—plasmids and bacterial artificial chromosomes, used in experiments with *E. coli*, and a vector used to clone large DNA segments in yeast.

Plasmids A plasmid is a circular DNA molecule that replicates separately from the host chromosome. The wide variety of naturally occurring bacterial plasmids range in size from 5,000 to 400,000 bp. Many of the plasmids found in bacterial populations are little more than molecular parasites, similar to viruses but with a more limited capacity to transfer from one cell to another. To survive in the host cell, plasmids incorporate several specialized sequences that enable them to make use of

the cell's resources for their own replication and gene expression.

Naturally occurring plasmids usually have a symbiotic role in the cell. They may provide genes that confer resistance to antibiotics or that perform new functions for the cell. For example, the Ti plasmid of *Agrobacterium tumefaciens* allows the host bacterium to colonize the cells of plants and make use of the plant's resources. The same properties that enable plasmids to grow and survive in a bacterial or eukaryotic host are useful to molecular biologists who want to engineer a vector for cloning a specific DNA segment. The classic *E. coli* plasmid pBR322, constructed in 1977, is a good example of a plasmid with features useful in almost all cloning vectors **(Fig. 9–3)**:

1. The plasmid pBR322 has an **origin of replication**, or **ori**, a sequence where replication is initiated by cellular enzymes (see Chapter 25). This sequence is required to propagate the plasmid. An associated regulatory system is present that limits replication to maintain pBR322 at a level of 10 to 20 copies per cell.
2. The plasmid contains genes that confer resistance to the antibiotics tetracycline (Tet^R) and ampicillin (Amp^R), allowing the selection of cells that contain the intact plasmid or a recombinant version of the plasmid (discussed below).
3. Several unique recognition sequences in pBR322 are targets for restriction endonucleases (*Pst*I, *Eco*RI, *Bam*HI, *Sal*I, and *Pvu*II), providing sites where the plasmid can be cut to insert foreign DNA.

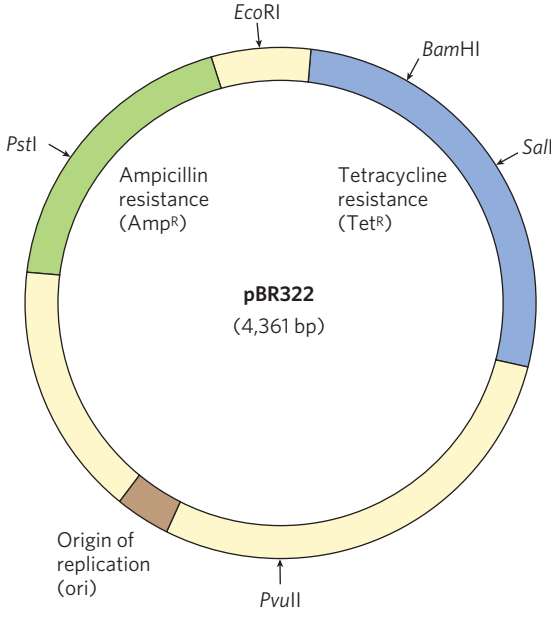


FIGURE 9–3 The constructed *E. coli* plasmid pBR322. Note the location of some important restriction sites—for *Pst*I, *Eco*RI, *Bam*HI, *Sal*I, and *Pvu*II; ampicillin- and tetracycline-resistance genes; and the replication origin (ori). Constructed in 1977, this was one of the early plasmids designed expressly for cloning in *E. coli*.

4. The small size of the plasmid (4,361 bp) facilitates its entry into cells and the biochemical manipulation of the DNA. This small size was generated simply by trimming away many DNA segments from a larger, parent plasmid—sequences that the molecular biologist does not need.

The replication origins inserted in common plasmid vectors were originally derived from naturally occurring plasmids. As in pBR322, each of these origins is regulated to maintain a particular plasmid copy number. Depending on the origin used, the plasmid copy number can vary from one to hundreds or thousands per cell, providing many options for investigators. Two different plasmids cannot function in the same cell if they use the same origin of replication, because the regulation of one will interfere with the replication of the other. Such plasmids are said to be incompatible. When a researcher wants to introduce two or more different plasmids into a bacterial cell, each plasmid must have a different replication origin.

In the laboratory, small plasmids can be introduced into bacterial cells by a process called **transformation**. The cells (often *E. coli*, but other bacterial species are also used) and plasmid DNA are incubated together at 0°C in a calcium chloride solution, then subjected to heat shock by rapidly shifting the temperature to between 37°C and 43°C. For reasons not well understood, some of the cells treated in this way take up the plasmid DNA. Some species of bacteria, such as *Acinetobacter baylyi*, are naturally competent for DNA uptake and do not require the calcium chloride–heat shock treatment. In an alternative method, cells incubated with the plasmid DNA are subjected to a high-voltage pulse. This approach, called **electroporation**, transiently renders the bacterial membrane permeable to large molecules.

Regardless of the approach, relatively few cells take up the plasmid DNA, so a method is needed to identify those that do. The usual strategy is to utilize one of two types of genes in the plasmid, referred to as selectable and screenable markers. **Selectable markers** either permit the growth of a cell (positive selection) or kill the cell (negative selection) under a defined set of conditions. The plasmid pBR322 provides examples of both positive and negative selection **(Fig. 9–4)**. A **screenable marker** is a gene encoding a protein that causes the cell to produce a colored or fluorescent molecule. Cells are not harmed when the gene is present, and the cells that carry the plasmid are easily identified by the colored or fluorescent colonies they produce.

Transformation of typical bacterial cells with purified DNA (never a very efficient process) becomes less successful as plasmid size increases, and it is difficult to clone DNA segments longer than about 15,000 bp when plasmids are used as the vector.

To illustrate the use of a plasmid as a cloning vector, consider a typical bacterial gene, that encoding a

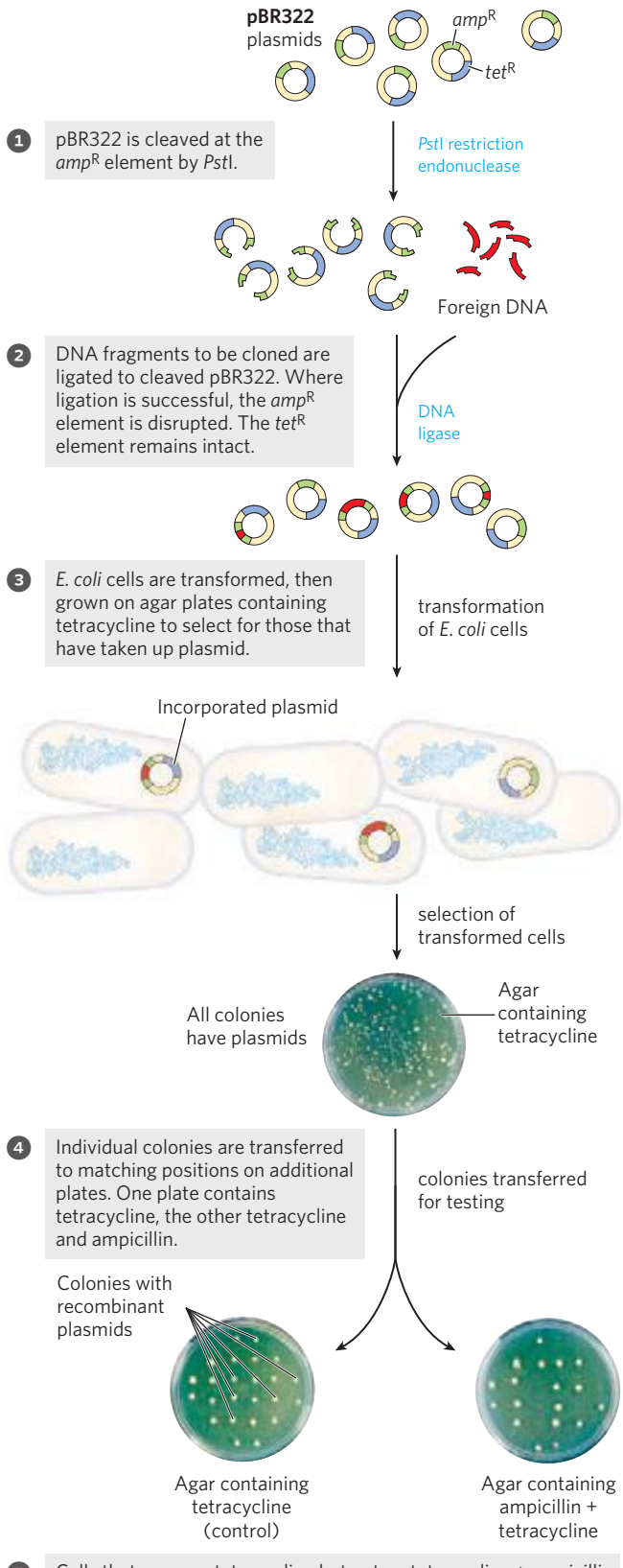


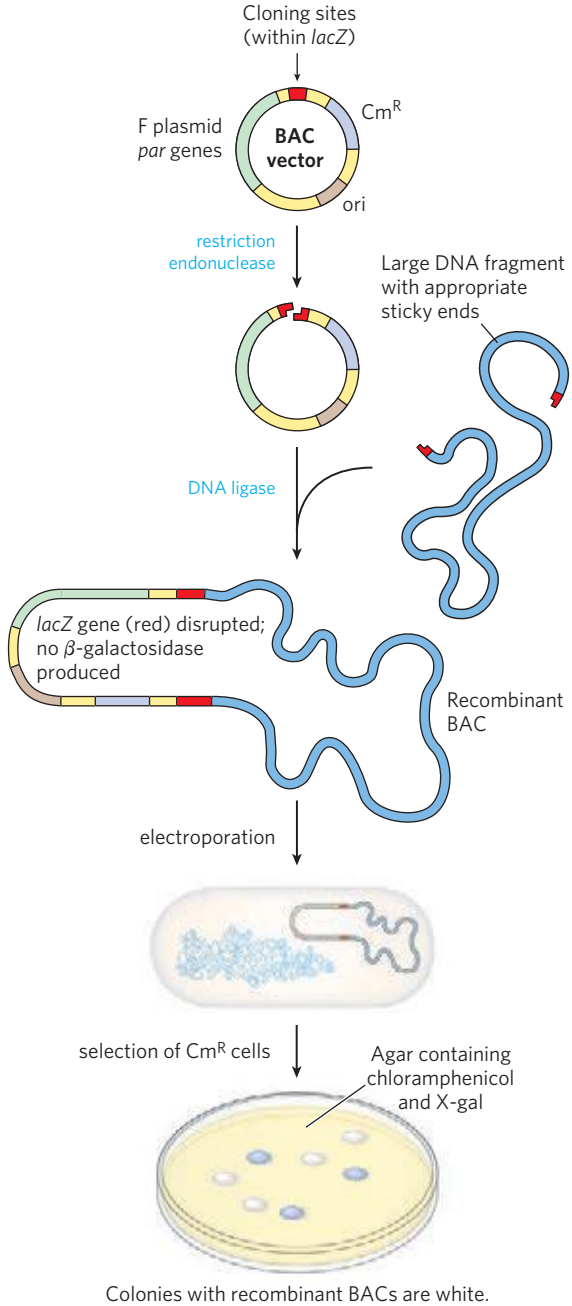
FIGURE 9–4 **Use of pBR322 to clone foreign DNA in *E. coli* and identify cells containing it.** Plasmid Cloning

recombinase called the RecA protein (see Chapter 25). In most bacteria, the gene encoding RecA is one of thousands of other genes on a chromosome millions of base pairs long. The *recA* gene is just over 1,000 bp long. A plasmid would be a good choice for cloning a gene of this size. As described later, the cloned gene can be altered in a variety of ways, and the gene variants can be expressed at high levels to enable purification of the encoded protein.

Bacterial Artificial Chromosomes Large-genome sequencing projects often require the cloning of much longer DNA segments than can typically be incorporated into standard plasmid cloning vectors such as pBR322. To meet this need, plasmid vectors have been developed with special features that allow the cloning of very long segments (typically 100,000 to 300,000 bp) of DNA. Once such large segments of cloned DNA have been added, these vectors are large enough to be thought of as chromosomes, and are known as **bacterial artificial chromosomes**, or **BACs (Fig. 9–5)**.

A BAC vector (without any cloned DNA inserted) is a relatively simple plasmid, generally not much larger than other plasmid vectors. To accommodate very long segments of cloned DNA, BAC vectors have stable origins of replication that maintain the plasmid at one or two copies per cell. The low copy number is useful in cloning large segments of DNA, because it limits the opportunities for unwanted recombination reactions that can unpredictably alter large cloned DNAs over time. BACs also include *par* genes, which encode proteins that direct the reliable distribution of the recombinant chromosomes to daughter cells at cell division, thereby increasing the likelihood of each daughter cell carrying one copy, even when few copies are present. The BAC vector includes both selectable and screenable markers. The BAC vector shown in Figure 9–5 contains a gene that confers resistance to the antibiotic chloramphenicol (Cm^R). Positive selection for vector-containing cells occurs on agar plates containing this antibiotic. A *lacZ* gene, required for production of the enzyme β-galactosidase, is a screenable marker that can reveal which cells contain plasmids—now chromosomes—that incorporate the cloned DNA segments. The β-galactosidase catalyzes the conversion of the colorless molecule 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) to a blue product. If the gene is intact and expressed, the colony containing it will be blue. If gene expression is disrupted by the introduction of a cloned DNA segment, the colony will be white.

Yeast Artificial Chromosomes As with *E. coli*, yeast genetics is a well-developed discipline. The genome of *Saccharomyces cerevisiae* contains only 14×10^6 bp (less than four times the size of the *E. coli* chromosome), and its entire sequence is known. Yeast is also very easy to maintain and grow on a large scale in the laboratory. Plasmid vectors have been constructed for yeast,



Colonies with recombinant BACs are white.

FIGURE 9–5 Bacterial artificial chromosomes (BACs) as cloning vectors. The vector is a relatively simple plasmid, with a replication origin (ori) that directs replication. The *par* genes, derived from a type of plasmid called an F plasmid, assist in the even distribution of plasmids to daughter cells at cell division. This increases the likelihood of each daughter cell carrying one copy of the plasmid, even when few copies are present. The low number of copies is useful in cloning large segments of DNA because it limits the opportunities for unwanted recombination reactions that can unpredictably alter large cloned DNAs over time. The BAC includes selectable markers. A *lacZ* gene (required for the production of the enzyme β-galactosidase) is situated in the cloning region such that it is inactivated by cloned DNA inserts. Introduction of recombinant BACs into cells by electroporation is promoted by the use of cells with an altered (more porous) cell wall. Recombinant DNAs are screened for resistance to the antibiotic chloramphenicol (Cm^R). Plates also contain X-gal, a substrate for β-galactosidase that yields a blue product. Colonies with active β-galactosidase and hence no DNA insert in the BAC vector turn blue; colonies without β-galactosidase activity—and thus with the desired DNA inserts—are white.

employing the same principles that govern the use of *E. coli* vectors. Convenient methods for moving DNA into and out of yeast cells permit the study of many aspects of eukaryotic cell biochemistry. Some recombinant plasmids incorporate multiple replication origins and other elements that allow them to be used in more than one species (e.g., in yeast and in *E. coli*). Such plasmids that can be propagated in cells of two or more species are called **shuttle vectors**.

Research on large genomes and the associated need for high-capacity cloning vectors led to the development of **yeast artificial chromosomes**, or **YACs (Fig. 9–6)**. YAC vectors contain all the elements needed to maintain a eukaryotic chromosome in the yeast nucleus: a yeast origin of replication, two selectable markers, and specialized sequences (derived from the telomeres and centromere) needed for stability and proper segregation of the chromosomes at cell division (see Chapter 24). In preparation for its use in cloning, the vector is propagated as a circular bacterial plasmid and then isolated and purified. Cleavage with a restriction endonuclease (*Bam*HI in Fig. 9–6) removes a length of DNA between two telomere sequences (TEL), leaving the telomeres at the ends of the linearized DNA. Cleavage at another internal site (by *Eco*RI in Fig. 9–6) divides the vector into two DNA segments, referred to as vector arms, each with a different selectable marker.

The genomic DNA to be cloned is prepared by partial digestion with restriction endonucleases to obtain a suitable fragment size. Genomic fragments are then separated by **pulsed field gel electrophoresis**, a variation of gel electrophoresis (see Fig. 3–18) that segregates very large DNA segments. DNA fragments of appropriate size (up to about 2×10^6 bp) are mixed with the prepared vector arms and ligated. The ligation mixture is then used to transform yeast cells (pretreated to partially degrade their cell walls) with these very large DNA molecules—which now have the structure and size to be considered yeast chromosomes. Culture on a medium that requires the presence of both selectable marker genes ensures the growth of only those yeast cells that contain an artificial chromosome with a large insert sandwiched between the two vector arms (Fig. 9–6). The stability of YAC clones increases with the length of the cloned DNA segment (up to a point). Those with inserts of more than 150,000 bp are nearly as stable as normal cellular chromosomes, whereas those with inserts less than 100,000 bp long are gradually lost during mitosis (so, generally, there are no yeast cell clones carrying only the two vector ends ligated together or vectors with only short inserts). YACs that lack a telomere at either end are rapidly degraded.

As with BACs, YAC vectors can be used to clone very long segments of DNA. In addition, the DNA cloned

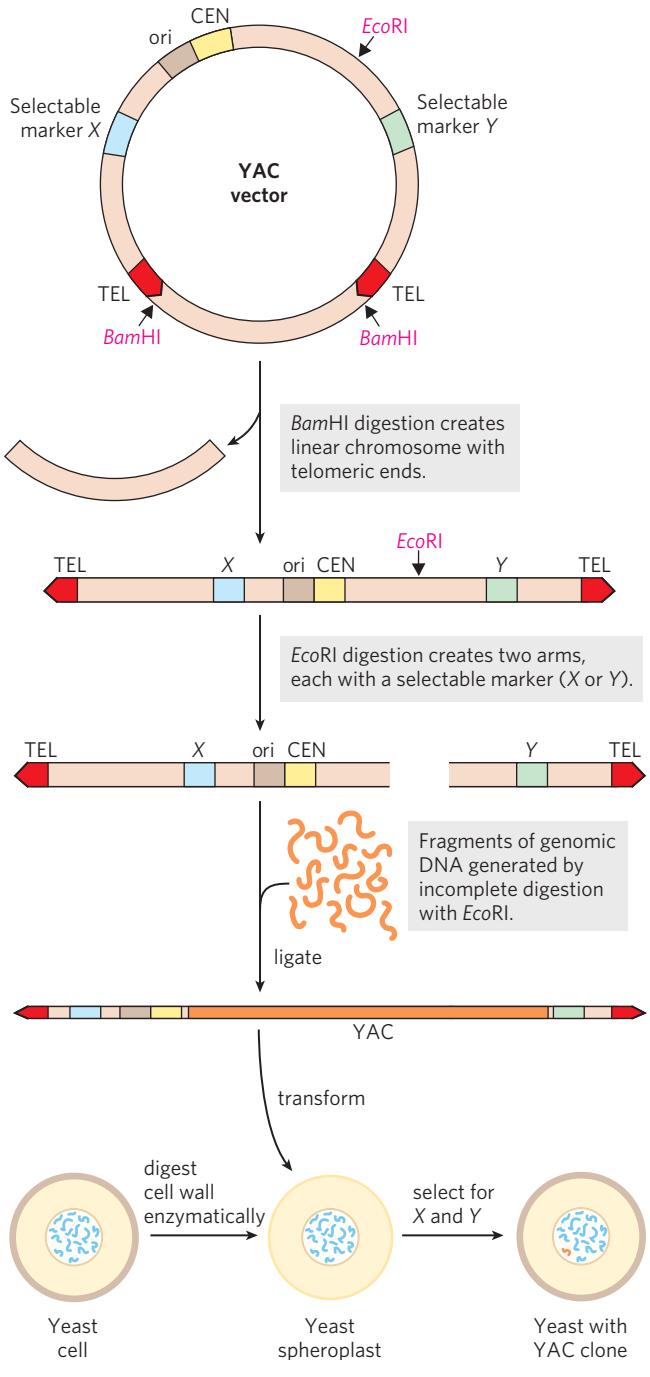


FIGURE 9–6 Construction of a yeast artificial chromosome (YAC). A YAC vector includes an origin of replication (ori), a centromere (CEN), two telomeres (TEL), and selectable markers (*X* and *Y*). Digestion with *Bam*HI and *Eco*RI generates two separate DNA arms, each with a telomeric end and one selectable marker. A large segment of DNA (e.g., up to 2×10^6 bp from the human genome) is ligated to the two arms to create a yeast artificial chromosome. The YAC transforms yeast cells (prepared by removal of the cell wall to form spheroplasts), and the cells are selected for *X* and *Y*; the surviving cells propagate the DNA insert.

in a YAC can be altered to study the function of specialized sequences in chromosome metabolism, mechanisms of gene regulation and expression, and many other problems in eukaryotic molecular biology.

Cloned Genes Can Be Expressed to Amplify Protein Production

Frequently, the product of a cloned gene, rather than the gene itself, is of primary interest—particularly when the protein has commercial, therapeutic, or research value. Biochemists use purified proteins for many purposes, including to elucidate protein function, study reaction mechanisms, generate antibodies to the proteins, reconstitute complex cellular activities in the test tube with purified components, and examine protein binding partners. With an increased understanding of the fundamentals of DNA, RNA, and protein metabolism and their regulation in a host organism such as *E. coli* or yeast, investigators can manipulate cells to express cloned genes in order to study their protein products. The general goal is to alter the sequences around a cloned gene to trick the host organism into producing the protein product of the gene, often at very high levels. This overexpression of a protein can make its subsequent purification much easier.

We'll use the expression of a eukaryotic protein in a bacterium as an example. Eukaryotic genes have surrounding sequences needed for their transcription and regulation in the cells they are derived from, but these sequences do not function in bacteria. Thus, eukaryotic genes lack the DNA sequence elements required for their controlled expression in bacterial cells—promoters (sequences that instruct RNA polymerase where to bind to initiate mRNA synthesis), ribosome-binding sites (sequences that allow translation of the mRNA to protein), and additional regulatory sequences. Therefore, appropriate bacterial regulatory sequences for transcription and translation must be inserted in the vector DNA at the correct positions relative to the eukaryotic gene. In some cases, cloned genes are so efficiently expressed that their protein product represents 10% or more of the cellular protein. At these concentrations, some foreign proteins can kill the host cell (usually *E. coli*), so expression of the cloned gene must be limited to the few hours before the planned harvesting of the cells.

Cloning vectors with the transcription and translation signals needed for the regulated expression of a cloned gene are called **expression vectors**. The rate of expression of the cloned gene is controlled by replacing the gene's normal promoter and regulatory sequences with more efficient and convenient versions supplied by the vector. Generally, a well-characterized promoter and its regulatory elements are positioned near several unique restriction sites for cloning, so that genes inserted at the restriction sites will be expressed from the regulated promoter elements **(Fig. 9–7)**. Some of these vectors incorporate other features, such as a bacterial ribosome-binding site to enhance translation of the mRNA derived from the gene (Chapter 27) or a transcription termination sequence (Chapter 26).

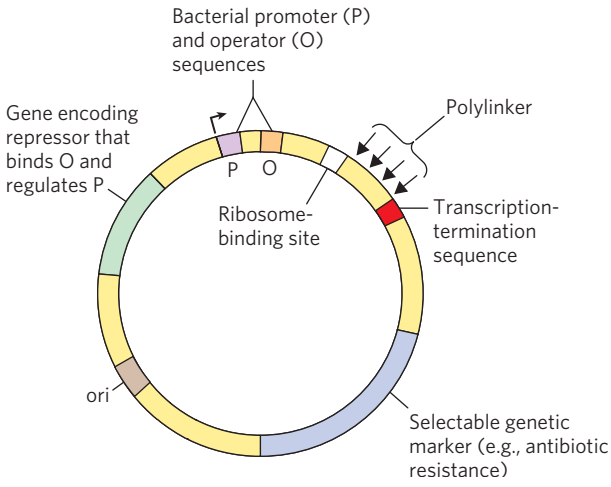


FIGURE 9–7 DNA sequences in a typical *E. coli* expression vector. The gene to be expressed is inserted into one of the restriction sites in the polylinker, near the promoter (P), with the end of the gene encoding the amino terminus of the protein positioned closest to the promoter. The promoter allows efficient transcription of the inserted gene, and the transcription-termination sequence sometimes improves the amount and stability of the mRNA produced. The operator (O) permits regulation by a repressor that binds to it. The ribosome-binding site provides sequence signals for the efficient translation of the mRNA derived from the gene. The selectable marker allows the selection of cells containing the recombinant DNA.

Many Different Systems Are Used to Express Recombinant Proteins

Every living organism has the capacity to express genes in its genomic DNA; thus, in principle, any organism can serve as a host to express proteins from a different (heterologous) species. Almost every sort of organism has, indeed, been used for this purpose, and each host type has a particular set of advantages and disadvantages.

Bacteria Bacteria, especially *E. coli*, remain the most common hosts for protein expression. The regulatory sequences that govern gene expression in *E. coli* and many other bacteria are well understood and can be harnessed to express cloned proteins at high levels. Bacteria are easy to store and grow in the laboratory, on inexpensive growth media. Efficient methods also exist to get DNA into bacteria and extract DNA from them. Bacteria can be grown in huge amounts in commercial fermenters, providing a rich source of the cloned protein. Problems do exist, however. When expressed in bacteria, some heterologous proteins do not fold correctly, and many do not undergo the covalent modifications or proteolytic cleavage that may be necessary for their activity. Certain features of a gene sequence also can make a particular gene difficult to express in bacteria. For example, intrinsically disordered regions are more common in eukaryotic proteins. Expressed in bacteria, many eukaryotic proteins aggregate into insoluble cellular precipitates called inclusion bodies. For these and many other reasons, some eukaryotic proteins are inactive when purified from bacteria or cannot be expressed at all. To help address some of these problems, new bacterial host strains are regularly being developed that include enhancements such as the engineered presence of eukaryotic protein chaperones or enzymes that modify eukaryotic proteins.

There are many specialized systems for expressing proteins in bacteria. The promoter and regulatory sequences associated with the lactose operon (see Chapter 28) are often fused to the gene of interest to direct transcription. The cloned gene will be transcribed when lactose is added to the growth medium. However, regulation in the lactose system is "leaky"; it is not turned off completely when lactose is absent—a potential problem if the product of the cloned gene is toxic to the host cells. Transcription from the Lac promoter is also not efficient enough for some applications.

An alternative system uses a promoter and RNA polymerase found in a bacterial virus called bacteriophage T7. If the cloned gene is fused to a T7 promoter, it is not transcribed by the *E. coli* RNA polymerase, but instead by the T7 RNA polymerase. The gene encoding this polymerase is separately cloned into the same cell in a construct that affords tight regulation (allowing controlled production of the T7 RNA polymerase). The polymerase is also very efficient and directs high levels of expression of most genes fused to the T7 promoter. This system has been used to express the RecA protein in bacterial cells **(Fig. 9–8)**.

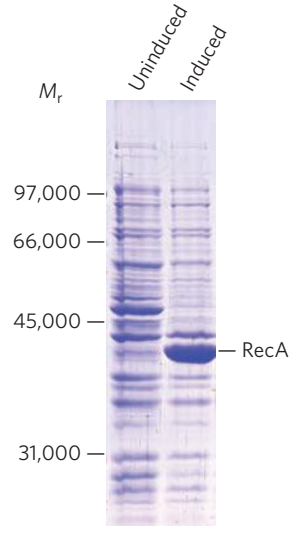


FIGURE 9–8 Regulated expression of RecA protein in a bacterial cell. The gene encoding the RecA protein, fused to a bacteriophage T7 promoter, is cloned into an expression vector. Under normal growth conditions (uninduced), no RecA protein appears. When the T7 RNA polymerase is induced in the cell, the *recA* gene is expressed, and large amounts of RecA protein are produced. The positions of standard molecular weight markers run on the same gel are indicated.

Yeast The yeast *Saccharomyces cerevisiae* is probably the best understood eukaryotic organism and one of the easiest to grow and manipulate in the laboratory. Like bacteria, this yeast can be grown on inexpensive media. Yeast have tough cell walls that are difficult to breach in order to introduce DNA vectors, so bacteria are more convenient for doing much of the genetic engineering and vector maintenance. This is why the yeast vector was first propagated in bacteria. Several excellent shuttle vectors exist for this purpose.

The principles underlying the expression of a protein in yeast are the same as those in bacteria. Cloned genes must be linked to promoters that can direct high-level expression in yeast. For example, the yeast *GAL1* and *GAL10* genes are under cellular regulation such that they are expressed when yeast cells are grown in media with galactose but shut down when the cells are grown in glucose. Thus, if a heterologous gene is expressed using the same regulatory sequences, the expression of that gene can be controlled simply by choosing an appropriate medium for cell growth.

Some of the same problems that accompany protein expression in bacteria also occur with yeast. Heterologous proteins may not fold properly, yeast may lack the enzymes needed to modify the proteins to their active forms, or the expression of proteins may be made difficult by certain features of the gene sequence. However, because *S. cerevisiae* is a eukaryote, the expression of eukaryotic genes (especially yeast genes) is sometimes more efficient in this host than in bacteria. Folding and modification of the products may also be more accurate than for proteins expressed in bacteria.

Insects and Insect Viruses Baculoviruses are insect viruses with double-stranded DNA genomes. When they infect their insect larval hosts, they act as parasites, killing the larvae and turning them into factories for virus production. Late in the infection process, the viruses produce large amounts of two proteins (p10 and polyhedrin), neither of which is needed for virus production in cultured insect cells. The genes for both of these proteins can be replaced with the gene of a heterologous protein. When the resulting recombinant virus is used to infect insect cells or larvae, the heterologous protein is often produced at very high levels—up to 25% of the total protein present at the end of the infection cycle.

Autographa californica multicapsid nucleopolyhedrovirus (AcMNPV) is the **baculovirus** most often used for protein expression. It has a large genome (134,000 bp), too large for direct cloning. Virus purification is also cumbersome. These problems have been solved by the creation of **bacmids**, large circular DNAs that include the entire baculovirus genome along with sequences that allow replication of the bacmid in *E. coli* **(Fig. 9–9)**. The gene of interest is cloned into a smaller plasmid and combined with the larger plasmid by site-specific recombination in vivo (see Fig. 25–37). The recombinant bacmid is then isolated and transfected into insect cells (the term **transfection** is used when the DNA used for transformation includes viral sequences and leads to viral replication), followed by recovery of the protein once the infection cycle is finished. A wide range of bacmid systems is available commercially. Baculovirus systems are not successful with all proteins. However, with these systems, insect cells sometimes successfully replicate the protein-modification patterns of higher eukaryotes and produce active, correctly modified eukaryotic proteins.

Mammalian Cells in Culture The most convenient way to introduce cloned genes into a mammalian cell is with viruses. This method takes advantage of the natural capacity of a virus to insert its DNA or RNA into a cell, and sometimes into the cellular chromosome. A variety of engineered mammalian viruses are available as vectors, including human adenoviruses and retroviruses. The gene of interest is cloned so that its expression is controlled by a virus promoter. The virus uses its natural infection mechanisms to introduce the recombinant genome into cells, where the cloned protein is expressed. These systems have the advantage that proteins can be expressed either transiently (if the viral DNA is maintained separately from the host cell genome and eventually degraded) or permanently (if the viral DNA is integrated into the host cell genome). With the correct choice of host cell, the proper post-translational modification of the protein to its active form can be ensured. However, the growth of mammalian cells in tissue culture is very expensive, and this technology is generally used to test the function of a protein in vivo rather than to produce a protein in large amounts.

Alteration of Cloned Genes Produces Altered Proteins

Cloning techniques can be used not only to overproduce proteins but to produce protein products altered, subtly or dramatically, from their native forms. Specific amino acids may be replaced individually by **site-directed mutagenesis**. This technique has greatly enhanced research on proteins by allowing investigators to make specific changes in the primary structure and examine the effects of these changes on the protein's folding, three-dimensional structure, and activity. This powerful approach to studying protein structure and function changes the amino acid sequence by altering the DNA sequence of the cloned gene. If appropriate restriction sites flank the sequence to be altered, researchers can simply remove a DNA segment and replace it with a

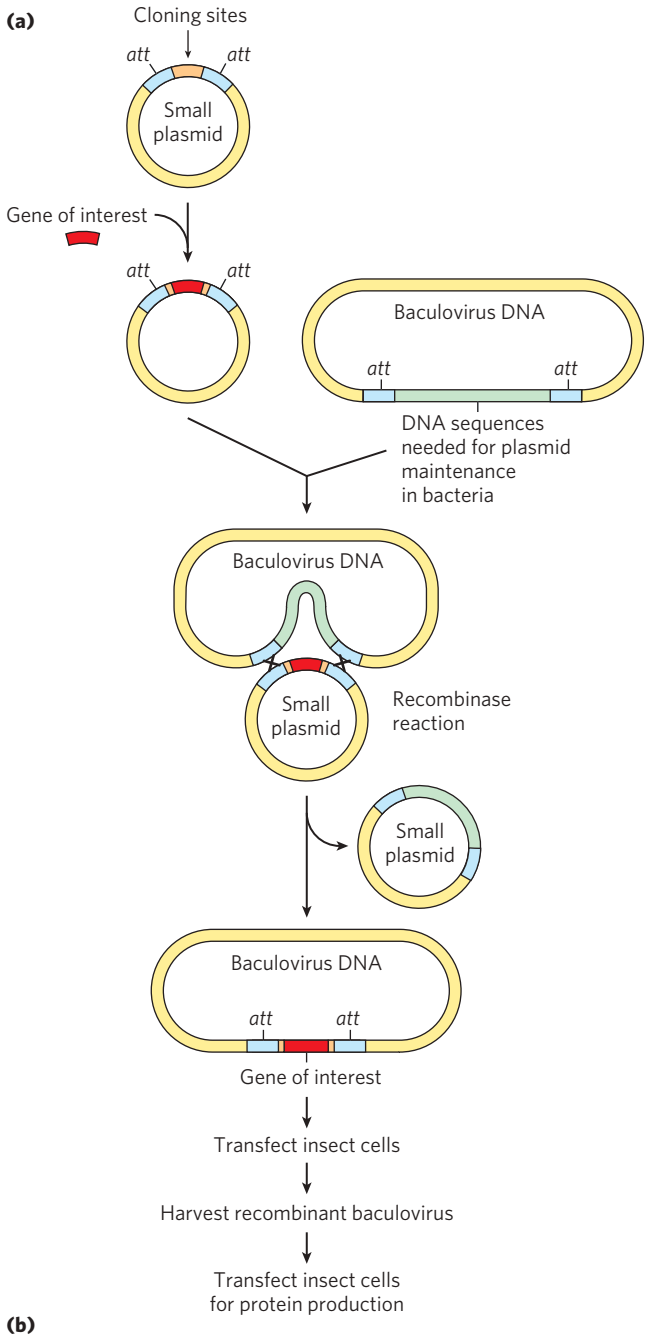


FIGURE 9–9 Cloning with baculoviruses. (a) Shown here is the construction of a typical vector used for protein expression in baculoviruses. The gene of interest is cloned into a small plasmid (left) between two sites (*att*) recognized by a site-specific recombinase, then introduced into the baculovirus vector by site-specific recombination (see Fig. 25-37). This generates a circular DNA product that is used to infect the cells of an insect larva. The gene of interest is expressed during the infection cycle, downstream of a promoter that normally expresses a baculovirus coat protein at very high levels. **(b)** The photographs show (left) an insect larva infected with a recombinant baculovirus vector expressing a protein that produces a red color and (right) an uninfected larva.

synthetic one, identical to the original except for the desired change **(Fig. 9–10a)**.

When suitably located restriction sites are not present, **oligonucleotide-directed mutagenesis** can create a specific DNA sequence change (Fig. 9–10b). Two short, complementary synthetic DNA strands, each with the desired base change, are annealed to opposite strands of the cloned gene within a suitable circular DNA vector. The mismatch of a single base pair in 30 to 40 bp does not prevent annealing. The two annealed oligonucleotides serve to prime DNA synthesis in both directions around the plasmid vector, creating two complementary strands that contain the mutation. After several cycles of selective amplification, using a technique called polymerase chain reaction (PCR; described on p. 327), the mutation-containing DNA predominates in the population and can be used to transform bacteria. Most of the transformed bacteria will have plasmids carrying the mutation. If necessary, the nonmutant template plasmid DNA can be selectively eliminated by cleavage with the restriction enzyme *Dpn*I. The template plasmid, usually isolated from wild-type *E. coli*, has a methylated A residue in every copy of the four-nucleotide palindrome GATC (called a dam site; see Fig. 25–21). The new DNA containing the mutation does not have methylated A residues because the replication is done in vitro. *Dpn*I selectively cleaves DNA at the sequence GATC only if the A residue in one or both strands is methylated—that is, it breaks down only the template.

For an example, we go back to the bacterial *recA* gene. The product of this gene, the RecA protein, has several activities (see Section 25.3). It binds to and forms a filamentous structure on DNA, aligns two DNAs of similar sequence, and hydrolyzes ATP. A particular amino acid residue in RecA (a 352 residue polypeptide), the Lys residue at position 72, is involved in ATP hydrolysis. By changing Lys^{72} to an Arg, a variant of RecA protein is created that will bind, but not hydrolyze, ATP (Fig. 9–10c). The engineering and purification of this variant RecA protein has facilitated research into the roles of ATP hydrolysis in the functioning of this protein.

Changes can be introduced into a gene that involve far more than one base pair. Large parts of a gene can be deleted by cutting out a segment with restriction

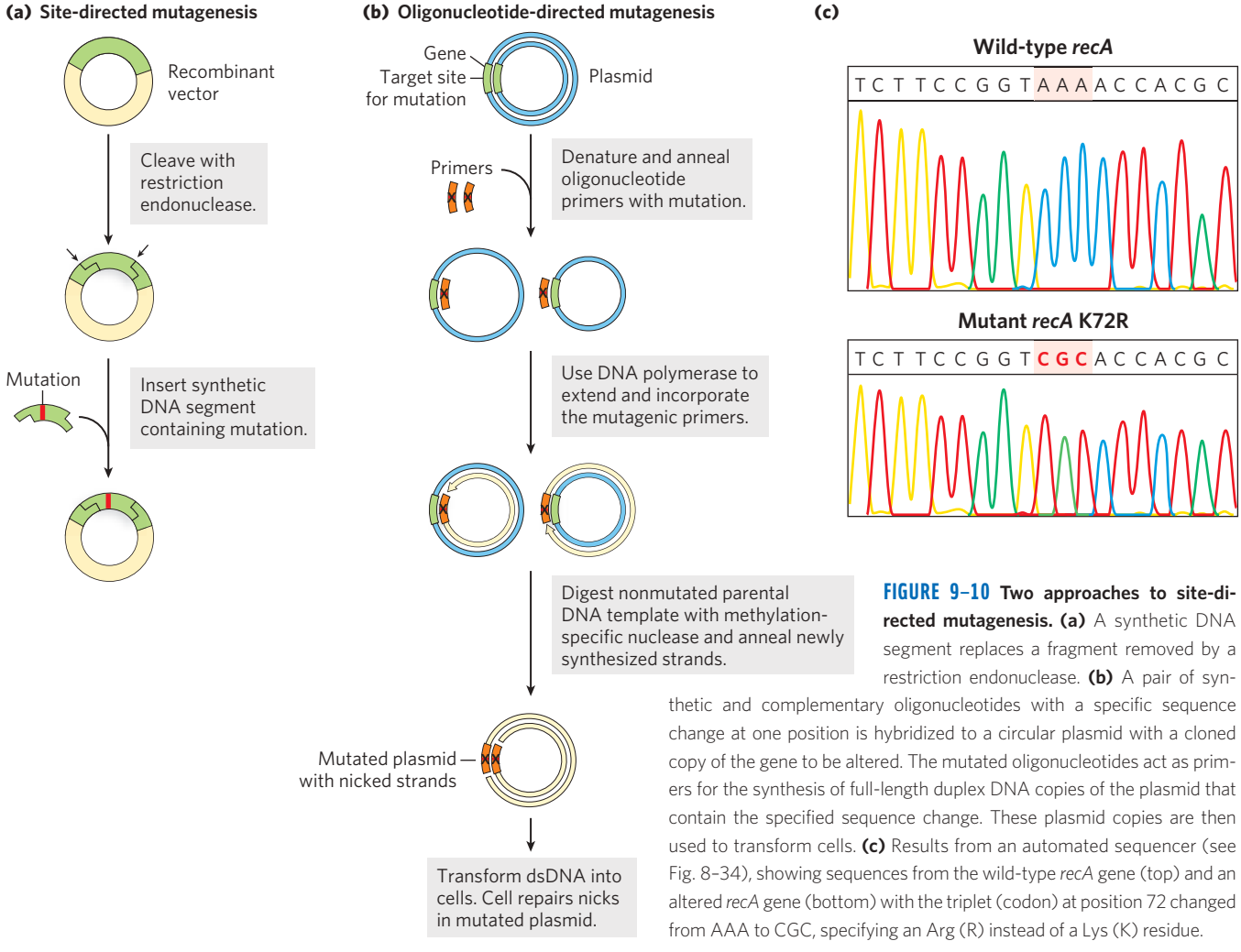


FIGURE 9–10 Two approaches to site-directed mutagenesis. (a) A synthetic DNA segment replaces a fragment removed by a restriction endonuclease. **(b)** A pair of synthetic and complementary oligonucleotides with a specific sequence change at one position is hybridized to a circular plasmid with a cloned copy of the gene to be altered. The mutated oligonucleotides act as primers for the synthesis of full-length duplex DNA copies of the plasmid that contain the specified sequence change. These plasmid copies are then used to transform cells. **(c)** Results from an automated sequencer (see Fig. 8–34), showing sequences from the wild-type *recA* gene (top) and an altered *recA* gene (bottom) with the triplet (codon) at position 72 changed from AAA to CGC, specifying an Arg (R) instead of a Lys (K) residue.

endonucleases and ligating the remaining portions to form a smaller gene. For example, if a protein has two domains, the gene segment encoding one of the domains can be removed to produce a protein that now has only one of the original two. Parts of two different genes can be ligated to create new combinations; the product of such a fused gene is called a **fusion protein**. Researchers have ingenious methods to bring about virtually any genetic alteration in vitro. After reintroducing the altered DNA into the cell, they can investigate the consequences of the alteration.

Terminal Tags Provide Handles for Affinity Purification

Affinity chromatography is one of the most efficient methods for purifying proteins (see Fig. 3–17c). Unfortunately, many proteins do not bind a ligand that can be conveniently immobilized on a column matrix. However, the gene for almost any protein can be altered to express a fusion protein that can be purified by affinity chromatography. The gene encoding the target protein is fused to a gene encoding a peptide or protein that binds a simple, stable ligand with high affinity and specificity. The peptide or protein used for this purpose is referred to as a **tag**. Tag sequences can be added to genes such that the resulting proteins have tags at their amino or carboxyl terminus. Table 9–3 lists some of the peptides or proteins commonly used as tags.

TABLE 9–3 Commonly Used Protein Tags

Tag protein/peptide	Molecular mass (kDa)	Immobilized ligand
Protein A	59	Fc portion of IgG
$(His)_6$	0.8	Ni^{2+}
Glutathione-*S*-transferase (GST)	26	Glutathione
Maltose-binding protein	41	Maltose
β-Galactosidase	116	*p*-Aminophenyl-β-D-thiogalactoside (TPEG)
Chitin-binding domain	5.7	Chitin

The general procedure can be illustrated by focusing on a system that uses the glutathione-*S*-transferase (GST) tag **(Fig. 9–11)**. GST is a small enzyme (M_r 26,000) that binds tightly and specifically to glutathione. When the GST gene sequence is fused to a target gene, the fusion protein acquires the capacity to bind glutathione. The fusion protein is expressed in a host organism such as a bacterium, and a crude extract is prepared. A column is filled with a porous matrix consisting of the ligand (glutathione) immobilized on microscopic beads of a stable polymer such as cross-linked agarose. As the crude extract percolates through this matrix, the fusion protein becomes immobilized by binding the glutathione. The other proteins in the extract are washed through the column and discarded. The interaction between GST and glutathione is tight but noncovalent, allowing the fusion protein to be gently eluted from the column with a solution containing either a higher concentration of salts or free glutathione to compete with the immobilized ligand for GST binding. The fusion protein is often obtained with good yield and high purity. In some commercially available systems, the tag can be entirely or largely removed from the purified fusion protein by a protease that cleaves a sequence near the junction between the target protein and its tag.

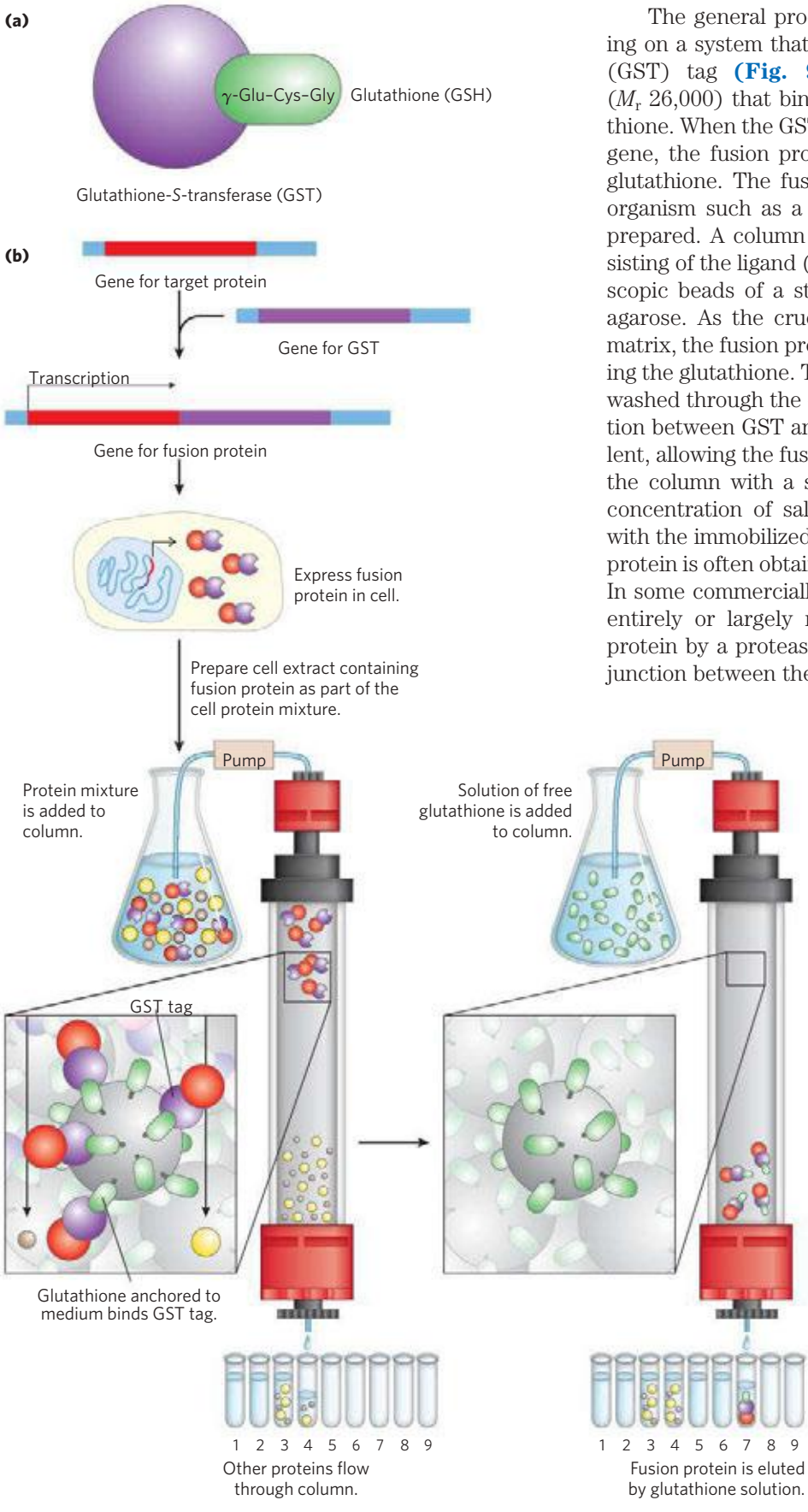


FIGURE 9–11 Use of tagged proteins in protein purification. (a) Glutathione-*S*-transferase (GST) is a small enzyme that binds glutathione (a glutamate residue to which a Cys-Gly dipeptide is attached at the carboxyl carbon of the Glu side chain, hence the abbreviation GSH). **(b)** The GST tag is fused to the carboxyl terminus of the protein by genetic engineering. The tagged protein is expressed in the cell and is present in the crude extract when the cells are lysed. The extract is subjected to affinity chromatography (see Fig. 3–17c) through a matrix with immobilized glutathione. The GST-tagged protein binds to the glutathione, retarding its migration through the column, while the other proteins are washed through rapidly. The tagged protein is subsequently eluted with a solution containing elevated salt concentration or free glutathione.

A shorter tag with widespread application consists of a simple sequence of six or more His residues. These histidine tags, or His tags, bind tightly and specifically to nickel ions. A chromatography matrix with immobilized Ni^{2+} can be used to quickly separate a His-tagged protein from other proteins in an extract. Some of the larger tags, such as maltose-binding protein, provide added stability and solubility, allowing the purification of cloned proteins that are otherwise inactive due to improper folding or insolubility.

Affinity chromatography using terminal tags is powerful and convenient. The tags have been successfully used in thousands of published studies; in many cases, the protein would be impossible to purify and study without the tag. However, even very small tags can affect the properties of the proteins they are attached to, thereby influencing the study results. For example, the tag may adversely affect protein folding. Even if the tag is removed by a protease, one or a few extra amino acid residues can remain behind on the target protein, which may or may not affect the protein's activity. The types of experiments to be carried out, and the results obtained from them, should always be evaluated with the aid of well-designed controls to assess any effect of a tag on protein function.

Gene Sequences Can Be Amplified with the Polymerase Chain Reaction

Genome projects worldwide generate international databases containing the complete genome sequences of hundreds of organisms and provide unprecedented access to gene sequence information. This, in turn, simplifies the process of cloning individual genes for more detailed analysis. If the sequence of at least the end portions of a DNA segment of interest is known, the number of copies of that DNA segment can be hugely amplified with the **polymerase chain reaction (PCR)**, a process conceived by Kary Mullis in 1983. The amplified DNA can then be cloned by the methods described earlier or can be used in a variety of analytical procedures.

The PCR procedure has an elegant simplicity. It relies on DNA polymerases, which synthesize DNA strands from deoxyribonucleotides, using a DNA template (Chapter 25). All DNA polymerases synthesize DNA in the 5′→3′ direction (see Fig. 8–33a). Further, DNA polymerases do not synthesize DNA de novo, but instead must add nucleotides to preexisting strands, referred to as primers. Two synthetic oligonucleotides are prepared, complementary to sequences on opposite strands of the target DNA at positions defining the ends of the segment to be amplified. The oligonucleotides serve as replication primers that can be extended by a DNA polymerase. The 3′ ends of the hybridized primers are oriented toward each other and positioned to prime DNA synthesis across the DNA segment **(Fig. 9–12a)**. Basic PCR requires four components: a DNA sample containing the segment to be amplified, the pair of synthetic oligonucleotide primers, deoxynucleoside triphosphates (dNTPs), and DNA polymerase. The reaction mixture is heated briefly to denature the DNA, separating the two strands. The mixture is cooled so that the primers can anneal to the DNA. The high concentration of primers increases the likelihood that they will anneal to each strand of the denatured DNA before the two DNA strands (present at a much lower concentration) can reanneal to each other. The primed segment is then replicated selectively by the DNA polymerase, using the pool of dNTPs. The cycle of heating, cooling, and replication is repeated 25 to 30 times over a few hours in an automated process, amplifying the DNA segment between the primers until it can be readily analyzed or cloned. Each cycle increases the amount of the DNA segment by a factor of 2, so the concentration of this DNA grows exponentially. After 20 cycles, the DNA segment has been amplified more than a millionfold (2^{20}); after 30 cycles, more than a billionfold. All other DNA in the sample remains unamplified. PCR uses a heat-stable DNA polymerase that remains active after every heating step and does not have to be replenished.

By careful design of the primers used for PCR, the amplified segment can be altered by the inclusion, at each end, of additional DNA not present in the chromosome that is being targeted. For example, restriction endonuclease cleavage sites can be included to facilitate the subsequent cloning of the amplified DNA (Fig. 9–12b).

This technology is highly sensitive: PCR can detect and amplify as little as one DNA molecule in almost any type of sample. The double-helical structure of DNA makes it a highly stable molecule, but DNA does degrade slowly over time (see Chapter 8). However, PCR has allowed the successful cloning of rare, undegraded DNA segments from samples more than 40,000 years old. Investigators have used the technique to clone DNA fragments from the mummified remains of humans and extinct animals, such as the woolly mammoth, creating the new fields of molecular archaeology and molecular paleontology. DNA from burial sites has been amplified by PCR and used to trace ancient human migrations. Epidemiologists can use PCR-enhanced DNA samples from human remains to trace the evolution of human pathogenic viruses. Thus, in addition to its usefulness for cloning DNA, PCR is a potent tool in forensic medicine (Box 9–1). It can be used to detect viral infections before they cause symptoms. For prospective parents from families with inherited genetic conditions of concern, PCR is an important tool for genetic counseling.

Given the extreme sensitivity of PCR methods, contamination of samples is a serious issue. In many applications, including forensic and ancient DNA tests, control samples ensure that the amplified DNA is not derived from the researcher or from contaminating bacteria.

Many specialized adaptations of PCR have increased the utility of the method. For example, sequences in RNA can be amplified if the first PCR cycle uses reverse transcriptase, an enzyme that works like DNA polymerase (see Fig. 8–33a), but uses RNA as a template

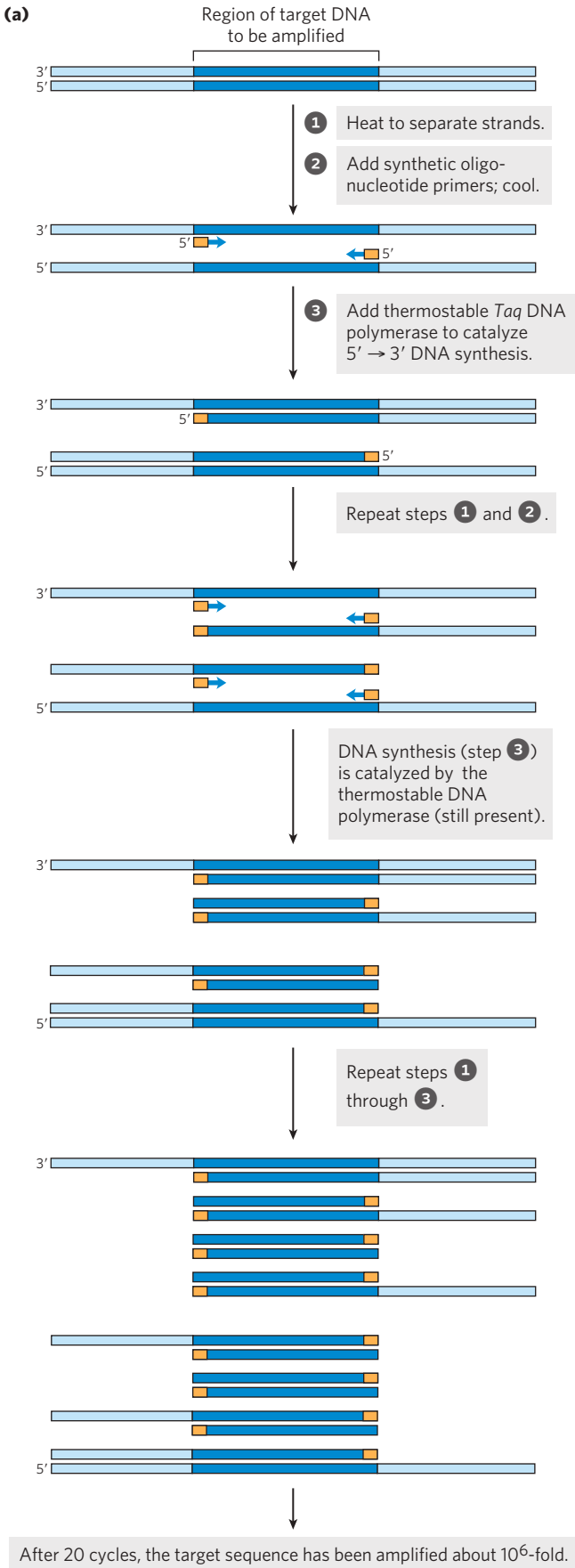
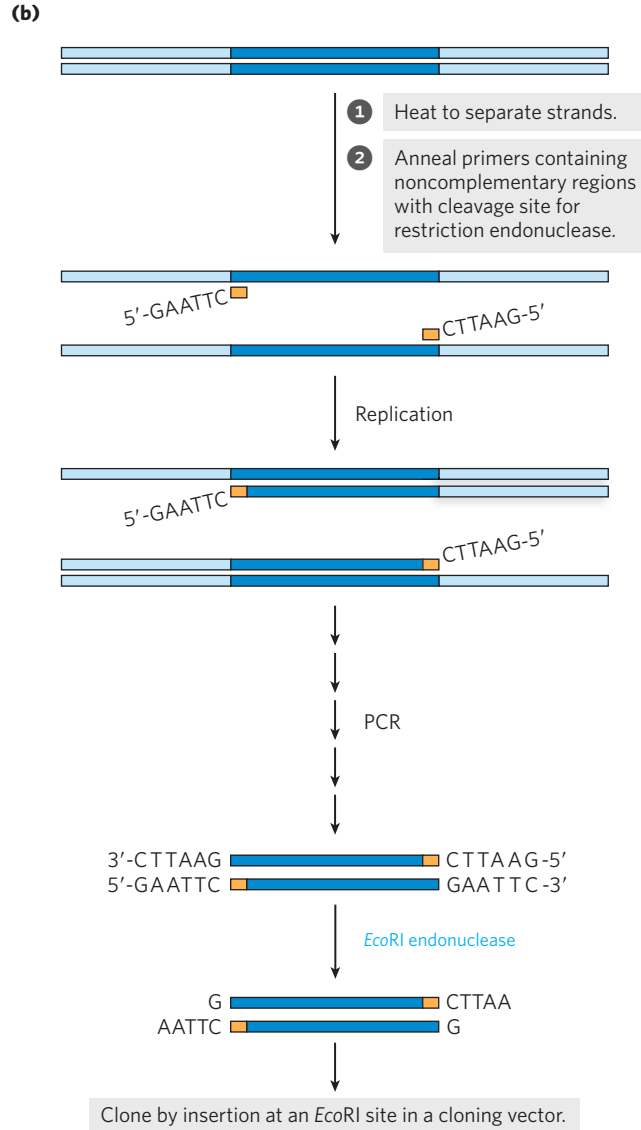


FIGURE 9–12 **Amplification of a DNA segment by the polymerase chain reaction (PCR). (a)** The PCR procedure has three steps. DNA strands are ❶ separated by heating, then ❷ annealed to an excess of short synthetic DNA primers (orange) that flank the region to be amplified (dark blue); ❸ new DNA is synthesized by polymerization catalyzed by DNA polymerase. The three steps are repeated for 25 or 30 cycles. The thermostable *Taq* DNA polymerase (from *Thermus aquaticus*, a bacterial species that grows in hot springs) is not denatured by the heating steps. **(b)** DNA amplified by PCR can be cloned. The primers can include noncomplementary ends that have a site for cleavage by a restriction endonuclease. Although these parts of the primers do not anneal to the target DNA, the PCR process incorporates them into the DNA that is amplified. Cleavage of the amplified fragments at these sites creates sticky ends, used in ligation of the amplified DNA to a cloning vector. **Polymerase Chain Reaction**



BOX 9–1 METHODS A Powerful Tool in Forensic Medicine

One of the most accurate methods for placing an individual at the scene of a crime is a fingerprint. The advent of recombinant DNA technology has made available a much more powerful tool: **DNA genotyping** (also called DNA fingerprinting or DNA profiling). The method is based on sequence polymorphisms—slight sequence differences among individuals: 1 in every 1,000 bp, on average. The use of these differences for forensic identification was first described by English geneticist Alec Jeffreys in 1985. Each difference from the prototype human genome sequence (the first one obtained) occurs in some fraction of the human population; every person has some differences from this prototype.

Modern forensic work focuses on differences in the lengths of **short tandem repeat (STR)** sequences. An STR is a short DNA sequence, repeated many times in tandem at a specific location in a chromosome; usually, the repeated sequence is 4 bp long. The STR loci most often used in DNA genotyping are short—4 to 50 repeats long (16 to 200 bp for tetranucleotide repeats)—and have multiple length variants in the human population. More than 20,000 tetranucleotide STR loci have been characterized and more than a million STRs of all types may be present in the human genome, accounting for about 3% of all human DNA.

The length of a particular STR in a given individual can be determined with the aid of the polymerase chain reaction (see Fig. 9–12). The use of PCR also makes the procedure sensitive enough to be applied to the very small samples of DNA often collected at crime scenes. The DNA sequences flanking STRs are unique to each STR locus and are identical (except for extremely rare mutations) in all humans. PCR primers are targeted to this flanking DNA and are designed to amplify the DNA across the STR (Fig. 1a). The length

(*continued on next page*)

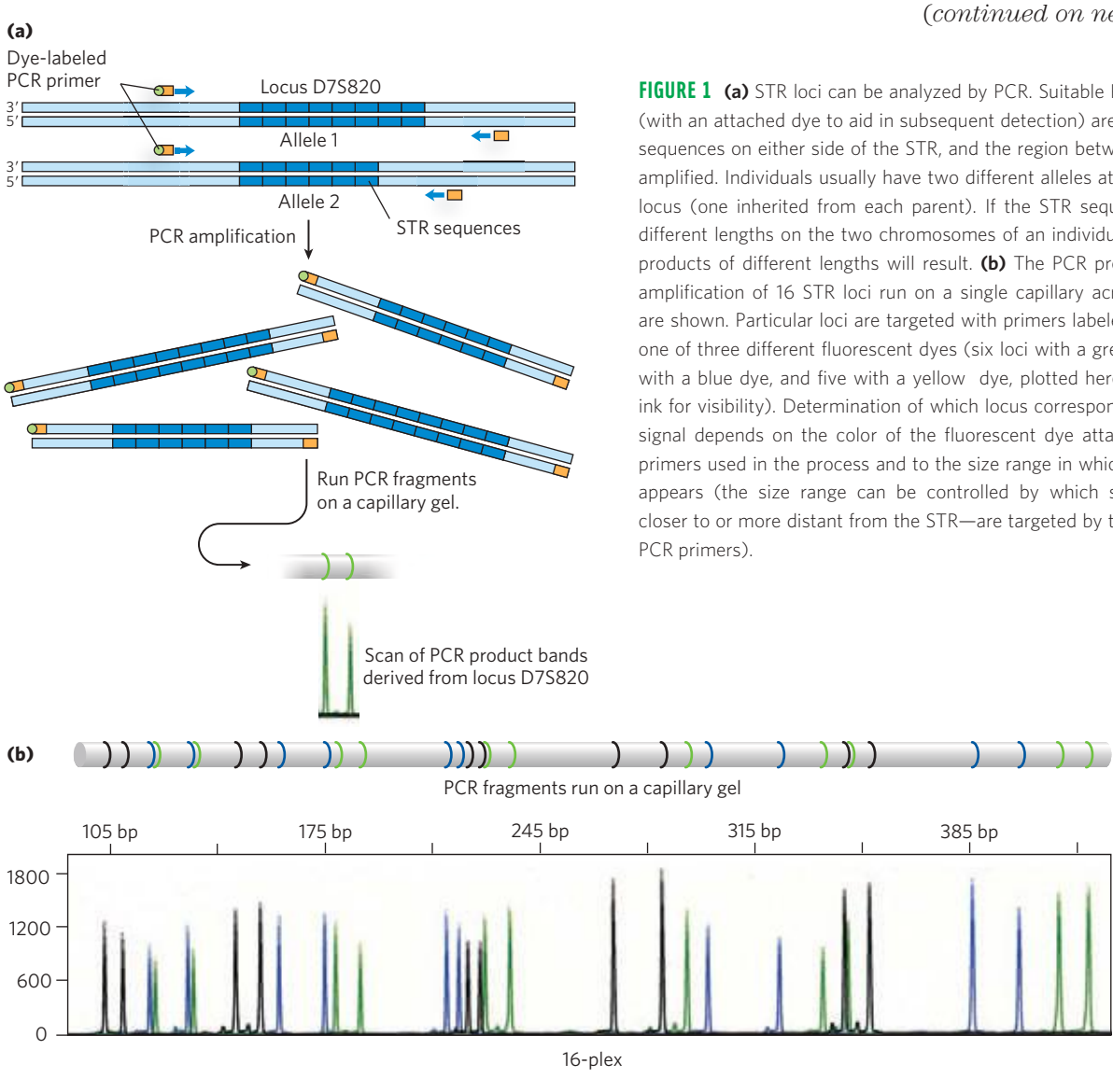


FIGURE 1 **(a)** STR loci can be analyzed by PCR. Suitable PCR primers (with an attached dye to aid in subsequent detection) are targeted to sequences on either side of the STR, and the region between them is amplified. Individuals usually have two different alleles at a particular locus (one inherited from each parent). If the STR sequences have different lengths on the two chromosomes of an individual, two PCR products of different lengths will result. **(b)** The PCR products from amplification of 16 STR loci run on a single capillary acrylamide gel are shown. Particular loci are targeted with primers labeled with only one of three different fluorescent dyes (six loci with a green dye, five with a blue dye, and five with a yellow dye, plotted here with black ink for visibility). Determination of which locus corresponds to which signal depends on the color of the fluorescent dye attached to the primers used in the process and to the size range in which the signal appears (the size range can be controlled by which sequences—closer to or more distant from the STR—are targeted by the designed PCR primers).

BOX 9–1 METHODS A Powerful Tool in Forensic Medicine (*Continued*)

of the PCR product then reflects the length of the STR in that sample. Because each human inherits one chromosome of each chromosome pair from each parent, the STR lengths on the two chromosomes are often different, generating two different STR lengths from one individual. The PCR products are subjected to electrophoresis on a very thin polyacrylamide gel in a capillary tube. The resulting bands are converted into a set of peaks that accurately reveal the size of each PCR fragment and so the length of the STR in the corresponding allele. Analysis of multiple STR loci can yield a profile that is unique to an individual (Fig. 1b). This is typically done with a commercially available kit that includes PCR primers unique to each locus, linked to colored dyes to help distinguish the different PCR products. PCR amplification enables investigators to obtain STR genotypes from less than 1 ng of partially degraded DNA, an amount that can be obtained from a single hair follicle, a drop of blood, a small semen sample, or samples that might be months or even many years old. When good STR genotypes are obtained, the chance of misidentification is less than 1 in 10^{18} (quintillion).

The successful forensic use of STR analysis required standardization, first attempted in the United Kingdom in 1995. The U.S. standard, called the Combined DNA Index System (CODIS), established in 1998, is based on 13 well-studied STR loci, which must be present in any DNA-typing experiment carried out in the United States (Table 1). The amelogenin gene is also used as a marker in the analyses. Present on the human sex chromosomes, this gene has a slightly different length on the X and Y chromosomes. PCR amplification across this gene thus generates different-size products that can reveal the sex of the DNA donor. By the beginning of 2010, the CODIS database contained more than 7 million STR genotypes and had assisted more than 100,000 forensic investigations.

DNA genotyping has been used to both convict and acquit suspects and to establish paternity with an extraordinary degree of certainty. The impact of these procedures on court cases will continue to grow as standards are refined and as international STR genotyping databases grow. Even very old mysteries can be solved. In 1996, DNA fingerprinting helped confirm the identification of the bones of the last Russian czar and his family, who were assassinated in 1918.

TABLE 1 Properties of the Loci Used for the CODIS Database

Locus	Chromosome	Repeat motif	Repeat length (range)*	Number of alleles seen†
CSF1PO	5	TAGA	5–16	20
FGA	4	CTTT	12.2–51.2	80
TH01	11	TCAT	3–14	20
TPOX	2	GAAT	4–16	15
VWA	12	[TCTG][TCTA]	10–25	28
D3S1358	3	[TCTG][TCTA]	8–21	24
D5S818	5	AGAT	7–18	15
D7S820	7	GATA	5–16	30
D8S1179	8	[TCTA][TCTG]	7–20	17
D13S317	13	TATC	5–16	17
D16S539	16	GATA	5–16	19
D18S51	18	AGAA	7–39.2	51
D21S11	21	[TCTA][TCTG]	12–41.2	82
Amelogenin‡	X, Y	Not applicable		

Source: Adapted from Butler, J.M. (2005) *Forensic DNA Typing*, 2nd edn, Academic Press, San Diego, p. 96.

*Repeat lengths observed in the human population. Partial or imperfect repeats can be included in some alleles.

†Number of different alleles observed as of 2005 in the human population. Careful analysis of a locus in many individuals is a prerequisite to its use in forensic DNA typing.

‡Amelogenin is a gene, of slightly different size on the X and Y chromosomes, that is used to establish gender.

(Fig. 9–12). After the DNA strand is made from the RNA template, the remaining cycles can be carried out with DNA polymerases, using standard PCR protocols. This **reverse transcriptase PCR (RT-PCR)** can be used, for example, to detect sequences derived from living cells (which are transcribing their DNA into RNA) as opposed to dead tissues.

PCR protocols can also be made quantitative for estimating the relative copy numbers of particular sequences in a sample. The approach is called **quantitative PCR**, or **qPCR**. If a DNA sequence is present in higher than usual amounts in a sample—for example, if certain genes are amplified in tumor cells—qPCR can reveal the increased representation of that sequence. In brief, the PCR is carried out in the presence of a probe that emits a fluorescent signal when the PCR product is present **(Fig. 9–13)**. If the sequence of interest is present at higher levels than other sequences in the sample, the PCR signal will reach a predetermined threshold faster. Reverse transcriptase PCR and real-time PCR can be combined to determine the relative concentrations of a particular RNA molecule in a cell under different environmental conditions.

Finally, PCR has facilitated the development of powerful new DNA sequencing procedures, as we shall see.

SUMMARY 9.1 Studying Genes and Their Products

- DNA cloning and genetic engineering involve the cleavage of DNA and assembly of DNA segments in new combinations—recombinant DNA.
- Cloning entails cutting DNA into fragments with enzymes; selecting and possibly modifying a fragment of interest; inserting the DNA fragment into a suitable cloning vector; transferring the vector with the DNA insert into a host cell for replication; and identifying and selecting cells that contain the DNA fragment.
- Key enzymes in gene cloning include restriction endonucleases (especially the type II enzymes) and DNA ligase.
- Cloning vectors include plasmids, and, for the longest DNA inserts, bacterial artificial chromosomes (BACs) and yeast artificial chromosomes (YACs).
- Genetic engineering techniques manipulate cells to express and/or alter cloned genes.
- Proteins or peptides can be attached to a protein of interest by altering its cloned gene, creating a fusion protein. The additional peptide segments can be used to detect the protein or to purify it using convenient affinity chromatography methods.
- The polymerase chain reaction (PCR) permits the amplification of chosen segments of DNA or RNA for detailed study or cloning.

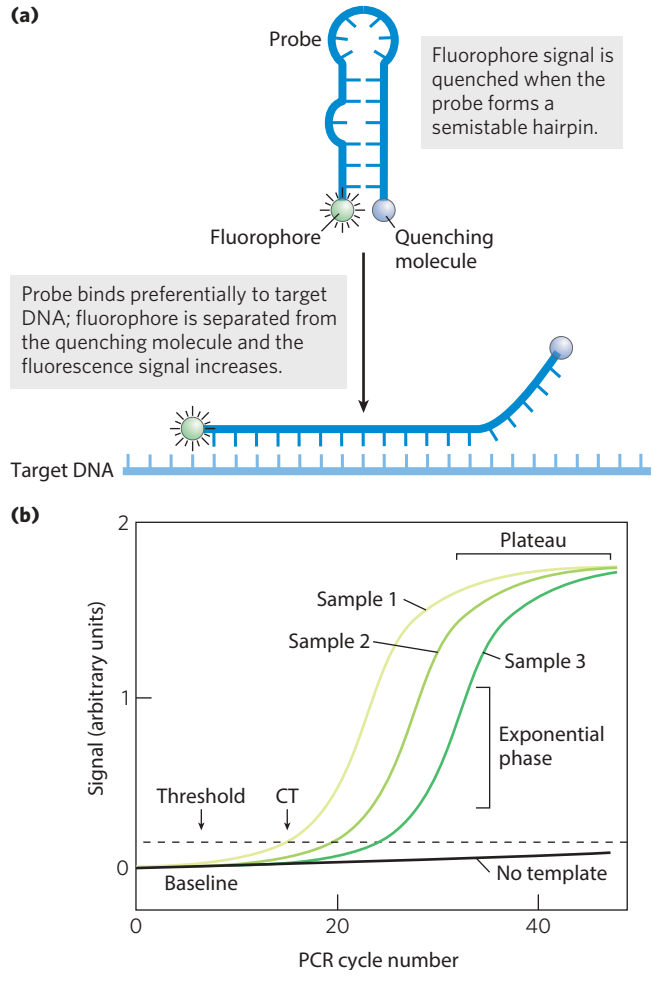


FIGURE 9–13 Quantitative PCR. PCR can be used quantitatively, by carefully monitoring the progress of a PCR amplification and determining when a DNA segment has been amplified to a specific threshold level. **(a)** The amount of PCR product present is determined by measuring the level of a fluorescent probe attached to a reporter oligonucleotide complementary to the DNA segment that is being amplified. Probe fluorescence is initially not detectable due to a fluorescence quencher attached to the same oligonucleotide. When the reporter oligonucleotide pairs with its complement in a copy of the amplified DNA segment, the fluorophore is separated from the quenching molecule and fluorescence results. **(b)** As the PCR reaction proceeds, the amount of the targeted DNA segment increases exponentially, and the fluorescent signal also increases exponentially as the oligonucleotide probes anneal to the amplified segments. After many PCR cycles, the signal reaches a plateau as one or more reaction components become exhausted. When a segment is present in greater amounts in one sample than another, its amplification reaches a defined threshold level earlier. The "No template" line follows the slow increase in background signal observed in a control that does not include added sample DNA. CT is the cycle number at which the threshold is first surpassed.

9.2 Using DNA-Based Methods to Understand Protein Function

Protein function can be described on three levels. **Phenotypic function** describes the effects of a protein on the entire organism. For example, the loss of the protein may lead to slower growth of the organism, an altered

development pattern, or even death. **Cellular function** is a description of the network of interactions a protein engages in at the cellular level. Identifying interactions with other proteins in the cell can help define the kinds of metabolic processes in which the protein participates. Finally, **molecular function** refers to the precise biochemical activity of a protein, including details such as the reactions an enzyme catalyzes or the ligands a receptor binds. The challenge of understanding the functions of the thousands of uncharacterized or poorly characterized proteins found in a typical cell has given rise to a wide variety of techniques. DNA-based methods make a critical contribution to this effort and can provide information on all three levels.

DNA Libraries Are Specialized Catalogs of Genetic Information

A **DNA library** is a collection of DNA clones, gathered for purposes of genome sequencing, gene discovery, or determination of gene/protein function. The library can take a variety of forms, depending on the source of the DNA and the ultimate purpose of the library.

The largest is a **genomic library**, produced when the complete genome of an organism is cleaved into thousands of fragments. All the fragments are cloned by insertion of each fragment into a cloning vector. This creates a complex mixture of recombinant vectors, each one with a different cloned fragment. The first step is *partial* digestion of the DNA by restriction endonucleases, such that any given sequence will appear in fragments of a limited range of sizes—a range compatible with the cloning vector to ensure that virtually all sequences are represented among the clones in the library. Fragments that are too large or too small for cloning are removed by centrifugation or electrophoresis. The cloning vector, such as a BAC or YAC, is cleaved with the same restriction endonuclease used to digest the DNA and ligated to the genomic DNA fragments. The ligated DNA mixture is then used to transform bacteria or yeast cells to produce a library of cells, each harboring a different recombinant DNA molecule. Ideally, all the DNA in the genome under study is represented in the library. Each transformed bacterium or yeast cell grows into a colony, or clone, of identical cells, each cell bearing the same recombinant plasmid, one of many represented in the overall library.

Efforts to define gene or protein function often make use of more specialized libraries. An example is a library that includes only those sequences of DNA that are *expressed*—that is, transcribed into RNA—in a given organism, or even just in certain cells or tissues. Such a library lacks the noncoding DNA that makes up a large portion of many eukaryotic genomes. The researcher first extracts mRNA from an organism, or from specific cells of an organism, and then prepares the **complementary DNAs (cDNAs)**. This multistep reaction, shown in **Figure 9–14**, relies on reverse transcriptase, which synthesizes DNA from a template RNA. The resulting double-stranded DNA fragments are inserted into a suitable vector and cloned, creating a population of clones called a **cDNA library**. The appearance of a gene for a particular protein in such a library implies that it is expressed in the cells and under the conditions used to generate the library.

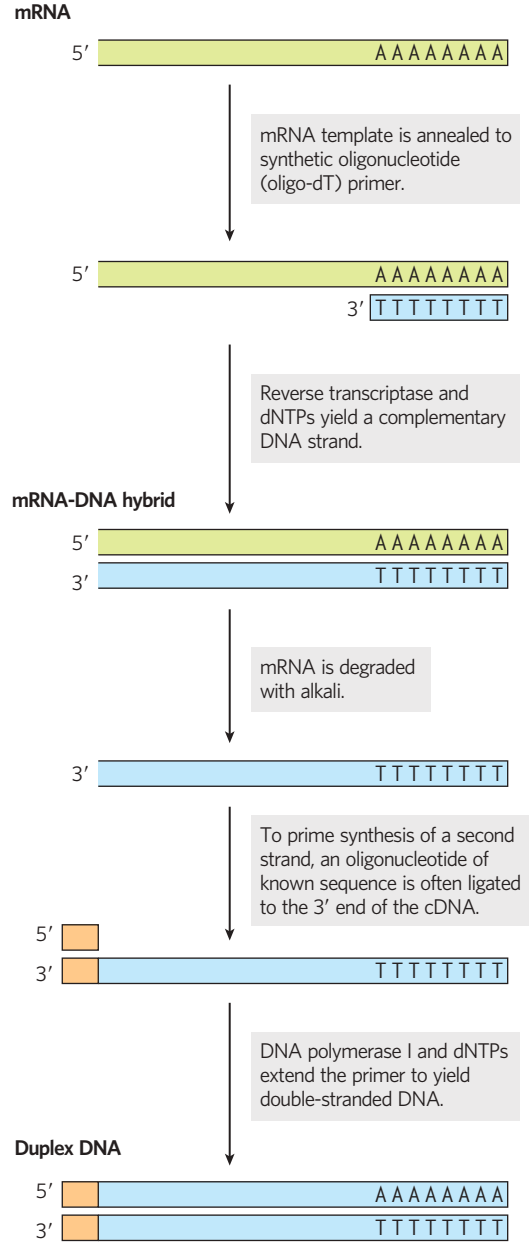


FIGURE 9–14 Building a cDNA library from mRNA. A cell's total mRNA content includes transcripts from thousands of genes, and the cDNAs generated from this mRNA are correspondingly heterogeneous. Reverse transcriptase can synthesize DNA on an RNA or a DNA template (see Fig. 26–32). To prime the synthesis of a second DNA strand, oligonucleotides of known sequence are ligated to the 3′ end of the first strand, and the double-stranded cDNA so produced is cloned into a plasmid.

Sequence or Structural Relationships Provide Information on Protein Function

One important reason to sequence many genomes is to provide a database that can be used to assign gene functions by genome comparisons, an enterprise referred to as **comparative genomics**. Sometimes a newly discovered gene is related by sequence homologies to a gene previously studied in another or the same species, and its function can be entirely or partly defined by that relationship. Genes that occur in different species but have a clear sequence and functional relationship to each other are called **orthologs**. Genes similarly related to each other within a single species are called **paralogs**. If the function of a gene has been characterized for one species, this information can be used to at least tentatively assign gene function to the ortholog found in the second species. The correlation is easiest to make when comparing genomes from relatively closely related species, such as mouse and human, although many clearly orthologous genes have been identified in species as distant as bacteria and humans. Sometimes even the order of genes on a chromosome is conserved over large segments of the genomes of closely related species **(Fig. 9–15)**. Conserved gene order, called **synteny**, provides additional evidence for an orthologous relationship between genes at identical locations within the related segments.

Alternatively, certain sequences associated with particular structural motifs (Chapter 4) may be identified within a protein. The presence of a structural motif may help to define molecular function by suggesting that a protein, say, catalyzes ATP hydrolysis, binds to DNA, or forms a complex with zinc ions. These relationships are determined with the aid of sophisticated computer programs, limited only by the current information on gene and protein structure and by our capacity to associate sequences with particular structural motifs.

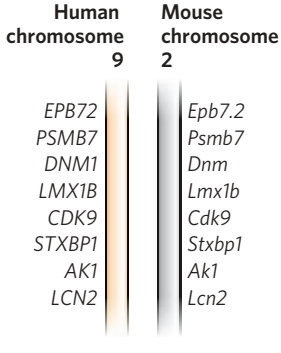


FIGURE 9–15 Synteny in the human and mouse genomes. Large segments of the two genomes have closely related genes aligned in the same order on the chromosomes. In these short segments of human chromosome 9 and mouse chromosome 2, the genes exhibit a very high degree of homology, as well as the same gene order. The different lettering schemes for the gene names simply reflect the different naming conventions in the two species.

Fusion Proteins and Immunofluorescence Can Localize Proteins in Cells

Often, an important clue to a gene product's function comes from determining its location within the cell. For example, a protein found exclusively in the nucleus could be involved in processes that are unique to that organelle, such as transcription, replication, or chromatin condensation. Researchers often engineer fusion proteins for the purpose of locating a protein in the cell or organism. Some of the most useful fusions involve the attachment of marker proteins that allow the investigator to determine the location by direct visualization or by immunofluorescence.

A particularly useful marker is the **green fluorescent protein (GFP)**. A target gene (coding the protein of interest) fused to the GFP gene generates a fusion protein that is highly fluorescent—it literally lights up when exposed to blue light—and can be visualized directly in a living cell. GFP is a protein derived from the jellyfish *Aequorea victoria* (see Box 12–3, Fig. 1). It has a β-barrel structure, and the fluorophore (the fluorescent component of the protein) is in the center of the barrel **(Fig. 9–16a)**. The fluorophore is derived from a rearrangement and oxidation of several amino acid residues. Because this reaction is autocatalytic and requires no other proteins or cofactors other than molecular oxygen (see Box 12–3, Fig. 3), GFP is readily cloned in an active form in almost any cell. Just a few molecules of this protein can be observed microscopically, allowing the study of its location and movements in a cell. Careful protein engineering, coupled with the isolation of related fluorescent proteins from other marine coelenterates, has made a wide range of these proteins available, in an array of colors (Fig. 9–16b) and other characteristics (brightness, stability). With this technology, for example, the protein GLR1 (a glutamate receptor of nervous tissue) has been visualized as a GLR1-GFP fusion protein in the nematode *Caenorhabditis elegans* (Fig. 9–16c).

In many cases, visualization of a GFP fusion protein in a live cell is not possible or practical or desirable. The GFP fusion protein may be inactive or may not be expressed at sufficient levels to allow visualization. In this case, **immunofluorescence** is an alternative approach for visualizing the endogenous (unaltered) protein. This approach requires the fixation (and thus death) of the cell. The protein of interest is sometimes expressed as a fusion protein with an **epitope tag**, a short protein sequence that is bound tightly by a well-characterized, commercially available antibody. Fluorescent molecules (fluorochromes) are attached to this antibody. More commonly, the target protein is unaltered. This protein is bound by an antibody that is specific for it. Next, a second antibody is added that binds specifically to the first one, and it is the second antibody that has the attached fluorochrome(s) **(Fig. 9–17a)**. A variation of this indirect visualization approach is to attach biotin molecules to the first antibody, then add

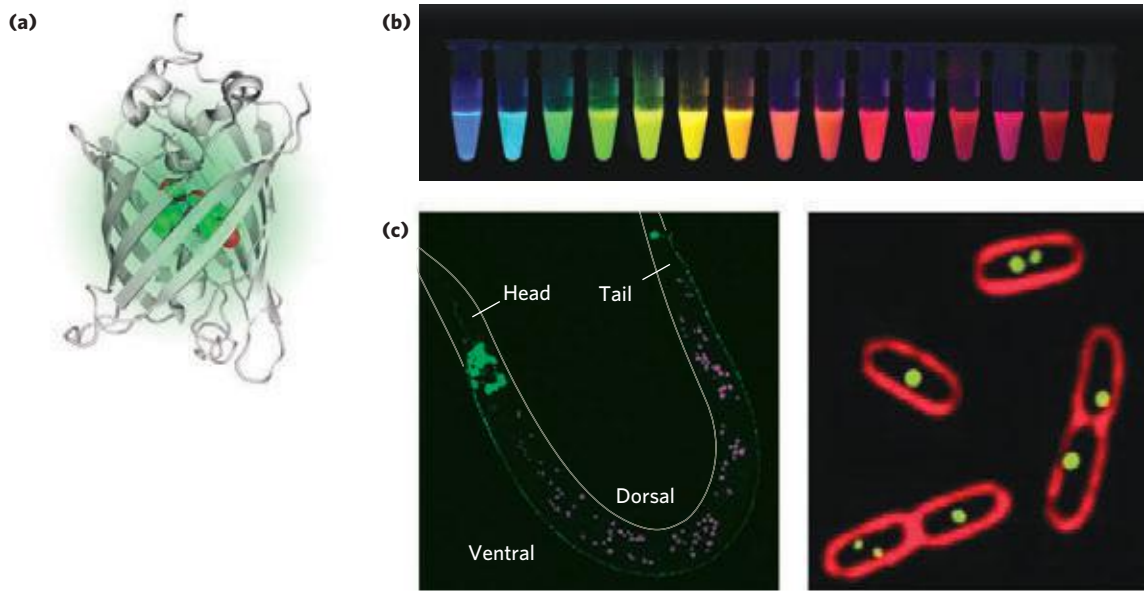


FIGURE 9–16 Green fluorescent protein (GFP). (a) The GFP protein (PDB ID 1GFL), derived from the jellyfish *Aequorea victoria*, has a β-barrel structure; the fluorophore (shown as a space-filling model) is in the center of the barrel. **(b)** Variants of GFP are now available in almost any color of the visible spectrum. **(c)** A GLR1-GFP fusion protein fluoresces bright green in *Caenorhabditis elegans*, a nematode worm (left). GLR1 is a glutamate receptor of nervous tissue. (Autofluorescing fat droplets are false colored in magenta.) The membranes of *E. coli* cells (right) are stained with a red fluorescent dye. The cells are expressing a protein that binds to a resident plasmid, fused to GFP. The green spots indicate the locations of plasmids.

streptavidin (a bacterial protein closely related to avidin, a protein that binds to biotin; see Table 5–1) complexed with fluorochromes. The interaction between biotin and streptavidin is one of the strongest and most specific known, and the potential to add multiple fluorochromes to each target protein gives this method great sensitivity.

Highly specialized cDNA libraries (Fig. 9–14) can be made by cloning cDNAs or cDNA fragments into a vector that fuses each cDNA sequence with the sequence for a marker called a reporter gene. The fused gene is often called a reporter construct. For example, all the genes in the library may be fused to the GFP gene **(Fig. 9–18)**. Each cell in the library expresses one of these fused genes. The cellular location of the product of any gene represented in the library will be revealed as foci of light in cells that express the appropriate fused gene at sufficient levels—assuming that the gene retains its normal function and location.

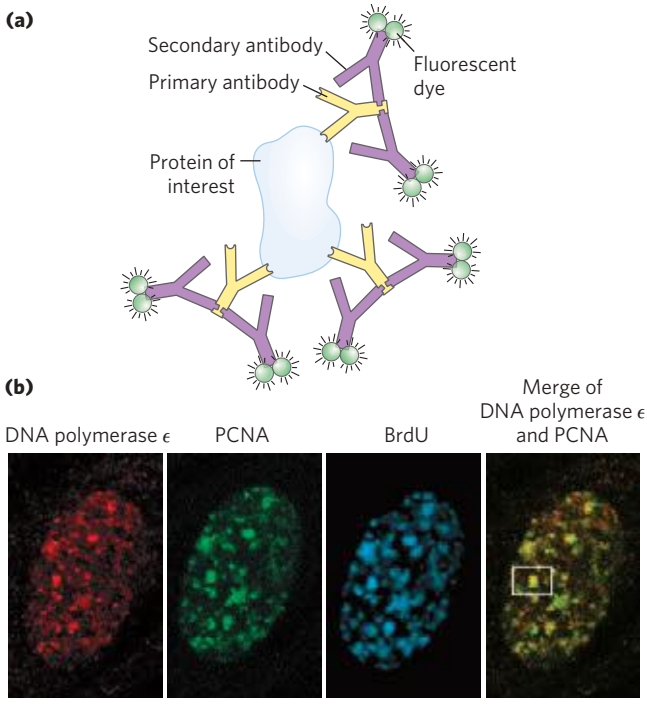


FIGURE 9–17 Indirect immunofluorescence. (a) The protein of interest is bound to a primary antibody, and a secondary antibody is added; this second antibody, with one or more attached fluorescent groups, binds to the first. Multiple secondary antibodies can bind the primary antibody, amplifying the signal. If the protein of interest is in the interior of the cell, the cell is fixed and permeabilized, and the two antibodies are added in succession. **(b)** The end result is an image in which bright spots indicate the location of the protein or proteins of interest in the cell. The images show a nucleus from a human fibroblast, successively stained with antibodies and fluorescent labels for DNA polymerase ε, for PCNA, an important polymerase accessory protein, and for bromo-deoxyuridine (BrdU), a nucleotide analog. The BrdU, added as a brief pulse, identifies regions undergoing active DNA replication. The patterns of staining show that DNA polymerase ε and PCNA co-localize to regions of active DNA synthesis. One such region is visible in the white box.

Protein–Protein Interactions Can Help Elucidate Protein Function

Another key to defining the function of a particular protein is to determine what it binds to. In the case of

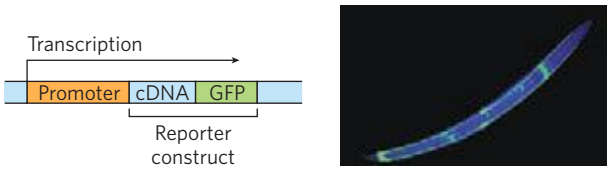


FIGURE 9–18 Specialized DNA libraries. Cloning of a cDNA next to the GFP gene creates a reporter construct. Transcription proceeds through the gene of interest (the inserted cDNA) and the reporter gene (here, GFP), and the mRNA transcript is expressed as a fusion protein. The GFP part of the protein is visible with the fluorescence microscope. Although only one example is shown, thousands of genes can be fused to GFP in similar constructs and stored in libraries in which each cell or organism in the library expresses a different protein fused to GFP. If the fusion protein is properly expressed, its location in the cell or organism can be assessed. The photograph shows a nematode worm containing a GFP fusion protein expressed only in the four "touch" neurons that run the length of its body.

protein-protein interactions, the association of a protein of unknown function with one whose function is known can provide useful and compelling "guilt by association." The techniques used in this effort are quite varied.

Purification of Protein Complexes With the construction of cDNA libraries in which each gene is fused to an epitope tag, investigators can precipitate the protein product of a gene by complexing it with the antibody that binds the epitope, a process called **immunoprecipitation (Fig. 9–19)**. If the tagged protein is expressed in cells, other proteins that bind to it precipitate with it. Identifying the associated proteins reveals some of the intracellular protein-protein interactions of the tagged protein. There are many variations of this process. For example, a crude extract of cells that express a tagged protein is added to a column containing immobilized antibody (see Fig. 3–17c for a description of affinity chromatography). The tagged protein binds to the antibody, and proteins that interact with the tagged protein are sometimes also retained on the column. The connection between the protein and the tag is cleaved with a specific protease, and the protein complexes are eluted from the column and analyzed. Researchers can use these methods to define complex networks of interactions within a cell. In principle, the chromatographic approach to analyzing protein-protein interactions can be used with any type of protein tag (His tag, GST, etc.) that can be immobilized on a suitable chromatographic medium.

The selectivity of this approach has been enhanced with **tandem affinity purification (TAP) tags**. Two consecutive tags are fused to a target protein, and the fusion protein is expressed in a cell **(Fig. 9–20)**. The first tag is protein A, a protein found at the surface of the bacterium *Staphylococcus aureus* that binds tightly to mammalian immunoglobulin G (IgG). The second tag is often a calmodulin-binding peptide. A crude extract containing the TAP-tagged fusion protein is passed through a column matrix with attached IgG antibodies that bind to protein A. Most of the unbound cellular proteins are

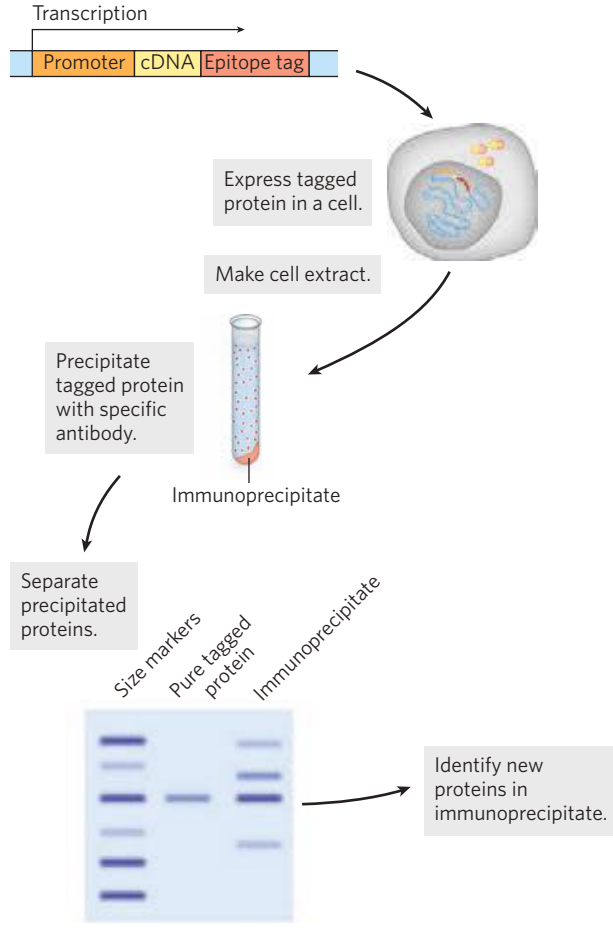


FIGURE 9–19 The use of epitope tags to study protein-protein interactions. The gene of interest is cloned next to a gene for an epitope tag, and the resulting fusion protein is precipitated by antibodies to the epitope. Any other proteins that interact with the tagged protein also precipitate, thereby helping to elucidate protein-protein interactions.

washed through the column, but proteins that normally interact with the target protein in the cell are retained. The first tag is then cleaved from the fusion protein with a highly specific protease, TEV protease, and the shortened fusion target protein and any proteins associated noncovalently with the target protein are eluted from the column. The eluent is then passed through a second column containing a matrix with attached calmodulin that binds the second tag. Loosely bound proteins are again washed from the column. After the second tag is cleaved, the target protein is eluted from the column with its associated proteins. The two consecutive purification steps eliminate any weakly bound contaminants. False positives are minimized, and protein interactions that persist through both steps are likely to be functionally significant.

Yeast Two-Hybrid Analysis A sophisticated genetic approach to defining protein-protein interactions is based on the properties of the Gal4 protein (Gal4p; see Fig. 28–30), which activates the transcription of *GAL* genes in yeast (genes encoding the enzymes of galactose metabolism). Gal4p has two domains: one that binds

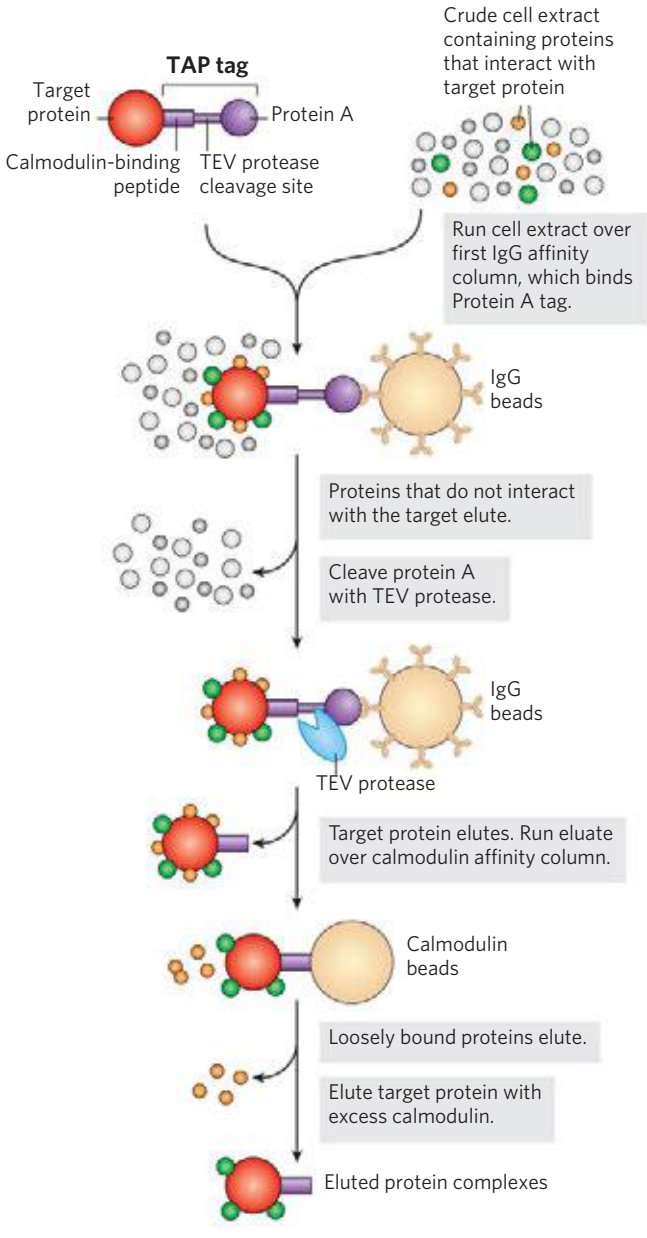


FIGURE 9–20 Tandem affinity purification (TAP) tags. A TAP-tagged protein and associated proteins are isolated by two consecutive affinity purifications, as described in the text.

a specific DNA sequence and another that activates RNA polymerase to synthesize mRNA from an adjacent gene. The two domains of Gal4p are stable when separated, but activation of RNA polymerase requires interaction with the activation domain, which in turn requires positioning by the DNA-binding domain. Hence, the domains must be brought together to function correctly.

In **yeast two-hybrid analysis**, the protein-coding regions of the genes to be analyzed are fused to the yeast gene for either the DNA-binding domain or the activation domain of Gal4p, and the resulting genes express a series of fusion proteins **(Fig. 9–21)**. If a protein fused to the DNA-binding domain interacts with a protein fused to the activation domain, transcription is activated. The reporter gene transcribed by this activation is generally one that yields a protein required for growth or an enzyme that catalyzes a reaction with a colored product. Thus, when grown on the proper medium, cells that contain a pair of interacting proteins are easily distinguished from those that do not. A library can be set up

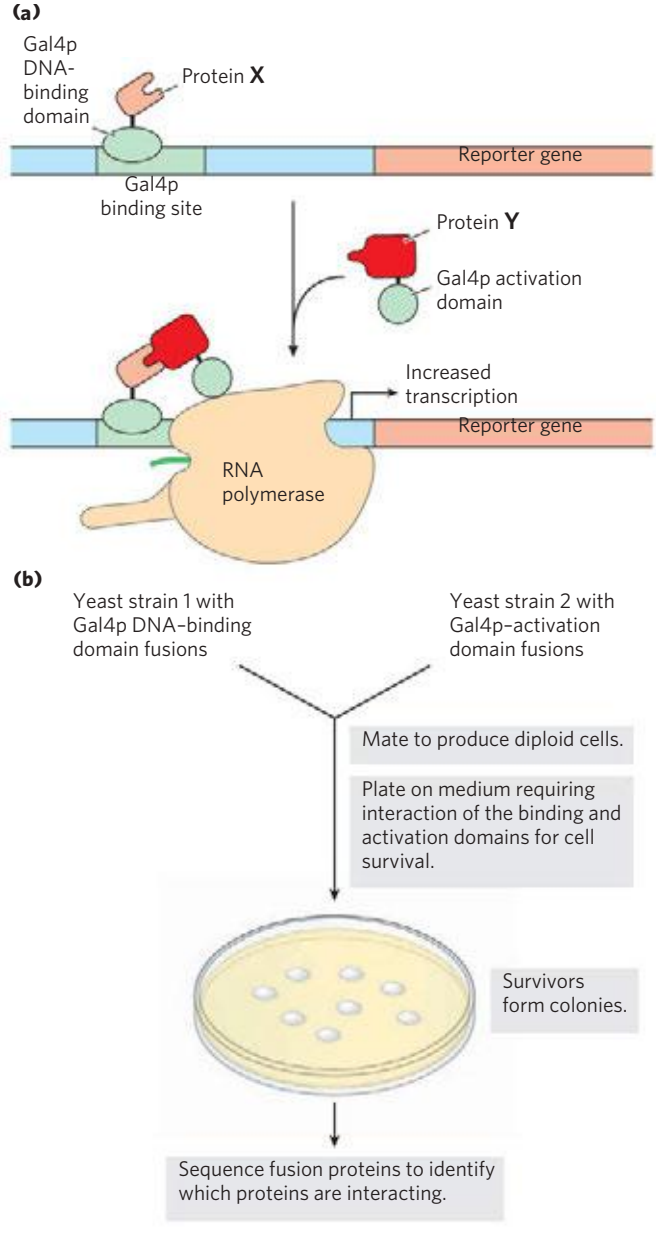


FIGURE 9–21 Yeast two-hybrid analysis. (a) The goal is to bring together the DNA-binding domain and the activation domain of the yeast Gal4 protein (Gal4p) through the interaction of two proteins, X and Y, to which one or other of the domains is fused. This interaction is accompanied by the expression of a reporter gene. **(b)** The two gene fusions are created in separate yeast strains, which are then mated. The mated mixture is plated on a medium on which the yeast cannot survive unless the reporter gene is expressed. Thus, all surviving colonies have interacting fusion proteins. Sequencing of the fusion proteins in the survivors reveals which proteins are interacting.

with a particular yeast strain in which each cell in the library has a gene fused to the Gal4p DNA-binding domain gene, and many such genes are represented in the library. In a second yeast strain, a gene of interest is fused to the gene for the Gal4p activation domain. The yeast strains are mated, and individual diploid cells are grown into colonies. The only cells that grow on the selective medium, or that produce the appropriate color, are those in which the gene of interest is binding to a partner, allowing transcription of the reporter gene. This allows for large-scale screening for cellular proteins that interact with the target protein. The interacting protein that is fused to the Gal4p DNA-binding domain present in a particular selected colony can be quickly identified by DNA sequencing of the fusion protein's gene. Some false positive results occur, due to the formation of multiprotein complexes.

These techniques for determining cellular localization and molecular interactions provide important clues to protein function. However, they do not replace classical biochemistry. They simply give researchers an expedited entrée into important new biological problems. When paired with the simultaneously evolving tools of biochemistry and molecular biology, the techniques described here are speeding the discovery not only of new proteins, but of new biological processes and mechanisms.

DNA Microarrays Reveal RNA Expression Patterns and Other Information

Major refinements of the technology underlying DNA libraries, PCR, and hybridization have come together in the development of **DNA microarrays**, which allow the rapid and simultaneous screening of many thousands of genes. DNA segments from genes of known sequence, a few dozen to hundreds of base pairs long, are amplified by PCR. Robotic devices then accurately deposit nanoliter quantities of the DNA solutions on a solid surface of just a few square centimeters, in a predesigned array, with each of the thousands of spots containing sequences derived from a particular gene. An alternative and increasingly common strategy is to synthesize DNA directly on the solid surface, using photolithography **(Fig. 9–22)**. The resulting array, often called a chip, may include sequences derived from every gene of a bacterial or yeast genome, or selected families of genes

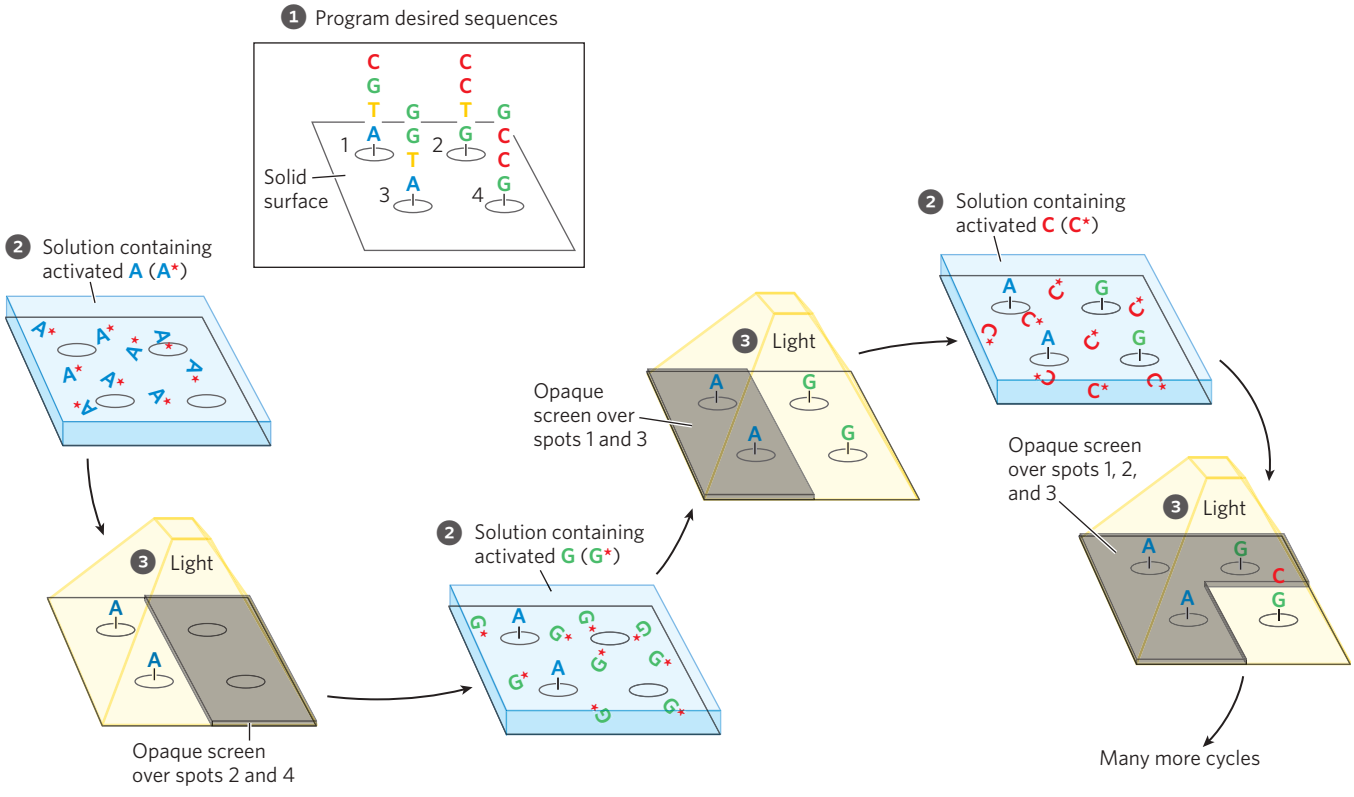
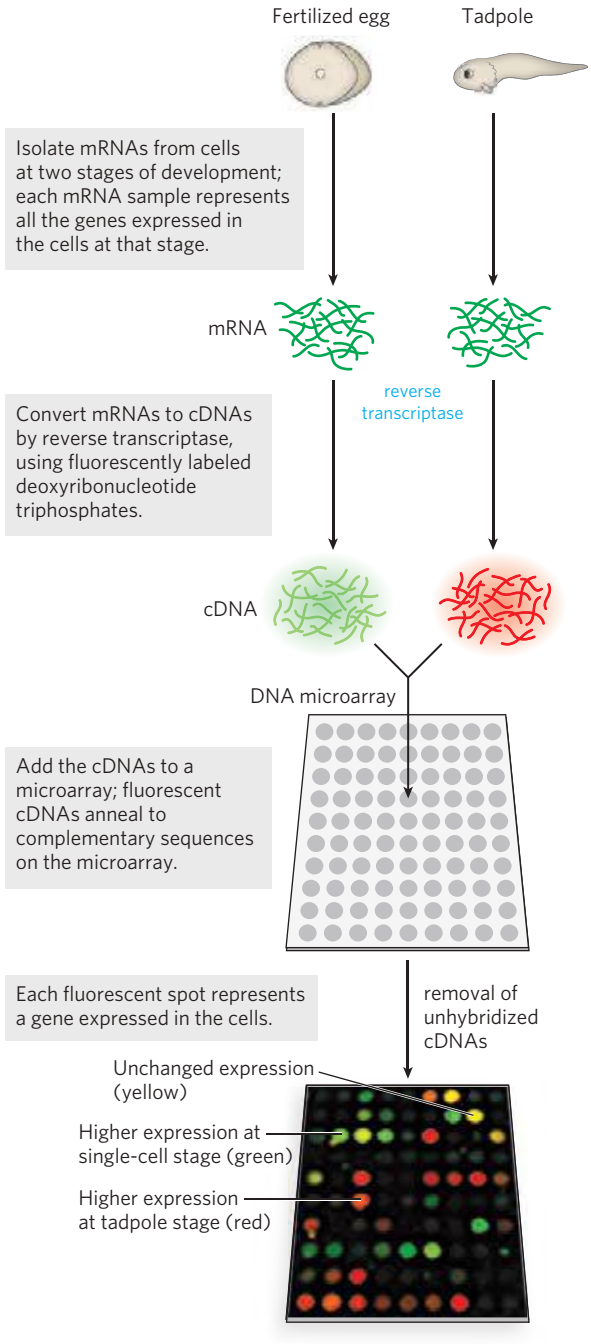


FIGURE 9–22 Photolithography to create a DNA microarray. ❶ A computer is programmed with the desired oligonucleotide sequences. ❷ The reactive groups, attached to a solid surface, are initially rendered inactive by photoactive blocking groups, which can be removed by a flash of light. An opaque screen blocks the light from some areas of the surface, preventing their activation. Other areas or "spots" are exposed. ❸ A solution containing one activated nucleotide (e.g., A*) is washed over the spots. The 5′ hydroxyl of the nucleotide is blocked to prevent unwanted reactions, and the nucleotide links to the surface groups at the appropriate spots through its 3′ hydroxyl. The surface is washed successively with solutions containing each remaining activated nucleotide (G*, C*, T*). The 5′-blocking groups on each nucleotide limit the reactions to addition of one nucleotide at a time, and these groups can also be removed by light. Once each spot has one nucleotide, a second nucleotide can be added to extend the nascent oligonucleotide at each spot, using screens and light to ensure that the correct nucleotides are added at each spot in the correct sequence. This continues until the required sequences are built up on each of the thousands of spots in a DNA microarray.

from a larger genome. Once constructed, the microarray can be probed with mRNAs or cDNAs from a particular cell type or cell culture to identify the genes being expressed in those cells.

A microarray can provide a snapshot of all the genes in an organism, informing the researcher about the genes that are expressed at a given stage in the organism's development or under a particular set of environmental conditions. For example, the total complement of mRNA can be isolated from cells at two different stages of development and converted to cDNA with reverse transcriptase. Fluorescently labeled deoxyribonucleotides can be used to make one cDNA sample fluoresce red, the other green **(Fig. 9–23)**. The cDNA from the two samples is mixed and used to probe the microarray. Each cDNA anneals to only one spot on the microarray, corresponding to the gene encoding the mRNA that gave rise to that cDNA. Spots that fluoresce green represent genes that produce mRNAs at higher levels at one developmental stage; those that fluoresce red represent genes expressed at higher levels at another stage. If a gene produces mRNAs that are equally abundant at both stages of development, the corresponding spot fluoresces yellow. By using a mixture of two samples to measure relative rather than absolute sequence abundance, the method corrects for variations in the amount of DNA originally deposited in each spot on the grid, as well as other possible inconsistencies among spots in the microarray. The spots that fluoresce provide a snapshot of all the genes being expressed in the cells at the moment they were harvested—gene expression examined on a genome-wide scale. For a gene of unknown function, the time and circumstances of its expression can provide important clues about its role in the cell. These technologies also reveal expression of many types of specialized RNAs, such as microRNAs (miRNAs; Chapter 26) and small interfering RNAs (siRNAs; Chapter 28).



An example of this technique is illustrated in **Figure 9–24**, which shows the dramatic results that microarray experiments can produce. Segments from each of the roughly 6,500 genes in the completely sequenced yeast genome were separately amplified by PCR, and each segment was deposited in a defined pattern to create the microarray. In a sense, this array provides a snapshot of the functioning of the entire yeast genome under one set of conditions.

SUMMARY 9.2 Using DNA-Based Methods to Understand Protein Function

- Proteins can be studied at the level of phenotypic, cellular, or molecular function.
- DNA libraries can be a prelude to many types of investigations that yield information about protein function.
- By fusing a gene of interest with genes that encode green fluorescent protein or epitope tags, researchers can visualize the cellular location of the gene product, either directly or by immunofluorescence.
- The interactions of a protein with other proteins or RNA can be investigated with epitope tags and immunoprecipitation or affinity chromatography.

FIGURE 9–23 A DNA microarray experiment. A microarray can be prepared from any known DNA sequence, from any source. Once the DNA is attached to a solid support, the microarray can be probed with other fluorescently labeled nucleic acids. Here, mRNA samples are collected from cells of a frog at two different stages of development.

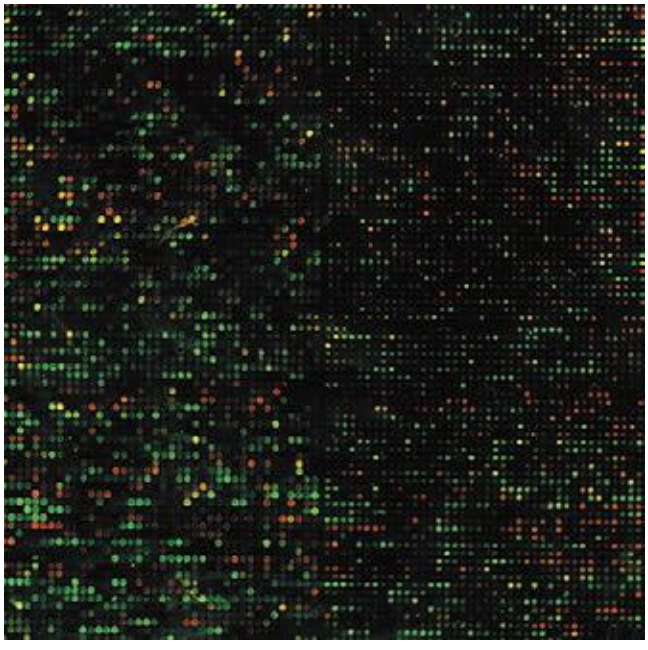


FIGURE 9–24 Enlarged image of a DNA microarray. Each glowing spot contains DNA from one of the roughly 6,500 genes of the yeast (*S. cerevisiae*) genome, with every gene represented in the array. The microarray has been probed with fluorescently labeled nucleic acid derived from mRNAs obtained when the cells were growing in culture (green) and 5 hours after they began to form spores (red). The green spots represent genes expressed at higher levels during growth; the red spots, genes expressed at higher levels during sporulation. The yellow spots represent genes that do not change their level of expression during sporulation. This image is enlarged; the microarray actually measures only 1.8 × 1.8 cm.

Yeast two-hybrid analysis probes molecular interactions in vivo.

- Microarrays can reveal expression patterns of genes that change in response to cellular stimuli, developmental stage, or conditions.

9.3 Genomics and the Human Story

DNA sequencing methods described in Chapter 8 led to the first complete genomic sequences of bacterial species in the 1990s. Two complete human genome sequences appeared in 2001. One of these was generated by a publicly funded effort led first by James Watson and later by Francis Collins. A parallel, private effort was led by Craig Venter. These accomplishments reflected more than a decade of intense effort coordinated in dozens of laboratories around the world, but they were just a beginning. DNA sequence databases now contain the complete genomic sequences of thousands of organisms of all types. We have only begun to effectively mine this vast informational resource.

Francis S. Collins

J. Craig Venter

Genomic Sequencing Is Aided by New Generations of DNA-Sequencing Methods

DNA-sequencing technologies continue to evolve. A complete human genome can now be sequenced in a day or two, a bacterial genome in a few hours. The day when a personal genomic sequence might be a routine part of one's medical record is within reach (Box 9–2). These advances have been made possible by approaches sometimes referred to as next-generation, or "next-gen," sequencing. The sequencing strategy is sometimes similar to and sometimes quite different from that used in the Sanger method described in Chapter 8. Innovations have allowed a miniaturization of the procedure, a massive increase in scale, and a corresponding decrease in cost.

A genomic sequence is generated in several steps. First, genomic sequences are broken at random locations by shearing to generate fragments that are a few hundred base pairs long. Synthetic oligonucleotides are ligated to the ends of all of the fragments, providing a known point of reference on every DNA molecule. The individual fragments are then immobilized on a solid surface, and each is amplified using PCR (Fig. 9–12). The solid surface is part of a channel that allows liquid solutions to flow over the samples. The result is a solid surface just a few centimeters across, with millions of DNA clusters attached, each cluster containing multiple copies of a single DNA sequence derived from one or another random genomic DNA fragment. The efficiency comes from sequencing all of these millions of clusters at the same time, with the data from each cluster captured and stored in a computer.

Two widely utilized next-generation sequencers use different strategies to accomplish the sequencing reactions. One of these, 454 sequencing (the numbers refer to a code used in the developmental phase of the technology and have no scientific meaning), uses a strategy called pyrosequencing in which the addition of nucleotides is detected with flashes of light **(Fig. 9–25)**. The four deoxynucleoside triphosphates (unaltered) are pulsed onto the reacting surface one at a time, in a repeating sequence. The nucleotide solution is retained on the surface just long enough for DNA polymerase to add that nucleotide to any cluster where it is complementary to the next template base in the sequence. Excess nucleotide is destroyed quickly with the enzyme apyrase before the next nucleotide pulse. When a specific nucleotide is successfully added to the strands of a cluster, pyrophosphate is released as a byproduct.

BOX 9–2 MEDICINE Personalized Genomic Medicine

When twins Noah and Alexis Beery were born in California, they exhibited symptoms that elicited a diagnosis of cerebral palsy. Treatments seemed to have no effect. Not satisfied with the diagnosis or the treatment, the twins' parents, Joe and Retta, took the twins, then age 5, to see a specialist in Michigan, who diagnosed them with a rare genetic condition called DOPA-responsive dystonia. A treatment regimen was devised that successfully suppressed the symptoms and allowed the twins to assume normal lives. However, at age 12, Alexis developed a severe cough and breathing difficulties that again seemed to threaten the child's survival. In one episode, paramedics had to revive her twice. The symptoms did not seem to be related to the dystonia. Might Noah be next? Frustrated and deeply worried, the twins' parents sought a complete genome sequence of both Noah and Alexis. This seemingly unusual step was a natural one for the Beery family. Joe was the chief information officer at Life Technologies, developers of sequencing technologies in use by many large DNA-sequencing centers. The cases of Noah and Alexis were taken up by Matthew Bainbridge and his team at the Baylor College of Medicine Human Genome Sequencing Center in Houston, Texas. The results proved decisive. The twins had mutations in their genomes that produced not only a deficiency in DOPA but also a potential deficiency in production of the hormone serotonin. A small adjustment in Alexis's therapy brought her life-threatening symptoms to an end, and the same therapy was given to her brother. Both siblings now lead normal lives.

The first draft human genome sequence was completed in 2001, after 12 years, at a cost of $3 billion. That cost has plummeted (Fig. 1), and newly completed human genomes are commonplace. The goal of a $1,000 human genome is on the horizon and promises to make this technology widely available. Since most genomic changes that affect human health are thought to be in protein-coding genes (an assumption that may be challenged in years to come), a cheaper alternative is simply to sequence the 1% of the genome that represents the coding regions (exons) of genes, or the **exome**.

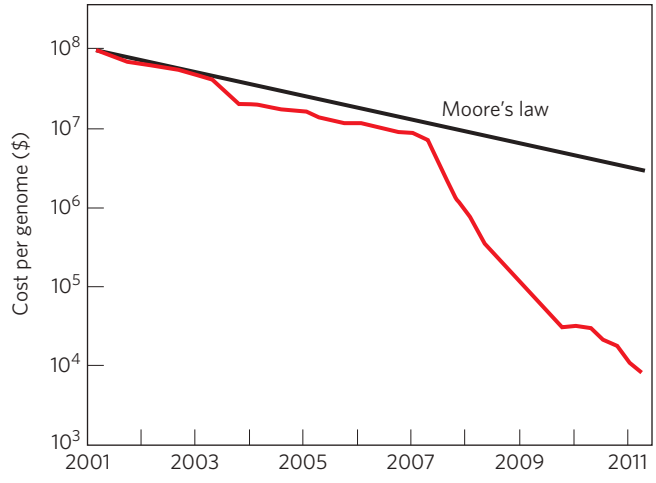


FIGURE 1 Since January 2008, the cost of human genome sequencing has been declining faster than the projected decline in the cost of processing data on computers (Moore's law).

That first human genome sequence came from a haploid genome, derived from a DNA amalgam from several different humans. Completion of a high-quality sequence in 2004 established this as the reference human genome. Subsequent completed human genome sequences, many from individual diploid genomes, have demonstrated how much individual genetic variation exists. Relative to the reference sequence, a typical human has about 3.5 million SNPs (about 60% of these are heterozygous, present on only one of two chromosomes) and another few hundred thousand differences in the form of small insertions and deletions and changes in repeat copy numbers. Only a small portion (5,000 to 10,000) of the SNPs affect the amino acid sequences of proteins encoded by genes.

This complexity ensures that, at least in the short term, successful diagnosis of a condition by whole genome sequencing will be the exception rather than the rule. However, few disciplines are advancing as rapidly as human genomics. The number of success stories is increasing rapidly as the technology becomes more widely available and the capacity of genomic analysis to recognize causative genetic changes improves.

Another enzyme in the solution bathing the surface is sulfurylase, which converts the pyrophosphate to ATP.

The appearance of ATP ultimately provides the signal that a nucleotide has been added to the DNA. Also present in the medium is an enzyme called luciferase and a substrate molecule, luciferin (luciferase is the enzyme that generates the flash of light produced by fireflies; see Box 13–1). When ATP is generated, luciferase catalyzes a reaction with luciferin that results in a tiny flash of light. When many tiny flashes occur in a cluster, the emitted light can be recorded in a captured image. For example, when dCTP is added to the solution, flashes will occur only at clusters where G is present in the template and C is the next nucleotide to be added to the growing DNA chain. If there is a string of 2, 3, or 4 G residues in the template, a similar number of C residues will be added to the growing strand in one cycle. This is recorded as a "flash" amplitude at that cluster that is 2, 3, or 4 times

greater than when only one C residue is added. Similarly, when dGTP is added, flashes occur at a different set of clusters, marking those as clusters where G is the next nucleotide added to the sequence. The length of DNA that can be reliably sequenced in a single cluster by this method—often referred to as the read length or "read"—is typically 400 to 500 nucleotides, as of early 2012, and is constantly improving.

The alternative method is the Illumina sequencer, which uses a technique known as reversible terminator sequencing **(Fig. 9–26)**. A special sequencing primer is added that is complementary to the oligonucleotides of known sequence that were ligated to the ends of the DNA fragments in each cluster (as described above). In addition, fluorescently labeled terminator nucleotides and DNA polymerase are added. The polymerase adds the appropriate nucleotide to the strands in each cluster, each type of nucleotide (A, T, G, or C) carrying a different fluorescent label. These terminator nucleotides have blocking groups attached to the 3′ ends that permit only one nucleotide addition to each strand. Next, lasers excite all the fluorescent labels, and an image of the entire surface reveals the color (and thus the identity of the base) added to each cluster. The fluorescent label and the blocking groups are then chemically or photolytically removed, in preparation for adding a new nucleotide to each cluster. The sequencing proceeds stepwise. Read lengths are shorter for this method, typically 100 to 200 nucleotides per cluster.

These technologies are modern manifestations of an approach to genomic sequencing that is sometimes called shotgun sequencing. Many copies of the genomic DNA are sheared to generate each set of fragments. Thus, a particular short segment of the genome may be present in dozens or even hundreds of different

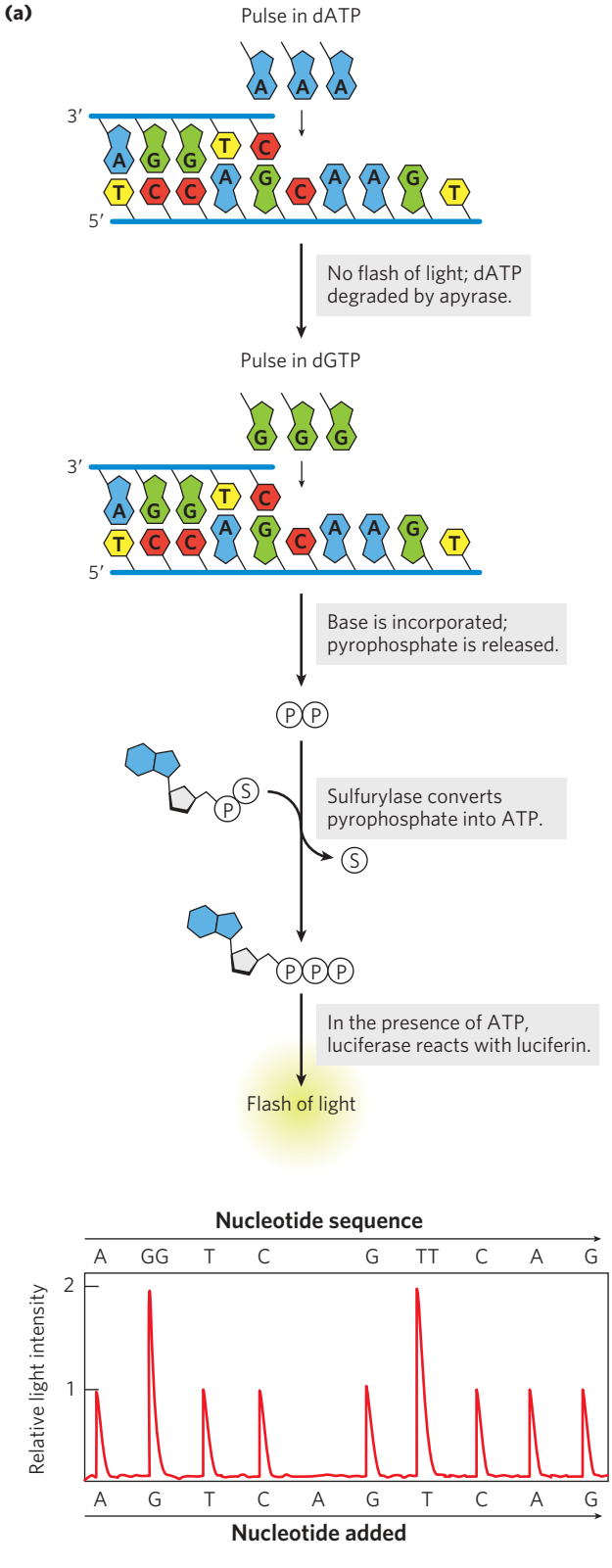


FIGURE 9–25 **Next-generation pyrosequencing. (a)** Pyrosequencing uses the enzymes sulfurylase (see Fig. 22–15) and luciferase (see Box 13–1) to detect nucleotide addition with flashes of light. The plot shows the light intensities observed during successive sequencing cycles for a DNA segment immobilized at a particular spot of a picotiter plate and (at top) the DNA nucleotide sequence derived from them. **(b)** An image of a very small part of one cycle of a 454 sequencing run. Each individual segment of DNA to be sequenced is attached to a tiny DNA capture bead, then amplified on the bead by PCR. Each bead is immersed in an emulsion and placed in a tiny (~29 μm) well on a picotiter plate. The reaction of luciferin and ATP with luciferase produces light flashes when a nucleotide is added to a particular DNA cluster in a particular well. Circles represent the same cluster over multiple cycles. In this case, reading the top (or bottom) circle from left to right across each row gives the sequence for that cluster.

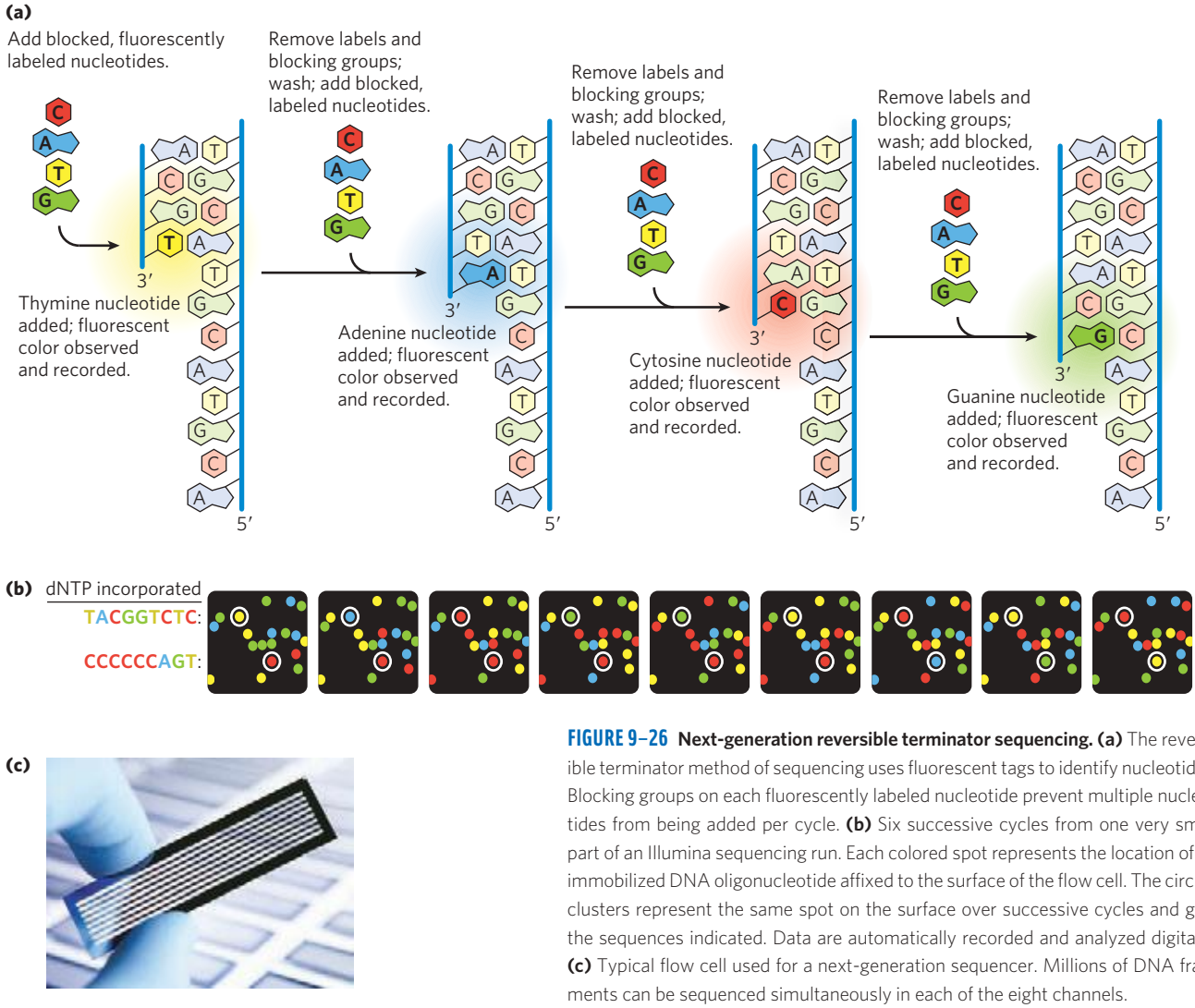


FIGURE 9–26 **Next-generation reversible terminator sequencing. (a)** The reversible terminator method of sequencing uses fluorescent tags to identify nucleotides. Blocking groups on each fluorescently labeled nucleotide prevent multiple nucleotides from being added per cycle. **(b)** Six successive cycles from one very small part of an Illumina sequencing run. Each colored spot represents the location of an immobilized DNA oligonucleotide affixed to the surface of the flow cell. The circled clusters represent the same spot on the surface over successive cycles and give the sequences indicated. Data are automatically recorded and analyzed digitally. **(c)** Typical flow cell used for a next-generation sequencer. Millions of DNA fragments can be sequenced simultaneously in each of the eight channels.

sequenced clusters. However, there is no landmark on an individual fragment to tell where in the genome it came from. Translating the sequences of these millions of fragments into a genomic sequence requires the computerized alignment of overlapping fragment sequences **(Fig. 9–27)**. The overlaps allow the computer to trace the sequence through a chromosome, from one fragment to another. This allows the assembly of long contiguous sequences called **contigs**. In a successful genomic sequencing exercise, many contigs can extend over millions of base pairs. Special strategies are needed to fill in the inevitable gaps and to deal with repetitive sequences.

The Human Genome Contains Genes and Many Other Types of Sequences

New and more-efficient DNA sequencing methods have produced an explosion of new genomic sequences. These are stored and made available in a range of publicly accessible databases. A good entrée to these resources can be found at the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov). The analysis of that growing wealth of genomic information is ongoing. What does our own genome, and its comparison with those of other organisms, tell us?

In some ways, we are not as complicated as we once imagined. Decades-old estimates that humans had about 100,000 genes within the approximately 3.2×10^9 bp of the human genome have been supplanted by the discovery that we have only about 25,000 protein-coding genes—less than twice the number in a fruit fly (13,600), not many more than in a nematode worm (20,000), and fewer than in a rice plant (38,000).

In other ways, however, we are more complex than we previously realized. The study of eukaryotic chromosome structure and genome sequences has revealed that many, if not most, eukaryotic genes contain one or more intervening segments of DNA that do not code for the amino acid sequence of the polypeptide product. These nontranslated inserts interrupt the otherwise colinear relationship between the gene's nucleotide

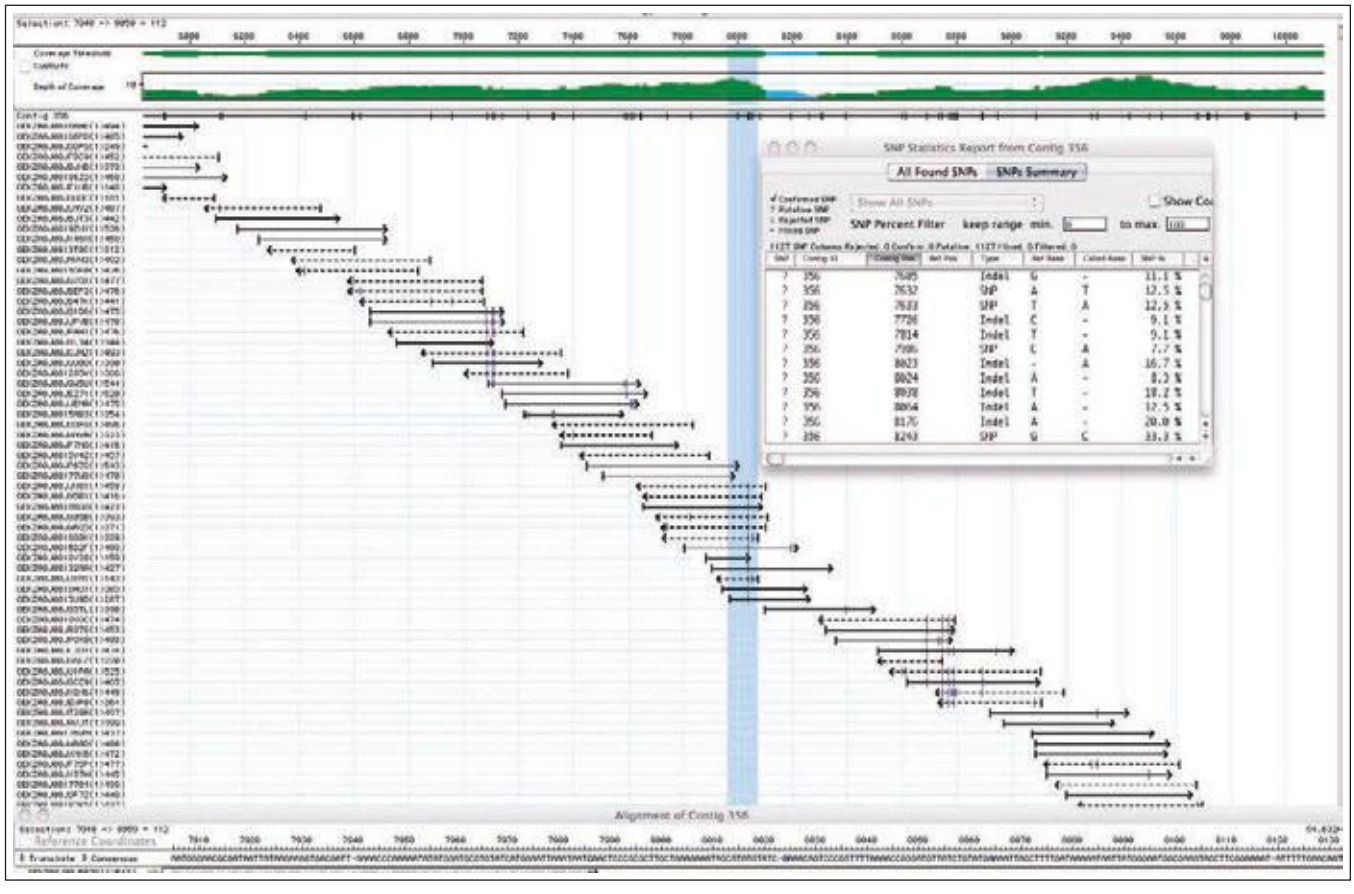


FIGURE 9–27 Sequence assembly. In a genomic sequence, each base pair of the genome is usually represented in several, often dozens, of the sequenced fragments, referred to as reads. Shown is a small part of the sequence of a new variant species of *E. coli*, with the reads generated by a 454 sequencer. The numbers at the top represent genomic base-pair positions, relative to an arbitrarily defined "0." The sequences all come from a particular long contig designated 356. The reads themselves are represented by horizontal arrows, with computer-assigned identifiers listed for each one at the left. DNA strand segments are sequenced at random, with sequences obtained from one strand (5′→3′, left to right) represented by solid arrows and sequences obtained from the other strand (5′→3′, right to left) represented by dashed lines. The latter sequences are automatically reported as their complement when they are merged with the overall dataset. The coverage threshold at the top is a measure of sequence quality. The wider green bar indicates sequences that have been obtained enough times to generate high confidence in the results. The depth of the coverage line indicates how many times a given base pair appears in a sequenced read. The vertical blue bar denotes a part of the sequence that is highlighted in the sequence line at the bottom of the figure. The SNP statistics report (inset) is a listing of positions where single nucleotide polymorphisms appear to be present in some of the reads. These putative SNPs are often checked by additional sequencing. They are indicated within the reads by thin, blue vertical slash marks within the horizontal lines for each read.

sequence and the amino acid sequence of the encoded polypeptide. Such nontranslated DNA segments are called **intervening sequences**, or **introns**, and the coding segments are called **exons** **(Fig. 9–28)**. Few bacterial genes contain introns. Introns are removed from a primary RNA transcript and the exons are spliced together to generate a transcript that can be translated contiguously into a protein product (see Chapter 26). An exon often (but not always) encodes a single domain of a larger, multidomain protein. Alternative modes of gene expression and RNA splicing permit the production of various combinations of exons, leading to the production of more than one protein from a single gene. Humans share many protein domain types with plants, worms, and flies, but we use these domains

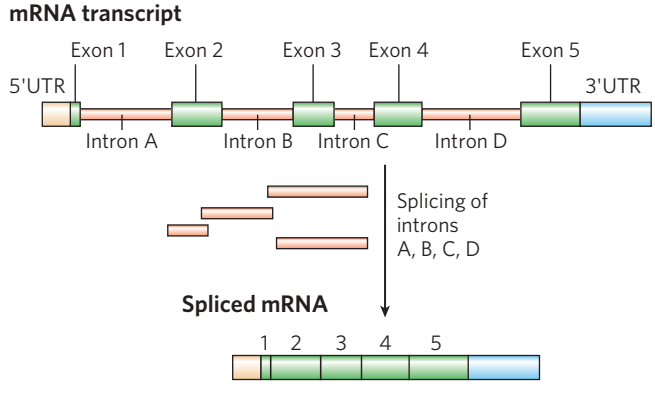


FIGURE 9–28 Introns and exons. This gene transcript contains five exons and four introns, along with 5′ and 3′ untranslated regions (5′UTR and 3′UTR). Splicing removes the introns to create an mRNA product for translation into protein.

in more complex arrangements and thereby generate more complex proteins.

In mammals and some other eukaryotes, the typical gene has a much higher proportion of intron DNA than exon DNA; in most cases, the function of introns is not clear. Only about 1.5% of human DNA is "coding" or exon DNA, carrying information for protein products **(Fig. 9–29a)**. However, when the much larger introns are included in the count, as much as 30% of the human genome consists of genes. Several efforts are underway to categorize the protein-coding genes by function (Fig. 9–29b).

The relative paucity of genes in the human genome leaves a lot of DNA unaccounted for. Much of the nongene DNA is in the form of repeated sequences of several kinds. Perhaps most surprising, about half the human genome is made up of moderately repeated sequences that are derived from transposable elements—segments of DNA, ranging from a few hundred to several thousand base pairs long, that can move from one location to another in the genome. Originally discovered in corn by Barbara McClintock, transposable elements, or **transposons**, are a kind of molecular parasite. They efficiently, and largely passively, make their home in the genomes of essentially every organism. Many transposons contain genes encoding the proteins that catalyze the transposition process itself, as described in more detail in Chapters 25 and 26. There are multiple classes of transposons in the human genome. Some are active, moving at a low frequency, but most are inactive, evolutionary relics altered by mutations.

Once the protein-coding genes (including exons and introns) and transposons are accounted for, perhaps 25% of the total DNA remains. The largest portion of this consists of unique sequences found between protein-coding genes. As described in Chapter 26, virtually all of these DNA segments are transcribed into RNA in at least some human cells. New classes of functional RNAs—encoded by genes whose existence was previously unsuspected—are being discovered at a rapid pace. Many genes encoding functional RNA are difficult to identify by automated methods, particularly when the RNA products have not been characterized. However, the RNA-coding genes are clearly a prominent feature of these otherwise uncharted genomic regions.

Another 3% or so of the human genome consists of highly repetitive sequences referred to as **simple-sequence repeats (SSRs)**. Generally less than 10 bp long, an SSR is sometimes repeated millions of times per cell and has identifiable functional importance in human cellular metabolism. The most prominent examples of SSR DNA occur in centromeres and telomeres (see Chapter 24). However, long repeats of simple sequences also occur throughout the genome.

What does all this information tell us about the similarities and differences among individual humans? Within the human population there are millions of single-base variations, called **single nucleotide polymorphisms**, or **SNPs** (pronounced "snips"). Each human differs from the next by, on average, 1 in every 1,000 bp. Many of these variations are in the form of SNPs, but a wide range of larger deletions, insertions, and small rearrangements occur in the human population as well. From these often subtle genetic differences comes the human variety we are all aware of—differences in hair color, stature, eyesight, allergies, foot size, and (to some unknown degree) behavior.

The process of genetic recombination and chromosomal segregation during meiosis tends to mix and match these small genetic variations so that different combinations of genes are inherited (see Chapter 25). When two such genetic variations are on different chromosomes, the particular variants that might be inherited by a given individual are a result of random chance. If the genetic variants lie on the same chromosome, the chance that they will be inherited together is an inverse function of the distance between them. Groups of SNPs and other genetic differences that are close together on the same chromosome are rarely affected by recombination and are usually inherited together; these groupings are known as **haplotypes**. Haplotypes provide

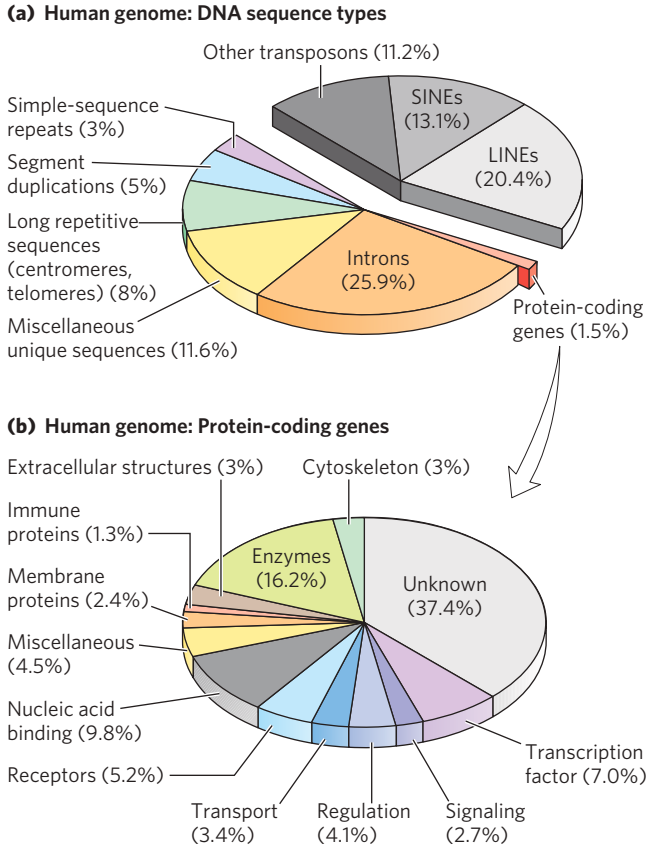


FIGURE 9–29 A snapshot of the human genome. (a) This pie chart shows the proportions of various types of sequences in our genome. Transposons, including SINEs and LINEs, are described in Chapters 25 and 26. **(b)** The approximately 25,000 protein-coding genes in the human genome can be classified by the type of protein encoded.

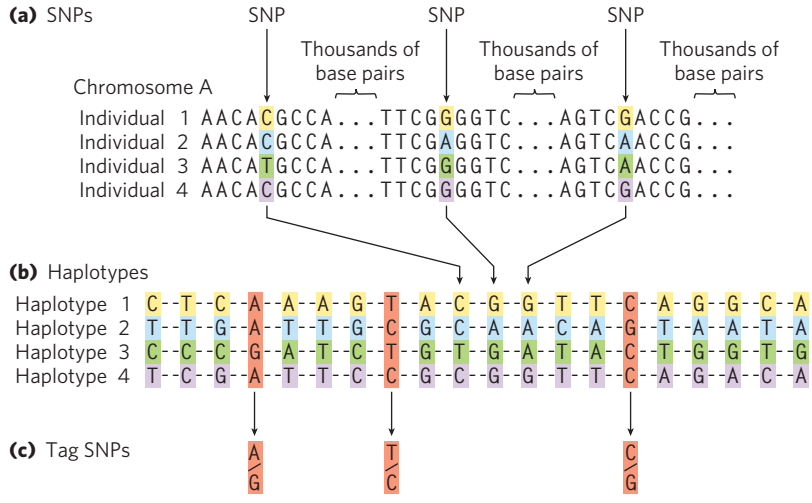


FIGURE 9–30 Haplotype identification. (a) Single nucleotide polymorphisms (SNPs), or positions in the human genome where the sequence varies from one individual to another, are identified in genomic samples, and **(b)** groups of SNPs that are relatively close to each other on a chromosome (within a few tens of thousands of base pairs) are compiled into a haplotype. The SNPs will vary in the overall human population, such as in the four fictitious individuals shown in panels (a) and (b). However, the SNPs chosen to define a haplotype are often the same in most individuals of a particular population. **(c)** Haplotype-defining SNPs (tag SNPs) can be used to simplify the process of identifying an individual's haplotype (by sequencing 3 defining SNPs instead of all 20). If the tag SNPs shown in light red are sequenced, the nucleotides A, T, and G at the three successive tag SNP positions might be characteristic of a population native to one location in northern Europe, whereas the nucleotides G, T, and C at these same three genomic positions might be found in a population in Asia. Multiple haplotypes of this kind are used to trace prehistoric human migrations (see Fig. 9–35).

convenient markers for certain human populations and individuals within populations.

Defining a haplotype requires several steps. First, positions that contain SNPs in the human population are identified in genomic DNA samples from multiple individuals **(Fig. 9–30a)**. Each SNP may be separated from the next by many thousands of base pairs. Second, SNPs that are relatively close to each other on a chromosome and thus usually inherited together are compiled into haplotypes (Fig. 9–30b). Each haplotype consists of the particular bases found at the various SNP positions in the defined haplotype. Finally, tag SNPs—a subset of the SNPs that define the entire haplotype—are chosen to uniquely identify each haplotype (Fig. 9–30c). By sequencing just these tag positions in genomic samples from human populations, researchers can quickly identify which haplotypes are present in each individual. Especially stable haplotypes exist in the mitochondrial genome (which, being inherited maternally, never undergoes meiotic recombination) and on the male Y chromosome (only 3% of which is homologous to the X chromosome and thus subject to recombination).

Genome Sequencing Informs Us about Our Humanity

A primary purpose of most genome sequencing projects is to identify conserved genetic elements of functional significance, such as conserved exon sequences, regulatory regions, and other genomic features such as centromeres and telomeres. One major purpose of sequencing the human genome is in identifying the differences between our genome and those of other organisms. Although the human genome is very closely related to other mammalian genomes over large segments of every chromosome, differences of a few percent in billions of base pairs add up to millions of genetic distinctions. Searching among these differences using comparative genomics techniques can reveal the molecular basis of human genetic diseases and can help identify genes, gene alterations, and other genomic features that are uniquely human and thus likely to contribute to definably human characteristics such as our large brain, language skills, tool-making ability, or bipedalism.

The genome sequences of our closest biological relatives, the chimpanzees and bonobos, offer some important clues and can illustrate the comparative process. Humans and chimpanzees shared a common ancestor about 7 million years ago. Genomic differences between the two species fall into two types: base-pair changes (SNPs) and larger genomic rearrangements of many types. SNPs in the protein-coding regions often result in amino acid changes that can be used to construct a phylogenetic tree **(Fig. 9–31a)**. Segments of chromosomes may become inverted over evolutionary time. The processes leading to such inversions are complicated and rare, but the human lineage has long segments of DNA that are inverted (relative to other primates) due to this process on chromosomes 1, 12, 15,

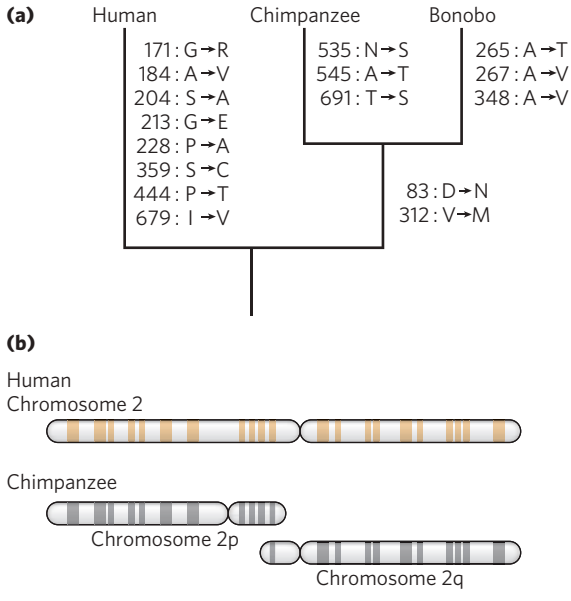


FIGURE 9–31 Genomic alterations in the human lineage. (a) This evolutionary tree for primates is derived from sequences for the progesterone receptor, which helps regulate many events in reproduction. The gene encoding this protein has undergone more evolutionary alterations than most. Amino acid changes associated uniquely with humans, chimpanzees, and bonobos are listed by residue number beside each branch. **(b)** The genes on chimpanzee chromosomes 2p and 2q are homologous to those on human chromosome 2, implying that two chromosomes fused into one at some point in the line leading to humans.

16, and 18. Chromosome fusions can also occur. In the human lineage, two chromosomes found in other primate lineages have been fused to form the single human chromosome 2 (Fig. 9–31b). The human lineage thus has 23 chromosome pairs rather than the 24 pairs typical of other primates. Once this fusion appeared in the line leading to humans, it would have represented a major barrier to interbreeding with other primates that lacked it.

If we ignore transposons and large chromosomal rearrangements, the published human and chimpanzee genomes differ by only 1.23% at the level of base pairs (compared with the 0.1% variance from one human to another). If we further ignore variations at positions where there is a known polymorphism in either the human or the chimpanzee population (these are unlikely to reflect a species-defining evolutionary change), the differences amount to about 1.06%, or about 1 in 100 bp. This small percentage translates into more than 30 million base-pair changes, some of which affect protein function and gene regulation. The genome rearrangements that help distinguish chimpanzees and humans include 5 million short insertions or deletions involving a few base pairs each, as well as a substantial number of larger insertions, deletions, inversions, or duplications that can involve many thousands of base pairs. When transposon insertions—a major source of genomic variance—are included, the differences between the human and chimpanzee genomes increase by about 90 million bp, representing another 3% of these genomes. In effect, each species has segments of DNA, constituting 40 to 45 million bp, that are entirely unique to that particular genome, with larger chromosomal insertions, duplications, and other rearrangements affecting more base pairs than do single-nucleotide changes. Thus, the total genomic difference between chimpanzees and humans amounts to about 4% of their genomes.

Sorting out which genomic distinctions are relevant to features that are uniquely human is a daunting task. If the two species share a common ancestor, then, assuming a similar rate of evolution in both lines, half the changes represent chimpanzee lineage changes and half represent human lineage changes. When you see a difference, how do you tell which variant was the one present in the common ancestor? One way is to compare both genome sequences with those of more distantly related organisms referred to as **outgroups**. Consider a locus, X, where there is a difference between the human and chimpanzee genomes **(Fig. 9–32)**. The lineage of the orangutan, an outgroup, diverged from that of chimps and humans prior to the chimpanzee-human common ancestor. If the sequence at locus X is identical in orangutans and chimpanzees, this sequence was probably present in the chimpanzee-human ancestor, and the sequence seen in humans is specific to the human lineage. Sequences that are identical in humans and orangutans can be eliminated as candidates for

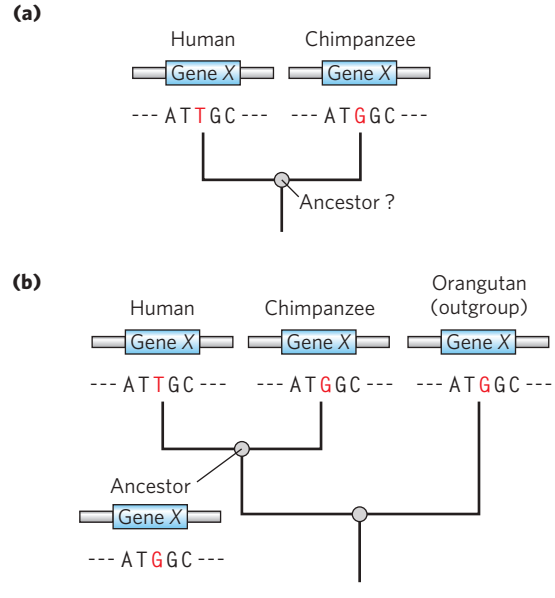


FIGURE 9–32 Determination of sequence alterations unique to one ancestral line. (a) Sequences from the same hypothetical gene in humans and chimpanzees are compared. The sequence of this gene in their last common ancestor is unknown. **(b)** The orangutan genome is used as an outgroup. The sequence of the orangutan gene is identical to the chimpanzee gene. This means that the mutation causing the difference between humans and chimpanzees almost certainly occurred in the line leading to modern humans, and the common ancestor of humans and chimpanzees (and orangutans) had the sequence now found in chimpanzees.

human-specific genomic features. The importance of comparisons with closely related outgroups has given rise to new efforts to sequence the genomes of orangutans, macaques, and many other primate species.

The search for the genetic underpinnings of special human characteristics, such as our enhanced brain function, can benefit from two complementary approaches. The first searches for genomic regions showing extreme changes from other primates. These could include multiple gene duplications or the addition of large genomic segments not present in other primates. The second approach looks at genes known to be involved in relevant human conditions. For brain function, for example, one would examine genes that, when mutated, contribute to cognitive or other mental disorders.

Notably, analyses of the human lineage have not detected an enrichment of genetic changes in protein-coding genes involved in brain development or size. In primates, most genes that function uniquely in the brain are even more highly conserved than genes functioning in other tissues. However, some differences in gene expression are observed. When changes in genomic regions related to gene regulation are analyzed, genes involved in neural development and nutrition are disproportionately affected. A variety of RNA-coding genes, some with expression concentrated in the brain, also show evidence of accelerated evolution **(Fig. 9–33)**. The many new classes of RNA that are being discovered (see Chapter 26) are likely to radically change our perspective on how evolution alters the workings of living systems. It is increasingly evident that which genes are expressed may not be as important as when, where, and how much they are expressed.

Genome Comparisons Help Locate Genes Involved in Disease

The Human Genome Project has fulfilled its potential for accelerating the discovery of genes underlying genetic diseases: over 1,600 human genetic diseases have now been mapped to particular genes. With use of a method called **linkage analysis**, the gene involved in a disease condition is mapped relative to well-characterized genetic polymorphisms that occur throughout the human genome. The search often begins with one or more large families that include several individuals affected by a particular disease over several generations. The most common approach is basically an exercise in phylogenetics (the study of evolutionary relatedness between groups of organisms) and is deeply rooted in concepts derived from evolutionary biology. We can illustrate by describing the search for one gene involved in Alzheimer disease. About 10% of all cases of this condition in the United States result from an inherited predisposition. Several different genes have been discovered that, when mutated, can lead to early onset of Alzheimer disease. One such gene (*PS1*) encodes the protein presenilin-1, and its discovery made heavy use of linkage analysis.

Just as haplotypes rely on SNPs that lie close together on a chromosome, linkage analysis involves a search for SNPs that lie close to a gene of interest. In studies of this type, researchers focus on families affected by the disease, and look for families in which DNA samples can be obtained from individuals of multiple generations. DNA samples are collected from both affected and unaffected family members. Researchers first localize the region associated with the disease to a specific chromosome, using sets (called panels) of genomic locations where common SNPs or other mapped genomic alterations occur in a significant proportion of the human population. Thus, many but not all humans will differ in the genomic sequence at these locations. Using a panel that includes several well-characterized SNP loci mapped to each chromosome, investigators compare the genotypes of individuals with and without the disease, focusing especially on close family members. For Alzheimer disease, two of the many family pedigrees used to search for this gene in the early 1990s are shown in **Figure 9–34**. Using such pedigrees, researchers look for the particular SNP variants that are inherited in the same or nearly the same pattern as the disease-causing gene. The responsible gene can gradually be localized to a single chromosome, if the inheritance of particular SNP variants on that chromosome is mirrored closely by the inheritance of the disease condition. More-detailed localization of a disease-causing gene on a chromosome relies on statistical methods to correlate the inheritance of additional, more closely spaced polymorphisms with the occurrence of the disease, focusing on a denser panel of polymorphisms known to occur on the chromosome of interest. The more closely a marker is located to a disease gene, the more likely it is

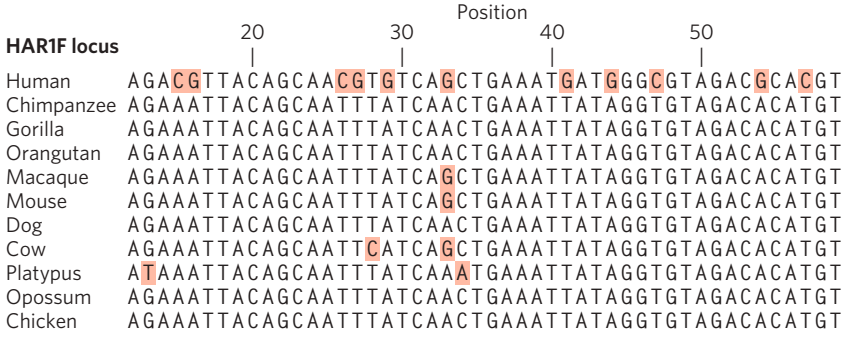


FIGURE 9–33 Accelerated evolution in some human genes. The HAR1F locus specifies a noncoding RNA that is highly conserved in vertebrates. In humans, the HAR1F gene exhibits an unusual number of mutations (shaded light red), providing evidence of accelerated evolution.

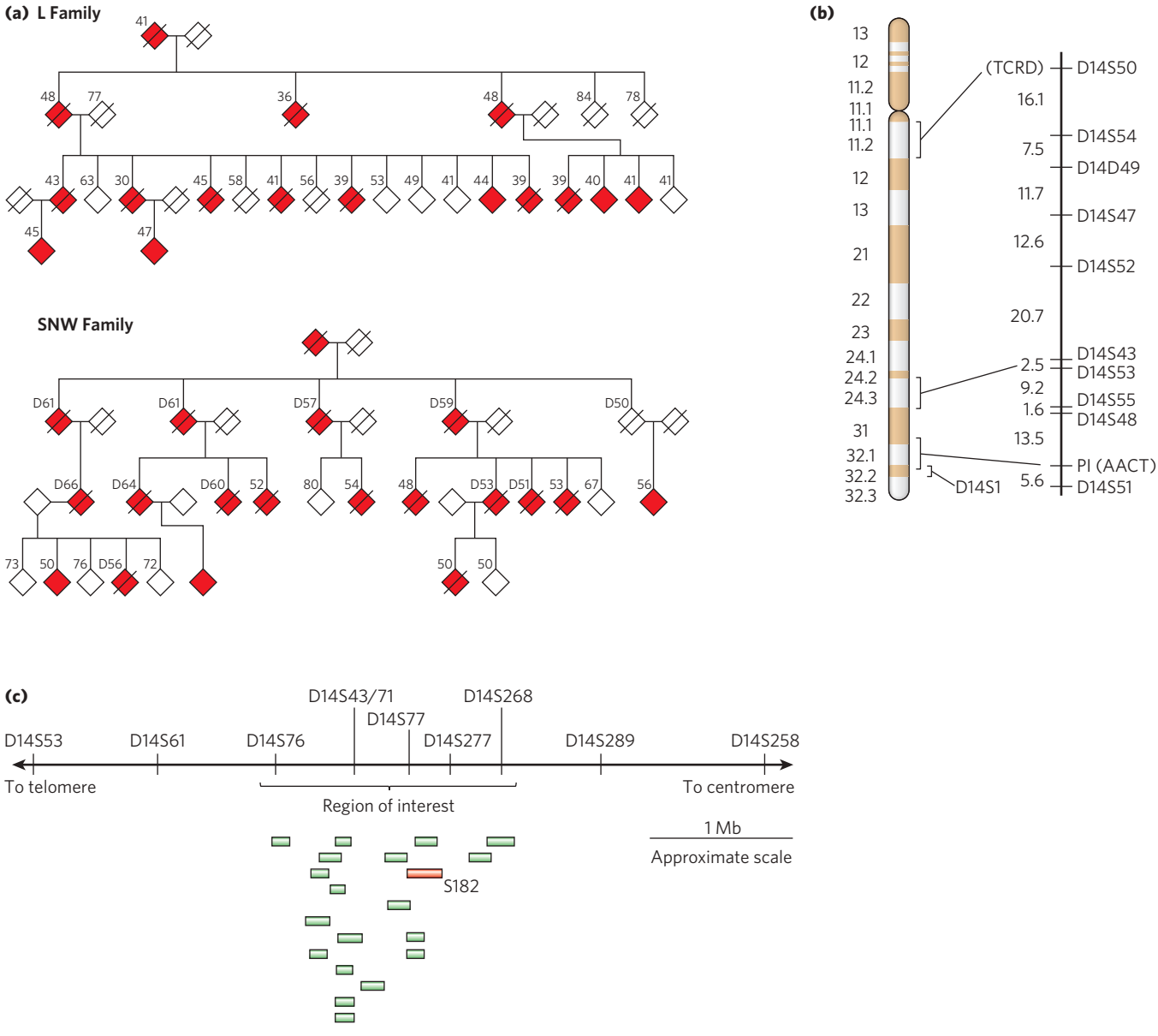


FIGURE 9–34 Linkage analysis in the discovery of disease genes. **(a)** These pedigrees for two families affected by early-onset Alzheimer disease are based on the data available at the time of the study. Red symbols represent affected individuals; slashes indicate deaths either before or soon after the study. The number above each symbol is the person's age, either at the time of the study or at time of death (indicated with a D). To protect family privacy, gender is not indicated. **(b)** Chromosome 14, with bands created by certain dyes. Chromosome marker positions are shown at the right, with the genetic distance between them presented in a genetic distance measurement called centimorgans, reflecting the frequency of recombination between them. *TCRD* (T-cell receptor delta) and *PI* (AACT (α1-antichymotrypsin)) are genes with alterations in the human population that were used as markers, along with SNPs, in chromosome mapping. **(c)** By comparisons of DNA from affected and unaffected family members, a region of interest that contains 19 expressed genes was eventually defined near marker D14S43. The gene labeled *S182* (red) encodes presenilin-1.

to be inherited along with that gene. This process can point to a region of the chromosome that contains the gene. In this example of Alzheimer disease, linkage analysis indicated that the disease-causing gene was somewhere near a SNP locus called D14S43 (Fig. 9–34c).

The final steps of the search for the disease gene again use the human genome databases. The local region containing the gene is examined and the genes within it are identified. DNA from many individuals, some who have the disease and some who do not, is sequenced over this region. This process gradually leads, with an increasing number of individuals analyzed, to the identification of gene variants consistently present in individuals with the disease state and not in unaffected individuals. The search can be aided by an understanding of the function of the genes in the target region, because particular metabolic pathways may be more likely than others to produce the disease state. In 1995, the chromosome 14 gene associated with Alzheimer disease was identified as gene *S182*. The product of this gene was given the name presenilin-1, and the gene itself was subsequently renamed *PS1*.

More complex are cases where a disease condition is caused by the presence of mutations in two different genes (neither of which, alone, causes the disease), or where a particular condition is enhanced by an otherwise innocuous mutation in another gene. Identifying the genes and mutations responsible for such digenic diseases is exceedingly difficult, and these diseases are sometimes possible to document only within small, isolated, and highly inbred populations.

Genome databases open alternative paths to the identification of disease genes, especially when biochemical information about the disease is known. In the case of Alzheimer disease, an accumulation of the amyloid β-protein in limbic and association cortices of the brain is at least partly responsible for the symptoms. Defects in presenilin-1 (and in a related protein, presenilin-2, encoded by a gene on chromosome 1) lead to elevated cortical levels of amyloid β-protein. Focused databases catalog such functional information on the protein products of genes and on protein-interaction networks, SNP locations, and other data, providing a streamlined path to the identification of candidate genes for a particular disease. A researcher with some knowledge about the kinds of enzymes or other proteins likely to contribute to a disease can use these databases to generate a list of genes known to encode proteins with relevant functions, additional uncharacterized genes with orthologous or paralogous relationships to the genes in this list, a list of proteins known to interact with the target proteins or orthologs in other organisms, and a map of gene positions. With data from selected family pedigrees, a short list of potentially relevant genes can often be determined rapidly.

These approaches are not limited to human diseases. The same methods can be used to identify the genes involved in diseases—or genes that produce desirable characteristics—in other animals and in plants. ■

Genome Sequences Inform Us about Our Past and Provide Opportunities for the Future

About 70,000 years ago, a small group of humans in Africa looked out across the Red Sea to Asia. Perhaps encouraged by some innovation in small boat construction, or driven by conflict or famine, or simply curious, they crossed the water barrier. That initial colonization, involving maybe 1,000 individuals, began a journey that did not stop until humans reached Tierra del Fuego (at the southern tip of South America) many thousands of years later. In the process, an established population from a previous hominid expansion into Eurasia, including *Homo neanderthalensis*, was displaced. The Neanderthals disappeared, just as *Homo erectus* and other hominid lines disappeared before them.

The story of how modern humans first appeared in Africa a few hundred thousand years ago, and their migrations as they eventually radiated out of Africa, is written in our DNA. Through the use of genomic sequences from multiple species, both primate and hominid evolution have come into sharper focus. Using haplotypes present in extant human populations, we can trace the migrations of our intrepid ancestors across the planet **(Fig. 9–35)**.

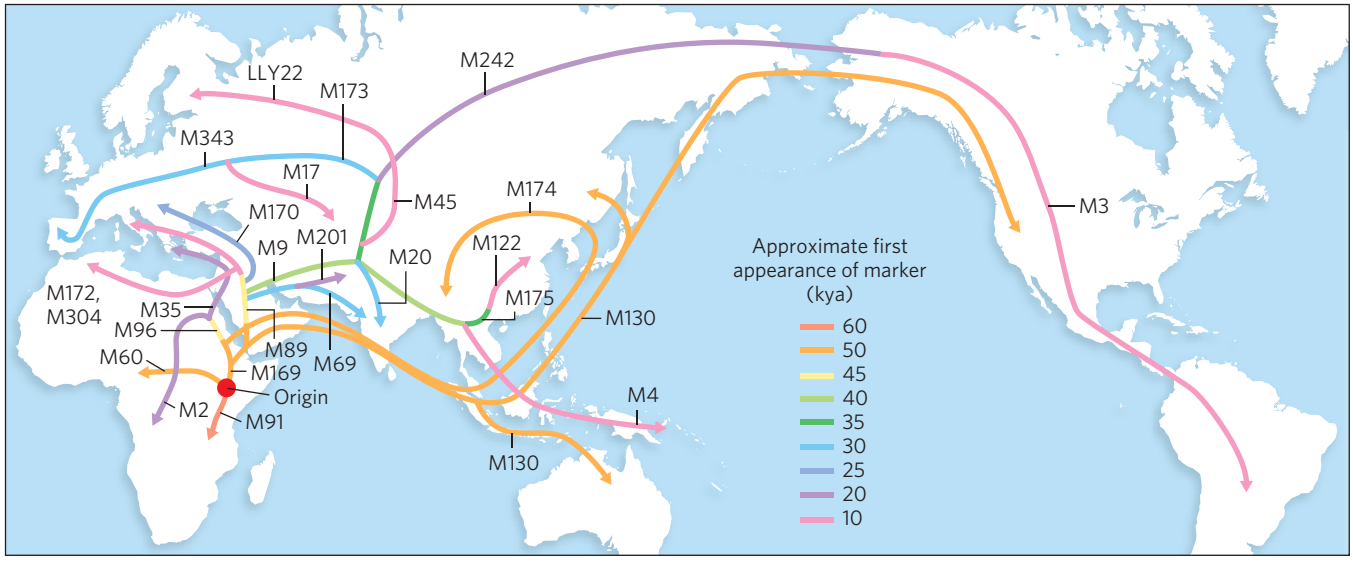


FIGURE 9–35 The paths of human migrations. When a small part of a human population migrates away from a larger group, it takes only part of the overall genetic diversity with it. Thus, some haplotypes are present in the migrating group but many are not. At the same time, mutations can create novel haplotypes over time. This map was generated from an analysis of genetic markers (defined haplotypes with M or LLY numbers) on the Y chromosome. The genetic samples were taken from indigenous populations long established at geographic points along the routes shown. Haplotypes that appear suddenly along a migration path, reflecting new changes (mutations) in particular SNP genomic locations in certain isolated populations, are called "founder events." These enable researchers to trace migrations from that point, as other populations that possess the new haplotype were probably descended from the founder population. The abbreviation *kya* means "thousand years ago."

BOX 9–3 Getting to Know the Neanderthals

Modern humans and Neanderthals coexisted in Europe and Asia as recently as 30,000 years ago. The human and Neanderthal ancestral populations diverged about 370,000 years ago, before the appearance of anatomically modern humans. Neanderthals used tools, lived in small groups, and buried their dead. Of the known hominid relatives of modern humans, Neanderthals are the closest. For hundreds of millennia, they inhabited large parts of Europe and western Asia (Fig. 1). If the chimpanzee genome can tell us something about what it is to be human, perhaps the Neanderthal genome can tell us more. Buried in the bones and remains taken from burial sites are fragments of Neanderthal genomic DNA. Technologies developed for use in forensic science (see Box 9–1) and ancient DNA studies have been combined to initiate a Neanderthal genome project.

This endeavor is unlike the genome projects aimed at extant species. The Neanderthal DNA is present in small amounts, and it is contaminated with DNA from other animals and bacteria. How does one get at it, and how can one be certain that the sequences really came from Neanderthals? The answers have been revealed by innovative applications of biotechnology. In essence, the small quantities of DNA fragments found in a Neanderthal bone or other remains are cloned into a library, and the cloned DNA segments are sequenced at random, contaminants and all. The sequencing results are compared with the existing human genome and chimpanzee genome databases. Segments derived from Neanderthal DNA are readily distinguished from segments derived from bacteria or insects by computerized analysis, because they have sequences closely related to human and chimpanzee DNA. Once a collection of Neanderthal DNA segments is sequenced, they can be used as probes to identify sequence fragments in ancient samples that overlap with these known fragments.

FIGURE 1 Neanderthals occupied much of Europe and western Asia until about 30,000 years ago. Major Neanderthal archaeological sites are shown here. (Note that the group was named for the site at Neanderthal in Germany.)

The Neanderthals were not simply displaced. Some mingling occurred. Using sensitive PCR-based methods, we now have a nearly complete sequence of the Neanderthal genome (Box 9–3). We know that about 4% of the human genomes of non-Africans are derived from Neanderthals. Some human populations also acquired genomic DNA from another recently discovered group, the Denisovans. Neanderthal DNA gave humans a more complex immune system, making us more resistant to infection but also a little more susceptible to autoimmune diseases. The story of our past is gradually taking shape as more human genomes, of those alive today and those who lived in past millennia, are being assembled.

The medical promise of personal genomic sequences grows as more genes underlying inherited diseases are defined. Knowledge of genomic sequences also provides the prospect of altering them. It is now commonplace to engineer the DNA sequences of organisms from bacteria and yeast to plants and mammals for research and commercial purposes. Efforts to cure inherited human diseases by human gene therapy have not yet lived up

The potential problem of contamination with the closely related modern human DNA can be controlled for by examining mitochondrial DNA. Human populations have readily identifiable haplotypes (distinctive sets of genomic differences; see Fig. 9–30) in their mitochondrial DNA, and analysis of Neanderthal samples has shown that Neanderthals' mitochondrial DNA has its own distinct haplotypes. The presence in the Neanderthal samples of some base-pair differences that are found in the chimpanzee database but not in the human database is more evidence that non-human hominid sequences are being found.

Completion of this challenging endeavor is on the horizon. The draft sequence of the Neanderthal genome unveiled in early 2009 covered more than 60% of the genomic sequences. A finished sequence will require just a little more time. The data provide evidence that modern humans and the Neanderthals who were the source of this DNA shared a common ancestor about 700,000 years ago (Fig. 2). Analysis of mitochondrial DNA suggests that the two groups continued on the same track, with some gene flow between them, for about 300,000 more years. The lines split with the appearance of anatomically modern humans, although evidence now exists for some intermingling of the lines somewhat later.

Expanded libraries of Neanderthal DNA from different sets of remains should eventually allow an analysis of Neanderthal genetic diversity, and perhaps Neanderthal migrations, providing a fascinating look at our hominid past.

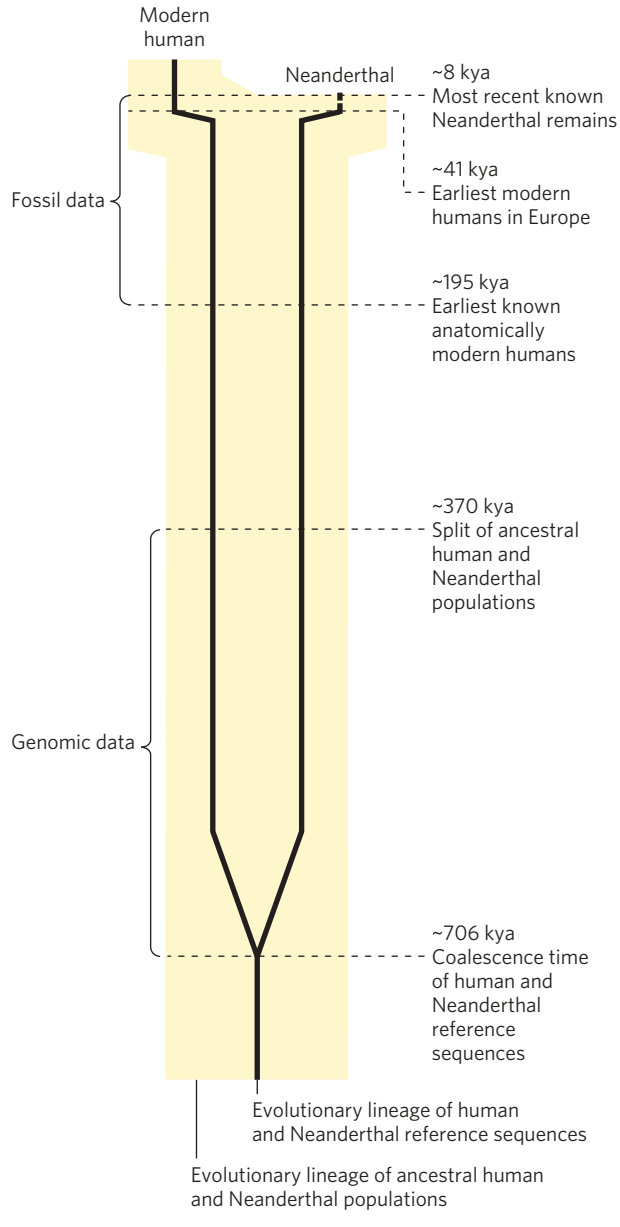


FIGURE 2 This timeline shows the divergence of human and Neanderthal genome sequences (black lines) and of ancestral human and Neanderthal populations (yellow screen). Genomic data provide evidence for some intermingling of the populations up to about 45,000 years ago. Key events in human evolution are noted.

to their potential, but technologies for gene delivery are constantly being improved. Few scientific disciplines will affect the future of our species more than modern genomics.

SUMMARY 9.3 Genomics and the Human Story

- Next-generation sequencing methods have vastly reduced the time required to generate complete genomic sequences.
- About 30% of the DNA in the human genome is in the exons and introns of genes encoding proteins. Nearly half of the DNA is derived from parasitic transposons. Much of the rest encodes RNAs of many types. Simple-sequence repeats make up the centromere and telomeres.
- The gene alterations that define humanity can be discerned in part through comparative genomics using other primates.
- Comparative genomics is also used to locate the gene alterations that define inherited diseases and can be used to study the evolution and migration of our human ancestors over millennia.

Key Terms

Terms in bold are defined in the glossary.

genomics 313
systems biology 313
cloning 314
vector 314
recombinant DNA 314
genetic engineering 314
restriction endonucleases 314
DNA ligases 314
plasmid 317
bacterial artificial chromosome (BAC) 319
yeast artificial chromosome (YAC) 320
expression vector 321
baculovirus 323
bacmid 323
site-directed mutagenesis 323
fusion protein 325
tag 325
polymerase chain reaction (PCR) 327
short tandem repeat (STR) 329
quantitative PCR (qPCR) 331
DNA library 332
genomic library 332
complementary DNA (cDNA) 332
cDNA library 332
comparative genomics 333
orthologs 333
paralogs 333
synteny 333
epitope tag 333
yeast two-hybrid analysis 336
DNA microarray 337
contig 342
single nucleotide polymorphism (SNP) 344
haplotype 344

Further Reading

General

Jackson, D.A., Symons, R.H., & 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.

The first recombinant DNA experiment linking DNA from two species.

Lobban, P.E. & Kaiser, A.D. (1973) Enzymatic end-to-end joining of DNA molecules. *J. Mol. Biol.* **78**, 453–471.

Report of the first recombinant DNA experiment.

Studying Genes and Their Products

Arnheim, N. & Erlich, H. (1992) Polymerase chain reaction strategy. *Annu. Rev. Biochem.* **61**, 131–156.

Foster, E.A., Jobling, M.A., Taylor, P.G., Donnelly, P., de Knijff, P., Mieremet, R., Zerjal, T., & Tyler-Smith, C. (1999) The Thomas Jefferson paternity case. *Nature* **397**, 32.

Last article of a series in an interesting case study of the uses of biotechnology to address historical questions.

Giepmans, B.N.G., Adams, S.R., Ellisman, M.H., & Tsien, R.Y. (2006) The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217–224.

Kayser, M. & de Knijff, P. (2011) Improving human forensics through advances in genetics, genomics, and molecular biology. *Nat. Rev. Genet.* **12**, 179–192.

Ståhl, P.L. & Lundeber, J. (2012) Toward the single-hour high-quality genome. *Annu. Rev. Biochem.* **81**, 359–378.

Zhao, J. & Grant, S.F.A. (2011) Advances in whole genome sequencing technology. *Curr. Pharm. Biotechnol.* **12**, 293–305.

Using DNA-Based Methods to Understand Protein Function

Budowle, B., Johnson, M.D., Fraser, C.M., Leighton, T.J., Murch, R.S., & Chakraborty, R. (2005) Genetic analysis and attribution of microbial forensics evidence. *Crit. Rev. Microbiol.* **31**, 233–254.

How biotechnology is used to fight bioterrorism.

Koonin, E.V. (2005) Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **39**, 309–338.

Good description of some comparative genomics basics.

Stoughton, R.B. (2005) Applications of DNA microarrays in biology. *Annu. Rev. Biochem.* **74**, 53–82.

Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *App. Microbiol. Biotechnol.* **72**, 211–222.

Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., Eisen, J.A., Heidelberg, K.B., Manning, G., Li, W., et al. (2007) The *Sorcerer II* global ocean sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**, e16.

One of a series of articles arising from an ambitious effort to sample microbial biodiversity throughout the world's oceans.

Genomics and the Human Story

Alkan, C., Sajjadian, S., & Eichler, E.E. (2011) Limitations of next-generation genome sequence assembly. *Nat. Methods.* **8**, 61–65.

Callaway, E. (2011) Ancient DNA reveals secrets of human history. *Nature* **476**, 136–137.

Carr, P.A. & Church, G.M. (2009) Genome engineering. *Nat. Biotechnol.* **27**, 1151–1162.

Carroll, S.B. (2003) Genetics and the making of *Homo sapiens*. *Nature* **422**, 849–857.

Gonzaga-Jauregui, C., Lupski, J.R., & Gibbs, R.A. (2012) Human genome sequencing in health and disease. *Annu. Rev. Med.* **63**, 35–61.

Green, E.D. & Guyer, M.S. (2011) Charting a course for genomic medicine from base pairs to bedside. *Nature* **470**, 204–213.

Kay, M.A. (2011) State-of-the-art gene-based therapies: the road ahead. *Nat. Rev. Genet.* **12**, 316–328.

Lander, E.S. (2011) Initial impact of the sequencing of the human genome. *Nature* **470**, 187–197.

Metzker, M.L. (2010) Sequencing technologies: the next generation. *Nat. Rev. Genet.* **11**, 31–46.

Perkel, J.M. (2011) Synthetic genomics: building a better bacterium. *Science* **331**, 1628–1630.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L., et al. (2010) Genetic history of an archaic hominin group from Denisova cave in Siberia. *Nature* **468**, 1053–1060.

Neanderthals were not the only other hominid group that humans encountered in their trek across Asia.

Stoneking, M. & Krause, J. (2011) Learning about human population history from ancient and modern genomes. *Nat. Rev. Genet.* **12**, 603–614.

Problems

1. Engineering Cloned DNA When joining two or more DNA fragments, a researcher can adjust the sequence at the junction in a variety of subtle ways, as seen in the following exercises.

(a) Draw the structure of each end of a linear DNA fragment produced by an *Eco*RI restriction digest (include those sequences remaining from the *Eco*RI recognition sequence).

(b) Draw the structure resulting from the reaction of this end sequence with DNA polymerase I and the four deoxynucleoside triphosphates (see Fig. 8–33).

(c) Draw the sequence produced at the junction that arises if two ends with the structure derived in (b) are ligated (see Fig. 25–16).

(d) Draw the structure produced if the structure derived in (a) is treated with a nuclease that degrades only single-stranded DNA.

(e) Draw the sequence of the junction produced if an end with structure (b) is ligated to an end with structure (d).

(f) Draw the structure of the end of a linear DNA fragment that was produced by a *Pvu*II restriction digest (include those sequences remaining from the *Pvu*II recognition sequence).

(g) Draw the sequence of the junction produced if an end with structure (b) is ligated to an end with structure (f).

(h) Suppose you can synthesize a short duplex DNA fragment with any sequence you desire. With this synthetic fragment and the procedures described in (a) through (g), design a protocol that would remove an *Eco*RI restriction site from a DNA molecule and incorporate a new *Bam*HI restriction site at approximately the same location. (See Fig. 9–2.)

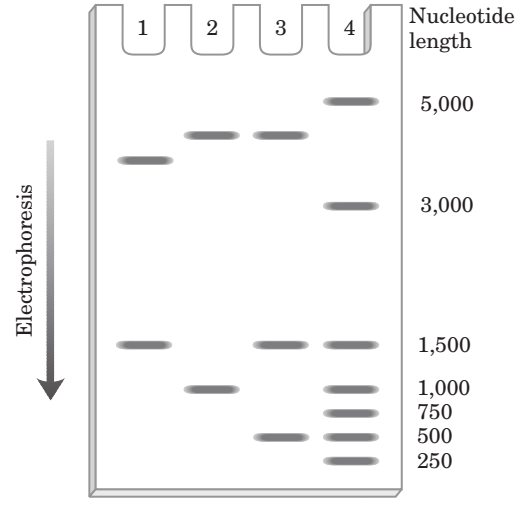
(i) Design four different short synthetic double-stranded DNA fragments that would permit ligation of structure (a) with a DNA fragment produced by a *Pst*I restriction digest. In one of these fragments, design the sequence so that the final junction contains the recognition sequences for both *Eco*RI and *Pst*I. In the second and third fragments, design the sequence so that the junction contains only the *Eco*RI and only the *Pst*I recognition sequence, respectively. Design the sequence of the fourth fragment so that neither the *Eco*RI nor the *Pst*I sequence appears in the junction.

2. Selecting for Recombinant Plasmids When cloning a foreign DNA fragment into a plasmid, it is often useful to insert the fragment at a site that interrupts a selectable marker (such as the tetracycline-resistance gene of pBR322). The loss of function of the interrupted gene can be used to identify clones containing recombinant plasmids with foreign DNA. With a bacteriophage λ vector it is not necessary to do this, yet one can easily distinguish vectors that incorporate large foreign DNA fragments from those that do not. How are these recombinant vectors identified?

3. DNA Cloning The plasmid cloning vector pBR322 (see Fig. 9–3) is cleaved with the restriction endonuclease *Pst*I. An isolated DNA fragment from a eukaryotic genome (also produced by *Pst*I cleavage) is added to the prepared vector and ligated. The mixture of ligated DNAs is then used to transform bacteria, and plasmid-containing bacteria are selected by growth in the presence of tetracycline.

(a) In addition to the desired recombinant plasmid, what other types of plasmids might be found among the transformed bacteria that are tetracycline-resistant? How can the types be distinguished?

(b) The cloned DNA fragment is 1,000 bp long and has an *Eco*RI site 250 bp from one end. Three different recombinant plasmids are cleaved with *Eco*RI and analyzed by gel electrophoresis, giving the patterns shown below. What does each pattern say about the cloned DNA? Note that in pBR322, the *Pst*I and *Eco*RI restriction sites are about 750 bp apart. The entire plasmid with no cloned insert is 4,361 bp. Size markers in lane 4 have the number of nucleotides noted.



4. Restriction Enzymes The partial sequence of one strand of a double-stranded DNA molecule is

5′ – – – GACGAAGTGCTGCAGAAAGTCCGCGTTATAGGCAT
GAATTCCTGAGG – – – 3′

The cleavage sites for the restriction enzymes *Eco*RI and *Pst*I are shown below.

*Eco*RI

```
      ↓  *
(5′) G A A T T C (3′)
     C T T A A G
        *  ↑
```

*Pst*I

```
         *  ↓
(5′) C T G C A G (3′)
     G A C G T C
     ↑  *
```

Write the sequence of *both strands* of the DNA fragment created when this DNA is cleaved with both *Eco*RI and *Pst*I. The top strand of your duplex DNA fragment should be derived from the strand sequence given above.

5. Designing a Diagnostic Test for a Genetic Disease Huntington disease (HD) is an inherited neurodegenerative disorder, characterized by the gradual, irreversible impairment of psychological, motor, and cognitive functions. Symptoms typically appear in middle age, but onset can occur at almost any age. The course of the disease can last 15 to 20 years. The molecular basis of the disease is becoming better understood. The genetic mutation underlying HD has been traced to a gene encoding a protein (M_r 350,000) of unknown function. In individuals who will not develop HD, a region of the gene that encodes the amino terminus of the protein has a sequence of CAG codons (for glutamine) that is repeated 6 to 39 times in succession. In individuals with adult-onset HD, this codon is typically repeated 40 to 55 times. In individuals with childhood-onset HD, this codon is repeated more than 70 times.

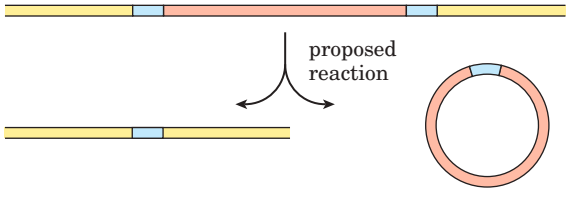
The length of this simple trinucleotide repeat indicates whether an individual will develop HD, and at approximately what age the first symptoms will occur.

A small portion of the amino-terminal coding sequence of the 3,143-codon HD gene is given below. The nucleotide sequence of the DNA is shown in black, the amino acid sequence corresponding to the gene is shown in blue, and the CAG repeat is shaded. Using Figure 27–7 to translate the genetic code, outline a PCR-based test for HD that could be carried out using a blood sample. Assume the PCR primer must be 25 nucleotides long. By convention, unless otherwise specified a DNA sequence encoding a protein is displayed with the coding strand (the sequence identical to the mRNA transcribed from the gene) on top such that it is read 5′ to 3′, left to right.

```
307 ATGGCGACCCTGGAAAAGCTGATGAAGGCCTTCGAGTCCCTCAAGTCCTTC
1   M  A  T  L  E  K  L  M  K  A  F  E  S  L  K  S  F

358 CAGCAGTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
18  Q  Q  F  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q

409 CAGCAGCAGCAGCAGCAGCAACAGCCGCCACCGCCGCCGCCGCCG
35  Q  Q  Q  Q  Q  Q  Q  Q  Q  P  P  P  P  P  P  P  P

460 CCGCCTCCTCAGCTTCCTCAGCCGCCGCCG
52  P  P  P  Q  L  P  Q  P  P  P
```

Source: The Huntington's Disease Collaborative Research Group. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983.

6. Using PCR to Detect Circular DNA Molecules In a species of ciliated protist, a segment of genomic DNA is sometimes deleted. The deletion is a genetically programmed reaction associated with cellular mating. A researcher proposes that the DNA is deleted in a type of recombination called site-specific recombination, with the DNA on either end of the segment joined together and the deleted DNA ending up as a circular DNA reaction product.



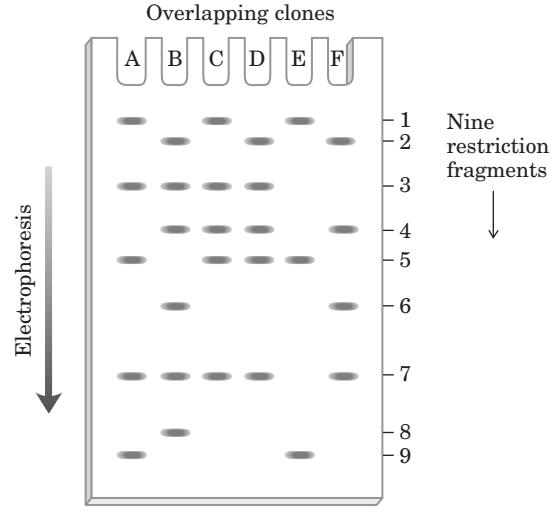
Suggest how the researcher might use the polymerase chain reaction (PCR) to detect the presence of the circular form of the deleted DNA in an extract of the protist.

7. Glowing Plants When grown in ordinary garden soil and watered normally, a plant engineered to express green fluorescent protein (see Fig. 9–16) will glow in the dark, whereas a plant engineered to express firefly luciferase will not. Explain these observations.

8. Designing PCR Primers One strand of a chromosomal DNA sequence is shown at the top of the next column. An investigator wants to amplify and isolate a DNA fragment defined by the segment shown in red, using the polymerase chain reaction. Design two PCR primers, each 20 nucleotides long, that can be used to amplify this DNA segment.

5′ – – – AATGCCGTCAGCCGATCTGCCTCGAGTCAATCGA
TGCTGGTAACTTGGGGTATAAAGCTTACCCATGGTATCG
TAGTTAGATTGATTGTTAGGTTCTTAGGTTTAGGTTTC
TGGTATTGGTTTAGGGTCTTTGATGCTATTAATTGTTTGG
TTTTGATTTGGTCTTTATATGGTTTATGTTTTAAGCCGGGT
TTTGTCTGGGATGGTTCGTCTGATGTGCGCGTAGCGTGCG
GCG – – – 3′

9. Mapping a Chromosome Segment A group of overlapping clones, designated A through F, is isolated from one region of a chromosome. Each of the clones is separately cleaved by a restriction enzyme and the pieces resolved by agarose gel electrophoresis, with the results shown below. There are nine different restriction fragments in this chromosomal region, with a subset appearing in each clone. Using this information, deduce the order of the restriction fragments in the chromosome.



10. Immunofluorescence In the more common protocol for immunofluorescence detection of cellular proteins, an investigator uses two antibodies. The first binds specifically to the protein of interest. The second is labeled with fluorochromes for easy visualization, and it binds to the first antibody. In principle, one could simply label the first antibody and skip one step. Why use two successive antibodies?

11. Yeast Two-Hybrid Analysis You are a researcher who has just discovered a new protein in a fungus. Design a yeast two-hybrid experiment to identify the other proteins in the fungal cell with which your protein interacts and explain how this could help you determine the function of your protein.

12. Use of Photolithography to Make a DNA Microarray Figure 9–22 shows the first steps in the process of making a DNA microarray, or DNA chip, using photolithography. Describe the remaining steps needed to obtain the desired sequences (a different four-nucleotide sequence on each of the four spots) shown in the first panel of the figure. After each step, give the resulting nucleotide sequence attached at each spot.

13. Genomic Sequencing In large-genome sequencing projects, the initial data usually reveal gaps where no sequence information has been obtained. To close the gaps, DNA primers

complementary to the 5′-ending strand (i.e., identical to the sequence of the 3′-ending strand) at the end of each contig are especially useful. Explain how these primers might be used.

14. Use of Outgroups in Comparative Genomics A hypothetical protein is found in orangutans, chimpanzees, and humans that has the following sequences (red indicates the amino acid residue differences):

Human: ATSAAG**Y**DEWEGGK**V**LIHL – – KLQNRGALL ELDIGAV

Orangutan: ATSAAG**W**DEWEGGK**V**LIHL**DG**KLQNRGALL ELDIGAV

Chimpanzee: ATSAAG**W**DEWEGGK**I**LIHL**DG**KLQNRGALL ELDIGAV

(Dashes indicate a deletion—the residues are missing in that sequence.)

What is the most likely sequence of the protein present in the last common ancestor of chimpanzees and humans?

15. Finding Disease Genes You are a gene hunter, trying to find the genetic basis for a rare inherited disease. Examination of six pedigrees of families affected by the disease provides inconsistent results. For two of the families, the disease is co-inherited with markers on chromosome 7. For the other four families, the disease is co-inherited with markers on chromosome 12. Explain how this might occur.

Data Analysis Problem

16. *Hinc*II: The First Restriction Endonuclease Discovery of the first restriction endonuclease to be of practical use was reported in two papers published in 1970. In the first paper, Smith and Wilcox described the isolation of an enzyme that cleaved double-stranded DNA. They initially demonstrated the enzyme's nuclease activity by measuring the decrease in viscosity of DNA samples treated with the enzyme.

(a) Why does treatment with a nuclease decrease the viscosity of a solution of DNA?

The authors determined whether the enzyme was an endo- or an exonuclease by treating ^{32}P-labeled DNA with the enzyme, then adding trichloroacetic acid (TCA). Under the conditions used in their experiment, single nucleotides would be TCA-soluble and oligonucleotides would precipitate.

(b) No TCA-soluble ^{32}P-labeled material formed on treatment of ^{32}P-labeled DNA with the nuclease. Based on this finding, is the enzyme an endo- or exonuclease? Explain your reasoning.

When a polynucleotide is cleaved, the phosphate usually is not removed but remains attached to the 5′ or 3′ end of the resulting DNA fragment. Smith and Wilcox determined the location of the phosphate on the fragment formed by the nuclease in the following steps:

1. Treat unlabeled DNA with the nuclease.
2. Treat a sample (A) of the product with γ-^{32}P-labeled ATP and polynucleotide kinase (which can attach the γ-phosphate of ATP to a 5′ OH but not to a 5′ phosphate or to a 3′ OH or 3′ phosphate). Measure the amount of ^{32}P incorporated into the DNA.
3. Treat another sample (B) of the product of step 1 with alkaline phosphatase (which removes phosphate groups from free 5′ and 3′ ends), followed by polynucleotide kinase and γ-^{32}P-labeled ATP. Measure the amount of ^{32}P incorporated into the DNA.

(c) Smith and Wilcox found that sample A had 136 counts/min of ^{32}P; sample B had 3,740 counts/min. Did the nuclease cleavage leave the phosphate on the 5′ or the 3′ end of the DNA fragments? Explain your reasoning.

(d) Treatment of bacteriophage T7 DNA with the nuclease gave approximately 40 specific fragments of various lengths. How is this result consistent with the enzyme's recognizing a specific sequence in the DNA as opposed to making random double-strand breaks?

At this point, there were two possibilities for the site-specific cleavage: the cleavage occurred either (1) at the site of recognition or (2) near the site of recognition but not within the sequence recognized. To address this issue, Kelly and Smith determined the sequence of the 5′ ends of the DNA fragments generated by the nuclease, in the following steps:

1. Treat phage T7 DNA with the enzyme.
2. Treat the resulting fragments with alkaline phosphatase to remove the 5′ phosphates.
3. Treat the dephosphorylated fragments with polynucleotide kinase and γ-^{32}P-labeled ATP to label the 5′ ends.
4. Treat the labeled molecules with DNases to break them into a mixture of mono-, di-, and trinucleotides.
5. Determine the sequence of the labeled mono-, di-, and trinucleotides by comparing them with oligonucleotides of known sequence on thin-layer chromatography.

The labeled products were identified as follows: mononucleotides: A and G; dinucleotides: (5′)ApA(3′) and (5′)GpA(3′); trinucleotides: (5′)ApApC(3′) and (5′)GpApC(3′).

(e) Which model of cleavage is consistent with these results? Explain your reasoning.

Kelly and Smith went on to determine the sequence of the 3′ ends of the fragments. They found a mixture of (5′)TpC(3′) and (5′)TpT(3′). They did not determine the sequence of any trinucleotides at the 3′ end.

(f) Based on these data, what is the recognition sequence for the nuclease and where in the sequence is the DNA backbone cleaved? Use Table 9–2 as a model for your answer.

References

Kelly, T.J. & Smith, H.O. (1970) A restriction enzyme from *Haemophilus influenzae*: II. Base sequence of the recognition site. *J. Mol. Biol.* **51**, 393–409.

Smith, H.O. & Wilcox, K.W. (1970) A restriction enzyme from *Haemophilus influenzae*: I. Purification and general properties. *J. Mol. Biol.* **51**, 379–391.

this page left intentionally blank

10

Lipids

Biological lipids are a chemically diverse group of compounds, the common and defining feature of which is their insolubility in water. The biological functions of the lipids are as diverse as their chemistry. Fats and oils are the principal stored forms of energy in many organisms. Phospholipids and sterols are major structural elements of biological membranes. Other lipids, although present in relatively small quantities, play crucial roles as enzyme cofactors, electron carriers, light-absorbing pigments, hydrophobic anchors for proteins, "chaperones" to help membrane proteins fold, emulsifying agents in the digestive tract, hormones, and intracellular messengers. This chapter introduces representative lipids of each type, organized according to their functional roles, with emphasis on their chemical structure and physical properties. Although we follow a functional organization for our discussion, the literally thousands of different lipids can also be organized into eight general categories of chemical structure (see Table 10–3). We discuss the energy-yielding oxidation of lipids in Chapter 17 and their synthesis in Chapter 21.

10.1 Storage Lipids

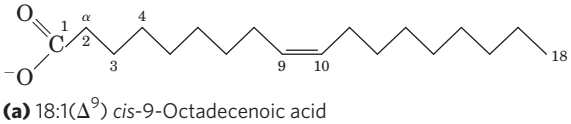
The fats and oils used almost universally as stored forms of energy in living organisms are derivatives of **fatty acids**. The fatty acids are hydrocarbon derivatives, at about the same low oxidation state (that is, as highly reduced) as the hydrocarbons in fossil fuels. The cellular oxidation of fatty acids (to CO_2 and H_2O), like the controlled, rapid burning of fossil fuels in internal combustion engines, is highly exergonic.

We introduce here the structures and nomenclature of the fatty acids most commonly found in living organisms. Two types of fatty acid–containing compounds, triacylglycerols and waxes, are described to illustrate the diversity of structure and physical properties in this family of compounds.

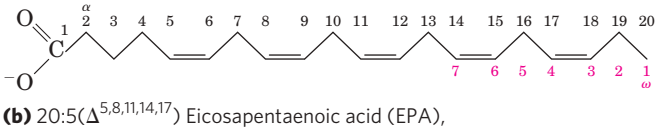
Fatty Acids Are Hydrocarbon Derivatives

Fatty acids are carboxylic acids with hydrocarbon chains ranging from 4 to 36 carbons long (C_4 to C_{36}). In some fatty acids, this chain is unbranched and fully saturated (contains no double bonds); in others the chain contains one or more double bonds (Table 10–1). A few contain three-carbon rings, hydroxyl groups, or methyl-group branches.

KEY CONVENTION: A simplified nomenclature for unbranched fatty acids specifies the chain length and number of double bonds, separated by a colon **(Fig. 10–1a)**; for example, the 16-carbon saturated palmitic acid is abbreviated 16:0, and the 18-carbon oleic acid, with one double bond, is 18:1. The positions of any double bonds are



(a) 18:1(Δ^9) *cis*-9-Octadecenoic acid



(b) 20:5($\Delta^{5,8,11,14,17}$) Eicosapentaenoic acid (EPA), an omega-3 fatty acid

FIGURE 10–1 Two conventions for naming fatty acids. (a) Standard nomenclature assigns the number 1 to the carboxyl carbon (C-1), and α to the carbon next to it. Each line segment of the zigzag represents a single bond between adjacent carbons. The position of any double bond(s) is indicated by Δ followed by a superscript number indicating the lower-numbered carbon in the double bond. **(b)** For polyunsaturated fatty acids (PUFAs), an alternative convention numbers the carbons in the opposite direction, assigning the number 1 to the methyl carbon at the other end of the chain; this carbon is also designated ω (omega; the last letter in the Greek alphabet). The positions of the double bonds are indicated relative to the ω carbon.

TABLE 10–1 Some Naturally Occurring Fatty Acids: Structure, Properties, and Nomenclature

Carbon skeleton	Structure*	Systematic name†	Common name (derivation)	Melting point (°C)	Solubility at 30 °C (mg/g solvent)	
					Water	Benzene
12:0	$CH_3(CH_2)_{10}COOH$	*n*-Dodecanoic acid	Lauric acid (Latin *laurus*, "laurel plant")	44.2	0.063	2,600
14:0	$CH_3(CH_2)_{12}COOH$	*n*-Tetradecanoic acid	Myristic acid (Latin *Myristica*, nutmeg genus)	53.9	0.024	874
16:0	$CH_3(CH_2)_{14}COOH$	*n*-Hexadecanoic acid	Palmitic acid (Latin *palma*, "palm tree")	63.1	0.0083	348
18:0	$CH_3(CH_2)_{16}COOH$	*n*-Octadecanoic acid	Stearic acid (Greek *stear*, "hard fat")	69.6	0.0034	124
20:0	$CH_3(CH_2)_{18}COOH$	*n*-Eicosanoic acid	Arachidic acid (Latin *Arachis*, legume genus)	76.5		
24:0	$CH_3(CH_2)_{22}COOH$	*n*-Tetracosanoic acid	Lignoceric acid (Latin *lignum*, "wood" + *cera*, "wax")	86.0		
16:1(Δ^9)	$CH_3(CH_2)_5CH{=}CH(CH_2)_7COOH$	*cis*-9-Hexadecenoic acid	Palmitoleic acid	1 to –0.5		
18:1(Δ^9)	$CH_3(CH_2)_7CH{=}CH(CH_2)_7COOH$	*cis*-9-Octadecenoic acid	Oleic acid (Latin *oleum*, "oil")	13.4		
18:2($\Delta^{9,12}$)	$CH_3(CH_2)_4CH{=}CHCH_2CH{=}CH(CH_2)_7COOH$	*cis*-,*cis*-9,12-Octadecadienoic acid	Linoleic acid (Greek *linon*, "flax")	1–5		
18:3($\Delta^{9,12,15}$)	$CH_3CH_2CH{=}CHCH_2CH{=}CHCH_2CH{=}CH(CH_2)_7COOH$	*cis*-,*cis*-,*cis*-9,12,15-Octadecatrienoic acid	α-Linolenic acid	−11		
20:4($\Delta^{5,8,11,14}$)	$CH_3(CH_2)_4CH{=}CHCH_2CH{=}CHCH_2CH{=}CHCH_2CH{=}CH(CH_2)_3COOH$	*cis*-,*cis*-,*cis*-,*cis*-5,8,11,14-Icosatetraenoic acid	Arachidonic acid	−49.5		

*All acids are shown in their nonionized form. At pH 7, all free fatty acids have an ionized carboxylate. Note that numbering of carbon atoms begins at the carboxyl carbon.

†The prefix *n*- indicates the "normal" unbranched structure. For instance, "dodecanoic" simply indicates 12 carbon atoms, which could be arranged in a variety of branched forms; "*n*-dodecanoic" specifies the linear, unbranched form. For unsaturated fatty acids, the configuration of each double bond is indicated; in biological fatty acids the configuration is almost always cis.

specified relative to the carboxyl carbon, numbered 1, by superscript numbers following Δ (delta); a 20-carbon fatty acid with one double bond between C-9 and C-10 (C-1 being the carboxyl carbon) and another between C-12 and C-13 is designated 20:2($\Delta^{9,12}$). ■

The most commonly occurring fatty acids have even numbers of carbon atoms in an unbranched chain of 12 to 24 carbons (Table 10–1). As we shall see in Chapter 21, the even number of carbons results from the mode of synthesis of these compounds, which involves successive condensations of two-carbon (acetate) units.

There is also a common pattern in the location of double bonds; in most monounsaturated fatty acids the double bond is between C-9 and C-10 (Δ^9), and the other double bonds of polyunsaturated fatty acids are generally Δ^{12} and Δ^{15}. (Arachidonic acid is an exception to this generalization.) The double bonds of polyunsaturated fatty acids are almost never conjugated (alternating single and double bonds, as in

—CH═CH—CH═CH—), but are separated by a methylene group: —CH═CH—CH_2—CH═CH— (Fig. 10–1b). In nearly all naturally occurring unsaturated fatty acids, the double bonds are in the cis configuration. Trans fatty acids are produced by fermentation in the rumen of dairy animals and are obtained from dairy products and meat.

KEY CONVENTION: The family of **polyunsaturated fatty acids (PUFAs)** with a double bond between the third and fourth carbon from the methyl end of the chain are of special importance in human nutrition. Because the physiological role of PUFAs is related more to the position of the first double bond near the *methyl* end of the chain than to the carboxyl end, an alternative nomenclature is sometimes used for these fatty acids. The carbon of the methyl group—that is, the carbon most distant from the carboxyl group—is called the ω (omega) carbon and is given the number 1 (Fig. 10–1b). In this convention, PUFAs with a double bond between C-3 and C-4 are called **omega-3 (ω-3) fatty acids**, and those with a double bond between C-6 and C-7 are **omega-6 (ω-6) fatty acids**. ■

Humans require but do not have the enzymatic capacity to synthesize the omega-3 PUFA α-linolenic acid (ALA; 18:3($\Delta^{9,12,15}$), in the standard convention), and must therefore obtain it in the diet. From ALA, humans can synthesize two other omega-3 PUFAs important in cellular function: eicosapentaenoic acid (EPA; 20:5($\Delta^{5,8,11,14,17}$), shown in Fig. 10–1b) and docosahexaenoic acid (DHA; 22:6($\Delta^{4,7,10,13,16,19}$)). An imbalance of omega-6 and omega-3 PUFAs in the diet is associated with an increased risk of cardiovascular disease. The optimal dietary ratio of omega-6 to omega-3 PUFAs is between 1:1 and 4:1, but the ratio in the diets of most North Americans is closer to 10:1 to 30:1. The "Mediterranean diet," which has been associated with lowered cardiovascular risk, is richer in omega-3 PUFAs, obtained in leafy vegetables (salads) and fish oils. The latter oils are especially rich in EPA and DHA, and fish oil supplements are often prescribed for individuals with a history of cardiovascular disease. ■

The physical properties of the fatty acids, and of compounds that contain them, are largely determined by the length and degree of unsaturation of the hydrocarbon chain. The nonpolar hydrocarbon chain accounts for the poor solubility of fatty acids in water. Lauric acid (12:0, M_r 200), for example, has a solubility in water of 0.063 mg/g—much less than that of glucose (M_r 180), which is 1,100 mg/g. The longer the fatty acyl chain and the fewer the double bonds, the lower is the solubility in water. The carboxylic acid group is polar (and ionized at neutral pH) and accounts for the slight solubility of short-chain fatty acids in water.

Melting points are also strongly influenced by the length and degree of unsaturation of the hydrocarbon chain. At room temperature (25 °C), the saturated fatty acids from 12:0 to 24:0 have a waxy consistency, whereas unsaturated fatty acids of these lengths are oily liquids. This difference in melting points is due to different degrees of packing of the fatty acid molecules **(Fig. 10–2)**. In the fully saturated compounds, free rotation around each carbon–carbon bond gives the hydrocarbon chain great flexibility; the most stable conformation is the fully extended form, in which the steric hindrance of neighboring atoms is minimized. These molecules can pack together tightly in nearly crystalline arrays, with atoms all along their lengths in van der Waals contact with the atoms of neighboring molecules. In unsaturated fatty acids, a cis double bond forces a kink in the hydrocarbon chain. Fatty acids with one or several such kinks cannot pack together as tightly as fully saturated fatty acids, and their interactions with each other are therefore weaker. Because less thermal energy is needed to disorder these poorly ordered arrays of unsaturated fatty acids, they have markedly lower melting points than saturated fatty acids of the same chain length (Table 10–1).

In vertebrates, free fatty acids (unesterified fatty acids, with a free carboxylate group) circulate in the blood bound noncovalently to a protein carrier, serum albumin. However, fatty acids are present in blood plasma mostly as carboxylic acid derivatives such as

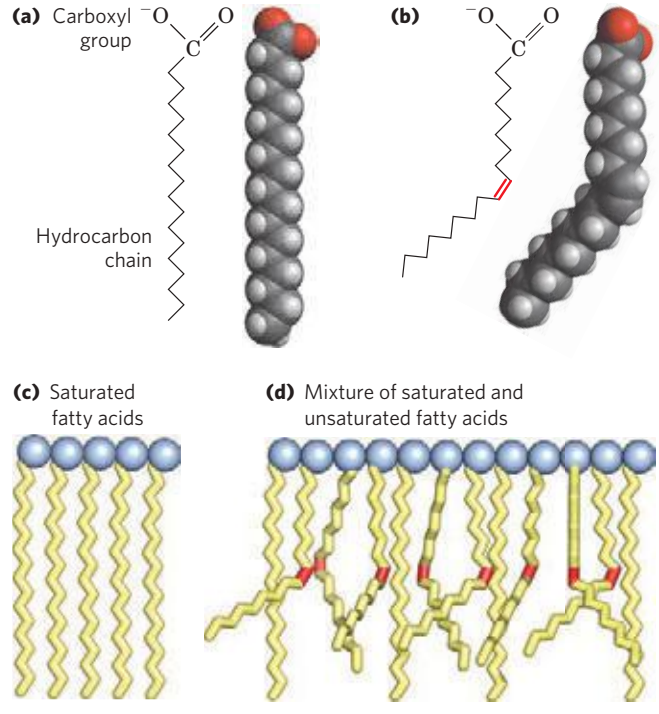


FIGURE 10–2 The packing of fatty acids into stable aggregates. The extent of packing depends on the degree of saturation. **(a)** Two representations of the fully saturated acid stearic acid, 18:0 (stearate at pH 7), in its usual extended conformation. **(b)** The cis double bond (red) in oleic acid, 18:1(Δ^9) (oleate), restricts rotation and introduces a rigid bend in the hydrocarbon tail. All other bonds in the chain are free to rotate. **(c)** Fully saturated fatty acids in the extended form pack into nearly crystalline arrays, stabilized by many hydrophobic interactions. **(d)** The presence of one or more fatty acids with cis double bonds (red) interferes with this tight packing and results in less stable aggregates.

esters or amides. Lacking the charged carboxylate group, these fatty acid derivatives are generally even less soluble in water than are the free fatty acids.

Triacylglycerols Are Fatty Acid Esters of Glycerol

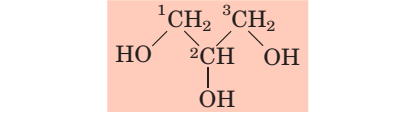
The simplest lipids constructed from fatty acids are the **triacylglycerols**, also referred to as triglycerides, fats, or neutral fats. Triacylglycerols are composed of three fatty acids each in ester linkage with a single glycerol **(Fig. 10–3)**. Those containing the same kind of fatty acid in all three positions are called simple triacylglycerols and are named after the fatty acid they contain. Simple triacylglycerols of 16:0, 18:0, and 18:1, for example, are tripalmitin, tristearin, and triolein, respectively. Most naturally occurring triacylglycerols are mixed; they contain two or three different fatty acids. To name these compounds unambiguously, the name and position of each fatty acid must be specified.

Because the polar hydroxyls of glycerol and the polar carboxylates of the fatty acids are bound in ester linkages, triacylglycerols are nonpolar, hydrophobic molecules, essentially insoluble in water. Lipids have lower specific gravities than water, which explains why mixtures of oil and water (oil-and-vinegar salad dressing, for example) have two phases: oil, with the lower specific gravity, floats on the aqueous phase.

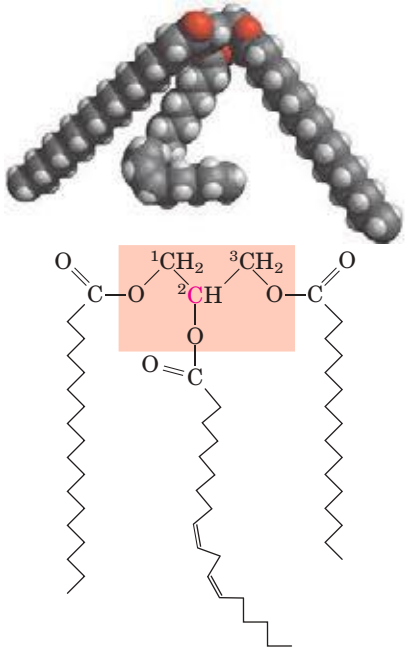
Triacylglycerols Provide Stored Energy and Insulation

In most eukaryotic cells, triacylglycerols form a separate phase of microscopic, oily droplets in the aqueous cytosol, serving as depots of metabolic fuel. In vertebrates, specialized cells called adipocytes, or fat cells, store large amounts of triacylglycerols as fat droplets that nearly fill the cell **(Fig. 10–4a)**. Triacylglycerols are also stored as oils in the seeds of many types of plants, providing energy and biosynthetic precursors during seed germination (Fig. 10–4b). Adipocytes and germinating seeds contain **lipases**, enzymes that catalyze the hydrolysis of stored triacylglycerols, releasing fatty acids for export to sites where they are required as fuel.

There are two significant advantages to using triacylglycerols as stored fuels, rather than polysaccharides such as glycogen and starch. First, the carbon atoms of fatty acids are more reduced than those of

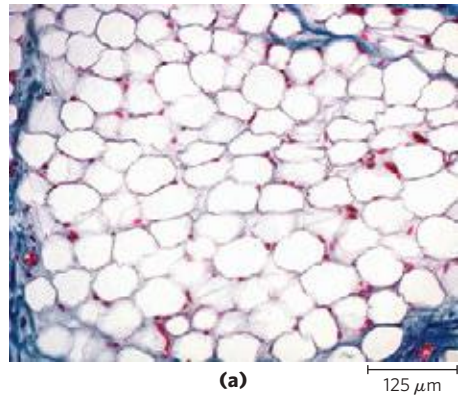


Glycerol



1-Stearoyl, 2-linoleoyl, 3-palmitoyl glycerol,
a mixed triacylglycerol

FIGURE 10–3 Glycerol and a triacylglycerol. The mixed triacylglycerol shown here has three different fatty acids attached to the glycerol backbone. When glycerol has different fatty acids at C-1 and C-3, C-2 is a chiral center (p. 17).



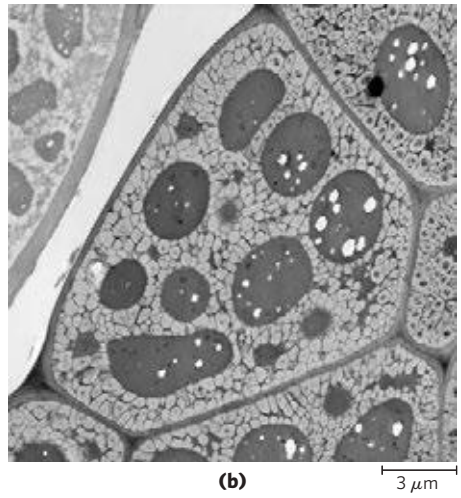


FIGURE 10–4 Fat stores in cells. (a) Cross section of human white adipose tissue. Each cell contains a fat droplet (white) so large that it squeezes the nucleus (stained red) against the plasma membrane. **(b)** Cross section of a cotyledon cell from a seed of the plant *Arabidopsis*. The large dark structures are protein bodies, which are surrounded by stored oils in the light-colored oil bodies.

sugars, and oxidation of triacylglycerols yields more than twice as much energy, gram for gram, as the oxidation of carbohydrates. Second, because triacylglycerols are hydrophobic and therefore unhydrated, the organism that carries fat as fuel does not have to carry the extra weight of water of hydration that is associated with stored polysaccharides (2 g per gram of polysaccharide). Humans have fat tissue (composed primarily of adipocytes) under the skin, in the abdominal cavity, and in the mammary glands. Moderately obese people with 15 to 20 kg of triacylglycerols deposited in their adipocytes could meet their energy needs for months by drawing on their fat stores. In contrast, the human body can store less than a day's energy supply in the form of glycogen. Carbohydrates such as glucose do offer certain advantages as quick sources of metabolic energy, one of which is their ready solubility in water.

In some animals, triacylglycerols stored under the skin serve not only as energy stores but as insulation against low temperatures. Seals, walruses, penguins, and other warm-blooded polar animals are amply padded with triacylglycerols. In hibernating animals (bears, for example), the huge fat reserves accumulated before hibernation serve the dual purposes of insulation and energy storage (see Box 17–1).

Partial Hydrogenation of Cooking Oils Produces Trans Fatty Acids

Most natural fats, such as those in vegetable oils, dairy products, and animal fat, are complex mixtures of simple and mixed triacylglycerols. These contain a variety of fatty acids differing in chain length and degree of saturation **(Fig. 10–5)**. Vegetable oils such as corn (maize) and olive oil are composed largely of triacylglycerols with unsaturated fatty acids and thus are liquids at room temperature. Triacylglycerols containing only saturated fatty acids, such as tristearin, the major component of beef fat, are white, greasy solids at room temperature.

When lipid-rich foods are exposed too long to the oxygen in air, they may spoil and become rancid. The unpleasant taste and smell associated with rancidity result from the oxidative cleavage of double bonds in unsaturated fatty acids, which produces aldehydes and carboxylic acids of shorter chain length and therefore higher volatility; these compounds pass readily through the air to your nose. To improve the shelf life of vegetable oils used in cooking, and to increase their stability at the high temperatures used in deep-frying, commercial vegetable oils are prepared by partial hydrogenation. This process converts many of the cis double bonds in the fatty acids to single bonds and increases the melting temperature of the oils so that they are more nearly solid at room temperature (margarine is produced from vegetable oil in this way). Partial hydrogenation has another, undesirable, effect: some cis double bonds are converted to trans double bonds. There is now strong evidence that dietary intake of trans fatty acids (often referred to simply as "trans fats") leads to a higher incidence of cardiovascular disease, and that avoiding these fats in the diet substantially reduces the risk of coronary heart disease. Dietary trans fatty acids raise the level of triacylglycerols and of LDL ("bad") cholesterol in the blood, and lower the level of HDL ("good") cholesterol, and these changes alone are enough to increase the risk of coronary heart disease. But trans fatty acids may have further adverse effects. They seem, for example, to increase the body's inflammatory response, which is another risk factor for heart disease. (See Chapter 21 for a description of LDL and HDL—low-density and high-density lipoprotein—cholesterol and their health effects.)

Many fast foods are deep-fried in partially hydrogenated vegetable oils and therefore contain high levels of trans fatty acids (Table 10–2). In view of the detrimental effects of these fats, some countries (Denmark, for example) and some cities (New York City and Philadelphia) severely restrict the use of partially hydrogenated oils in restaurants. French fries prepared in a chain fast-food restaurant in Denmark now contain almost no detectable trans fatty acids, whereas the same product prepared in the United States contains 5 to 10 g of trans fatty acids per serving (Table 10–2). The deleterious effects of trans fats occur at intakes of 2 to 7 g/day (20 to 60 kcal in a daily caloric intake of 2,000 kcal; note that a nutritional Calorie is the equivalent of the kilocalorie used by chemists and biochemists, so a 2,000 Calorie diet is the equivalent of a 2,000 kcal diet). A single serving of french fries in a U.S. restaurant may contain this amount of trans fatty acid! Many other prepared foods, baked goods, and snacks on the shelves of supermarkets have comparably high levels of trans fats. ■

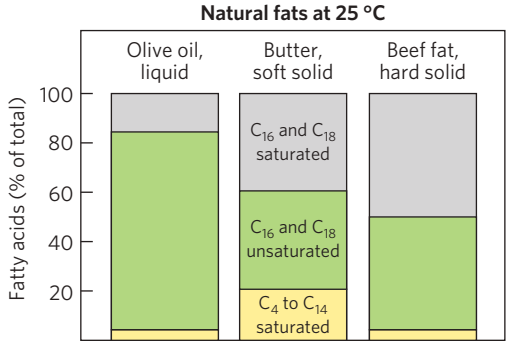


FIGURE 10–5 Fatty acid composition of three food fats. Olive oil, butter, and beef fat consist of mixtures of triacylglycerols, differing in their fatty acid composition. The melting points of these fats—and hence their physical state at room temperature (25 °C)—are a direct function of their fatty acid composition. Olive oil has a high proportion of long-chain (C_{16} and C_{18}) unsaturated fatty acids, which accounts for its liquid state at 25 °C. The higher proportion of long-chain (C_{16} and C_{18}) saturated fatty acids in butter increases its melting point, so butter is a soft solid at room temperature. Beef fat, with an even higher proportion of long-chain saturated fatty acids, is a hard solid.

TABLE 10–2 Trans Fatty Acids in Some Typical Fast Foods and Snacks

	Trans fatty acid content	
	In a typical serving (g)	As % of total fatty acids
French fries	4.7–6.1	28–36
Breaded fish burger	5.6	28
Breaded chicken nuggets	5.0	25
Pizza	1.1	9
Corn tortilla chips	1.6	22
Doughnut	2.7	25
Muffin	0.7	14
Chocolate bar	0.2	2

Source: Adapted from Table 1 in Mozaffarian, D., Katan, M.B., Ascherio, P.H., Stampfer, M.J., & Willet, W.C. (2006). Trans fatty acids and cardiovascular disease. *N. Engl. J. Med.* **354**, 1604–1605.

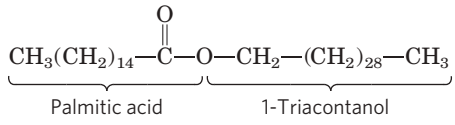
Note: All data for foods prepared with partially hydrogenated vegetable oil in the United States in 2002.

Waxes Serve as Energy Stores and Water Repellents

Biological waxes are esters of long-chain (C_{14} to C_{36}) saturated and unsaturated fatty acids with long-chain (C_{16} to C_{30}) alcohols **(Fig. 10–6)**. Their melting points (60 to 100 °C) are generally higher than those of triacylglycerols. In plankton, the free-floating microorganisms at the bottom of the food chain for marine animals, waxes are the chief storage form of metabolic fuel.

Waxes also serve a diversity of other functions related to their water-repellent properties and their firm consistency. Certain skin glands of vertebrates secrete waxes to protect hair and skin and keep it pliable, lubricated, and waterproof. Birds, particularly waterfowl, secrete waxes from their preen glands to keep their feathers water-repellent. The shiny leaves of holly, rhododendrons, poison ivy, and many tropical plants are coated with a thick layer of waxes, which prevents excessive evaporation of water and protects against parasites.

Biological waxes find a variety of applications in the pharmaceutical, cosmetic, and other industries. Lanolin (from lamb's wool), beeswax (Fig. 10–6), carnauba wax (from a Brazilian palm tree), and wax extracted from spermaceti oil (from whales) are widely used in the manufacture of lotions, ointments, and polishes.



(a)

(b)

FIGURE 10–6 Biological wax. (a) Triacontanoylpalmitate, the major component of beeswax, is an ester of palmitic acid with the alcohol triacontanol. **(b)** A honeycomb, constructed of beeswax, is firm at 25 °C and completely impervious to water. The term "wax" originates in the Old English *weax*, meaning "the material of the honeycomb."

SUMMARY 10.1 Storage Lipids

- Lipids are water-insoluble cellular components, of diverse structure, that can be extracted from tissues by nonpolar solvents.
- Almost all fatty acids, the hydrocarbon components of many lipids, have an even number of carbon atoms (usually 12 to 24); they are either saturated or unsaturated, with double bonds almost always in the cis configuration.
- Triacylglycerols contain three fatty acid molecules esterified to the three hydroxyl groups of glycerol. Simple triacylglycerols contain only one type of fatty acid; mixed triacylglycerols, two or three types. Triacylglycerols are primarily storage fats; they are present in many foods.
- Partial hydrogenation of vegetable oils in the food industry converts some cis double bonds to the trans configuration. Trans fatty acids in the diet are an important risk factor for coronary heart disease.

10.2 Structural Lipids in Membranes

The central architectural feature of biological membranes is a double layer of lipids, which acts as a barrier to the passage of polar molecules and ions. Membrane lipids are amphipathic: one end of the molecule is hydrophobic, the other hydrophilic. Their hydrophobic interactions with each other and their hydrophilic interactions with water direct their packing into sheets called membrane bilayers. In this section we describe five general types of membrane lipids: glycerophospholipids, in which the hydrophobic regions are composed

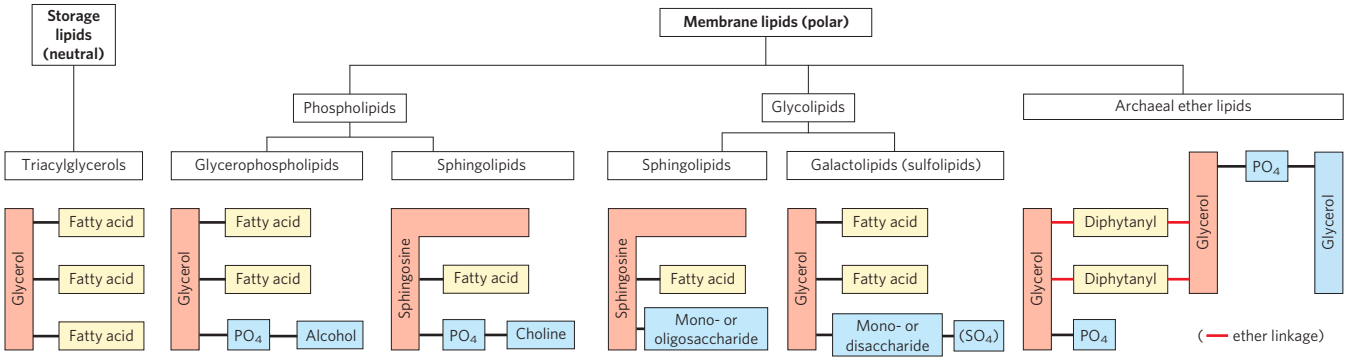


FIGURE 10–7 Some common types of storage and membrane lipids. All the lipid types shown here have either glycerol or sphingosine as the backbone (light red screen), to which are attached one or more long-chain alkyl groups (yellow) and a polar head group (blue). In triacylglycerols, glycerophospholipids, galactolipids, and sulfolipids, the alkyl groups are fatty acids in ester linkage. Sphingolipids contain a single fatty acid, in amide linkage to the sphingosine backbone. The membrane lipids of archaea are variable; that shown here has two very long, branched alkyl chains, each end in ether linkage with a glycerol moiety. In phospholipids the polar head group is joined through a phosphodiester, whereas glycolipids have a direct glycosidic linkage between the head-group sugar and the backbone glycerol.

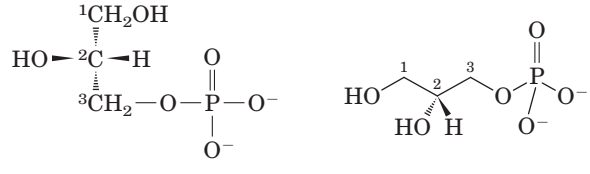
of two fatty acids joined to glycerol; galactolipids and sulfolipids, which also contain two fatty acids esterified to glycerol, but lack the characteristic phosphate of phospholipids; archaeal tetraether lipids, in which two very long alkyl chains are ether-linked to glycerol at both ends; sphingolipids, in which a single fatty acid is joined to a fatty amine, sphingosine; and sterols, compounds characterized by a rigid system of four fused hydrocarbon rings.

The hydrophilic moieties in these amphipathic compounds may be as simple as a single —OH group at one end of the sterol ring system, or they may be much more complex. In glycerophospholipids and some sphingolipids, a polar head group is joined to the hydrophobic moiety by a phosphodiester linkage; these are the **phospholipids**. Other sphingolipids lack phosphate but have a simple sugar or complex oligosaccharide at their polar ends; these are the **glycolipids** **(Fig. 10–7)**. Within these groups of membrane lipids, enormous diversity results from various combinations of fatty acid "tails" and polar "heads." The arrangement of these lipids in membranes, and their structural and functional roles therein, are considered in the next chapter.

Glycerophospholipids Are Derivatives of Phosphatidic Acid

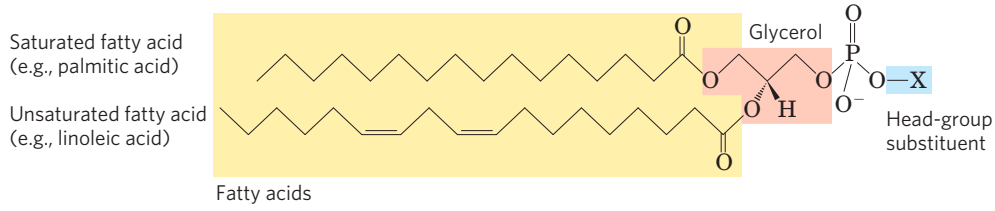
Glycerophospholipids, also called phosphoglycerides, are membrane lipids in which two fatty acids are attached in ester linkage to the first and second carbons of glycerol, and a highly polar or charged group is attached through a phosphodiester linkage to the third carbon. Glycerol is prochiral; it has no asymmetric carbons, but attachment of phosphate at one end converts it into a chiral compound, which can be correctly named either L-glycerol 3-phosphate, D-glycerol 1-phosphate, or *sn*-glycerol 3-phosphate **(Fig. 10–8)**. Glycerophospholipids are named as derivatives of the parent compound, phosphatidic acid **(Fig. 10–9)**, according to the polar alcohol in the head group. Phosphatidylcholine and phosphatidylethanolamine have choline and ethanolamine as their polar head groups, for example. In all these compounds, the head group is joined to glycerol through a phosphodiester bond, in which the phosphate group bears a negative charge at neutral pH. The polar alcohol may be negatively charged (as in phosphatidylinositol 4,5-bisphosphate), neutral (phosphatidylserine), or positively charged (phosphatidylcholine, phosphatidylethanolamine). As we shall see in Chapter 11, these charges contribute greatly to the surface properties of membranes.

The fatty acids in glycerophospholipids can be any of a wide variety, so a given phospholipid (phosphatidylcholine,



L-Glycerol 3-phosphate (*sn*-glycerol 3-phosphate)

FIGURE 10–8 L-Glycerol 3-phosphate, the backbone of phospholipids. Glycerol itself is not chiral, as it has a plane of symmetry through C-2. However, glycerol is prochiral—it can be converted to a chiral compound by adding a substituent such as phosphate to either of the —CH_2OH groups. One unambiguous nomenclature for glycerol phosphate is the D, L system (described on p. 78), in which the isomers are named according to their stereochemical relationships to glyceraldehyde isomers. By this system, the stereoisomer of glycerol phosphate found in most lipids is correctly named either L-glycerol 3-phosphate or D-glycerol 1-phosphate. Another way to specify stereoisomers is the *sn* (stereospecific *n*umbering) system, in which C-1 is, by definition, the group of the prochiral compound that occupies the pro-S position. The common form of glycerol phosphate in phospholipids is, by this system, *sn*-glycerol 3-phosphate (in which C-2 has the R configuration). In archaea, the glycerol in lipids has the other configuration; it is D-glycerol 3-phosphate.



Name of glycerophospholipid	Name of X—O	Formula of X	Net charge (at pH 7)
Phosphatidic acid	—	—H	−2
Phosphatidylethanolamine	Ethanolamine	NH_3^+	0
Phosphatidylcholine	Choline	N^+	0
Phosphatidylserine	Serine		−1
Phosphatidylglycerol	Glycerol		−1
Phosphatidylinositol 4,5-bisphosphate	*myo*-Inositol 4,5-bisphosphate		−4*
Cardiolipin	Phosphatidyl-glycerol		−2

FIGURE 10–9 Glycerophospholipids. The common glycerophospholipids are diacylglycerols linked to head-group alcohols through a phosphodiester bond. Phosphatidic acid, a phosphomonoester, is the parent compound. Each derivative is named for the head-group alcohol (X), with the prefix "phosphatidyl-." In cardiolipin, two phosphatidic acids share a single glycerol (R^1 and R^2 are fatty acyl groups). *Note that the phosphate esters in phosphatidylinositol 4,5-bisphosphate each have a charge of about −1.5; one of their −OH groups is only partially ionized at pH 7.

for example) may consist of several molecular species, each with its unique complement of fatty acids. The distribution of molecular species is specific for different organisms, different tissues of the same organism, and different glycerophospholipids in the same cell or tissue. In general, glycerophospholipids contain a C_{16} or C_{18} saturated fatty acid at C-1 and a C_{18} or C_{20} unsaturated fatty acid at C-2. With few exceptions, the biological significance of the variation in fatty acids and head groups is not yet understood.

Some Glycerophospholipids Have Ether-Linked Fatty Acids

Some animal tissues and some unicellular organisms are rich in **ether lipids**, in which one of the two acyl chains is attached to glycerol in ether, rather than ester, linkage. The ether-linked chain may be saturated, as in the alkyl ether lipids, or may contain a double bond between C-1 and C-2, as in **plasmalogens (Fig. 10–10)**. Vertebrate heart tissue is uniquely enriched in ether lipids; about

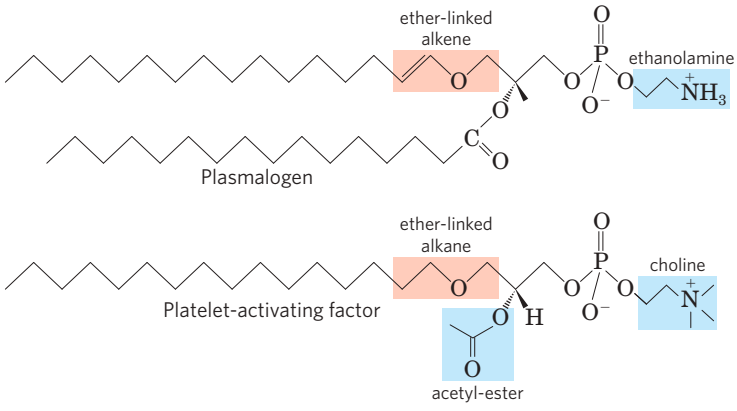


FIGURE 10–10 Ether lipids. Plasmalogens have an ether-linked alkenyl chain where most glycerophospholipids have an ester-linked fatty acid (compare Fig. 10–9). Platelet-activating factor has a long ether-linked alkyl chain at C-1 of glycerol, but C-2 is ester-linked to acetic acid, which makes the compound much more water-soluble than most glycerophospholipids and plasmalogens. The head-group alcohol is ethanolamine in plasmalogens and choline in platelet-activating factor.

half of the heart phospholipids are plasmalogens. The membranes of halophilic bacteria, ciliated protists, and certain invertebrates also contain high proportions of ether lipids. The functional significance of ether lipids in these membranes is unknown; perhaps their resistance to the phospholipases that cleave ester-linked fatty acids from membrane lipids is important in some roles.

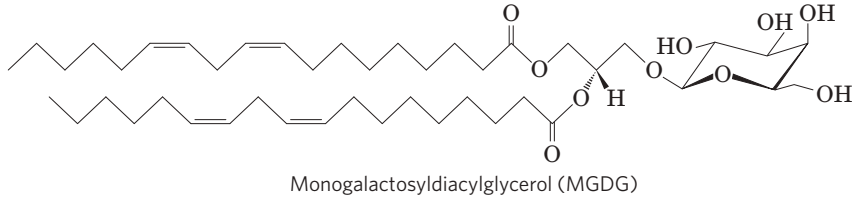
At least one ether lipid, **platelet-activating factor**, is a potent molecular signal. It is released from leukocytes called basophils and stimulates platelet aggregation and the release of serotonin (a vasoconstrictor) from platelets. It also exerts a variety of effects on liver, smooth muscle, heart, uterine, and lung tissues and plays an important role in inflammation and the allergic response. ■

Chloroplasts Contain Galactolipids and Sulfolipids

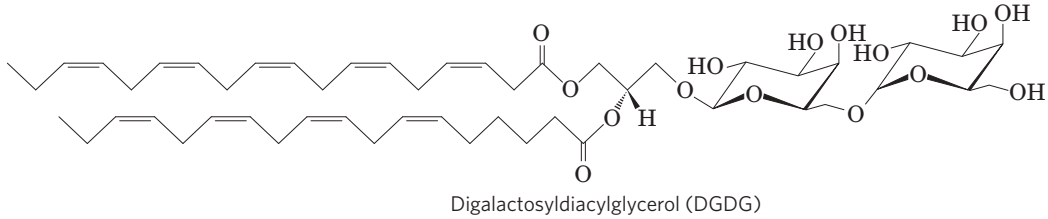
The second group of membrane lipids are those that predominate in plant cells: the **galactolipids**, in which one or two galactose residues are connected by a glycosidic linkage to C-3 of a 1,2-diacylglycerol **(Fig. 10–11**; see also Fig. 10–7). Galactolipids are localized in the thylakoid membranes (internal membranes) of chloroplasts; they make up 70% to 80% of the total membrane lipids of a vascular plant, and are therefore probably the most abundant membrane lipids in the biosphere. Phosphate is often the limiting plant nutrient in soil, and perhaps the evolutionary pressure to conserve phosphate for more critical roles favored plants that made phosphate-free lipids. Plant membranes also contain sulfolipids, in which a sulfonated glucose residue is joined to a diacylglycerol in glycosidic linkage. The sulfonate group bears a negative charge like that of the phosphate group in phospholipids.

Archaea Contain Unique Membrane Lipids

Some archaea that live in ecological niches with extreme conditions—high temperatures (boiling water), low pH, high ionic strength, for example—have membrane lipids containing long-chain (32 carbons) branched hydrocarbons linked at each end to glycerol **(Fig. 10–12)**. These linkages are through ether bonds, which are much more stable to hydrolysis at low pH and high temperature than are the ester bonds found in the lipids of bacteria and eukaryotes. In their fully extended form, these archaeal lipids are twice the length of phospholipids and sphingolipids, and can span the full width of the plasma membrane. At each end of the extended molecule is a polar head consisting of glycerol linked to either phosphate or



Monogalactosyldiacylglycerol (MGDG)



Digalactosyldiacylglycerol (DGDG)

FIGURE 10–11 Two galactolipids of chloroplast thylakoid membranes. In monogalactosyldiacylglycerols (MGDGs) and digalactosyldiacylglycerols (DGDGs) the acyl groups are both polyunsaturated and the head groups are uncharged.

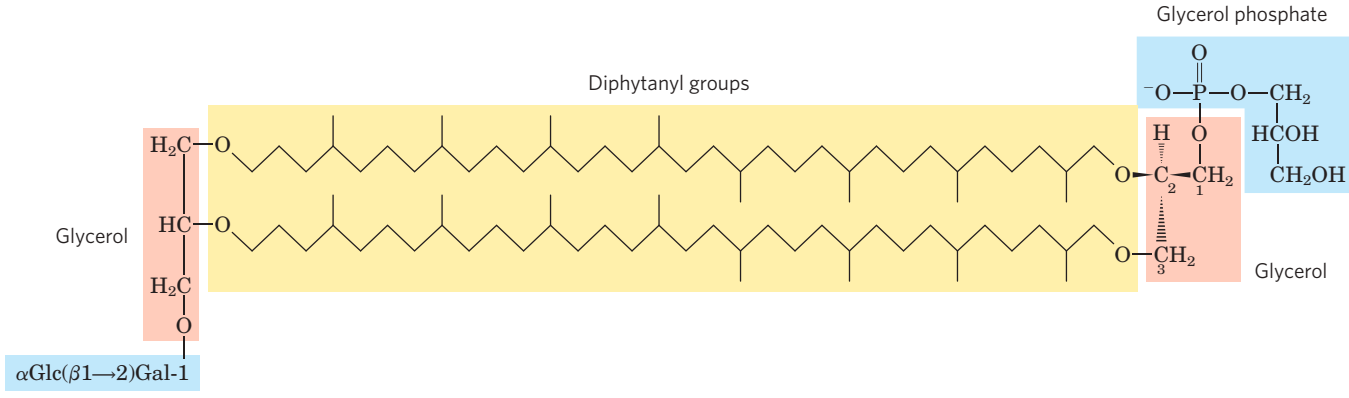


FIGURE 10–12 An unusual membrane lipid found only in some archaea. In this diphytanyl tetraether lipid, the diphytanyl moieties (yellow) are long hydrocarbons composed of eight five-carbon isoprene groups condensed end-to-end (on the condensation of isoprene units, see Fig. 21–36; also, compare the diphytanyl groups with the 20-carbon phytol side chain of chlorophylls in Fig. 19–49a). In this extended form, the diphytanyl groups are about twice the length of a 16-carbon fatty acid typically found in the membrane lipids of bacteria and eukaryotes. The glycerol moieties in the archaeal lipids are in the R configuration, in contrast to those of bacteria and eukaryotes, which have the S configuration. Archaeal lipids differ in the substituents on the glycerols. In the molecule shown here, one glycerol is linked to the disaccharide α-glucopyranosyl-(1→2)-β-galactofuranose; the other glycerol is linked to a glycerol phosphate head group.

sugar residues. The general name for these compounds, glycerol dialkyl glycerol tetraethers (GDGTs), reflects their unique structure. The glycerol moiety of the archaeal lipids is not the same stereoisomer as that in the lipids of bacteria and eukaryotes; the central carbon is in the R configuration in archaea, in the S configuration in bacteria and eukaryotes (Fig. 10–8).

Sphingolipids Are Derivatives of Sphingosine

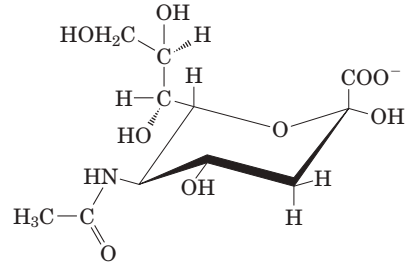
Sphingolipids, the fourth large class of membrane lipids, also have a polar head group and two nonpolar tails, but unlike glycerophospholipids and galactolipids they contain no glycerol. Sphingolipids are composed of one molecule of the long-chain amino alcohol sphingosine (also called 4-sphingenine) or one of its derivatives, one molecule of a long-chain fatty acid, and a polar head group that is joined by a glycosidic linkage in some cases and a phosphodiester in others **(Fig. 10–13)**.

Carbons C-1, C-2, and C-3 of the sphingosine molecule are structurally analogous to the three carbons of glycerol in glycerophospholipids. When a fatty acid is attached in amide linkage to the —NH_2 on C-2, the resulting compound is a **ceramide**, which is structurally similar to a diacylglycerol. Ceramide is the structural parent of all sphingolipids.

There are three subclasses of sphingolipids, all derivatives of ceramide but differing in their head groups: sphingomyelins, neutral (uncharged) glycolipids, and gangliosides. **Sphingomyelins** contain phosphocholine or phosphoethanolamine as their polar head group and are therefore classified along with glycerophospholipids as phospholipids (Fig. 10–7). Indeed, sphingomyelins resemble phosphatidylcholines in their general properties and three-dimensional structure, and in having no net charge on their head groups **(Fig. 10–14)**. Sphingomyelins are present in the plasma membranes of animal cells and are especially prominent in myelin, a membranous sheath that surrounds and insulates the axons of some neurons—thus the name "sphingomyelins."

Glycosphingolipids, which occur largely in the outer face of plasma membranes, have head groups with one or more sugars connected directly to the —OH at C-1 of the ceramide moiety; they do not contain phosphate. **Cerebrosides** have a single sugar linked to ceramide; those with galactose are characteristically found in the plasma membranes of cells in neural tissue, and those with glucose in the plasma membranes of cells in nonneural tissues. **Globosides** are glycosphingolipids with two or more sugars, usually D-glucose, D-galactose, or *N*-acetyl-D-galactosamine. Cerebrosides and globosides are sometimes called **neutral glycolipids**, as they have no charge at pH 7.

Gangliosides, the most complex sphingolipids, have oligosaccharides as their polar head groups and one or more residues of *N*-acetylneuraminic acid (Neu5Ac), a sialic acid (often simply called "sialic acid"), at the termini. Sialic acid gives gangliosides the negative charge at pH 7 that distinguishes them from globosides. Gangliosides with one sialic acid residue are in the GM (*M* for mono-) series, those with two are in the GD (*D* for di-) series, and so on (GT, three sialic acid residues; GQ, four).



N-Acetylneuraminic acid (a sialic acid) (Neu5Ac)

Johann Thudichum, 1829–1901

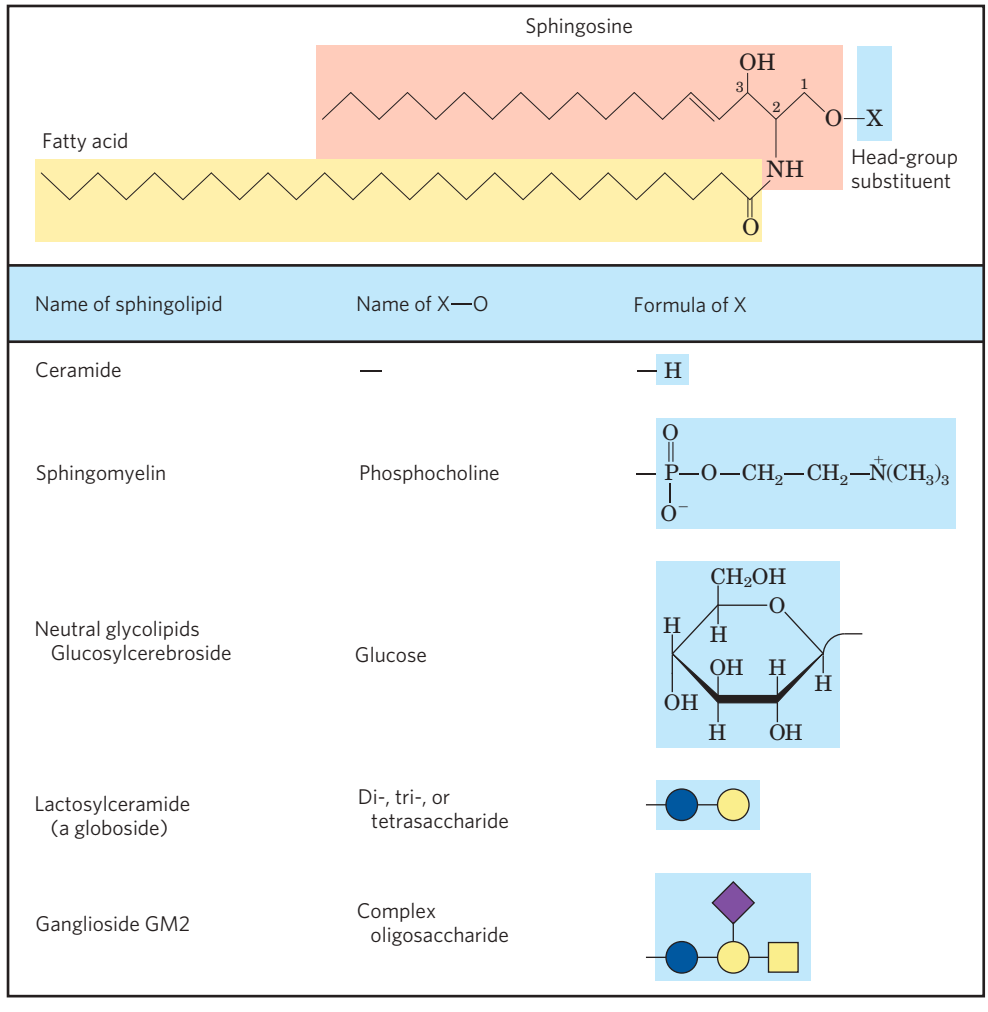


FIGURE 10–13 Sphingolipids. The first three carbons at the polar end of sphingosine are analogous to the three carbons of glycerol in glycerophospholipids. The amino group at C-2 bears a fatty acid in amide linkage. The fatty acid is usually saturated or monounsaturated, with 16, 18, 22, or 24 carbon atoms. Ceramide is the parent compound for this group. Other sphingolipids differ in the polar head group (X) attached at C-1. Gangliosides have very complex oligosaccharide head groups. Standard symbols for sugars are used in this figure, as shown in Table 7–1.

Sphingolipids at Cell Surfaces Are Sites of Biological Recognition

When sphingolipids were discovered more than a century ago by the physician-chemist Johann Thudichum, their biological role seemed as enigmatic as the Sphinx, for which he therefore named them. In humans, at least 60 different sphingolipids have been identified in cellular membranes. Many of these are especially prominent in the plasma membranes of neurons, and some are clearly recognition sites on the cell surface, but a specific function for only a few sphingolipids has been

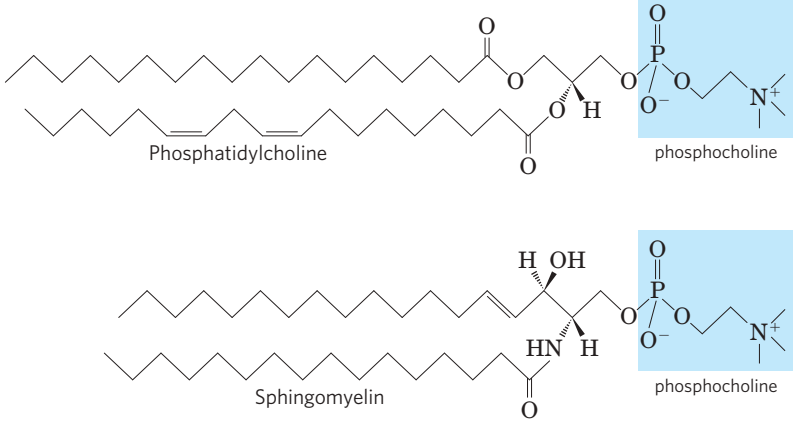


FIGURE 10–14 The molecular structures of two types of membrane lipid classes are similar. Phosphatidylcholine (a glycerophospholipid) and sphingomyelin (a sphingolipid) have similar dimensions and physical properties, but presumably play different roles in membranes.

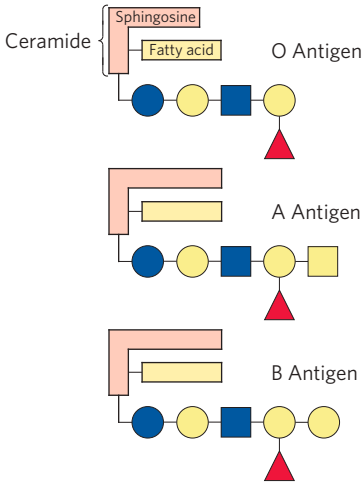


FIGURE 10–15 Glycosphingolipids as determinants of blood groups. The human blood groups (O, A, B) are determined in part by the oligosaccharide head groups of these glycosphingolipids. The same three oligosaccharides are also found attached to certain blood proteins of individuals of blood types O, A, and B, respectively. Standard symbols for sugars are used here (see Table 7–1).

discovered thus far. The carbohydrate moieties of certain sphingolipids define the human blood groups and therefore determine the type of blood that individuals can safely receive in blood transfusions **(Fig. 10–15)**.

Gangliosides are concentrated in the outer surface of cells, where they present points of recognition for extracellular molecules or surfaces of neighboring cells. The kinds and amounts of gangliosides in the plasma membrane change dramatically during embryonic development. Tumor formation induces the synthesis of a new complement of gangliosides, and very low concentrations of a specific ganglioside have been found to induce differentiation of cultured neuronal tumor cells. Investigation of the biological roles of diverse gangliosides remains fertile ground for future research.

Phospholipids and Sphingolipids Are Degraded in Lysosomes

Most cells continually degrade and replace their membrane lipids. For each hydrolyzable bond in a glycerophospholipid, there is a specific hydrolytic enzyme in the lysosome **(Fig. 10–16)**. Phospholipases of the A type remove one of the two fatty acids, producing a lysophospholipid. (These esterases do not attack the ether link of plasmalogens.) Lysophospholipases remove the remaining fatty acid.

Gangliosides are degraded by a set of lysosomal enzymes that catalyze the stepwise removal of sugar units, finally yielding a ceramide. A genetic defect in any of these hydrolytic enzymes leads to the accumulation of gangliosides in the cell, with severe medical consequences (Box 10–1).

Sterols Have Four Fused Carbon Rings

Sterols are structural lipids present in the membranes of most eukaryotic cells. The characteristic structure of

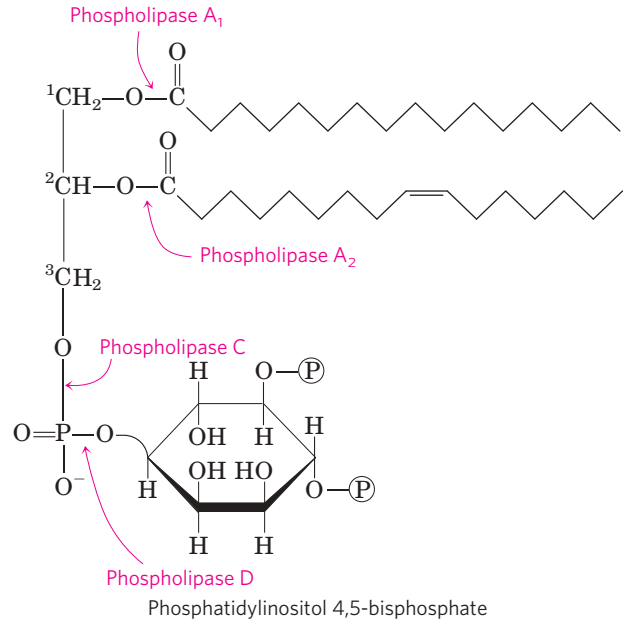


FIGURE 10–16 The specificities of phospholipases. Phospholipases A_1 and A_2 hydrolyze the ester bonds of intact glycerophospholipids at C-1 and C-2 of glycerol, respectively. When one of the fatty acids has been removed by a type A phospholipase, the second fatty acid is removed by a lysophospholipase (not shown). Phospholipases C and D each split one of the phosphodiester bonds in the head group. Some phospholipases act on only one type of glycerophospholipid, such as phosphatidylinositol 4,5-bisphosphate (shown here) or phosphatidylcholine; others are less specific.

this fifth group of membrane lipids is the steroid nucleus, consisting of four fused rings, three with six carbons and one with five **(Fig. 10–17)**. The steroid nucleus is almost planar and is relatively rigid; the fused rings do not allow rotation about C—C bonds. **Cholesterol**, the major sterol in animal tissues, is amphipathic, with a polar head group (the hydroxyl group at C-3) and a nonpolar hydrocarbon body (the steroid nucleus and the hydrocarbon side chain at C-17), about as long as a 16-carbon fatty acid in its extended form. Similar sterols are found in other eukaryotes: stigmasterol in plants and ergosterol in fungi, for example. Bacteria cannot synthesize sterols; a few bacterial species, however, can

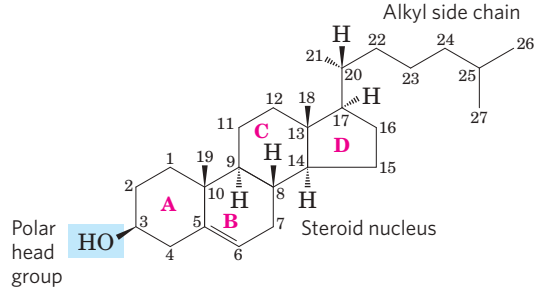


FIGURE 10–17 Cholesterol. In this chemical structure of cholesterol, the rings are labeled A through D to simplify reference to derivatives of the steroid nucleus; the carbon atoms are numbered in blue. The C-3 hydroxyl group (shaded blue) is the polar head group. For storage and transport of the sterol, this hydroxyl group condenses with a fatty acid to form a sterol ester.

BOX 10–1 MEDICINE Abnormal Accumulations of Membrane Lipids: Some Inherited Human Diseases

The polar lipids of membranes undergo constant metabolic turnover, the rate of their synthesis normally counterbalanced by the rate of breakdown. The breakdown of lipids is promoted by hydrolytic enzymes in lysosomes, each enzyme capable of hydrolyzing a specific bond. When sphingolipid degradation is impaired by a defect in one of these enzymes (Fig. 1), partial breakdown products accumulate in the tissues, causing serious disease.

For example, Niemann-Pick disease is caused by a rare genetic defect in the enzyme sphingomyelinase, which cleaves phosphocholine from sphingomyelin. Sphingomyelin accumulates in the brain, spleen, and liver. The disease becomes evident in infants and causes mental retardation and early death. More common is Tay-Sachs disease, in which ganglioside GM2 accumulates in the brain and spleen (Fig. 2) owing to lack of the enzyme hexosaminidase A. The symptoms of Tay-Sachs disease are progressive developmental retardation, paralysis, blindness, and death by the age of 3 or 4 years.

Genetic counseling can predict and avert many inheritable diseases. Tests on prospective parents can detect abnormal enzymes, then DNA testing can determine the exact nature of the defect and the risk it poses for offspring. Once a pregnancy occurs, fetal cells obtained by sampling a part of the placenta (chorionic villus sampling) or the fluid surrounding the fetus (amniocentesis) can be tested in the same way.

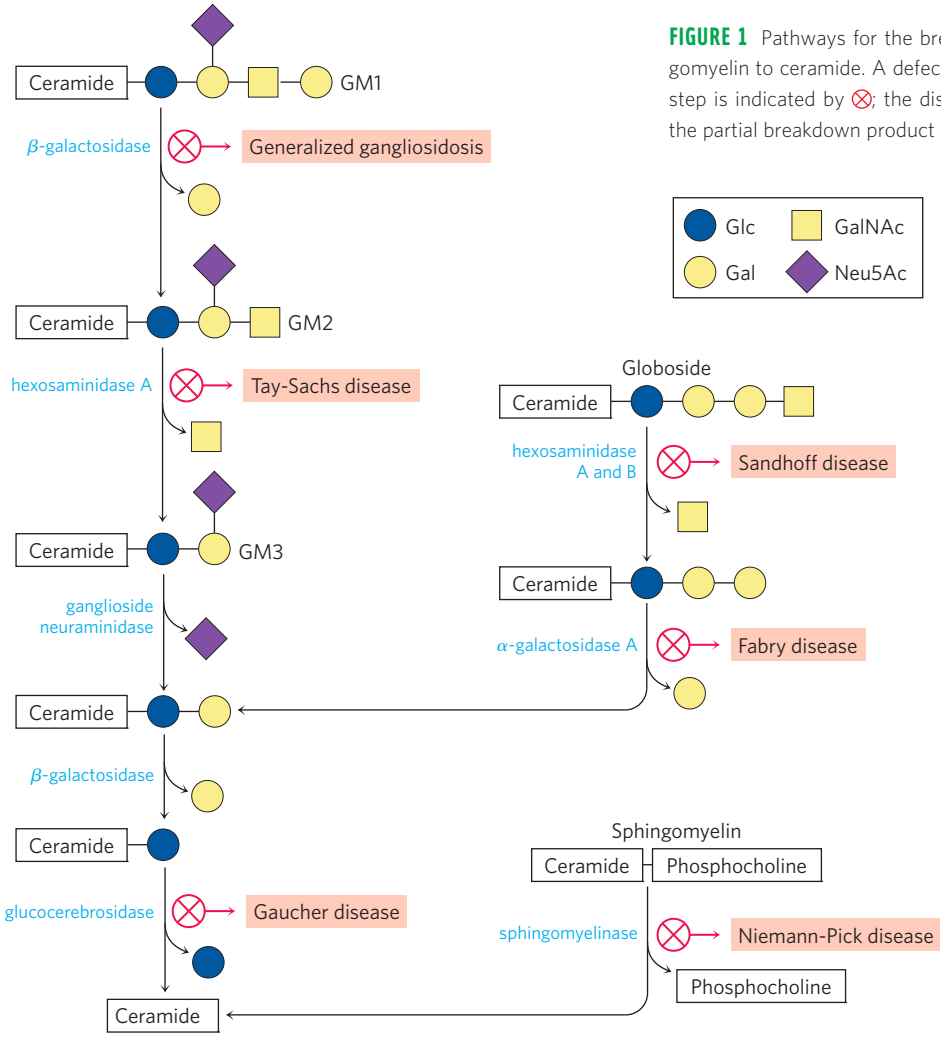


FIGURE 1 Pathways for the breakdown of GM1, globoside, and sphingomyelin to ceramide. A defect in the enzyme hydrolyzing a particular step is indicated by ⊗; the disease that results from accumulation of the partial breakdown product is noted.

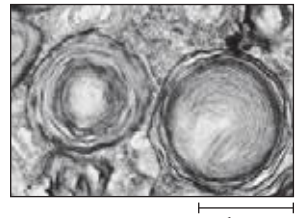
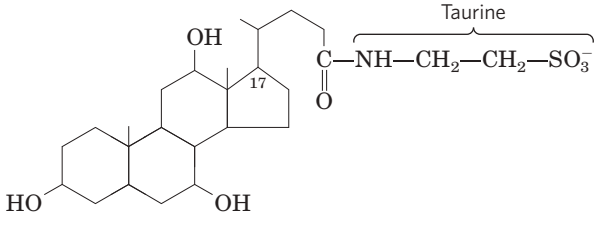


FIGURE 2 Electron micrograph of a portion of a brain cell from an infant with Tay-Sachs disease, obtained post mortem, showing abnormal ganglioside deposits in the lysosomes.

incorporate exogenous sterols into their membranes. The sterols of all eukaryotes are synthesized from simple five-carbon isoprene subunits, as are the fat-soluble vitamins, quinones, and dolichols described in Section 10.3.

In addition to their roles as membrane constituents, the sterols serve as precursors for a variety of products with specific biological activities. Steroid hormones, for example, are potent biological signals that regulate gene

expression. **Bile acids** are polar derivatives of cholesterol that act as detergents in the intestine, emulsifying dietary fats to make them more readily accessible to digestive lipases.



Taurocholic acid
(a bile acid)

We return to cholesterol and other sterols in later chapters, to consider the structural role of cholesterol in biological membranes (Chapter 11), signaling by steroid hormones (Chapter 12), and the remarkable biosynthetic pathway to cholesterol and transport of cholesterol by lipoprotein carriers (Chapter 21).

SUMMARY 10.2 Structural Lipids in Membranes

- The polar lipids, with polar heads and nonpolar tails, are major components of membranes. The most abundant are the glycerophospholipids, which contain fatty acids esterified to two of the hydroxyl groups of glycerol, and a second alcohol, the head group, esterified to the third hydroxyl of glycerol via a phosphodiester bond. Other polar lipids are the sterols.
- Glycerophospholipids differ in the structure of their head group; common glycerophospholipids are phosphatidylethanolamine and phosphatidylcholine. The polar heads of the glycerophospholipids are charged at pH near 7.
- Chloroplast membranes are rich in galactolipids, composed of a diacylglycerol with one or two linked galactose residues, and sulfolipids, diacylglycerols with a linked sulfonated sugar residue and thus a negatively charged head group.
- Some archaea have unique membrane lipids, with long-chain alkyl groups ether-linked to glycerol at both ends and with sugar residues and/or phosphate joined to the glycerol to provide a polar or charged head group. These lipids are stable under the harsh conditions in which these archaea live.
- The sphingolipids contain sphingosine, a long-chain aliphatic amino alcohol, but no glycerol. Sphingomyelin has, in addition to phosphoric acid and choline, two long hydrocarbon chains, one contributed by a fatty acid and the other by sphingosine. Three other classes of sphingolipids are cerebrosides, globosides, and gangliosides, which contain sugar components.
- Sterols have four fused rings and a hydroxyl group. Cholesterol, the major sterol in animals, is both a structural component of membranes and precursor to a wide variety of steroids.

10.3 Lipids as Signals, Cofactors, and Pigments

The two functional classes of lipids considered thus far (storage lipids and structural lipids) are major cellular components; membrane lipids make up 5% to 10% of the dry mass of most cells, and storage lipids more than 80% of the mass of an adipocyte. With some important exceptions, these lipids play a *passive* role in the cell; lipid fuels are stored until oxidized by enzymes, and membrane lipids form impermeable barriers around cells and cellular compartments. Another group of lipids, present in much smaller amounts, have *active* roles in the metabolic traffic as metabolites and messengers. Some serve as potent signals—as hormones, carried in the blood from one tissue to another, or as intracellular messengers generated in response to an extracellular signal (hormone or growth factor). Others function as enzyme cofactors in electron-transfer reactions in chloroplasts and mitochondria, or in the transfer of sugar moieties in a variety of glycosylation reactions. A third group consists of lipids with a system of conjugated double bonds: pigment molecules that absorb visible light. Some of these act as light-capturing pigments in vision and photosynthesis; others produce natural colorations, such as the orange of pumpkins and carrots and the yellow of canary feathers. Finally, a very large group of volatile lipids produced in plants serve as signals that pass through the air, allowing plants to communicate with each other, and to invite animal friends and deter foes. We describe in this section a few representatives of these biologically active lipids. In later chapters, their synthesis and biological roles are considered in more detail.

Phosphatidylinositols and Sphingosine Derivatives Act as Intracellular Signals

Phosphatidylinositol and its phosphorylated derivatives act at several levels to regulate cell structure and metabolism. Phosphatidylinositol 4,5-bisphosphate (Fig. 10–16) in the cytoplasmic (inner) face of plasma membranes serves as a reservoir of messenger molecules that are released inside the cell in response to extracellular signals interacting with specific surface receptors. Extracellular signals such as the hormone vasopressin activate a specific phospholipase C in the membrane, which hydrolyzes phosphatidylinositol 4,5-bisphosphate to release two products that act as intracellular messengers: inositol 1,4,5-trisphosphate (IP_3), which is water-soluble, and diacylglycerol, which remains associated with the plasma membrane. IP_3 triggers release of Ca^{2+} from the endoplasmic reticulum, and the combination of

diacylglycerol and elevated cytosolic Ca^{2+} activates the enzyme protein kinase C. By phosphorylating specific proteins, this enzyme brings about the cell's response to the extracellular signal. This signaling mechanism is described more fully in Chapter 12 (see Fig. 12–10).

Inositol phospholipids also serve as points of nucleation for supramolecular complexes involved in signaling or in exocytosis. Certain signaling proteins bind specifically to phosphatidylinositol 3,4,5-trisphosphate in the plasma membrane, initiating the formation of multienzyme complexes at the membrane's cytosolic surface. Formation of phosphatidylinositol 3,4,5-trisphosphate in response to extracellular signals therefore brings the proteins together in signaling complexes at the surface of the plasma membrane (see Fig. 12–16).

Membrane sphingolipids also can serve as sources of intracellular messengers. Both ceramide and sphingomyelin (Fig. 10–13) are potent regulators of protein kinases, and ceramide or its derivatives are involved in the regulation of cell division, differentiation, migration, and programmed cell death (also called apoptosis; see Chapter 12).

Eicosanoids Carry Messages to Nearby Cells

Eicosanoids are paracrine hormones, substances that act only on cells near the point of hormone synthesis instead of being transported in the blood to act on cells in other tissues or organs. These fatty acid derivatives have a variety of dramatic effects on vertebrate tissues. They are involved in reproductive function; in the inflammation, fever, and pain associated with injury or disease; in the formation of blood clots and the regulation of blood pressure; in gastric acid secretion; and in various other processes important in human health or disease.

All eicosanoids are derived from arachidonic acid ($20:4(\Delta^{5,8,11,14})$) **(Fig. 10–18)**, the 20-carbon polyunsaturated fatty acid from which they take their general name (Greek *eikosi*, "twenty"). There are three classes of eicosanoids: prostaglandins, thromboxanes, and leukotrienes.

Prostaglandins (PG) contain a five-carbon ring originating from the chain of arachidonic acid. Their name derives from the prostate gland, the tissue from which they were first isolated by Bengt Samuelsson and Sune Bergström. Two groups of prostaglandins were originally defined: PGE (*e*ther-soluble) and PGF (*f*osfat (Swedish for phosphate) buffer–soluble). Each group contains numerous subtypes, named PGE_1, PGE_2, PGF_1, and so forth. Prostaglandins have an array of functions. Some stimulate contraction of the smooth muscle of the uterus during menstruation and labor. Others affect blood flow to specific organs, the wake-sleep cycle, and the responsiveness of certain tissues to hormones such as epinephrine and glucagon. Prostaglandins in a third group elevate body temperature (producing fever) and cause inflammation and pain.

The **thromboxanes** have a six-membered ring containing an ether. They are produced by platelets (also called thrombocytes) and act in the formation of blood clots and the reduction of blood flow to the site of a clot. As shown by John Vane, the nonsteroidal antiinflammatory drugs (NSAIDs)—aspirin, ibuprofen, and

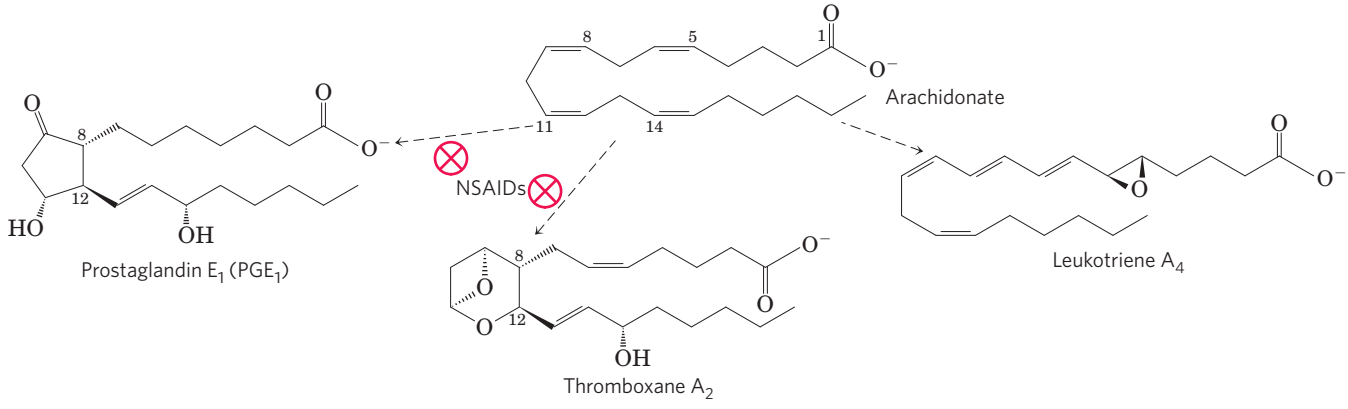


FIGURE 10–18 Arachidonic acid and some eicosanoid derivatives. Arachidonic acid (arachidonate at pH 7) is the precursor of eicosanoids, including the prostaglandins, thromboxanes, and leukotrienes. In prostaglandin E_1, C-8 and C-12 of arachidonate are joined to form the characteristic five-membered ring. In thromboxane A_2, the C-8 and C-12 are joined and an oxygen atom is added to form the six-membered ring. Leukotriene A_4 has a series of three conjugated double bonds. Nonsteroidal antiinflammatory drugs (NSAIDs) such as aspirin and ibuprofen block the formation of prostaglandins and thromboxanes from arachidonate by inhibiting the enzyme cyclooxygenase (prostaglandin H_2 synthase).

John Vane, Sune Bergström, and Bengt Samuelsson

meclofenamate, for example—inhibit the enzyme prostaglandin H_2 synthase (also called cyclooxygenase, or COX), which catalyzes an early step in the pathway from arachidonate to prostaglandins and thromboxanes (Fig. 10–18; see also Fig. 21–15).

Leukotrienes, first found in leukocytes, contain three conjugated double bonds. They are powerful biological signals. For example, leukotriene D_4, derived from leukotriene A_4, induces contraction of the smooth muscle lining the airways to the lung. Overproduction of leukotrienes causes asthmatic attacks, and leukotriene synthesis is one target of antiasthmatic drugs such as prednisone. The strong contraction of the smooth muscle of the lungs that occurs during anaphylactic shock is part of the potentially fatal allergic reaction in individuals hypersensitive to bee stings, penicillin, or other agents. ■

Steroid Hormones Carry Messages between Tissues

Steroids are oxidized derivatives of sterols; they have the sterol nucleus but lack the alkyl chain attached to ring D of cholesterol, and they are more polar than cholesterol. Steroid hormones move through the bloodstream (on protein carriers) from their site of production to target tissues, where they enter cells, bind to highly specific receptor proteins in the nucleus, and trigger changes in gene expression and thus metabolism. Because hormones have very high affinity for their receptors, very low concentrations of hormones (nanomolar or less) are sufficient to produce responses in target tissues. The major groups of steroid hormones are the male and female sex hormones and the hormones produced by the adrenal cortex, cortisol and aldosterone **(Fig. 10–19)**. Prednisone and prednisolone are steroid drugs with potent antiinflammatory activities, mediated in part by the inhibition of arachidonate release by phospholipase A_2 and consequent inhibition of the synthesis of leukotrienes, prostaglandins, and thromboxanes. These drugs have a variety of medical applications, including the treatment of asthma and rheumatoid arthritis. ■

Vascular plants contain the steroidlike brassinolide (Fig. 10–19), a potent growth regulator that increases the rate of stem elongation and affects the orientation of cellulose microfibrils in the cell wall during growth.

Vascular Plants Produce Thousands of Volatile Signals

Plants produce literally thousands of different lipophilic compounds, volatile substances that are used to attract pollinators, to repel herbivores, to attract organisms that defend the plant against herbivores, and to communicate with other plants. Jasmonate, for example (see Fig. 12–33), derived from the fatty acid 18:3($\Delta^{9,12,15}$) in membrane lipids, triggers the plant's defenses in response to insect-inflicted damage. The methyl ester of jasmonate gives the characteristic fragrance of jasmine oil, which is

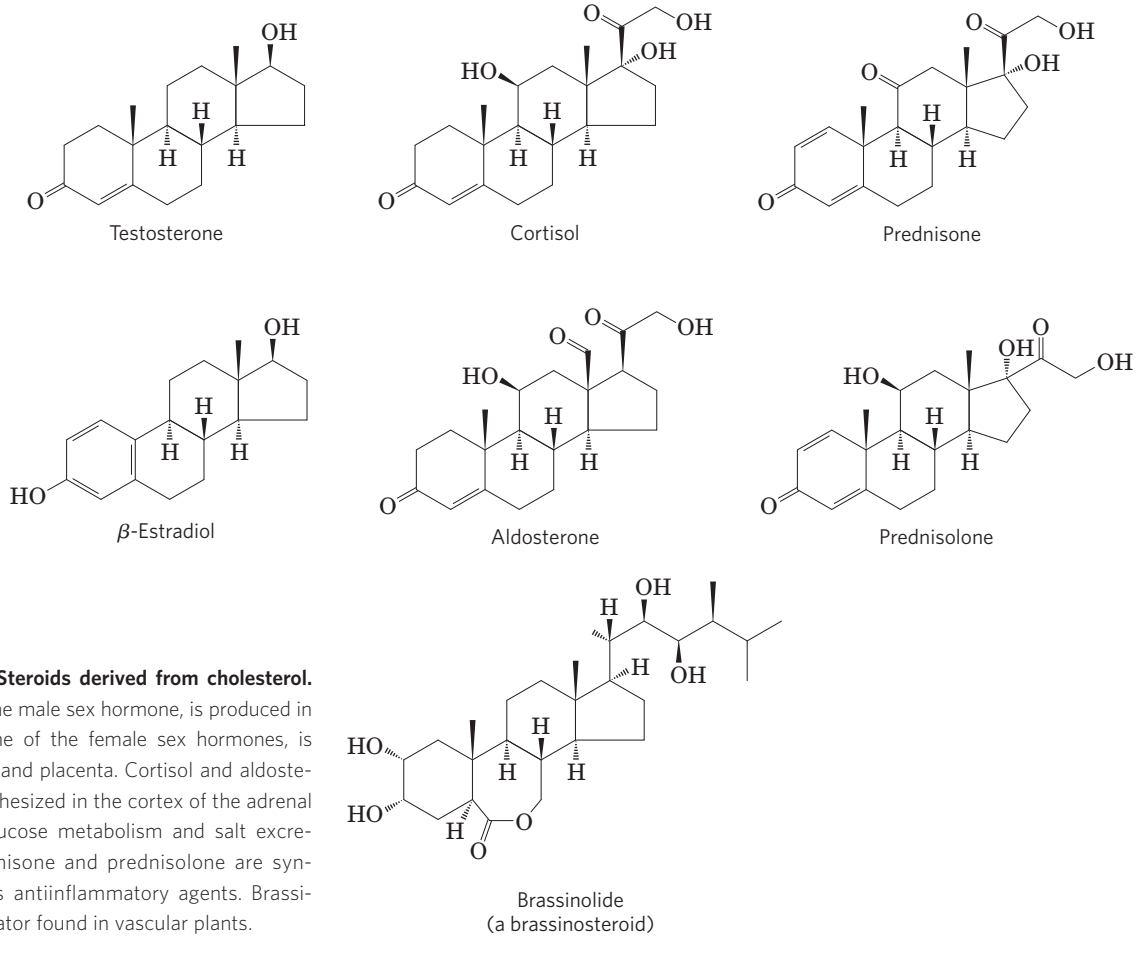
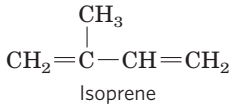


FIGURE 10–19 Steroids derived from cholesterol. Testosterone, the male sex hormone, is produced in the testes. Estradiol, one of the female sex hormones, is produced in the ovaries and placenta. Cortisol and aldosterone are hormones synthesized in the cortex of the adrenal gland; they regulate glucose metabolism and salt excretion, respectively. Prednisone and prednisolone are synthetic steroids used as antiinflammatory agents. Brassinolide is a growth regulator found in vascular plants.

widely used in the perfume industry. Many of the plant volatiles are derived from fatty acids, or from compounds made by the condensation of five-carbon isoprene units; these include geraniol (the characteristic scent of geraniums), β-pinene (pine trees), limonene (limes), menthol, and carvone (see Fig. 1–24a), to name but a few.



Isoprene

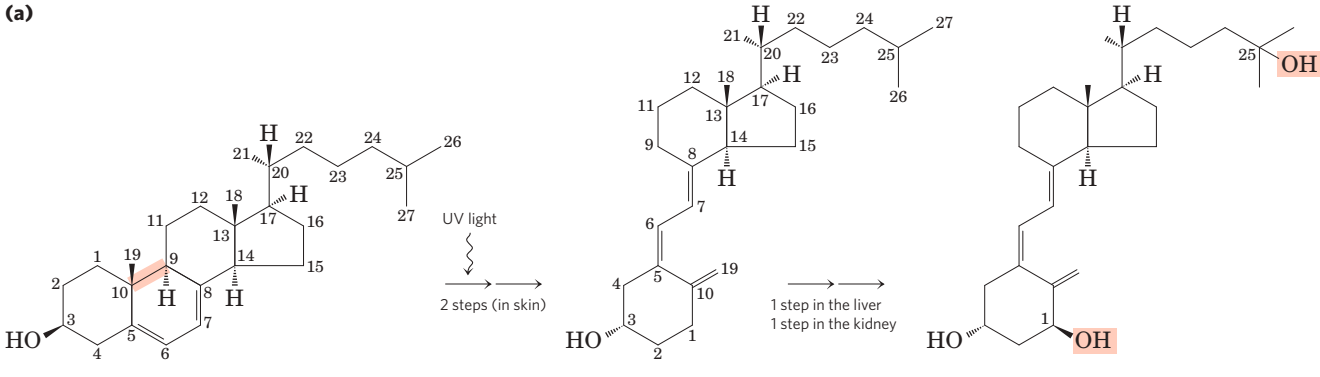
Vitamins A and D Are Hormone Precursors

During the first third of the twentieth century, a major focus of research in physiological chemistry was the identification of **vitamins**, compounds that are essential to the health of humans and other vertebrates but cannot be synthesized by these animals and must therefore be obtained in the diet. Early nutritional studies identified two general classes of such compounds: those soluble in nonpolar organic solvents (fat-soluble vitamins) and those that could be extracted from foods with aqueous solvents (water-soluble vitamins). Eventually the fat-soluble group was resolved into the four vitamin groups A, D, E, and K, all of which are isoprenoid compounds synthesized by the condensation of multiple isoprene units. Two of these (D and A) serve as hormone precursors.

Vitamin D_3, also called **cholecalciferol**, is normally formed in the skin from 7-dehydrocholesterol in a photochemical reaction driven by the UV component of sunlight **(Fig. 10–20a)**. Vitamin D_3 is not itself biologically active, but it is converted by enzymes in the liver and kidney to 1α,25-dihydroxyvitamin D_3 (calcitriol), a hormone that regulates calcium uptake in the intestine and calcium levels in kidney and bone. Deficiency of vitamin D leads to defective bone formation and the disease rickets, for which administration of vitamin D produces a dramatic cure (Fig. 10–20b). Vitamin D_2 (ergocalciferol) is a commercial product formed by UV irradiation of the ergosterol of yeast. Vitamin D_2 is structurally similar to D_3, with slight modification to the side chain attached to the sterol D ring. Both have the same biological effects, and D_2 is commonly added to milk and butter as a dietary supplement. Like steroid hormones, the product of vitamin D metabolism, 1α,25-dihydroxyvitamin D_3, regulates gene expression by interacting with specific nuclear receptor proteins (pp. 1182–1183).

Vitamin A (retinol), in its various forms, functions as a hormone and as the visual pigment of the vertebrate eye **(Fig. 10–21)**. Acting through receptor proteins in the cell nucleus, the vitamin A derivative retinoic acid regulates gene expression in the development of epithelial tissue, including skin. Retinoic acid is the active ingredient in the drug tretinoin (Retin-A), used in the treatment of severe acne and wrinkled skin. Retinal, another vitamin A derivative, is the pigment that initiates the response of rod and cone cells of the retina to light, producing a

(a)



7-Dehydrocholesterol — Cholecalciferol (vitamin D_3) — 1α,25-Dihydroxyvitamin D_3 (calcitriol)

FIGURE 10–20 Vitamin D_3 production and metabolism. **(a)** Cholecalciferol (vitamin D_3) is produced in the skin by UV irradiation of 7-dehydrocholesterol, which breaks the bond shaded light red. In the liver, a hydroxyl group is added at C-25; in the kidney, a second hydroxylation at C-1 produces the active hormone, 1α,25-dihydroxyvitamin D_3. This hormone regulates the metabolism of Ca^{2+} in kidney, intestine, and bone. **(b)** Dietary vitamin D prevents rickets, a disease once common in cold climates where heavy clothing blocks the UV component of sunlight necessary for the production of vitamin D_3 in skin. In this detail from a large mural by John Steuart Curry, *The Social Benefits of Biochemical Research* (1943), the people and animals on the left show the effects of poor nutrition, including the bowed legs of a boy with classical rickets. On the right are the people and animals made healthier with the "social benefits of research," including the use of vitamin D to prevent and treat rickets. This mural is in the Department of Biochemistry at the University of Wisconsin–Madison.


(b)

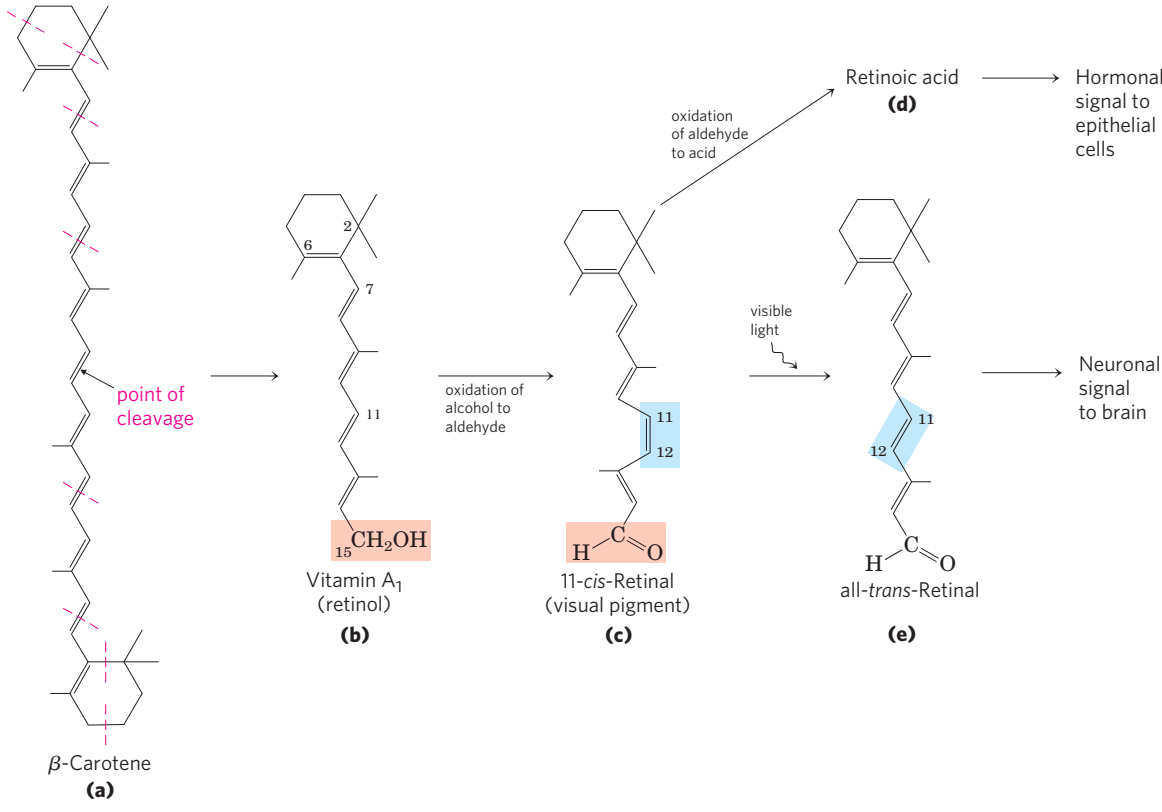


FIGURE 10–21 Vitamin A_1 and its precursor and derivatives. (a) β-Carotene is the precursor of vitamin A_1. Isoprene structural units are set off by dashed red lines (see p. 373). Cleavage of β-carotene yields two molecules of vitamin A_1 (retinol) **(b)**. Oxidation at C-15 converts retinol to the aldehyde, retinal **(c)**, and further oxidation produces retinoic acid **(d)**, a hormone that regulates gene expression. Retinal combines with the protein opsin to form rhodopsin (not shown), a visual pigment widespread in nature. In the dark, retinal of rhodopsin is in the 11-*cis* form (c). When a rhodopsin molecule is excited by visible light, the 11-*cis*-retinal undergoes a series of photochemical reactions that convert it to all-*trans*-retinal **(e)**, forcing a change in the shape of the entire rhodopsin molecule. This transformation in the rod cell of the vertebrate retina sends an electrical signal to the brain that is the basis of visual transduction, a topic we address in more detail in Chapter 12.

neuronal signal to the brain. This role of retinal is described in detail in Chapter 12.

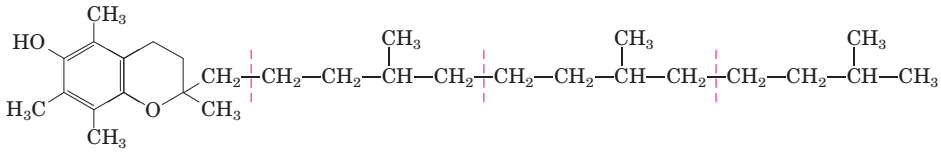
Vitamin A was first isolated from fish liver oils; liver, eggs, whole milk, and butter are also good dietary sources. In vertebrates, β-carotene, the pigment that gives carrots, sweet potatoes, and other yellow vegetables their characteristic color, can be enzymatically converted to vitamin A. Deficiency of vitamin A leads to a variety of symptoms in humans, including dryness of the skin, eyes, and mucous membranes; retarded development and growth; and night blindness, an early symptom commonly used in diagnosing vitamin A deficiency. ■

Vitamins E and K and the Lipid Quinones Are Oxidation-Reduction Cofactors

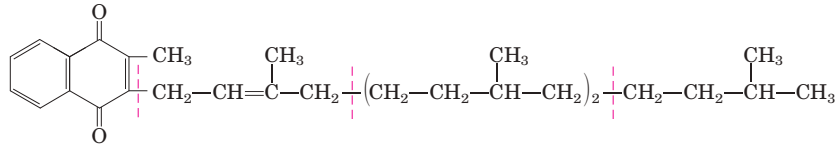
Vitamin E is the collective name for a group of closely related lipids called **tocopherols**, all of which contain a substituted aromatic ring and a long isoprenoid side chain **(Fig. 10–22a)**. Because they are hydrophobic, tocopherols associate with cell membranes, lipid deposits, and lipoproteins in the blood. Tocopherols are biological antioxidants. The aromatic ring reacts with and destroys the most reactive forms of oxygen radicals and other free radicals, protecting unsaturated fatty acids from oxidation and preventing oxidative damage to membrane lipids, which can cause cell fragility. Tocopherols are found in eggs and vegetable oils and are especially abundant in wheat germ. Laboratory animals fed diets depleted of vitamin E develop scaly skin, muscular weakness and wasting, and sterility. Vitamin E deficiency in humans is very rare; the principal symptom is fragile erythrocytes.

The aromatic ring of **vitamin K** (Fig. 10–22b) undergoes a cycle of oxidation and reduction during the formation of active prothrombin, a blood plasma protein essential in blood clotting. Prothrombin is a proteolytic enzyme that splits peptide bonds in the blood protein fibrinogen to convert it to fibrin, the insoluble fibrous protein that holds blood clots together (see Fig. 6–39). Henrik Dam and Edward A. Doisy independently discovered that vitamin K deficiency slows blood clotting, which can be fatal. Vitamin K deficiency is very uncommon in humans, aside from a small percentage of infants who suffer from hemorrhagic disease of the newborn, a potentially fatal disorder. In the United States, newborns are routinely given a 1 mg injection of vitamin K. Vitamin K_1 (phylloquinone) is found in green plant leaves; a related form, vitamin K_2 (menaquinone), is formed by bacteria living in the vertebrate intestine.

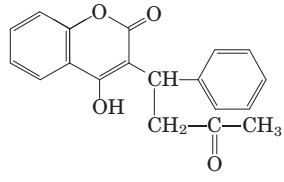
(a) Vitamin E: an antioxidant



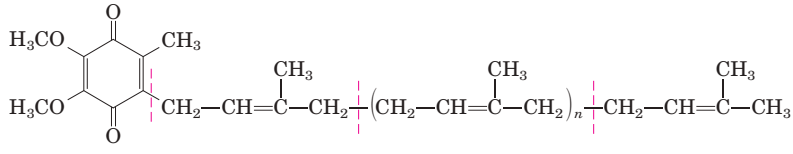
(b) Vitamin K_1: a blood-clotting cofactor (phylloquinone)



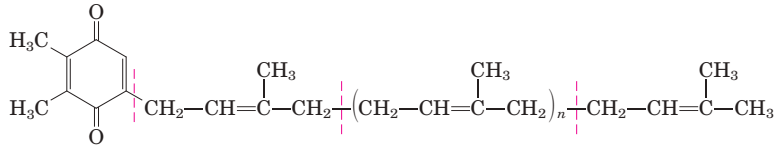
(c) Warfarin: a blood anticoagulant



(d) Ubiquinone: a mitochondrial electron carrier (coenzyme Q) (n = 4 to 8)



(e) Plastoquinone: a chloroplast electron carrier (n = 4 to 8)



(f) Dolichol: a sugar carrier (n = 9 to 22)

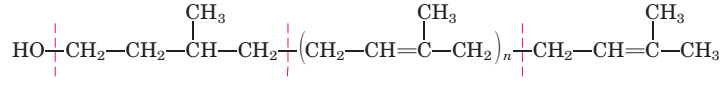


FIGURE 10–22 Some other biologically active isoprenoid compounds or derivatives. Units derived from isoprene are set off by dashed red lines. In most mammalian tissues, ubiquinone (also called coenzyme Q) has 10 isoprene units. Dolichols of animals have 17 to 21 isoprene units (85 to 105 carbon atoms), bacterial dolichols have 11, and those of plants and fungi have 14 to 24.

Henrik Dam, 1895–1976

Edward A. Doisy, 1893–1986

Warfarin (Fig. 10–22c) is a synthetic compound that inhibits the formation of active prothrombin. It is particularly poisonous to rats, causing death by internal bleeding. Ironically, this potent rodenticide is also an invaluable anticoagulant drug for treating humans at risk for excessive blood clotting, such as surgical patients and those with coronary thrombosis. ■

Ubiquinone (also called coenzyme Q) and plastoquinone (Fig. 10–22d, e) are isoprenoids that function as lipophilic electron carriers in the oxidation-reduction reactions that drive ATP synthesis in mitochondria and chloroplasts, respectively. Both ubiquinone and plastoquinone can accept either one or two electrons and either one or two protons (see Fig. 19–3).

Dolichols Activate Sugar Precursors for Biosynthesis

During assembly of the complex carbohydrates of bacterial cell walls, and during the addition of polysaccharide units to certain proteins (glycoproteins) and lipids (glycolipids) in eukaryotes, the sugar units to be added are chemically activated by attachment to isoprenoid alcohols called **dolichols** (Fig. 10–22f). These compounds have strong hydrophobic interactions with membrane lipids, anchoring the attached sugars to the membrane, where they participate in sugar-transfer reactions.

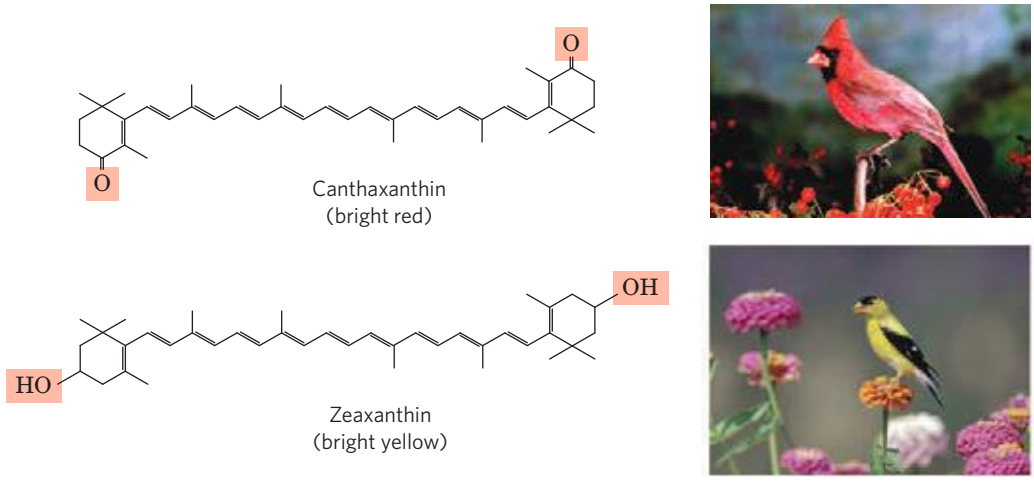


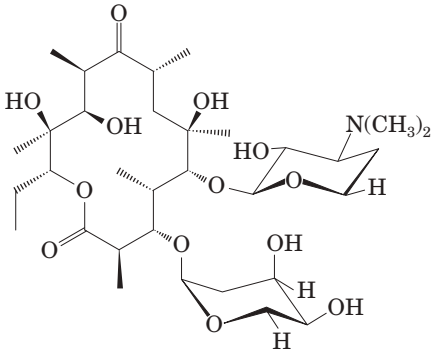
FIGURE 10–23 Lipids as pigments in plants and bird feathers. Compounds with long conjugated systems absorb light in the visible region of the spectrum. Subtle differences in the chemistry of these compounds produce pigments of strikingly different colors. Birds acquire the pigments that color their feathers red or yellow by eating plant materials that contain carotenoid pigments, such as canthaxanthin and zeaxanthin. The differences in pigmentation between male and female birds are the result of differences in intestinal uptake and processing of carotenoids.

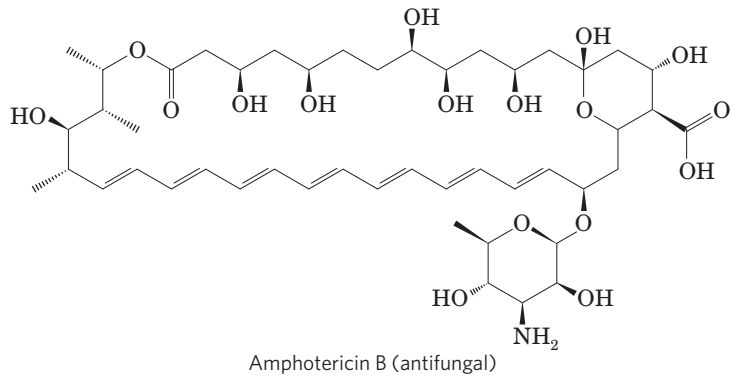
Many Natural Pigments Are Lipidic Conjugated Dienes

Conjugated dienes have carbon chains with alternating single and double bonds. Because this structural arrangement allows the delocalization of electrons, the compounds can be excited by low-energy electromagnetic radiation (visible light), giving them colors visible to humans and other animals. Carotene (Fig. 10–21) is yellow-orange; similar compounds give bird feathers their striking reds, oranges, and yellows **(Fig. 10–23)**. Like sterols, steroids, dolichols, vitamins A, E, D, and K, ubiquinone, and plastoquinone, these pigments are synthesized from five-carbon isoprene derivatives; the biosynthetic pathway is described in detail in Chapter 21.

Polyketides Are Natural Products with Potent Biological Activities

Polyketides are a diverse group of lipids with biosynthetic pathways (Claisen condensations) similar to those for fatty acids. They are **secondary metabolites**, compounds that are not central to an organism's metabolism but that serve some subsidiary function that gives their producers an advantage in some ecological niche. Many polyketides find use in medicine as antibiotics (erythromycin), antifungals (amphotericin B), or inhibitors of cholesterol synthesis (lovastatin) **(Fig. 10–24)**.





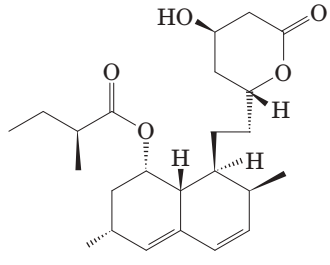


FIGURE 10–24 Three polyketide natural products used in human medicine.

SUMMARY 10.3 Lipids as Signals, Cofactors, and Pigments

- Some types of lipids, although present in relatively small quantities, play critical roles as cofactors or signals.
- Phosphatidylinositol bisphosphate is hydrolyzed to yield two intracellular messengers, diacylglycerol and inositol 1,4,5-trisphosphate. Phosphatidylinositol 3,4,5-trisphosphate is a nucleation point for supramolecular protein complexes involved in biological signaling.
- Prostaglandins, thromboxanes, and leukotrienes (the eicosanoids), derived from arachidonate, are extremely potent hormones.
- Steroid hormones, such as the sex hormones, are derived from sterols. They serve as powerful biological signals, altering gene expression in target cells.
- Vitamins D, A, E, and K are fat-soluble compounds made up of isoprene units. All play essential roles in the metabolism or physiology of animals. Vitamin D is precursor to a hormone that regulates calcium metabolism. Vitamin A furnishes the visual pigment of the vertebrate eye and is a regulator of gene expression during epithelial cell growth. Vitamin E functions in the protection of membrane lipids from oxidative damage, and vitamin K is essential in the blood-clotting process.
- Ubiquinones and plastoquinones, also isoprenoid derivatives, are electron carriers in mitochondria and chloroplasts, respectively.
- Dolichols activate and anchor sugars to cellular membranes; the sugar groups are then used in the synthesis of complex carbohydrates, glycolipids, and glycoproteins.
- Lipidic conjugated dienes serve as pigments in flowers and fruits and give bird feathers their striking colors.
- Polyketides are natural products widely used in medicine.

10.4 Working with Lipids

Because lipids are insoluble in water, their extraction and subsequent fractionation require the use of organic solvents and some techniques not commonly used in the purification of water-soluble molecules such as proteins and carbohydrates. In general, complex mixtures of lipids are separated by differences in polarity or solubility in nonpolar solvents. Lipids that contain ester- or amide-linked fatty acids can be hydrolyzed by treatment with acid or alkali or with specific hydrolytic enzymes (phospholipases, glycosidases) to yield their components for analysis. Some methods commonly used in lipid analysis are shown in **Figure 10–25** and discussed below.

Lipid Extraction Requires Organic Solvents

Neutral lipids (triacylglycerols, waxes, pigments, and so forth) are readily extracted from tissues with ethyl ether, chloroform, or benzene, solvents that do not permit lipid clustering driven by hydrophobic interactions. Membrane lipids are more effectively extracted

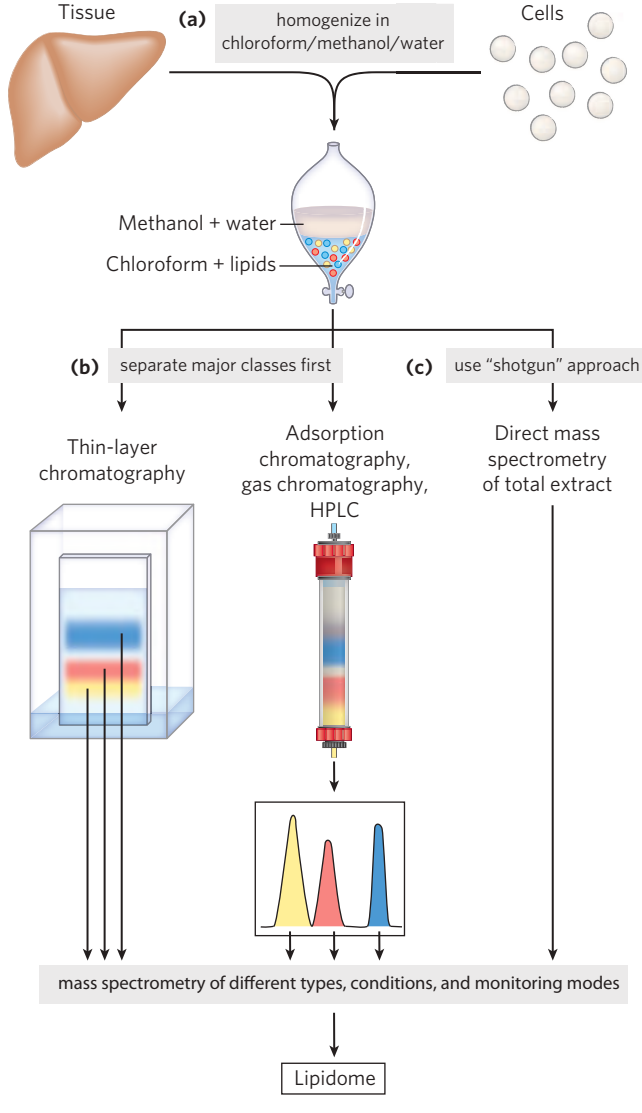


FIGURE 10-25 Common procedures in the extraction, separation, and identification of cellular lipids. (a) Tissue is homogenized in a chloroform/methanol/water mixture, which on addition of water and removal of unextractable sediment by centrifugation yields two phases. **(b)** Major classes of extracted lipids in the chloroform phase may first be separated by thin-layer chromatography (TLC), in which lipids are carried up a silica gel–coated plate by a rising solvent front, less-polar lipids traveling farther than more-polar or charged lipids, or by adsorption chromatography on a column of silica gel, through which solvents of increasing polarity are passed. For example, column chromatography with appropriate solvents can be used to separate closely related lipid species such as phosphatidylserine, phosphatidylglycerol, and phosphatidylinositol. Once separated, each lipid's complement of fatty acids can be determined by mass spectrometry. **(c)** Alternatively, in the "shotgun" approach, an unfractionated extract of lipids can be directly subjected to high-resolution mass spectrometry of different types and under different conditions to determine the total composition of all the lipids: the lipidome.

by more polar organic solvents, such as ethanol or methanol, which reduce the hydrophobic interactions among lipid molecules while also weakening the hydrogen bonds and electrostatic interactions that bind membrane lipids to membrane proteins. A commonly used extractant is a mixture of chloroform, methanol, and water, initially in volume proportions (1:2:0.8) that are miscible, producing a single phase. After tissue is homogenized in this solvent to extract all lipids, more water is added to the resulting extract and the mixture separates into two phases, methanol/water (top phase) and chloroform (bottom phase). The lipids remain in the chloroform layer, and the more polar molecules such as proteins and sugars partition into the methanol/water layer (Fig. 10–25a).

Adsorption Chromatography Separates Lipids of Different Polarity

Complex mixtures of tissue lipids can be fractionated by chromatographic procedures based on the different polarities of each class of lipid (Fig. 10–25b). In adsorption chromatography, an insoluble, polar material such as silica gel (a form of silicic acid, $Si(OH)_4$) is packed into a glass column, and the lipid mixture (in chloroform solution) is applied to the top of the column. (In high-performance liquid chromatography, the column is of smaller diameter and solvents are forced through the column under high pressure.) The polar lipids bind tightly to the polar silicic acid, but the neutral lipids pass directly through the column and emerge in the first chloroform wash. The polar lipids are then eluted, in order of increasing polarity, by washing the column with solvents of progressively higher polarity. Uncharged but polar lipids (cerebrosides, for example) are eluted with acetone, and very polar or charged lipids (such as glycerophospholipids) are eluted with methanol.

Thin-layer chromatography on silicic acid employs the same principle (Fig. 10–25b). A thin layer of silica gel is spread onto a glass plate, to which it adheres. A small sample of lipids dissolved in chloroform is applied near one edge of the plate, which is dipped in a shallow container of an organic solvent or solvent mixture; the entire setup is enclosed in a chamber saturated with the solvent vapor. As the solvent rises on the plate by capillary action, it carries lipids with it. The less polar lipids move farthest, as they have less tendency to bind to the silicic acid. The separated lipids can be detected by spraying the plate with a dye (rhodamine) that fluoresces when associated with lipids, or by exposing the plate to iodine fumes. Iodine reacts reversibly with the double bonds in fatty acids, such that lipids containing unsaturated fatty acids develop a yellow or brown color. Several other spray reagents are also useful in detecting specific lipids. For subsequent analysis, regions containing separated lipids can be scraped from the plate and the lipids recovered by extraction with an organic solvent.

Gas-Liquid Chromatography Resolves Mixtures of Volatile Lipid Derivatives

Gas-liquid chromatography separates volatile components of a mixture according to their relative tendencies to dissolve in the inert material packed in the chromatography column or to volatilize and move through the column, carried by a current of an inert gas such as helium. Some lipids are naturally volatile, but most must first be derivatized to increase their volatility (that is, lower their boiling point). For an analysis of the fatty acids in a sample of phospholipids, the lipids are first transesterified: heated in a methanol/HCl or methanol/NaOH mixture to convert fatty acids esterified to glycerol into their methyl esters. These fatty acyl methyl esters are then loaded onto the gas-liquid chromatography column, and the column is heated to volatilize the compounds. Those fatty acyl esters most soluble in the column material partition into (dissolve in) that material; the less soluble lipids are carried by the stream of inert gas and emerge first from the column. The order of elution depends on the nature of the solid adsorbant in the column and on the boiling point of the components of the lipid mixture. Using these techniques, mixtures of fatty acids of various chain lengths and various degrees of unsaturation can be completely resolved.

Specific Hydrolysis Aids in Determination of Lipid Structure

Certain classes of lipids are susceptible to degradation under specific conditions. For example, all ester-linked fatty acids in triacylglycerols, phospholipids, and sterol esters are released by mild acid or alkaline treatment, and somewhat harsher hydrolysis conditions release amide-bound fatty acids from sphingolipids. Enzymes that specifically hydrolyze certain lipids are also useful in the determination of lipid structure. Phospholipases A, C, and D (Fig. 10–16) each split particular bonds in phospholipids and yield products with characteristic solubilities and chromatographic behaviors. Phospholipase C, for example, releases a water-soluble phosphoryl alcohol (such as phosphocholine from phosphatidylcholine) and a chloroform-soluble diacylglycerol, each of which can be characterized separately to determine the structure of the intact phospholipid. The combination of specific hydrolysis with characterization of the products by thin-layer, gas-liquid, or high-performance liquid chromatography often allows determination of a lipid structure.

Mass Spectrometry Reveals Complete Lipid Structure

To establish unambiguously the length of a hydrocarbon chain or the position of double bonds, mass spectrometric analysis of lipids or their volatile derivatives is invaluable. The chemical properties of similar lipids (for example, two fatty acids of similar length unsaturated at different positions, or two isoprenoids with different numbers of isoprene units) are very much alike, and their order of

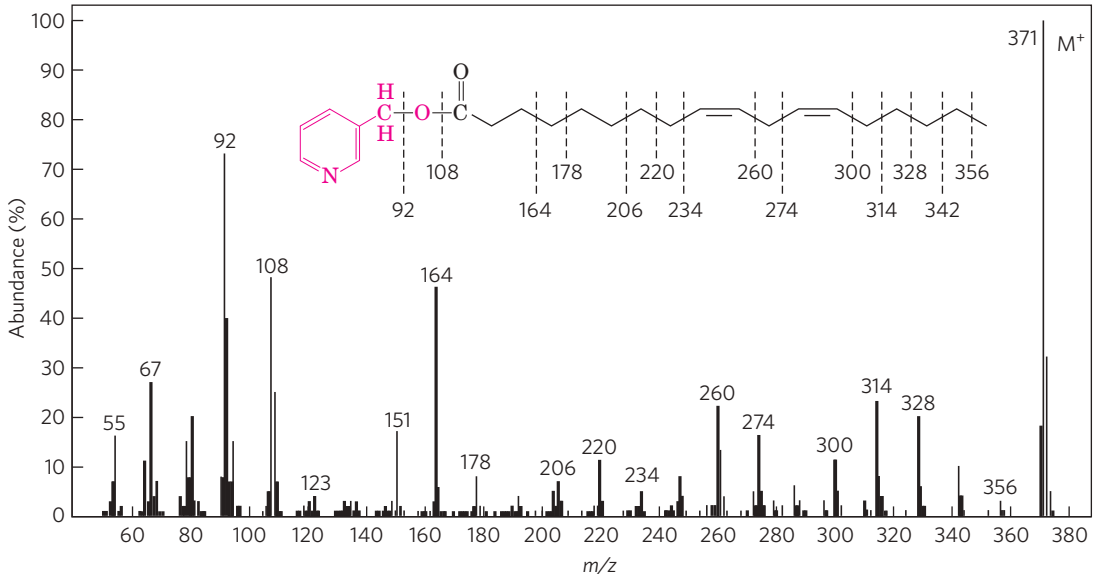


FIGURE 10–26 Determination of fatty acid structure by mass spectrometry. The fatty acid is first converted to a derivative that minimizes migration of the double bonds when the molecule is fragmented by electron bombardment. The derivative shown here is a picolinyl ester of linoleic acid—18:2($\Delta^{9,12}$) (M_r 371)—in which the alcohol is picolinol (red). When bombarded with a stream of electrons, this molecule is volatilized and converted to a parent ion (M^+; M_r 371), in which the N atom bears the positive charge, and a series of smaller fragments produced by breakage of C—C bonds in the fatty acid. The mass spectrometer separates these charged fragments according to their mass/charge ratio (m/z). (To review the principles of mass spectrometry, see pp. 100–102.)

The prominent ions at m/z = 92, 108, 151, and 164 contain the pyridine ring of the picolinol and various fragments of the carboxyl group, showing that the compound is indeed a picolinyl ester. The molecular ion, M^+ (m/z = 371), confirms the presence of a C_{18} fatty acid with two double bonds. The uniform series of ions 14 atomic mass units (u) apart represents loss of each successive methyl and methylene group from the methyl end of the acyl chain (beginning at C-18; the right end of the molecule as shown here), until the ion at m/z = 300 is reached. This is followed by a gap of 26 u for the carbons of the terminal double bond, at m/z = 274; a further gap of 14 u for the C-11 methylene group, at m/z = 260; and so forth. By this means the entire structure is determined, although these data alone do not reveal the configuration (cis or trans) of the double bonds.

elution from the various chromatographic procedures often does not distinguish between them. When the eluate from a chromatography column is sampled by mass spectrometry, however, the components of a lipid mixture can be simultaneously separated and identified by their unique pattern of fragmentation **(Fig. 10–26)**. With the increased resolution of mass spectrometry, it is possible to identify individual lipids in very complex mixtures without first fractionating the lipids in a crude extract. This "shotgun" method (Fig. 10–25c) avoids losses during preliminary separation of lipid subclasses, and it is faster.

Lipidomics Seeks to Catalog All Lipids and Their Functions

In exploring the biological role of lipids in cells and tissues, it is important to know which lipids are present and in what proportions, and to know how this lipid composition changes with embryonic development, disease, or drug treatment. As lipid biochemists have become aware of the thousands of different naturally occurring lipids, they have proposed a new nomenclature system, with the aim of making it easier to compile and search databases of lipid composition. The system places each lipid in one of eight chemical groups (Table 10–3) designated by two letters. Within these groups, finer distinctions are indicated by numbered classes and subclasses. For example, all glycerophosphocholines are GP01; the subgroup of glycerophosphocholines with two fatty acids in ester linkage is designated GP0101; with one fatty acid ether-linked at position 1 and one in ester linkage at position 2, this becomes GP0102. Specific fatty acids are designated by numbers that give every lipid its own unique identifier, so that each individual lipid, including lipid types not yet discovered, can be unambiguously described in terms of a 12-character identifier. One factor used in this classification is the nature of the biosynthetic precursor. For example, prenol lipids (dolichols and vitamins E and K, for example) are formed from isoprenyl precursors. Polyketides include some natural products, many toxic, with biosynthetic pathways related to those for fatty acids. The eight chemical categories in Table 10–3 do not coincide perfectly with the divisions according to biological function that we have used in this chapter. For example, the structural lipids of membranes include both glycerophospholipids and sphingolipids, separate categories in Table 10–3. Each method of categorization has its advantages.

The application of mass spectrometric techniques with high throughput and high resolution can provide quantitative catalogs of all the lipids present in a specific cell type under particular conditions—the **lipidome**—and of the ways in which the lipidome changes with differentiation, disease such as cancer, or drug treatment.

TABLE 10–3 Eight Major Categories of Biological Lipids

Category	Category code	Examples
Fatty acids	FA	Oleate, stearoyl-CoA, palmitoylcarnitine
Glycerolipids	GL	Di- and triacylglycerols
Glycerophospholipids	GP	Phosphatidylcholine, phosphatidylserine, phosphatidylethanolamine
Sphingolipids	SP	Sphingomyelin, ganglioside GM2
Sterol lipids	ST	Cholesterol, progesterone, bile acids
Prenol lipids	PR	Farnesol, geraniol, retinol, ubiquinone
Saccharolipids	SL	Lipopolysaccharide
Polyketides	PK	Tetracycline, erythromycin, aflatoxin B_1

An animal cell contains more than a thousand different lipid species, each presumably having a specific function. These functions are known for a growing number of lipids, but the still largely unexplored lipidome offers a rich source of new problems for the next generation of biochemists and cell biologists to solve.

SUMMARY 10.4 Working with Lipids

- In the determination of lipid composition, the lipids are first extracted from tissues with organic solvents and separated by thin-layer, gas-liquid, or high-performance liquid chromatography.
- Phospholipases specific for one of the bonds in a phospholipid can be used to generate simpler compounds for subsequent analysis.
- Individual lipids are identified by their chromatographic behavior, their susceptibility to hydrolysis by specific enzymes, or mass spectrometry.
- High-resolution mass spectrometry allows the analysis of crude mixtures of lipids without prefractionation—the "shotgun" approach.
- Lipidomics combines powerful analytical techniques to determine the full complement of lipids in a cell or tissue (the lipidome) and to assemble annotated databases that allow comparisons between lipids of different cell types and under different conditions.

Key Terms

Terms in bold are defined in the glossary.

fatty acid 357
polyunsaturated fatty acid (PUFA) 359
triacylglycerol 360
lipases 360
phospholipid 363
glycolipid 363
glycerophospholipid 363
ether lipid 364
plasmalogen 364
galactolipid 365
sphingolipid 366
ceramide 366
sphingomyelin 366
glycosphingolipid 366
cerebroside 366
globoside 366
ganglioside 366
sterol 368
cholesterol 368
prostaglandin 371
thromboxane 371
leukotriene 372
vitamin 373
vitamin D_3 373
cholecalciferol 373
vitamin A (retinol) 373
vitamin E 374
tocopherol 374
vitamin K 374
dolichol 375
polyketide 376
lipidome 379

Further Reading

General

Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H., Jr., Murphy, R.C., Raetz, C.R.H., Russell, D.W., Seyama, Y., Shaw, W., et al. (2005) A comprehensive classification system for lipids. *J. Lipid Res.* **46**, 839–862.

A new system of nomenclature for biological lipids, separating them into eight major categories. The definitive reference on lipid classification.

Gurr, M.I., Harwood, J.L., & Frayn, K.N. (2002) *Lipid Biochemistry: An Introduction,* 5th edn, Blackwell Science Ltd., Oxford.

A good general resource on lipid structure and metabolism, at the intermediate level.

Lipid Maps. *Nature* Lipidomics Gateway, www.lipidmaps.org.

Tutorials and lectures on lipid structure and function, lipidomics methods, and databases on lipid structures and properties.

Vance, J.E. & Vance, D.E. (eds). (2008) *Biochemistry of Lipids, Lipoproteins, and Membranes,* 5th edn, Elsevier Science Publishing Co., Inc., New York.

An excellent collection of reviews on various aspects of lipid structure, biosynthesis, and function.

Lipids as Nutrients

Angerer, P. & von Schacky, C. (2000) Omega-3 polyunsaturated fatty acids and the cardiovascular system. *Curr. Opin. Lipidol.* **11**, 57–63.

Covington, M.B. (2004) Omega-3 fatty acids. *Am. Fam. Physician* **70**, 133–140.

Succinct statement of the findings that omega-3 fatty acids reduce the risk of cardiovascular disease.

de Logeril, M., Salen, P., Martin, J.L., Monjaud, I., Delaye, J., & Mamelle, N. (1999) Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after

myocardial infarction: final report of the Lyon Diet Heart Study. *Circulation* **99**, 779–785.

Lavie, C.J., Milani, R.V., Mehra, M.R., & Ventura, H.O. (2009) Omega-3 polyunsaturated fatty acids and cardiovascular diseases. *J. Am. Coll. Cardiol.* **54**, 585–594.

Mello, M.M. (2009) New York City's war on fat. *N. Engl. J. Med.* **360**, 2015–2020.

Legal and ethical aspects of the ban on trans fatty acids in restaurants.

Mozaffarian, D., Katan, M.B., Ascherio, P.H., Stampfer, M.J., & Willet, W.C. (2006) Trans fatty acids and cardiovascular disease. *N. Engl. J. Med.* **354**, 1601–1613.

A summary of the evidence that dietary trans fatty acids predispose to coronary heart disease.

Structural Lipids in Membranes

Bogdanov, M. & Dowhan, W. (1999) Lipid-assisted protein folding. *J. Biol. Chem.* **274**, 36,827–36,830.

A minireview of the role of membrane lipids in the folding of membrane proteins.

Dowhan, W. (1997) Molecular basis for membrane phospholipid diversity: why are there so many lipids? *Annu. Rev. Biochem.* **66**, 199–232.

Jacquemet, A., Barbeau, J., Lemiegre, L., & Benvegnu, T. (2009) Archaeal tetraether bipolar lipids: structures, functions and applications. *Biochimie* **91**, 711–717.

Valle, D., Beaudet, A.L., Vogelstein, B., Kinzler, K.W., Antonarakis, S.E., & Ballabio. A. (eds). (2006) *Scriver's Online Metabolic & Molecular Bases of Inherited Disease*, www.ommbid.com.

This classic medical encyclopedia, last published in 2001 as a four-volume set, is now maintained and kept updated online. It contains definitive descriptions of the clinical, biochemical, and genetic aspects of hundreds of human metabolic diseases—an authoritative source and fascinating reading. Part 16: Lysosomal Disorders includes 24 articles on various disorders of lipid metabolism.

Lipids as Signals, Cofactors, and Pigments

Bell, R.M., Exton, J.H., & Prescott, S.M. (eds). (1996) *Lipid Second Messengers*, Handbook of Lipid Research, Vol. 8, Plenum Press, New York.

Berkner, K.L. & Runge, K.W. (2004) The physiology of vitamin K nutriture and vitamin K-dependent protein function in atherosclerosis. *J. Thromb. Haemost.* **2**, 2118–2132.

Binkley, N.C. & Suttie, J.W. (1995) Vitamin K nutrition and osteoporosis. *J. Nutr.* **125**, 1812–1821.

Brigelius-Flohé, R. & Traber, M.G. (1999) Vitamin E: function and metabolism. *FASEB J.* **13**, 1145–1155.

Chojnacki, T. & Dallner, G. (1988) The biological role of dolichol. *Biochem. J.* **251**, 1–9.

Clouse, S.D. (2002) Brassinosteroid signal transduction: clarifying the pathway from ligand perception to gene expression. *Mol. Cell* **10**, 973–982.

DeLuca, H.F. (2008) Evolution of our understanding of vitamin D. *Nutr. Rev.* **66**, S73–S87.

Dicke, M., van Loon, J.J.A., & Soler, R. (2009) Chemical complexity of volatiles from plants induced by multiple attack. *Nat. Chem. Biol.* **5**, 317–324.

Holick, M.F. (2007) Vitamin D deficiency. *N. Engl. J. Med.* **357**, 266–281.

James, D.J., Khodthong, C., Kowalchyk, J.A., & Martin, T.F. (2010) Phosphatidylinositol 4,5-bisphosphate regulation of SNARE function in membrane fusion mediated by CAPS. *Adv. Enzyme Regul.* **50**, 62–70.

Jones M.B., Rosenberg, J.N., Betenbaugh, M.J., & Krag, S.S. (2009) Structure and synthesis of polyisoprenoids used in *N*-glycosylation across the three domains of life. *Biochim. Biophys. Acta* **1790**, 485–494.

Intermediate-level review of role of dolichols in glycosylation.

Lee, D. (2007) *Nature's Palette: The Science of Plant Color*, University of Chicago Press, Chicago, IL.

Fascinating intermediate-level book about lipids as biological pigments.

Rosen, H., Gonzalez-Cabrera, P.J., Sanna, M.G., & Brown, S. (2009) Sphingosine 1-phosphate receptor signaling. *Annu. Rev. Biochem.* **78**, 743–768.

Suttie, J.W. (1993) Synthesis of vitamin K-dependent proteins. *FASEB J.* **7**, 445–452.

Describes the biochemical basis for the requirement of vitamin K in blood clotting and the importance of carboxylation in the synthesis of the blood-clotting protein thrombin.

Weber, H. (2002) Fatty acid-derived signals in plants. *Trends Plant Sci.* **7**, 217–224.

Wymann, M.P. & Schneiter, R. (2008) Lipid signaling in disease. *Nat. Rev. Mol. Cell Biol.* **9**, 162–176.

Working with Lipids

Christie, W.W. (1998) Gas chromatography–mass spectrometry methods for structural analysis of fatty acids. *Lipids* **33**, 343–353.

A detailed description of the methods used to obtain data such as those presented in Figure 10–26.

Christie, W.W. (2003) *Lipid Analysis*, 3rd edn, The Oily Press, Bridgwater, England.

Dennis, E.A., Deems, R.A., Harkewicz, R., Quehenberger, O., Brown, H.A., Milne, S.B., Myers, D.S., Glass, C.K., Hardiman, G., Reichart, D., et al. (2010) A mouse macrophage lipidome. *J. Biol. Chem.* **285**, 39,976–39,985.

Research paper describing variations in the total macrophage lipidome.

Harkewicz, R. & Dennis, E.A. (2011) Applications of mass spectrometry to lipids and membranes. *Annu. Rev. Biochem.* **80**, 301–325.

Advanced discussion of mass spectrometry and lipidomics.

Murphy, R.C. & Gasskell, S.J. (2011) New applications of mass spectrometry in lipid analysis. *J. Biol. Chem.* **286**, 25,427–25,433.

Watson, A.D. (2006) Lipidomics: a global approach to lipid analysis in biological systems. *J. Lipid Res.* **47**, 2101–2111.

A short, intermediate-level review of the classes of lipids, the methods for extracting and separating them, and mass spectrometric means for identifying and quantifying all lipids in a given cell, tissue, or organelle.

Problems

1. Operational Definition of Lipids How is the definition of "lipid" different from the types of definitions used for other biomolecules that we have considered, such as amino acids, nucleic acids, and proteins?

2. Melting Points of Lipids The melting points of a series of 18-carbon fatty acids are: stearic acid, 69.6 °C; oleic acid, 13.4 °C; linoleic acid, –5 °C ; and linolenic acid, –11 °C.

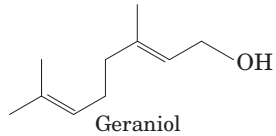
(a) What structural aspect of these 18-carbon fatty acids can be correlated with the melting point?

(b) Draw all the possible triacylglycerols that can be constructed from glycerol, palmitic acid, and oleic acid. Rank them in order of increasing melting point.

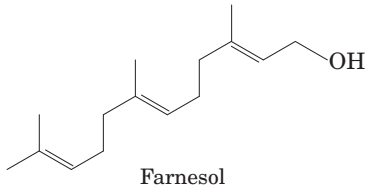
(c) Branched-chain fatty acids are found in some bacterial membrane lipids. Would their presence increase or decrease the fluidity of the membranes (that is, give them a lower or higher melting point)? Why?

3. Preparation of Béarnaise Sauce During the preparation of béarnaise sauce, egg yolks are incorporated into melted butter to stabilize the sauce and avoid separation. The stabilizing agent in the egg yolks is lecithin (phosphatidylcholine). Suggest why this works.

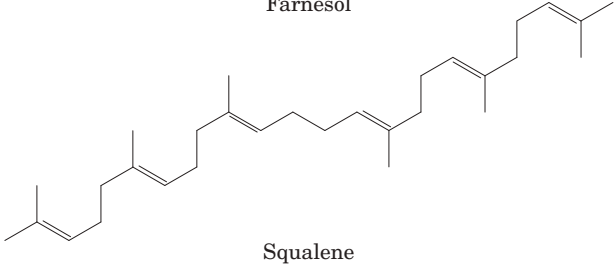
4. Isoprene Units in Isoprenoids Geraniol, farnesol, and squalene are called isoprenoids, because they are synthesized from five-carbon isoprene units. In each compound, circle the five-carbon units representing isoprene units (see Fig. 10–22).



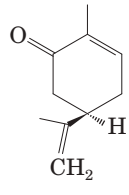
Geraniol



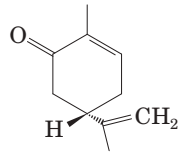
Farnesol


Squalene

5. Naming Lipid Stereoisomers The two compounds below are stereoisomers of carvone with quite different properties; the one on the left smells like spearmint, and that on the right, like caraway. Name the compounds using the RS system.

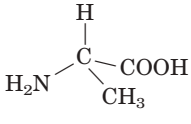


Spearmint

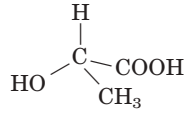


Caraway

6. RS Designations for Alanine and Lactate Draw (using wedge-bond notation) and label the (*R*) and (*S*) isomers of 2-aminopropanoic acid (alanine) and 2-hydroxypropanoic acid (lactic acid).



2-Aminopropanoic acid (alanine)



2-Hydroxypropanoic acid (lactic acid)

7. Hydrophobic and Hydrophilic Components of Membrane Lipids A common structural feature of membrane lipids is their amphipathic nature. For example, in phosphatidylcholine, the two fatty acid chains are hydrophobic and the phosphocholine head group is hydrophilic. For each of the following membrane lipids, name the components that serve as the hydrophobic and hydrophilic units: (a) phosphatidylethanolamine; (b) sphingomyelin; (c) galactosylcerebroside; (d) ganglioside; (e) cholesterol.

8. Structure of Omega-6 Fatty Acid Draw the structure of the omega-6 fatty acid 16:1.

9. Catalytic Hydrogenation of Vegetable Oils Catalytic hydrogenation, used in the food industry, converts double bonds in the fatty acids of the oil triacylglycerols to $—CH_2—CH_2—$. How does this affect the physical properties of the oils?

10. Alkali Lability of Triacylglycerols A common procedure for cleaning the grease trap in a sink is to add a product that contains sodium hydroxide. Explain why this works.

11. Deducing Lipid Structure from Composition Compositional analysis of a certain lipid shows that it has exactly one mole of fatty acid per mole of inorganic phosphate. Could this be a glycerophospholipid? A ganglioside? A sphingomyelin?

12. Deducing Lipid Structure from Molar Ratio of Components Complete hydrolysis of a glycerophospholipid yields glycerol, two fatty acids (16:1(Δ^9) and 16:0), phosphoric acid, and serine in the molar ratio 1:1:1:1:1. Name this lipid and draw its structure.

13. Impermeability of Waxes What property of the waxy cuticles that cover plant leaves makes the cuticles impermeable to water?

14. The Action of Phospholipases The venom of the Eastern diamondback rattler and the Indian cobra contains phospholipase A_2, which catalyzes the hydrolysis of fatty acids at the C-2 position of glycerophospholipids. The phospholipid breakdown product of this reaction is lysolecithin (lecithin is phosphatidylcholine). At high concentrations, this and other lysophospholipids act as detergents, dissolving the membranes of erythrocytes and lysing the cells. Extensive hemolysis may be life-threatening.

(a) All detergents are amphipathic. What are the hydrophilic and hydrophobic portions of lysolecithin?

(b) The pain and inflammation caused by a snake bite can be treated with certain steroids. What is the basis of this treatment?

(c) Though the high levels of phospholipase A_2 in venom can be deadly, this enzyme is necessary for a variety of normal metabolic processes. What are these processes?

15. Lipids in Blood Group Determination We note in Figure 10–15 that the structure of glycosphingolipids determines the blood groups A, B, and O in humans. It is also true that glycoproteins determine blood groups. How can both statements be true?

16. Intracellular Messengers from Phosphatidylinositols When the hormone vasopressin stimulates cleavage of phosphatidylinositol 4,5-bisphosphate by hormone-sensitive

phospholipase C, two products are formed. What are they? Compare their properties and their solubilities in water, and predict whether either would diffuse readily through the cytosol.

17. Storage of Fat-Soluble Vitamins In contrast to water-soluble vitamins, which must be part of our daily diet, fat-soluble vitamins can be stored in the body in amounts sufficient for many months. Suggest an explanation for this difference.

18. Hydrolysis of Lipids Name the products of mild hydrolysis with dilute NaOH of (a) 1-stearoyl-2,3-dipalmitoylglycerol; (b) 1-palmitoyl-2-oleoylphosphatidylcholine.

19. Effect of Polarity on Solubility Rank the following in order of increasing solubility in water: a triacylglycerol, a diacylglycerol, and a monoacylglycerol, all containing only palmitic acid.

20. Chromatographic Separation of Lipids A mixture of lipids is applied to a silica gel column, and the column is then washed with increasingly polar solvents. The mixture consists of phosphatidylserine, phosphatidylethanolamine, phosphatidylcholine, cholesteryl palmitate (a sterol ester), sphingomyelin, palmitate, *n*-tetradecanol, triacylglycerol, and cholesterol. In what order will the lipids elute from the column? Explain your reasoning.

21. Identification of Unknown Lipids Johann Thudichum, who practiced medicine in London about 100 years ago, also dabbled in lipid chemistry in his spare time. He isolated a variety of lipids from neural tissue, and characterized and named many of them. His carefully sealed and labeled vials of isolated lipids were rediscovered many years later.

(a) How would you confirm, using techniques not available to Thudichum, that the vials labeled "sphingomyelin" and "cerebroside" actually contain these compounds?

(b) How would you distinguish sphingomyelin from phosphatidylcholine by chemical, physical, or enzymatic tests?

22. Ninhydrin to Detect Lipids on TLC Plates Ninhydrin reacts specifically with primary amines to form a purplish-blue product. A thin-layer chromatogram of rat liver phospholipids is sprayed with ninhydrin, and the color is allowed to develop. Which phospholipids can be detected in this way?

Data Analysis Problem

23. Determining the Structure of the Abnormal Lipid in Tay-Sachs Disease Box 10–1, Figure 1, shows the pathway of breakdown of gangliosides in healthy (normal) individuals and individuals with certain genetic diseases. Some of the data on which the figure is based were presented in a paper by Lars Svennerholm (1962). Note that the sugar Neu5Ac, *N*-acetylneuraminic acid, represented in the Box 10–1 figure as ◆, is a sialic acid.

Svennerholm reported that "about 90% of the monosialogangliosides isolated from normal human brain" consisted of a compound with ceramide, hexose, *N*-acetylgalactosamine, and *N*-acetylneuraminic acid in the molar ratio 1:3:1:1.

(a) Which of the gangliosides (GM1 through GM3 and globoside) in Box 10–1, Figure 1, fits this description? Explain your reasoning.

(b) Svennerholm reported that 90% of the gangliosides from a patient with Tay-Sachs had a molar ratio (of the same four components given above) of 1:2:1:1. Is this consistent with the Box 10–1 figure? Explain your reasoning.

To determine the structure in more detail, Svennerholm treated the gangliosides with neuraminidase to remove the *N*-acetylneuraminic acid. This resulted in an asialoganglioside that was much easier to analyze. He hydrolyzed it with acid, collected the ceramide-containing products, and determined the molar ratio of the sugars in each product. He did this for both the normal and the Tay-Sachs gangliosides. His results are shown below.

Ganglioside	Ceramide	Glucose	Galactose	Galactosamine
Normal				
Fragment 1	1	1	0	0
Fragment 2	1	1	1	0
Fragment 3	1	1	1	1
Fragment 4	1	1	2	1
Tay-Sachs				
Fragment 1	1	1	0	0
Fragment 2	1	1	1	0
Fragment 3	1	1	1	1

(c) Based on these data, what can you conclude about the structure of the normal ganglioside? Is this consistent with the structure in Box 10–1? Explain your reasoning.

(d) What can you conclude about the structure of the Tay-Sachs ganglioside? Is this consistent with the structure in Box 10–1? Explain your reasoning.

Svennerholm also reported the work of other researchers who "permethylated" the normal asialoganglioside. Permethylation is the same as exhaustive methylation: a methyl group is added to every free hydroxyl group on a sugar. They found the following permethylated sugars: 2,3,6-trimethylglycopyranose; 2,3,4,6-tetramethylgalactopyranose; 2,4,6-trimethylgalactopyranose; and 4,6-dimethyl-2-deoxy-2-aminogalactopyranose.

(e) To which sugar of GM1 does each of the permethylated sugars correspond? Explain your reasoning.

(f) Based on all the data presented so far, what pieces of information about normal ganglioside structure are missing?

Reference

Svennerholm, L. (1962) The chemical structure of normal human brain and Tay-Sachs gangliosides. *Biochem. Biophys. Res. Comm.* **9**, 436–441.

this page left intentionally blank

11

Biological Membranes and Transport

The first cell probably came into being when a membrane formed, enclosing a small volume of aqueous solution and separating it from the rest of the universe. Membranes define the external boundaries of cells and control the molecular traffic across that boundary **(Fig. 11–1)**; in eukaryotic cells, they divide the internal space into discrete compartments to segregate processes and components. They organize complex reaction sequences and are central to both biological energy conservation and cell-to-cell communication. The biological activities of membranes flow from their remarkable physical properties. Membranes are flexible, self-sealing, and selectively permeable to polar solutes. Their flexibility permits the shape changes that accompany cell growth and movement (such as amoeboid movement). With their ability to break and reseal, two membranes can fuse, as in exocytosis, or a single membrane-enclosed compartment can undergo fission to yield two sealed compartments, as in endocytosis or cell division, without creating gross leaks through cellular surfaces. Because membranes are selectively permeable, they retain certain compounds and ions within cells and within specific cellular compartments while excluding others.

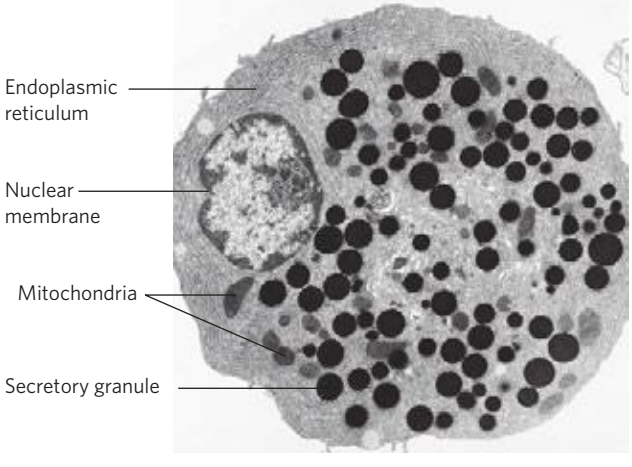


FIGURE 11–1 Biological membranes. This electron micrograph of a thin-sectioned exocrine pancreas cell shows several compartments made of or bounded by membranes: the endoplasmic reticulum, mitochondria, secretory granules, and the nuclear membrane.

Membranes are not merely passive barriers. They include an array of proteins specialized for promoting or catalyzing various cellular processes. At the cell surface, transporters move specific organic solutes and inorganic ions across the membrane; receptors sense extracellular signals and trigger molecular changes in the cell; adhesion molecules hold neighboring cells together. Within the cell, membranes organize cellular processes such as the synthesis of lipids and certain proteins, and the energy transductions in mitochondria and chloroplasts. Because membranes consist of just two layers of molecules, they are very thin—essentially two-dimensional. Intermolecular collisions are far more probable in this two-dimensional space than in three-dimensional space, so the efficiency of enzyme-catalyzed processes organized within membranes is vastly increased.

In this chapter we first describe the composition of cellular membranes and their chemical architecture—the molecular structures that underlie their biological functions. Next, we consider the remarkable dynamic features of membranes, in which lipids and proteins move relative to each other. Cell adhesion, endocytosis, and the membrane fusion accompanying neurotransmitter secretion illustrate the dynamic roles of membrane proteins. We then turn to the protein-mediated passage of solutes across membranes via transporters and ion channels. In later chapters we discuss the roles of membranes in signal transduction (Chapters 12 and 23), energy transduction (Chapter 19), lipid synthesis (Chapter 21), and protein synthesis (Chapter 27).

11.1 The Composition and Architecture of Membranes

One approach to understanding membrane function is to study membrane composition—to determine, for example, which components are common to all membranes and which are unique to membranes with specific functions. So before describing membrane structure and function, we consider the molecular components of membranes: proteins and polar lipids, which account for almost all the mass of biological membranes, and carbohydrates, present as part of glycoproteins and glycolipids.

Each Type of Membrane Has Characteristic Lipids and Proteins

The relative proportions of protein and lipid vary with the type of membrane (Table 11–1), reflecting the diversity of biological roles. For example, certain neurons have a myelin sheath—an extended plasma membrane that wraps around the cell many times and acts as a passive electrical insulator. The myelin sheath consists primarily of lipids, whereas the plasma membranes of bacteria and the membranes of mitochondria and chloroplasts, the sites of many enzyme-catalyzed processes, contain more protein than lipid (in mass per total mass).

For studies of membrane composition, the first task is to isolate a selected membrane. When eukaryotic cells are subjected to mechanical shear, their plasma membranes are torn and fragmented, releasing cytoplasmic components and membrane-bounded organelles such as mitochondria, chloroplasts, lysosomes, and nuclei. Plasma membrane fragments and intact organelles can be isolated by techniques described in Chapter 1 (see Fig. 1–8) and in Worked Example 2–1 (p. 57).

Cells clearly have mechanisms to control the kinds and amounts of membrane lipid they synthesize and to target specific lipids to particular organelles. Each kingdom, each species, each tissue or cell type, and the organelles of each cell type have a characteristic set of membrane lipids. Plasma membranes, for example, are enriched in cholesterol and contain no detectable cardiolipin **(Fig. 11–2)**; mitochondrial membranes are very low in cholesterol and sphingolipids, but they contain phosphatidylglycerol and cardiolipin, which are synthesized within the mitochondria. In all but a few cases, the functional significance of these combinations is not yet known.

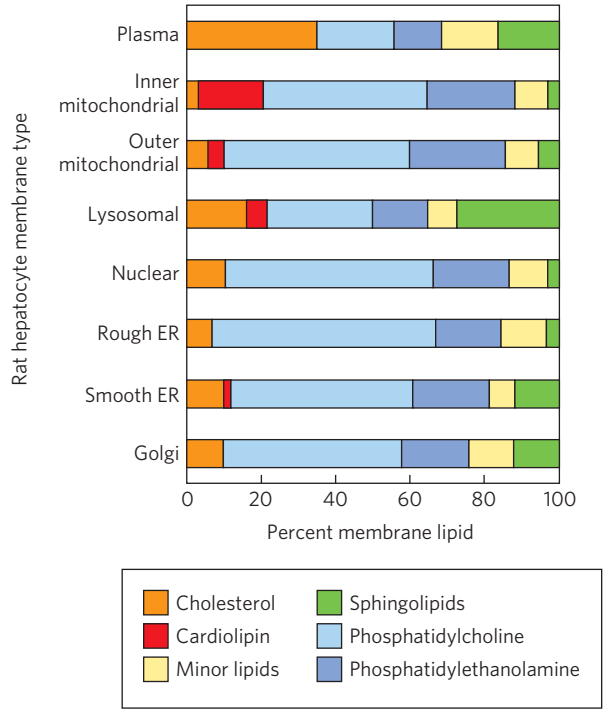


FIGURE 11–2 Lipid composition of the plasma membrane and organelle membranes of a rat hepatocyte. The functional specialization of each membrane type is reflected in its unique lipid composition. Cholesterol is prominent in plasma membranes but barely detectable in mitochondrial membranes. Cardiolipin is a major component of the inner mitochondrial membrane but not of the plasma membrane. Phosphatidylserine, phosphatidylinositol, and phosphatidylglycerol are relatively minor components of most membranes but serve critical functions; phosphatidylinositol and its derivatives, for example, are important in signal transductions triggered by hormones. Sphingolipids, phosphatidylcholine, and phosphatidylethanolamine are present in most membranes but in varying proportions. Glycolipids, which are major components of the chloroplast membranes of plants, are virtually absent from animal cells.

TABLE 11–1 Major Components of Plasma Membranes in Various Organisms

	Components (% by weight)				
	Protein	Phospholipid	Sterol	Sterol type	Other lipids
Human myelin sheath	30	30	19	Cholesterol	Galactolipids, plasmalogens
Mouse liver	45	27	25	Cholesterol	—
Maize leaf	47	26	7	Sitosterol	Galactolipids
Yeast	52	7	4	Ergosterol	Triacylglycerols, steryl esters
Paramecium (ciliated protist)	56	40	4	Stigmasterol	—
E. coli	75	25	0	—	—

Note: Values do not add up to 100% in every case because there are components other than protein, phospholipids, and sterol; plants, for example, have high levels of glycolipids.

The protein composition of membranes from different sources varies even more widely than their lipid composition, reflecting functional specialization. In addition, some membrane proteins are covalently linked to oligosaccharides. For example, in glycophorin, a glycoprotein of the erythrocyte plasma membrane, 60% of the mass consists of complex oligosaccharides covalently attached to specific amino acid residues. Ser, Thr, and Asn residues are the most common points of attachment (see Fig. 7–30). The sugar moieties of surface glycoproteins influence the folding of the proteins as well as their stability and intracellular destination, and they play a significant role in the specific binding of ligands to glycoprotein surface receptors (see Fig. 7–37).

Some membrane proteins are covalently attached to one or more lipids, which serve as hydrophobic anchors that hold the proteins to the membrane, as we shall see.

All Biological Membranes Share Some Fundamental Properties

Membranes are impermeable to most polar or charged solutes, but permeable to nonpolar compounds. They are 5 to 8 nm (50 to 80 Å) thick when proteins protruding on both sides are included and appear trilaminar when viewed in cross section with the electron microscope. The combined evidence from electron microscopy and studies of chemical composition, as well as physical studies of permeability and the motion of individual protein and lipid molecules within membranes, led to the development of the **fluid mosaic model** for the structure of biological membranes **(Fig. 11–3)**. Phospholipids form a bilayer in which the nonpolar regions of the lipid molecules in each layer face the core of the bilayer and their polar head groups face outward, interacting with the aqueous phase on either side. Proteins are embedded in this bilayer sheet, held by hydrophobic interactions between the membrane lipids and hydrophobic domains in the proteins. Some proteins protrude from only one side of the membrane; others have domains exposed on both sides. The orientation of proteins in the bilayer is asymmetric, giving the membrane "sidedness": the protein domains exposed on one side of the bilayer are different from those exposed on the other side, reflecting functional asymmetry. The individual lipid and protein units in a membrane form a fluid mosaic with a pattern that, unlike a mosaic of ceramic tile and mortar, is free to change constantly. The membrane mosaic is fluid because most of the interactions among its components are noncovalent, leaving individual lipid and protein molecules free to move laterally in the plane of the membrane.

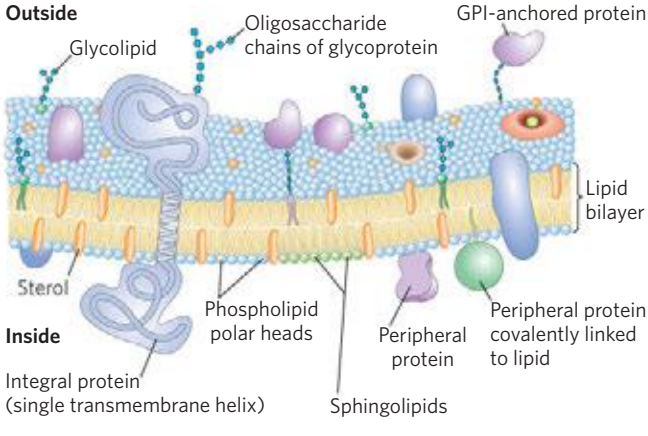


FIGURE 11–3 Fluid mosaic model for plasma membrane structure. The fatty acyl chains in the interior of the membrane form a fluid, hydrophobic region. Integral proteins float in this sea of lipid, held by hydrophobic interactions with their nonpolar amino acid side chains. Both proteins and lipids are free to move laterally in the plane of the bilayer, but movement of either from one leaflet of the bilayer to the other is restricted. The carbohydrate moieties attached to some proteins and lipids of the plasma membrane are exposed on the extracellular surface.

We now look at some of these features of the fluid mosaic model in more detail and consider the experimental evidence that supports the basic model but has necessitated its refinement in several ways.

A Lipid Bilayer Is the Basic Structural Element of Membranes

Glycerophospholipids, sphingolipids, and sterols are virtually insoluble in water. When mixed with water, they spontaneously form microscopic lipid aggregates, clustering together, with their hydrophobic moieties in contact with each other and their hydrophilic groups interacting with the surrounding water. This clustering reduces the amount of hydrophobic surface exposed to water and thus minimizes the number of molecules in the shell of ordered water at the lipid-water interface (see Fig. 2–7), resulting in an increase in entropy. Hydrophobic interactions among lipid molecules provide the thermodynamic driving force for the formation and maintenance of these clusters.

Depending on the precise conditions and the nature of the lipids, three types of lipid aggregate can form when amphipathic lipids are mixed with water **(Fig. 11–4)**. **Micelles** are spherical structures that contain anywhere from a few dozen to a few thousand amphipathic molecules. These molecules are arranged with their hydrophobic regions aggregated in the interior, where water is excluded, and their hydrophilic head groups at the surface, in contact with water. Micelle formation is favored when the cross-sectional area of the head group is greater than that of the acyl side chain(s), as in free fatty acids, lysophospholipids (phospholipids lacking one fatty acid), and detergents such as sodium dodecyl sulfate (SDS; p. 94).

A second type of lipid aggregate in water is the **bilayer**, in which two lipid monolayers (leaflets) form a two-dimensional sheet. Bilayer formation is favored if the cross-sectional areas of the head group and acyl side chain(s) are similar, as in glycerophospholipids and

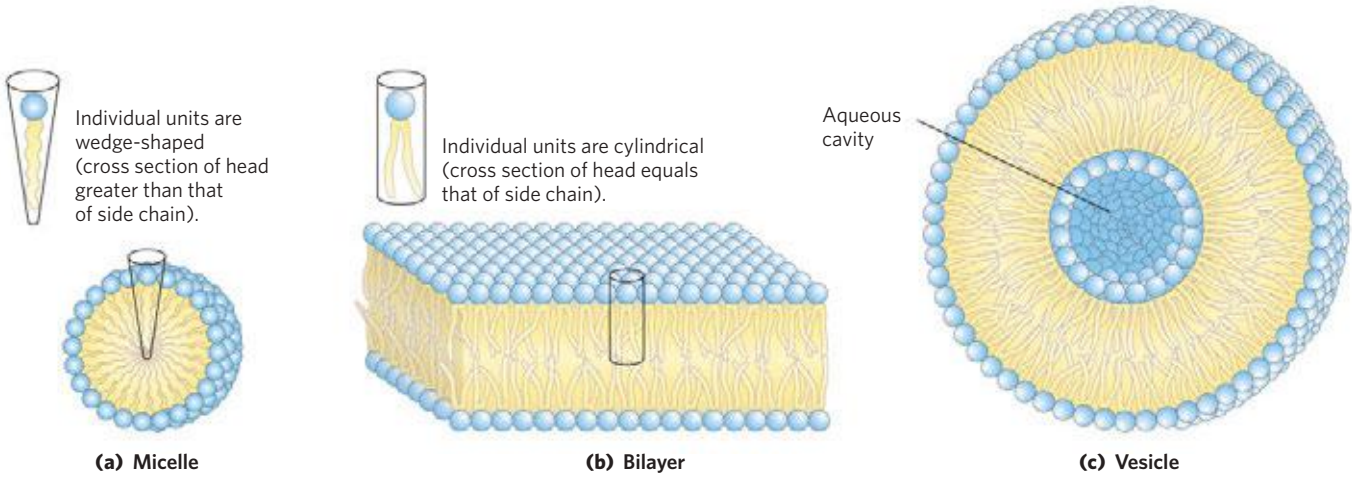


FIGURE 11–4 Amphipathic lipid aggregates that form in water. (a) In micelles, the hydrophobic chains of the fatty acids are sequestered at the core of the sphere. There is virtually no water in the hydrophobic interior. **(b)** In an open bilayer, all acyl side chains except those at the edges of the sheet are protected from interaction with water. **(c)** When a two-dimensional bilayer folds on itself, it forms a closed bilayer, a three-dimensional hollow vesicle (liposome) enclosing an aqueous cavity.

sphingolipids. The hydrophobic portions in each monolayer, excluded from water, interact with each other. The hydrophilic head groups interact with water at each surface of the bilayer. Because the hydrophobic regions at its edges (Fig. 11–4b) are in contact with water, the bilayer sheet is relatively unstable and spontaneously folds back on itself to form a hollow sphere, a **vesicle** (Fig. 11–4c). The continuous surface of vesicles eliminates exposed hydrophobic regions, allowing bilayers to achieve maximal stability in their aqueous environment. Vesicle formation also creates a separate aqueous compartment. It is likely that the precursors to the first living cells resembled lipid vesicles, their aqueous contents segregated from their surroundings by a hydrophobic shell.

The lipid bilayer is 3 nm (30 Å) thick. The hydrocarbon core, made up of the —CH_2— and —CH_3 of the fatty acyl groups, is about as nonpolar as decane, and vesicles formed in the laboratory from pure lipids (liposomes) are essentially impermeable to polar solutes, as is the lipid bilayer of biological membranes (although biological membranes, as we shall see, are permeable to solutes for which they have specific transporters).

Plasma membrane lipids are asymmetrically distributed between the two monolayers of the bilayer, although the asymmetry, unlike that of membrane proteins, is not absolute. In the plasma membrane of the erythrocyte, for example, choline-containing lipids (phosphatidylcholine and sphingomyelin) are typically found in the outer (extracellular, or exoplasmic) leaflet **(Fig. 11–5)**, whereas phosphatidylserine, phosphatidylethanolamine, and the phosphatidylinositols are much more common in the inner (cytoplasmic) leaflet. The flow of membrane components from the endoplasmic reticulum through the Golgi apparatus and to the plasma membrane via transport vesicles is accompanied by changes in lipid composition and disposition across the bilayer **(Fig. 11–6)**. Phosphatidylcholine is the principal phospholipid in the lumenal monolayer of the Golgi membrane, but in transport vesicles phosphatidylcholine has been largely replaced by sphingolipids and cholesterol, which, on fusion of transport vesicles with the plasma membrane, make up the majority of the lipids in the outer monolayer of the plasma membrane.

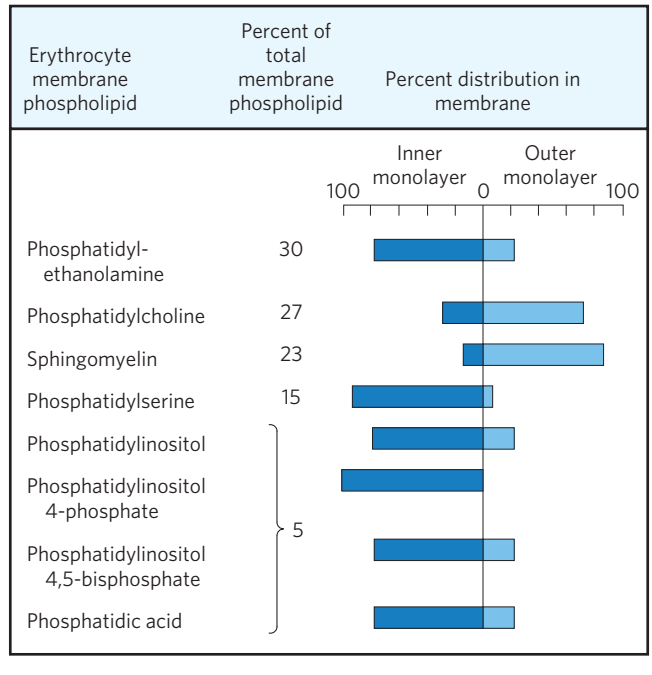


FIGURE 11–5 Asymmetric distribution of phospholipids between the inner and outer monolayers of the erythrocyte plasma membrane. The distribution of a specific phospholipid is determined by treating the intact cell with phospholipase C, which cannot reach lipids in the inner monolayer (leaflet) but removes the head groups of lipids in the outer monolayer. The proportion of each head group released provides an estimate of the fraction of each lipid in the outer monolayer.

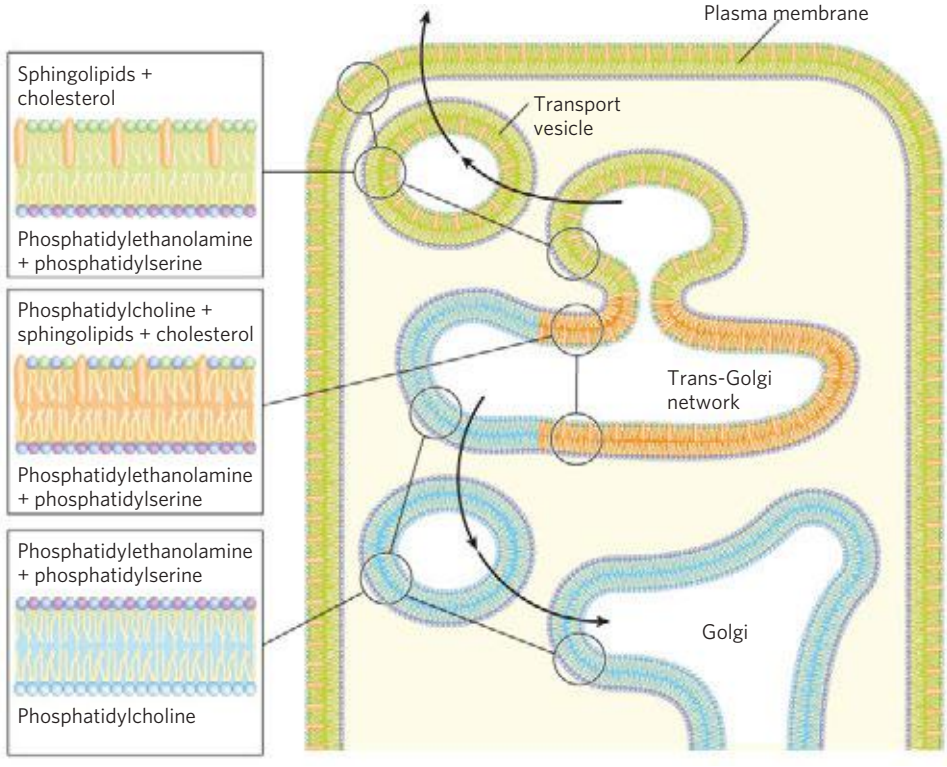


FIGURE 11–6 The distribution of lipids in the membranes of a typical cell. Each membrane has its own characteristic composition, and the two monolayers of a given membrane may differ in composition as well.

Changes in the distribution of lipids between plasma membrane leaflets have biological consequences. For example, only when the phosphatidylserine in the plasma membrane moves into the outer leaflet is a platelet able to play its role in formation of a blood clot. For many other cell types, phosphatidylserine exposure on the outer surface marks a cell for destruction by programmed cell death. The transbilayer movement of phospholipid molecules is catalyzed and regulated by specific proteins (see Fig. 11–17).

Three Types of Membrane Proteins Differ in Their Association with the Membrane

Integral membrane proteins are very firmly associated with the lipid bilayer and are removable only by agents that interfere with hydrophobic interactions, such as detergents, organic solvents, or denaturants **(Fig. 11–7)**. **Peripheral membrane proteins** associate with the membrane through electrostatic interactions and hydrogen bonding with the hydrophilic domains of integral proteins and with the polar head

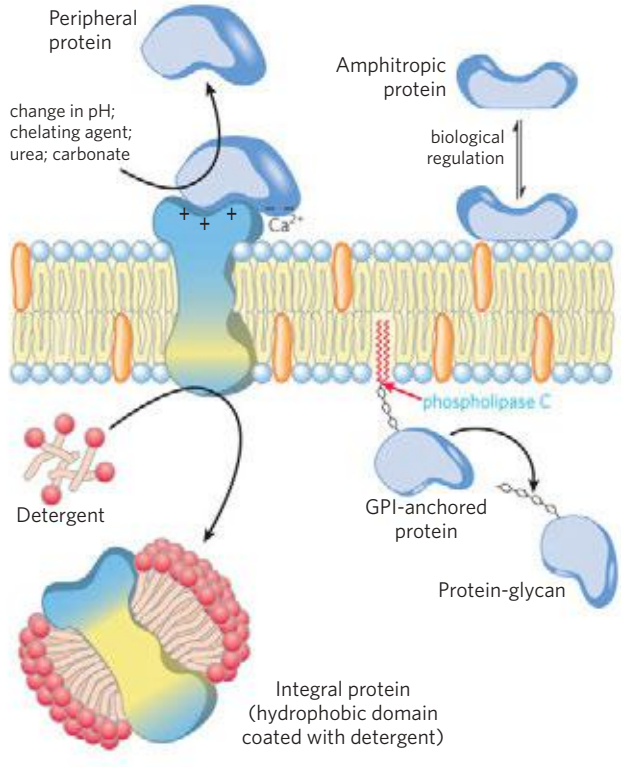


FIGURE 11–7 Peripheral, integral, and amphitropic proteins. Membrane proteins can be operationally distinguished by the conditions required to release them from the membrane. Most peripheral proteins are released by changes in pH or ionic strength, removal of Ca^{2+} by a chelating agent, or addition of urea or carbonate. Integral proteins are extractable with detergents, which disrupt the hydrophobic interactions with the lipid bilayer and form micelle-like clusters around individual protein molecules. Integral proteins covalently attached to a membrane lipid, such as a glycosyl phosphatidylinositol (GPI; see Fig. 11–15), can be released by treatment with phospholipase C. Amphitropic proteins are sometimes associated with membranes and sometimes not, depending on some type of regulatory process such as reversible palmitoylation.

groups of membrane lipids. They can be released by relatively mild treatments that interfere with electrostatic interactions or break hydrogen bonds; a commonly used agent is carbonate at high pH. **Amphitropic proteins** are found both in the cytosol and in association with membranes. Their affinity for membranes results in some cases from the protein's noncovalent interaction with a membrane protein or lipid, and in other cases from the presence of one or more lipids covalently attached to the amphitropic protein (see Fig. 11–15). Generally, the reversible association of amphitropic proteins with the membrane is regulated; for example, phosphorylation or ligand binding can force a conformational change in the protein, exposing a membrane-binding site that was previously inaccessible.

Many Membrane Proteins Span the Lipid Bilayer

Membrane protein topology (the localization of protein domains relative to the lipid bilayer) can be determined with reagents that react with protein side chains but cannot cross membranes—polar chemical reagents that react with primary amines of Lys residues, for example, or enzymes such as trypsin that cleave proteins but cannot cross the membrane. The human erythrocyte is convenient for such studies because it has no membrane-bounded organelles; the plasma membrane is the only membrane present. If a membrane protein in an intact erythrocyte reacts with a membrane-impermeant reagent, that protein must have at least one domain exposed on the outer (extracellular) face of the membrane. Trypsin cleaves extracellular domains but does not affect domains buried within the bilayer or exposed on the inner surface only, unless the plasma membrane is broken to make these domains accessible to the enzyme.

Experiments with such topology-specific reagents show that the erythrocyte glycoprotein **glycophorin** spans the plasma membrane. Its amino-terminal domain (bearing the carbohydrate chains) is on the outer surface and is cleaved by trypsin. The carboxyl terminus protrudes on the inside of the cell, where it cannot react with impermeant reagents. Both the amino-terminal and carboxyl-terminal domains contain many polar or charged amino acid residues and are therefore hydrophilic. However, a segment in the center of the protein (residues 75 to 93) contains mainly hydrophobic amino acid residues, suggesting that glycophorin has a transmembrane segment arranged as shown in **Figure 11–8**.

These noncrystallographic experiments also revealed that the orientation of glycophorin in the membrane is asymmetric: its amino-terminal segment is always on the outside. Similar studies of other membrane proteins show that each has a specific orientation in the bilayer, giving the membrane a distinct sidedness. For glycophorin, and for all other glycoproteins of the plasma membrane, the glycosylated domains

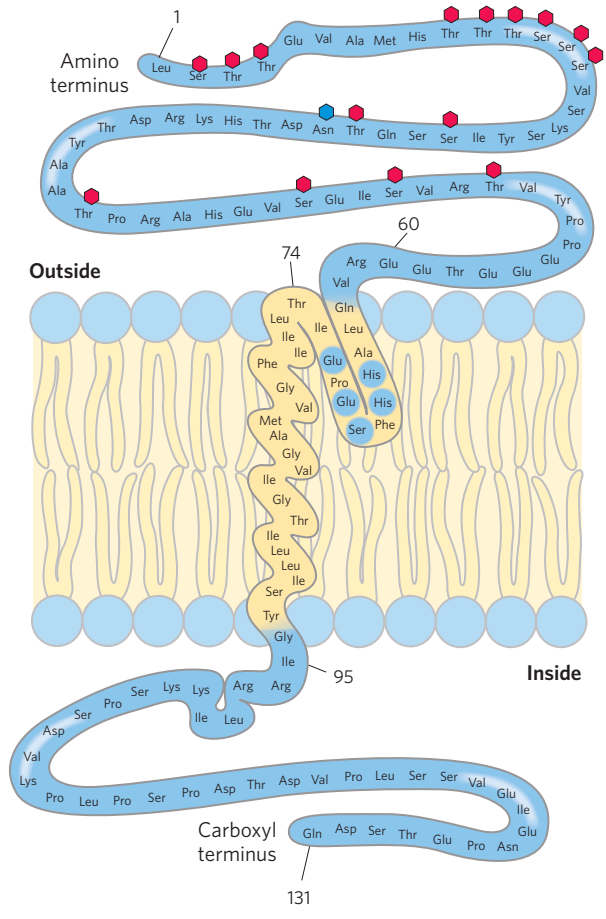


FIGURE 11–8 Transbilayer disposition of glycophorin in an erythrocyte. One hydrophilic domain, containing all the sugar residues, is on the outer surface, and another hydrophilic domain protrudes from the inner face of the membrane. Each red hexagon represents a tetrasaccharide (containing two Neu5Ac (sialic acid), Gal, and GalNAc) *O*-linked to a Ser or Thr residue; the blue hexagon represents an oligosaccharide *N*-linked to an Asn residue. The relative size of the oligosaccharide units is larger than shown here. A segment of 19 hydrophobic residues (residues 75 to 93) forms an α helix that traverses the membrane bilayer (see Fig. 11-12a). The segment from residues 64 to 74 has some hydrophobic residues and probably penetrates the outer face of the lipid bilayer, as shown.

are invariably found on the extracellular face of the bilayer. As we shall see, the asymmetric arrangement of membrane proteins results in functional asymmetry. All the molecules of a given ion pump, for example, have the same orientation in the membrane and pump ions in the same direction.

Integral Proteins Are Held in the Membrane by Hydrophobic Interactions with Lipids

The firm attachment of integral proteins to membranes is the result of hydrophobic interactions between membrane lipids and hydrophobic domains of the protein. Some proteins have a single hydrophobic sequence in the middle (as in glycophorin) or at the amino or carboxyl terminus. Others have multiple hydrophobic sequences,

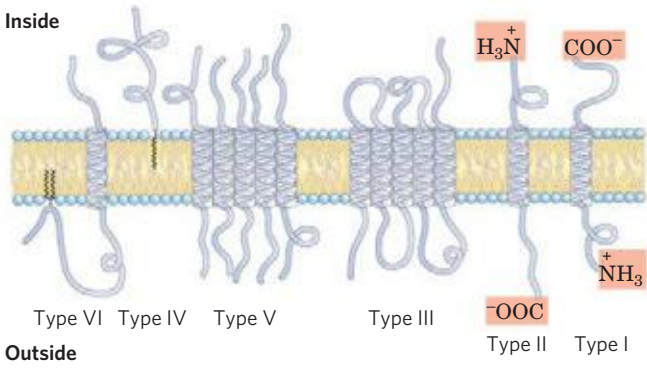


FIGURE 11–9 Integral membrane proteins. For known proteins of the plasma membrane, the spatial relationships of protein domains to the lipid bilayer fall into six categories. Types I and II have a single transmembrane helix; the amino-terminal domain is outside the cell in type I proteins and inside in type II. Type III proteins have multiple transmembrane helices in a single polypeptide. In type IV proteins, transmembrane domains of several different polypeptides assemble to form a channel through the membrane. Type V proteins are held to the bilayer primarily by covalently linked lipids (see Fig. 11–15), and type VI proteins have both transmembrane helices and lipid anchors.

In this figure, and in figures throughout the book, we represent transmembrane protein segments in their most likely conformations: as α helices of six to seven turns. Sometimes these helices are shown simply as cylinders. As relatively few membrane protein structures have been deduced by x-ray crystallography, our representation of the extramembrane domains is arbitrary and not necessarily to scale.

each of which, when in the α-helical conformation, is long enough to span the lipid bilayer **(Fig. 11–9)**.

One of the best-studied membrane-spanning proteins, bacteriorhodopsin, has seven very hydrophobic internal sequences and crosses the lipid bilayer seven times. Bacteriorhodopsin is a light-driven proton pump densely packed in regular arrays in the purple membrane of the bacterium *Halobacterium salinarum*. X-ray crystallography reveals a structure with seven α-helical segments, each traversing the lipid bilayer, connected by nonhelical loops at the inner and outer face of the membrane **(Fig. 11–10)**. In the amino acid sequence of bacteriorhodopsin, seven segments of about 20 hydrophobic residues can be identified, each forming an α helix that spans the bilayer. The seven helices are clustered together and oriented not quite perpendicular to the bilayer plane, a pattern that (as we shall see in Chapter 12) is a common motif in membrane proteins involved in signal reception. Hydrophobic interactions between the nonpolar amino acids and the fatty acyl groups of the membrane lipids firmly anchor the protein in the membrane.

Crystallized membrane proteins solved (i.e., their molecular structure deduced) by crystallography often include molecules of phospholipids, which are presumed to be positioned in the crystals as they are in the native membranes. Many of these phospholipid molecules lie on the protein surface, their head groups interacting with polar amino acid residues at the inner

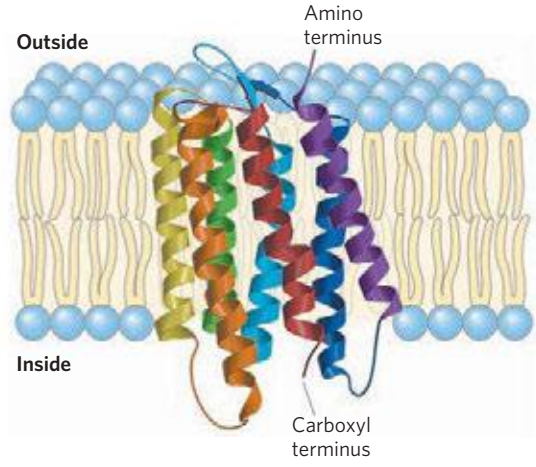


FIGURE 11–10 Bacteriorhodopsin, a membrane-spanning protein. (PDB ID 2AT9) The single polypeptide chain folds into seven hydrophobic α helices, each of which traverses the lipid bilayer roughly perpendicular to the plane of the membrane. The seven transmembrane helices are clustered, and the space around and between them is filled with the acyl chains of membrane lipids. The light-absorbing pigment retinal (see Fig. 10–21) is buried deep in the membrane in contact with several of the helical segments (not shown). The helices are colored to correspond with the hydropathy plot in Figure 11–12b.

and outer membrane–water interfaces and their side chains associated with nonpolar residues. These **annular lipids** form a bilayer shell (annulus) around the protein, oriented roughly as expected for phospholipids in a bilayer **(Fig. 11–11)**. Other phospholipids are found at the interfaces between monomers of multisubunit membrane proteins, where they form a "grease seal." Yet others are embedded deep within a membrane protein, often with their head groups well below the plane of the bilayer. For example, succinate dehydrogenase (Complex II, found in mitochondria; see Fig. 19–10) has several deeply embedded phospholipid molecules.

The Topology of an Integral Membrane Protein Can Sometimes Be Predicted from Its Sequence

Determination of the three-dimensional structure of a membrane protein—that is, its topology—is generally much more difficult than determining its amino acid sequence, either directly or by gene sequencing. The amino acid sequences are known for thousands of membrane proteins, but relatively few three-dimensional structures have been established by crystallography or NMR spectroscopy. The presence of unbroken sequences of more than 20 hydrophobic residues in a membrane protein is commonly taken as evidence that these sequences traverse the lipid bilayer, acting as hydrophobic anchors or forming transmembrane channels. Virtually all integral proteins have at least one such sequence. Application of this logic to entire genomic sequences leads to the conclusion that in many species, 20% to 30% of all proteins are integral membrane proteins.

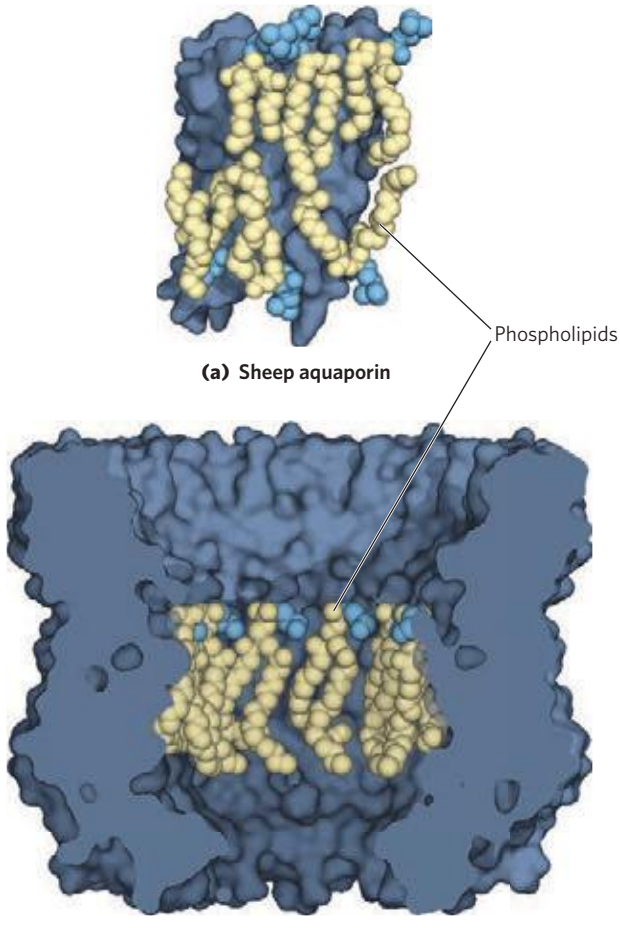


FIGURE 11–11 Lipid annuli associated with two integral membrane proteins. (a) The crystal structure of sheep aquaporin (PDB ID 2B6O), a transmembrane water channel, includes a shell of phospholipids positioned with their head groups (blue) at the expected positions on the inner and outer membrane surfaces and their hydrophobic acyl chains (gold) intimately associated with the surface of the protein exposed to the bilayer. The lipid forms a "grease seal" around the protein, which is depicted as a dark blue surface representation. **(b)** The crystal structure of the V_o integral protein complex of the V-type Na^+ ATPase from *Enterococcus hirae* (PDB ID 2BL2) has 10 identical subunits, each with four transmembrane helices, surrounding a central cavity filled with phosphatidylglycerol (PG). Here five of the subunits have been cut away to reveal the PG molecules associated with each subunit around the interior of this structure.

What can we predict about the secondary structure of the membrane-spanning portions of integral proteins? An α-helical sequence of 20 to 25 residues is just long enough to span the thickness (30 Å) of the lipid bilayer (recall that the length of an α helix is 1.5 Å (0.15 nm) per amino acid residue). A polypeptide chain surrounded by lipids, having no water molecules with which to hydrogen-bond, will tend to form α helices or β sheets, in which intrachain hydrogen bonding is maximized. If the side chains of all amino acids in a helix are nonpolar, hydrophobic interactions with the surrounding lipids further stabilize the helix.

Several simple methods of analyzing amino acid sequences yield reasonably accurate predictions of secondary structure for transmembrane proteins. The relative polarity of each amino acid has been determined experimentally by measuring the free-energy change accompanying the movement of that amino acid side chain from a hydrophobic solvent into water. This free energy of transfer, which can be expressed as a **hydropathy index** (see Table 3–1), ranges from very exergonic for charged or polar residues to very endergonic for amino acids with aromatic or aliphatic hydrocarbon side chains. The overall hydropathy index (hydrophobicity) of a sequence of amino acids is estimated by summing the free energies of transfer for the residues in the sequence. To scan a polypeptide sequence for potential membrane-spanning segments, an investigator calculates the hydropathy index for successive segments (called windows) of a given size, from 7 to 20 residues. For a window of seven residues, for example, the average indices for residues 1 to 7, 2 to 8, 3 to 9, and so on are plotted as in **Figure 11–12** (plotted for the

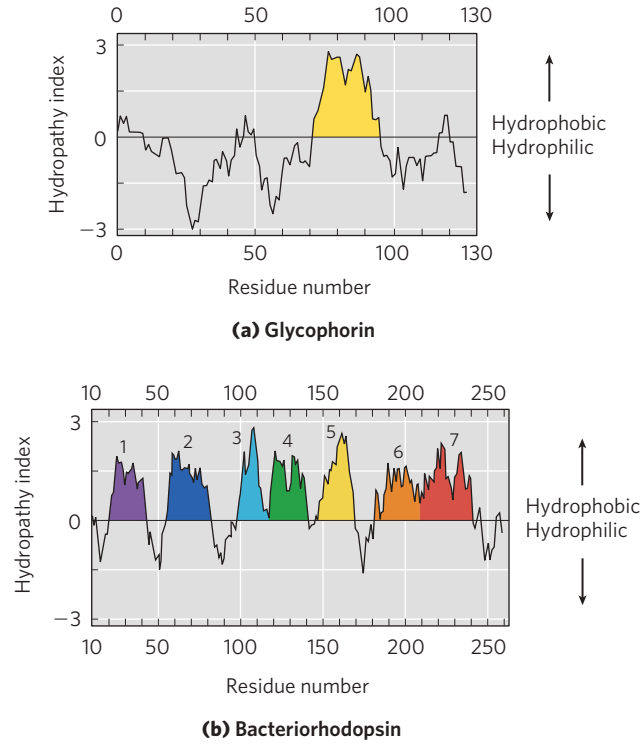


FIGURE 11–12 Hydropathy plots. Average hydropathy index (see Table 3–1) is plotted against residue number for two integral membrane proteins. The hydropathy index for each amino acid residue in a sequence of defined length, or "window," is used to calculate the average hydropathy for that window. The horizontal axis shows the residue number in the middle of the window. **(a)** Glycophorin from human erythrocytes has a single hydrophobic sequence between residues 75 and 93 (yellow); compare this with Figure 11–8. **(b)** Bacteriorhodopsin, known from independent physical studies to have seven transmembrane helices (see Fig. 11–10), has seven hydrophobic regions. Note, however, that the hydropathy plot is ambiguous in the region of segments 6 and 7. X-ray crystallography has confirmed that this region has two transmembrane segments.

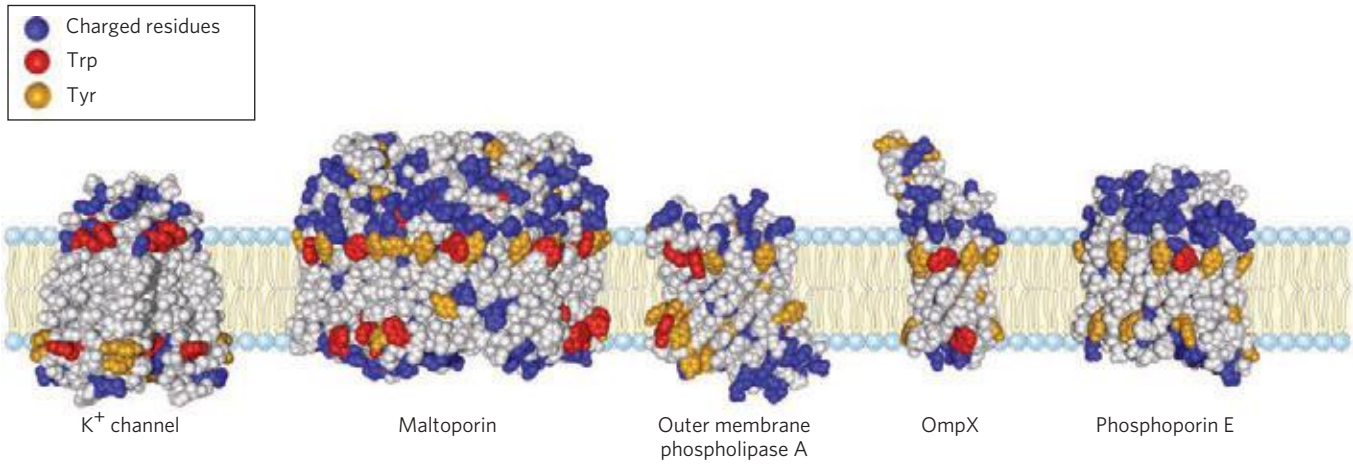


FIGURE 11–13 Tyr and Trp residues of membrane proteins clustering at the water-lipid interface. The detailed structures of these five integral membrane proteins are known from crystallographic studies. The K^+ channel (PDB ID 1BL8) is from the bacterium *Streptomyces lividans* (see Fig. 11–47); maltoporin (PDB ID 1AF6), outer membrane phospholipase A (OmpLA, PDB ID 1QD5), OmpX (PDB ID 1QJ9), and phosphoporin E (PDB ID 1PHO) are proteins of the outer membrane of *E. coli*. Residues of Tyr and Trp are found predominantly where the nonpolar region of acyl chains meets the polar head group region. Charged residues (Lys, Arg, Glu, Asp) are found almost exclusively in the aqueous phases.

middle residue in each window—residue 4 for residues 1 to 7, for example). A region with more than 20 residues of high hydropathy index is presumed to be a transmembrane segment. When the sequences of membrane proteins of known three-dimensional structure are scanned in this way, we find a reasonably good correspondence between predicted and known membrane-spanning segments. Hydropathy analysis predicts a single hydrophobic helix for glycophorin (Fig. 11–12a) and seven transmembrane segments for bacteriorhodopsin (Fig. 11–12b)—in agreement with experimental studies.

On the basis of their amino acid sequences and hydropathy plots, many of the transport proteins described in this chapter are believed to have multiple membrane-spanning helical regions—that is, they are type III or type IV integral proteins (Fig. 11–9). When predictions are consistent with chemical studies of protein localization (such as those described above for glycophorin and bacteriorhodopsin), the assumption that hydrophobic regions correspond to membrane-spanning domains is much better justified.

A further remarkable feature of many transmembrane proteins of known structure is the presence of Tyr and Trp residues at the interface between lipid and water **(Fig. 11–13)**. The side chains of these residues apparently serve as membrane interface anchors, able to interact simultaneously with the central lipid phase and the aqueous phases on either side of the membrane. Another generalization about amino acid location relative to the bilayer is described by the **positive-inside rule**: the positively charged Lys, His, and Arg residues of membrane proteins occur more commonly on the cytoplasmic face of membranes.

Not all integral membrane proteins are composed of transmembrane α helices. Another structural motif common in bacterial membrane proteins is the **β barrel** (see Fig. 4–18b), in which 20 or more transmembrane segments form β sheets that line a cylinder **(Fig. 11–14)**. The same factors that favor α-helix formation in the hydrophobic interior of a lipid bilayer also stabilize β barrels: when no water molecules are available to hydrogen-bond with the carbonyl oxygen and nitrogen of the peptide bond, maximal intrachain hydrogen bonding gives the most stable conformation. Planar β sheets do not maximize these interactions and are generally not found in the membrane interior; β barrels allow all possible hydrogen bonds and are apparently common among membrane proteins. **Porins**, proteins that allow certain polar solutes to cross the outer membrane of gram-negative bacteria such as *E. coli*, have many-stranded β barrels lining the polar transmembrane passage. The outer membranes of mitochondria and chloroplasts also contain a variety of β barrels.

A polypeptide is more extended in the β conformation than in an α helix; just seven to nine residues of β conformation are needed to span a membrane.

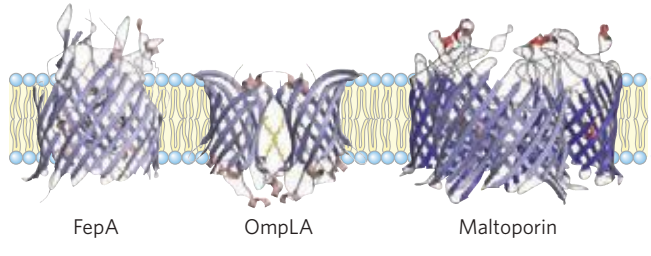


FIGURE 11–14 Membrane proteins with β-barrel structure. Three proteins of the *E. coli* outer membrane are shown, viewed in the plane of the membrane. FepA (PDB ID 1FEP), involved in iron uptake, has 22 membrane-spanning β strands. OmpLA (derived from PDB ID 1QD5), a phospholipase, is a 12-stranded β barrel that exists as a dimer in the membrane. Maltoporin (derived from PDB ID 1MAL), a maltose transporter, is a trimer; each monomer consists of 16 β strands.

Recall that in the β conformation, alternating side chains project above and below the sheet (see Fig. 4–6). In β strands of membrane proteins, every second residue in the membrane-spanning segment is hydrophobic and interacts with the lipid bilayer; aromatic side chains are commonly found at the lipid-protein interface. The other residues may or may not be hydrophilic. The hydropathy plot is not useful in predicting transmembrane segments for proteins with β barrel motifs, but as the database of known β-barrel motifs increases, sequence-based predictions of transmembrane β conformations have become feasible. For example, sequence analysis has correctly predicted that some outer membrane proteins of gram-negative bacteria (Fig. 11–14) contain β barrels.

Covalently Attached Lipids Anchor Some Membrane Proteins

Some membrane proteins contain one or more covalently linked lipids, which may be of several types: long-chain fatty acids, isoprenoids, sterols, or *g*lycosylated derivatives of *p*hosphatidyl*i*nositol (GPIs; **Fig. 11–15**). The attached lipid provides a hydrophobic anchor that inserts into the lipid bilayer and holds the protein at the membrane surface. The strength of the hydrophobic interaction between a bilayer and a single hydrocarbon chain linked to a protein is barely enough to anchor the protein securely, but many proteins have more than one attached lipid moiety. Other interactions, such as ionic attractions between positively charged Lys residues in the protein and negatively charged lipid head groups, probably contribute to the stability of the attachment. The association of these lipid-linked proteins with the membrane is certainly weaker than that for integral membrane proteins and is, at least in the case of cysteine palmitoylation, reversible.

Beyond merely anchoring a protein to the membrane, the attached lipid may have a more specific role. In the plasma membrane, proteins with GPI anchors are exclusively on the outer face and are clustered in certain regions, as discussed later in the chapter (p. 399), whereas other types of lipid-linked proteins (with farnesyl or geranylgeranyl groups attached; Fig. 11–15) are exclusively on the inner face. In polarized epithelial cells (such as intestinal epithelial cells; see Fig. 11–43), in which apical and basal surfaces have different roles, GPI-anchored proteins are directed specifically to the apical surface. Attachment of a specific lipid to a newly synthesized membrane protein therefore has a targeting function, directing the protein to its correct membrane location.

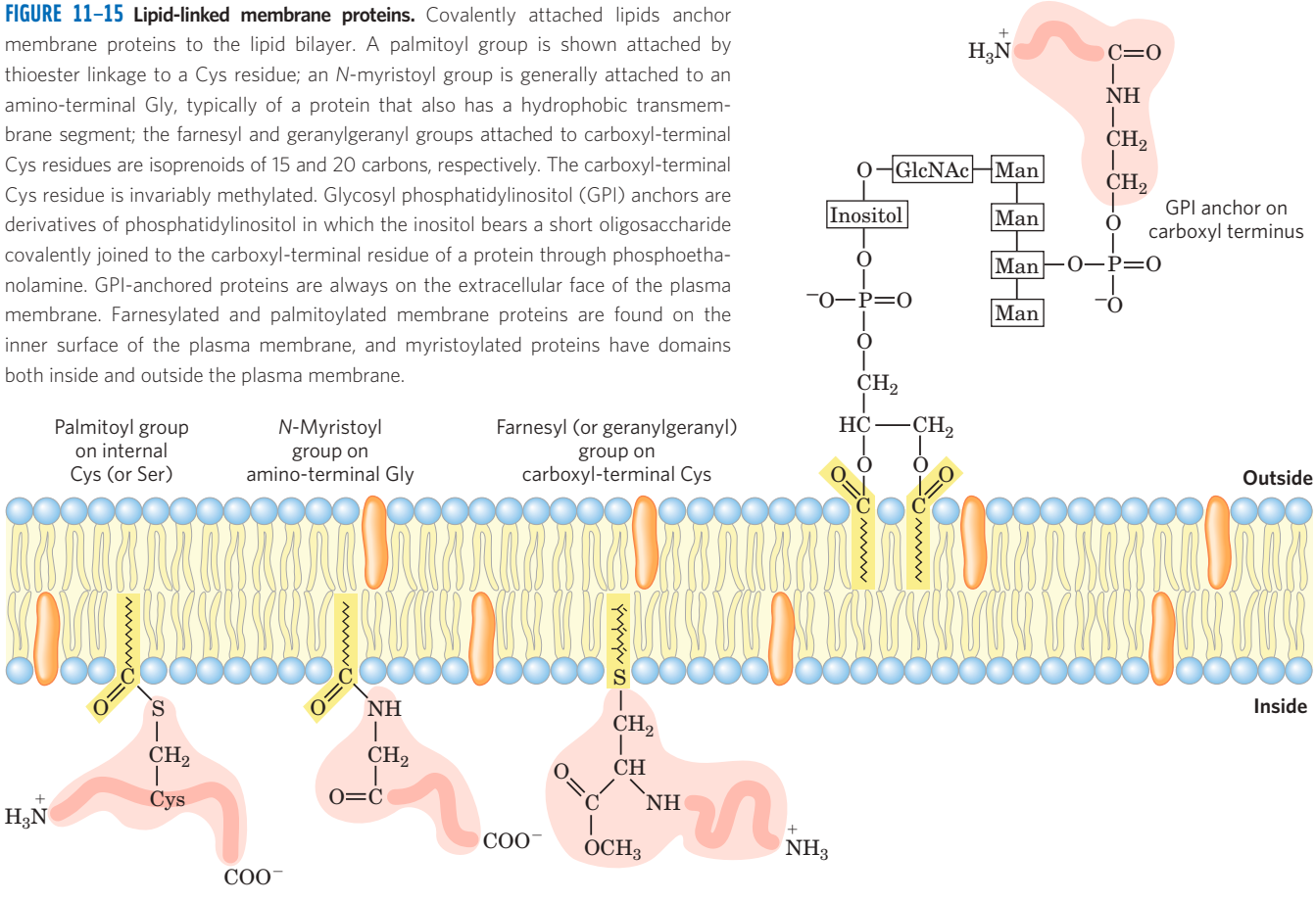


FIGURE 11–15 Lipid-linked membrane proteins. Covalently attached lipids anchor membrane proteins to the lipid bilayer. A palmitoyl group is shown attached by thioester linkage to a Cys residue; an *N*-myristoyl group is generally attached to an amino-terminal Gly, typically of a protein that also has a hydrophobic transmembrane segment; the farnesyl and geranylgeranyl groups attached to carboxyl-terminal Cys residues are isoprenoids of 15 and 20 carbons, respectively. The carboxyl-terminal Cys residue is invariably methylated. Glycosyl phosphatidylinositol (GPI) anchors are derivatives of phosphatidylinositol in which the inositol bears a short oligosaccharide covalently joined to the carboxyl-terminal residue of a protein through phosphoethanolamine. GPI-anchored proteins are always on the extracellular face of the plasma membrane. Farnesylated and palmitoylated membrane proteins are found on the inner surface of the plasma membrane, and myristoylated proteins have domains both inside and outside the plasma membrane.

SUMMARY 11.1 The Composition and Architecture of Membranes

- Biological membranes define cellular boundaries, divide cells into discrete compartments, organize complex reaction sequences, and act in signal reception and energy transformations.
- Membranes are composed of lipids and proteins in varying combinations particular to each species, cell type, and organelle. The lipid bilayer is the basic structural unit.
- Peripheral membrane proteins are loosely associated with the membrane through electrostatic interactions and hydrogen bonds or by covalently attached lipid anchors. Integral proteins associate firmly with membranes by hydrophobic interactions between the lipid bilayer and their nonpolar amino acid side chains, which are oriented toward the outside of the protein molecule. Amphitropic proteins associate reversibly with membranes.
- Many membrane proteins span the lipid bilayer several times, with hydrophobic sequences of about 20 amino acid residues forming transmembrane α helices. Multistranded β barrels are also common in integral proteins in bacterial membranes. Tyr and Trp residues of transmembrane proteins are commonly found at the lipid-water interface.
- The lipids and proteins of membranes are inserted into the bilayer with specific sidedness; thus membranes are structurally and functionally asymmetric. Plasma membrane glycoproteins are always oriented with the oligosaccharide-bearing domain on the extracellular surface.

11.2 Membrane Dynamics

One remarkable feature of all biological membranes is their flexibility—their ability to change shape without losing their integrity and becoming leaky. The basis for this property is the noncovalent interactions among lipids in the bilayer and the mobility allowed to individual lipids because they are not covalently anchored to one another. We turn now to the dynamics of membranes: the motions that occur and the transient structures allowed by these motions.

Acyl Groups in the Bilayer Interior Are Ordered to Varying Degrees

Although the lipid bilayer structure is stable, its individual phospholipid molecules have much freedom of motion **(Fig. 11–16)**, depending on the temperature and the lipid composition. Below normal physiological temperatures, the lipids in a bilayer form a semisolid **liquid-ordered (L_o) state**, in which all types of motion of individual lipid molecules are strongly constrained;

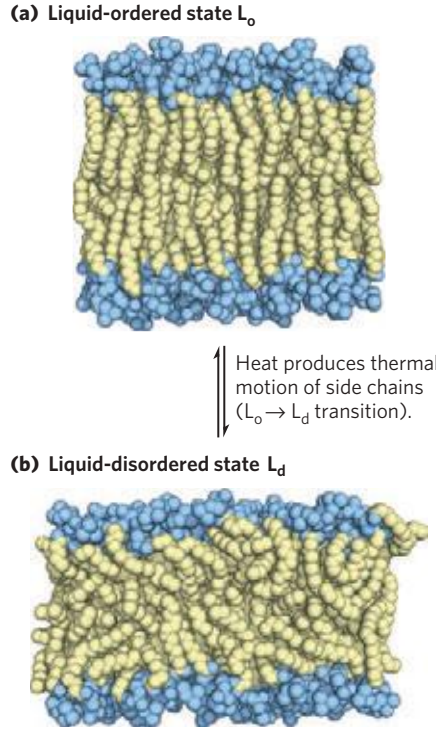


FIGURE 11–16 Two extreme states of bilayer lipids. (a) In the liquid-ordered (L_o) state, polar head groups are uniformly arrayed at the surface, and the acyl chains are nearly motionless and packed with regular geometry. **(b)** In the liquid-disordered (L_d) state, or fluid state, acyl chains undergo much thermal motion and have no regular organization. The state of membrane lipids in biological membranes is maintained somewhere between these extremes.

the bilayer is paracrystalline (Fig. 11–16a). Above physiological temperatures, individual hydrocarbon chains of fatty acids are in constant motion produced by rotation about the carbon–carbon bonds of the long acyl side chains and by lateral diffusion of individual lipid molecules in the plane of the bilayer. This is the **liquid-disordered (L_d) state** (Fig. 11–16b). In the transition from the L_o state to the L_d state, the general shape and dimensions of the bilayer are maintained; what changes is the degree of motion (lateral and rotational) allowed to individual lipid molecules.

At temperatures in the physiological range for a mammal (about 20 to 40 °C), long-chain saturated fatty acids (such as 16:0 and 18:0) tend to pack into an L_o gel phase, but the kinks in unsaturated fatty acids (see Fig. 10–2) interfere with packing, favoring the L_d state. Shorter-chain fatty acyl groups have the same effect. The sterol content of a membrane (which varies greatly with organism and organelle; Table 11–1) is another important determinant of lipid state. Sterols (such as cholesterol) have paradoxical effects on bilayer fluidity: they interact with phospholipids containing unsaturated fatty acyl chains, compacting them and constraining their motion in bilayers. Sterol association with sphingolipids and phospholipids with long, saturated fatty acyl chains tends, rather, to fluidize the bilayer, which, without

TABLE 11–2 Fatty Acid Composition of *E. coli* Cells Cultured at Different Temperatures

	Percentage of total fatty acids*			
	10 °C	20 °C	30 °C	40 °C
Myristic acid (14:0)	4	4	4	8
Palmitic acid (16:0)	18	25	29	48
Palmitoleic acid (16:1)	26	24	23	9
Oleic acid (18:1)	38	34	30	12
Hydroxymyristic acid	13	10	10	8
Ratio of unsaturated to saturated†	2.9	2.0	1.6	0.38

Source: Data from Marr, A.G. & Ingraham, J.L. (1962) Effect of temperature on the composition of fatty acids in *Escherichia coli*. *J. Bacteriol.* 84, 1260.

*The exact fatty acid composition depends not only on growth temperature but on growth stage and growth medium composition.

†Ratios calculated as the total percentage of 16:1 plus 18:1 divided by the total percentage of 14:0 plus 16:0. Hydroxymyristic acid was omitted from this calculation.

cholesterol, would adopt the L_o state. In biological membranes composed of a variety of phospholipids and sphingolipids, cholesterol tends to associate with sphingolipids and to form regions in the L_o state surrounded by cholesterol-poor regions in the L_d state (see the discussion of membrane rafts below).

Cells regulate their lipid composition to achieve a constant membrane fluidity under various growth conditions. For example, bacteria synthesize more unsaturated fatty acids and fewer saturated ones when cultured at low temperatures than when cultured at higher temperatures (Table 11–2). As a result of this adjustment in lipid composition, membranes of bacteria cultured at high or low temperatures have about the same degree of fluidity. This is presumably essential for the function of many proteins—enzymes, transporters, and receptors—that act within the lipid bilayer.

Transbilayer Movement of Lipids Requires Catalysis

At physiological temperatures, transbilayer—or "flip-flop"—diffusion of a lipid molecule from one leaflet of the bilayer to the other **(Fig. 11–17a)** occurs very slowly if at all in most membranes, although lateral diffusion *in the plane* of the bilayer is very rapid (Fig. 11–17b). Transbilayer movement requires that a polar or charged head group leave its aqueous environment and move into the hydrophobic interior of the bilayer, a process with a large, positive free-energy change. There are, however, situations in which such movement is essential. For example, in the ER, membrane glycerophospholipids are synthesized on the cytosolic surface, whereas sphingolipids are synthesized or modified on the lumenal surface. To get from their site of synthesis to their eventual point of deposition, these lipids must undergo flip-flop diffusion.

The asymmetric disposition of lipid types in the bilayer predicts the existence of flippases, floppases, and scramblases (Fig. 11–17c), which facilitate the transbilayer movement of lipids, providing a path that is energetically more favorable and much faster than the uncatalyzed movement. The combination of asymmetric biosynthesis of membrane lipids, very slow uncatalyzed flip-flop diffusion, and the presence of selective, energy-dependent lipid translocators could account for the transbilayer asymmetry in lipid composition shown in Figure 11–5. Besides contributing to this asymmetry of composition, the energy-dependent transport of lipids to one bilayer leaflet may, by creating a larger surface on one side of the bilayer, be important in generating the membrane curvature essential in the budding of vesicles.

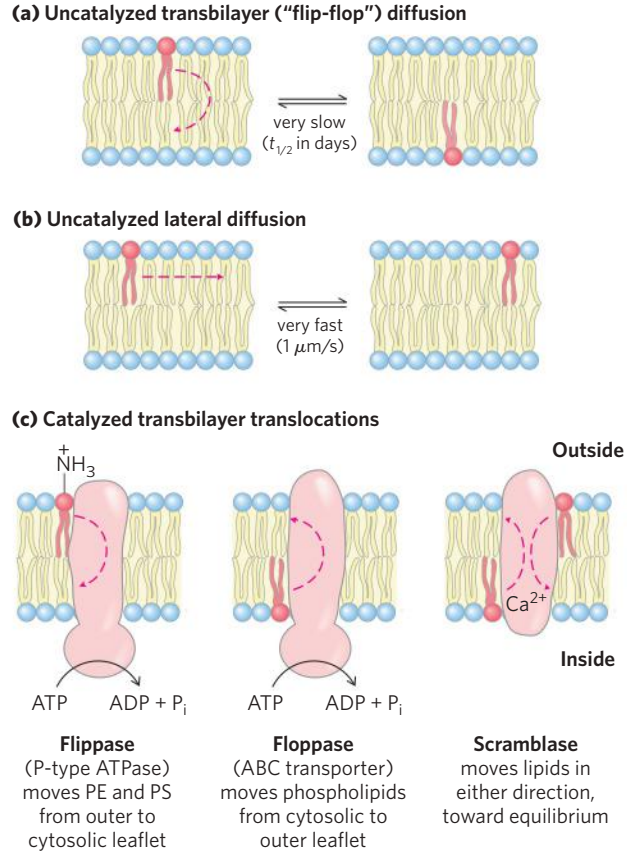


FIGURE 11–17 **Motion of single phospholipids in a bilayer. (a)** Uncatalyzed movement from one leaflet to the other is very slow, but **(b)** lateral diffusion within the leaflet is very rapid, requiring no catalysis. **(c)** Three types of phospholipid translocaters in the plasma membrane. PE is phosphatidylethanolamine; PS is phosphatidylserine.

Flippases catalyze translocation of the *amino*-phospholipids phosphatidylethanolamine and phosphatidylserine from the extracellular to the cytosolic leaflet of the plasma membrane, contributing to the asymmetric distribution of phospholipids: phosphatidylethanolamine and phosphatidylserine primarily in the cytosolic leaflet, and the sphingolipids and phosphatidylcholine in the outer leaflet. Keeping phosphatidylserine out of the extracellular leaflet is important: its exposure on the outer surface triggers apoptosis (programmed cell death; see Chapter 12) and engulfment by macrophages that carry phosphatidylserine receptors. Flippases also act in the ER, where they move newly synthesized phospholipids from their site of synthesis in the cytosolic leaflet to the lumenal leaflet. Flippases consume about one ATP per molecule of phospholipid translocated, and they are structurally and functionally related to the P-type ATPases (active transporters) described on page 410.

Two other types of lipid-translocating activities are known but less well characterized. **Floppases** move plasma membrane phospholipids from the cytosolic to the extracellular leaflet and like flippases are ATP-dependent. Floppases are members of the ABC transporter family described on page 413, all of which actively transport hydrophobic substrates outward across the plasma membrane. **Scramblases** are proteins that move any membrane phospholipid across the bilayer down its concentration gradient (from the leaflet where it has a higher concentration to the leaflet where it has a lower concentration); their activity is not dependent on ATP. Scramblase activity leads to controlled randomization of the head-group composition on the two faces of the bilayer. The activity rises sharply with an increase in cytosolic Ca^{2+} concentration, which may result from cell activation, cell injury, or apoptosis; as noted above, exposure of phosphatidylserine on the outer surface marks a cell for apoptosis and engulfment by macrophages. Finally, a group of proteins that act primarily to move phosphatidylinositol lipids across lipid bilayers, the phosphatidylinositol transfer proteins, are believed to have important roles in lipid signaling and membrane trafficking.

Lipids and Proteins Diffuse Laterally in the Bilayer

Individual lipid molecules can move laterally in the plane of the membrane by changing places with neighboring lipid molecules; that is, they undergo Brownian movement within the bilayer (Fig. 11–17b), which can be quite rapid. A molecule in the outer leaflet of the erythrocyte plasma membrane, for example, can diffuse laterally so fast that it circumnavigates the erythrocyte in seconds. This rapid lateral diffusion in the plane of the bilayer tends to randomize the positions of individual molecules in a few seconds.

Lateral diffusion can be shown experimentally by attaching fluorescent probes to the head groups of lipids and using fluorescence microscopy to follow the probes over time **(Fig. 11–18)**. In one technique, a small region ($5\ \mu m^2$) of a cell surface with fluorescence-tagged lipids

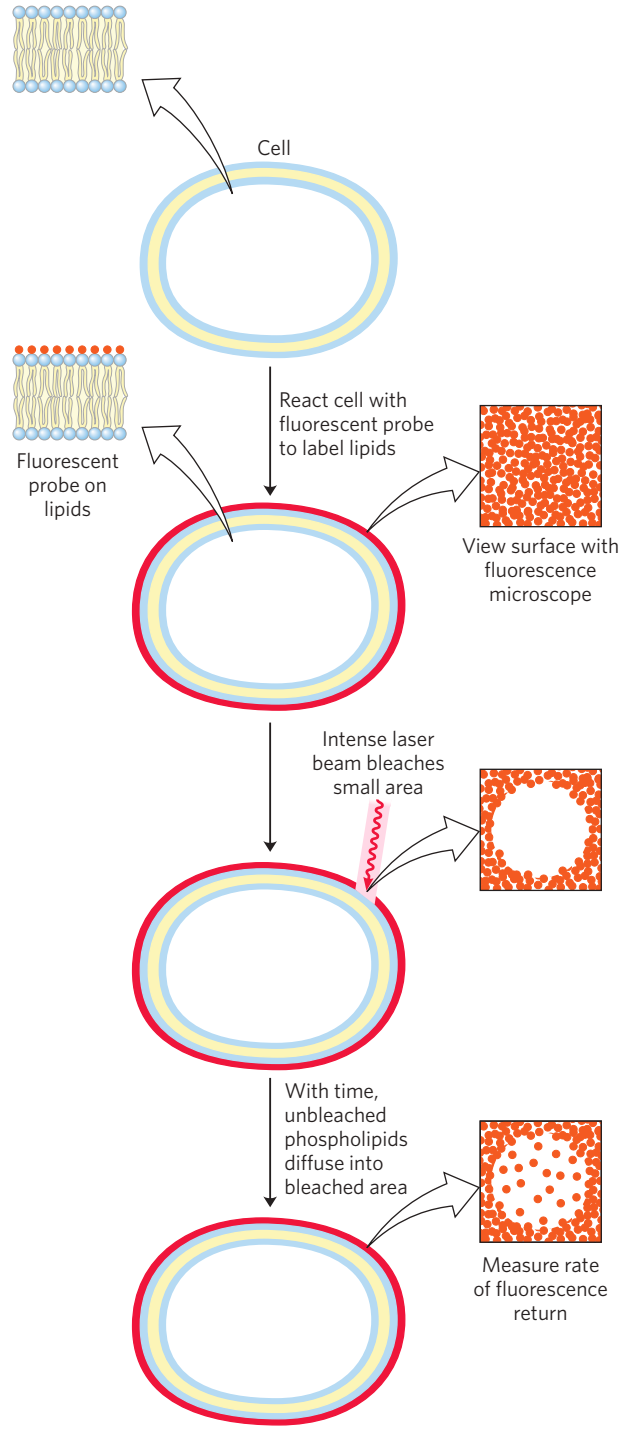


FIGURE 11–18 Measurement of lateral diffusion rates of lipids by fluorescence recovery after photobleaching (FRAP). Lipids in the outer leaflet of the plasma membrane are labeled by reaction with a membrane-impermeant fluorescent probe (red) so that the surface is uniformly labeled when viewed with a fluorescence microscope. A small area is bleached by irradiation with an intense laser beam and becomes nonfluorescent. With the passage of time, labeled lipid molecules diffuse into the bleached region, and it again becomes fluorescent. Researchers can track the time course of fluorescence return and determine a diffusion coefficient for the labeled lipid. The diffusion rates are typically high; a lipid moving at this speed could circumnavigate an *E. coli* cell in one second. (The FRAP method can also be used to measure lateral diffusion of membrane proteins.)

is bleached by intense laser radiation so that the irradiated patch no longer fluoresces when viewed with less-intense (nonbleaching) light in the fluorescence microscope. However, within milliseconds, the region recovers its fluorescence as unbleached lipid molecules diffuse into the bleached patch and bleached lipid molecules diffuse away from it. The rate of *f*luorescence *r*ecovery *a*fter *p*hotobleaching, or **FRAP**, is a measure of the rate of lateral diffusion of the lipids. Using the FRAP technique, researchers have shown that some membrane lipids diffuse laterally at rates of up to 1 μm/s.

Another technique, single particle tracking, allows one to follow the movement of a *single* lipid molecule in the plasma membrane on a much shorter time scale. Results from these studies confirm rapid lateral diffusion within small, discrete regions of the cell surface and show that movement from one such region to a nearby region ("hop diffusion") is inhibited; membrane lipids behave as though corralled by fences that they can occasionally cross by hop diffusion **(Fig. 11–19)**.

Many membrane proteins move as if afloat in a sea of lipids. Like membrane lipids, these proteins are free to diffuse laterally in the plane of the bilayer and are in constant motion, as shown by the FRAP technique with fluorescence-tagged surface proteins. Some membrane proteins associate to form large aggregates ("patches") on the surface of a cell or organelle in which individual protein molecules do not move relative to one another; for example, acetylcholine receptors form dense, near-crystalline patches on neuronal plasma membranes at synapses. Other membrane proteins are anchored to

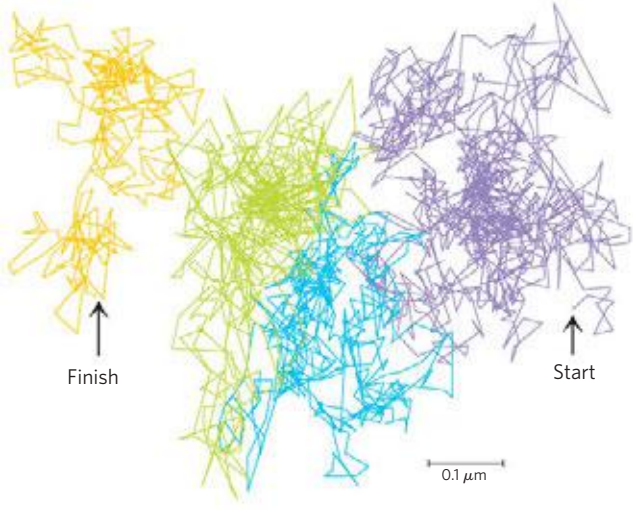


FIGURE 11–19 Hop diffusion of individual lipid molecules. The motion of a single fluorescently labeled lipid molecule in a cell surface is recorded on video by fluorescence microscopy, with a time resolution of 25 μs (equivalent to 40,000 frames/s). The track shown here represents a molecule followed for 56 ms (2,250 frames); the trace begins in the purple area and continues through blue, green, and orange. The pattern of movement indicates rapid diffusion within a confined region (about 250 nm in diameter, shown by a single color), with occasional hops into an adjoining region. This finding suggests that the lipids are corralled by molecular fences that they occasionally jump.

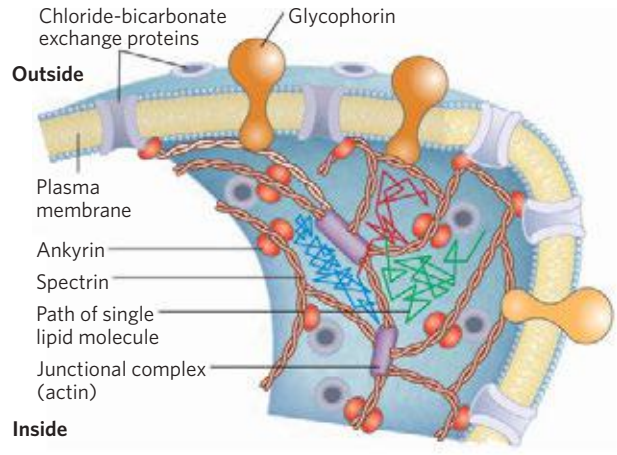


FIGURE 11–20 Restricted motion of the erythrocyte chloride-bicarbonate exchanger and glycophorin. The proteins span the membrane and are tethered to spectrin, a cytoskeletal protein, by another protein, ankyrin, limiting their lateral mobility. Ankyrin is anchored in the membrane by a covalently bound palmitoyl side chain (see Fig. 11–15). Spectrin, a long, filamentous protein, is cross-linked at junctional complexes containing actin. A network of cross-linked spectrin molecules attached to the cytoplasmic face of the plasma membrane stabilizes the membrane, making it resistant to deformation. This network of anchored membrane proteins may form the "corral" suggested by the experiment shown in Figure 11–19; the lipid tracks shown here are confined to different regions defined by the tethered membrane proteins. Occasionally a lipid molecule (green track) jumps from one corral to another (blue track), then another (red track).

internal structures that prevent their free diffusion. In the erythrocyte membrane, both glycophorin and the chloride-bicarbonate exchanger (p. 407) are tethered to spectrin, a filamentous cytoskeletal protein **(Fig. 11–20)**. One possible explanation for the pattern of lateral diffusion of lipid molecules shown in Figure 11–19 is that membrane proteins immobilized by their association with spectrin form the "fences" that define the regions of relatively unrestricted lipid motion.

Sphingolipids and Cholesterol Cluster Together in Membrane Rafts

We have seen that diffusion of membrane lipids from one bilayer leaflet to the other is very slow unless catalyzed and that the different lipid species of the plasma membrane are asymmetrically distributed in the two leaflets of the bilayer (Fig. 11–5). Even within a single leaflet, the lipid distribution is not uniform. Glycosphingolipids (cerebrosides and gangliosides), which typically contain long-chain saturated fatty acids, form transient clusters in the outer leaflet that largely exclude glycerophospholipids, which typically contain one unsaturated fatty acyl group and a shorter saturated acyl group. The long, saturated acyl groups of sphingolipids can form more compact, more stable associations with the long ring system of cholesterol than can the shorter, often unsaturated, chains of phospholipids. The cholesterol-sphingolipid **microdomains** in the outer monolayer of the plasma

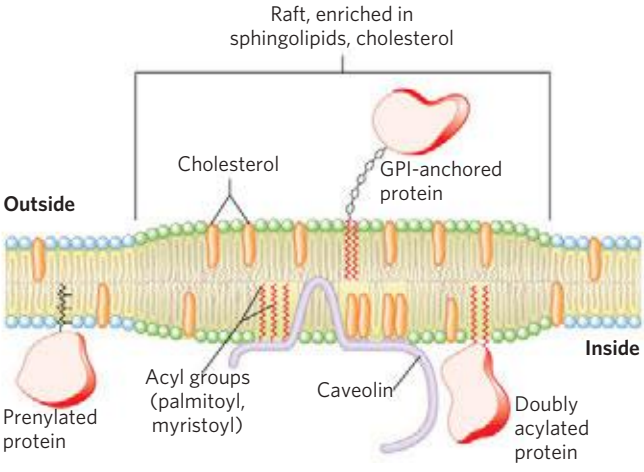


FIGURE 11–21 Membrane microdomains (rafts). Stable associations of sphingolipids and cholesterol in the outer leaflet produce a microdomain, slightly thicker than other membrane regions, that is enriched with specific types of membrane proteins. GPI-anchored proteins are prominent in the outer leaflet of these rafts, and proteins with one or several covalently attached long-chain acyl groups are common in the inner leaflet. Inwardly curved rafts called caveolae are especially enriched in the protein caveolin (see Fig. 11–22). Proteins with attached prenyl groups (such as Ras; see Box 12–2) tend to be excluded from rafts.

membrane are slightly thicker and more ordered (less fluid) than neighboring microdomains rich in phospholipids and are more difficult to dissolve with nonionic detergents; they behave like liquid-ordered sphingolipid **rafts** adrift on an ocean of liquid-disordered phospholipids **(Fig. 11–21)**.

These lipid rafts are remarkably enriched in two classes of integral membrane proteins: those anchored to the membrane by two covalently attached long-chain saturated fatty acids attached through Cys residues (two palmitoyl groups or a palmitoyl and a myristoyl group) and **GPI-anchored proteins** (Fig. 11–15). Presumably these lipid anchors, like the long, saturated acyl chains of sphingolipids, form more stable associations with the cholesterol and long acyl groups in rafts than with the surrounding phospholipids. (It is notable that other lipid-linked proteins, those with covalently attached isoprenyl groups such as farnesyl, are *not* preferentially associated with the outer leaflet of sphingolipid/cholesterol rafts (Fig. 11–21).) The "raft" and "sea" domains of the plasma membrane are not rigidly separated; membrane proteins can move into and out of lipid rafts on a time scale of seconds. But in the shorter time scale (microseconds) more relevant to many membrane-mediated biochemical processes, many of these proteins reside primarily in a raft.

We can estimate the fraction of the cell surface occupied by rafts from the fraction of the plasma membrane that resists detergent solubilization, which can be as high as 50% in some cases: the rafts cover half of the ocean. Indirect measurements in cultured fibroblasts suggest a diameter of roughly 50 nm for an individual raft, which corresponds to a patch containing a few thousand sphingolipids and perhaps 10 to 50 membrane proteins. Because most cells express more than 50 different kinds of plasma membrane proteins, it is likely that a single raft contains only a subset of membrane proteins and that this segregation of membrane proteins is functionally significant. For a process that involves interaction of two membrane proteins, their presence in a single raft would hugely increase the likelihood of their collision. Certain membrane receptors and signaling proteins, for example, seem to be segregated together in membrane rafts. Experiments show that signaling through these proteins can be disrupted by manipulations that deplete the plasma membrane of cholesterol and destroy lipid rafts.

Caveolin is an integral membrane protein with two globular domains connected by a hairpin-shaped hydrophobic domain, which binds the protein to the cytoplasmic leaflet of the plasma membrane. Three palmitoyl groups attached to the carboxyl-terminal globular domain further anchor it to the membrane. Caveolin (actually, a family of related caveolins) forms dimers and associates with cholesterol-rich regions in the membrane, and the presence of caveolin dimers forces the associated lipid bilayer to curve inward, forming **caveolae** ("little caves") in the surface of the cell **(Fig. 11–22)**. Caveolae are unusual rafts: they involve *both* leaflets of the bilayer—the cytoplasmic leaflet, from which the caveolin globular domains project, and the extracellular leaflet, a typical sphingolipid/cholesterol raft with associated GPI-anchored proteins. Caveolae are implicated in a variety of cellular functions, including membrane trafficking within cells and the transduction of external signals into cellular responses. The receptors for insulin and other growth factors, as well as certain GTP-binding proteins and protein kinases associated with transmembrane signaling, seem to be localized in rafts and perhaps in caveolae. We discuss some possible roles of rafts in signaling in Chapter 12.

Membrane Curvature and Fusion Are Central to Many Biological Processes

Caveolin is not unique in its ability to induce curvature in membranes. Changes of curvature are central to one of the most remarkable features of biological membranes: their ability to undergo fusion with other membranes without losing their continuity. Although membranes are stable, they are by no means static. Within the eukaryotic endomembrane system (which includes the nuclear membrane, endoplasmic reticulum, Golgi complex, and various small vesicles), the membranous compartments constantly reorganize. Vesicles bud from the ER to carry newly synthesized lipids and proteins to other organelles and to the plasma membrane. Exocytosis, endocytosis, cell division, fusion of egg and sperm cells, and entry of a membrane-enveloped virus into its host cell all involve membrane reorganization in which the fundamental operation is fusion of two membrane

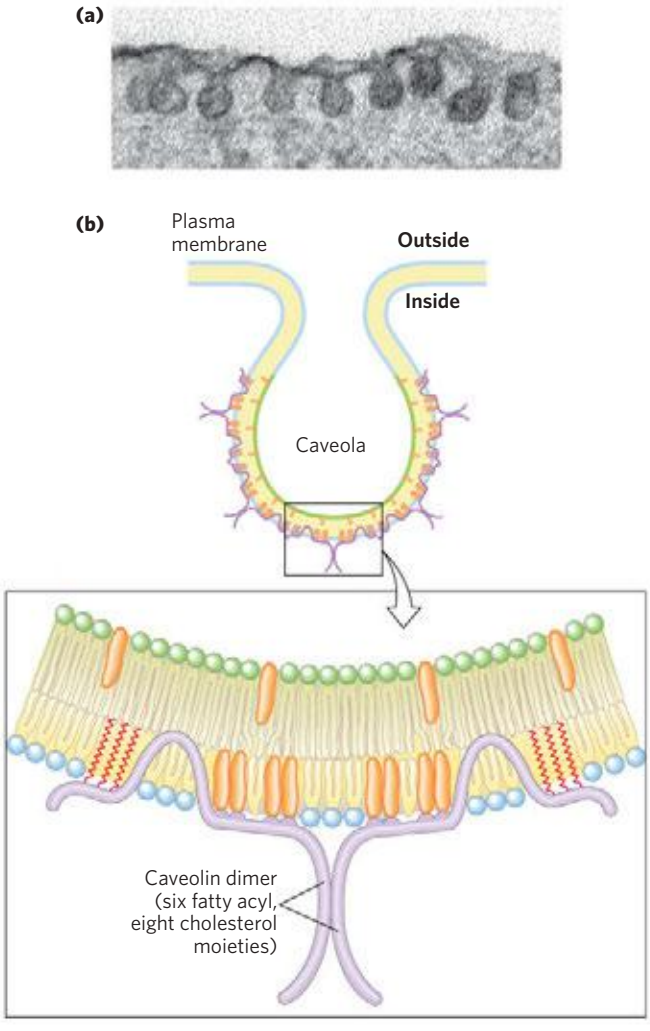


FIGURE 11–22 Caveolin forces inward curvature of a membrane. Caveolae are small invaginations in the plasma membrane, as seen in **(a)** an electron micrograph of an adipocyte that is surface-labeled with an electron-dense marker. **(b)** Cartoon showing the location and role of caveolin in causing inward membrane curvature. Each caveolin monomer has a central hydrophobic domain and three long-chain acyl groups (red), which hold the molecule to the inside of the plasma membrane. When several caveolin dimers are concentrated in a small region (a raft), they force a curvature in the lipid bilayer, forming a caveola. Cholesterol molecules in the bilayer are shown in orange.

segments without loss of continuity **(Fig. 11–23)**. Most of these processes begin with a local increase in membrane curvature. A protein that is intrinsically curved may force curvature in a bilayer by binding to it **(Fig. 11–24)**; the binding energy provides the driving force for the increase in bilayer curvature. Alternatively, multiple subunits of a scaffold protein may assemble into curved supramolecular complexes and stabilize curves that spontaneously form in the bilayer. For example, a superfamily of proteins containing **BAR domains** (named for the first three members of the family to be identified: *B*IN1, *a*mphiphysin, and *R*VS167) can assemble into a crescent-shaped scaffold that binds to the membrane surface, forcing or favoring membrane curvature.

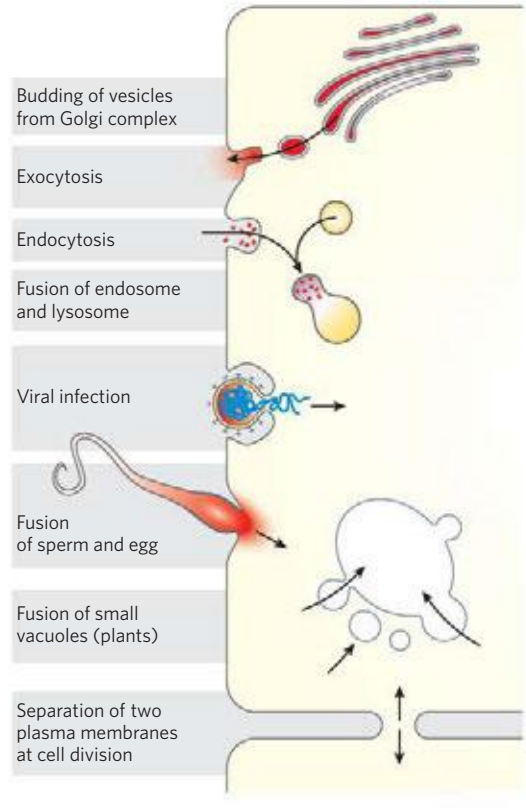


FIGURE 11–23 Membrane fusion. The fusion of two membranes is central to a variety of cellular processes involving organelles and the plasma membrane.

BAR domains consist of coiled coils that form long, thin, curved dimers with a positively charged concave surface that tends to form ionic interactions with the negatively charged head groups of membrane phospholipids (Fig. 11–24). Some of these BAR proteins also have a helical region that inserts into one leaflet of the bilayer, expanding its area relative to the other leaflet and thereby forcing curvature.

Specific fusion of two membranes requires that (1) they recognize each other; (2) their surfaces become closely apposed, which requires the removal of water molecules normally associated with the polar head groups of lipids; (3) their bilayer structures become locally disrupted, resulting in fusion of the outer leaflet of each membrane (hemifusion); and (4) their bilayers fuse to form a single continuous bilayer. The fusion occurring in receptor-mediated endocytosis, or regulated secretion, also requires that (5) the process is triggered at the appropriate time or in response to a specific signal. Integral proteins called **fusion proteins** mediate these events, bringing about specific recognition and a transient local distortion of the bilayer structure that favors membrane fusion. (Note that these fusion proteins are unrelated to the products encoded by two fused genes, also called fusion proteins, discussed in Chapter 9.)

A well-studied example of membrane fusion is that occurring at synapses, when intracellular vesicles loaded

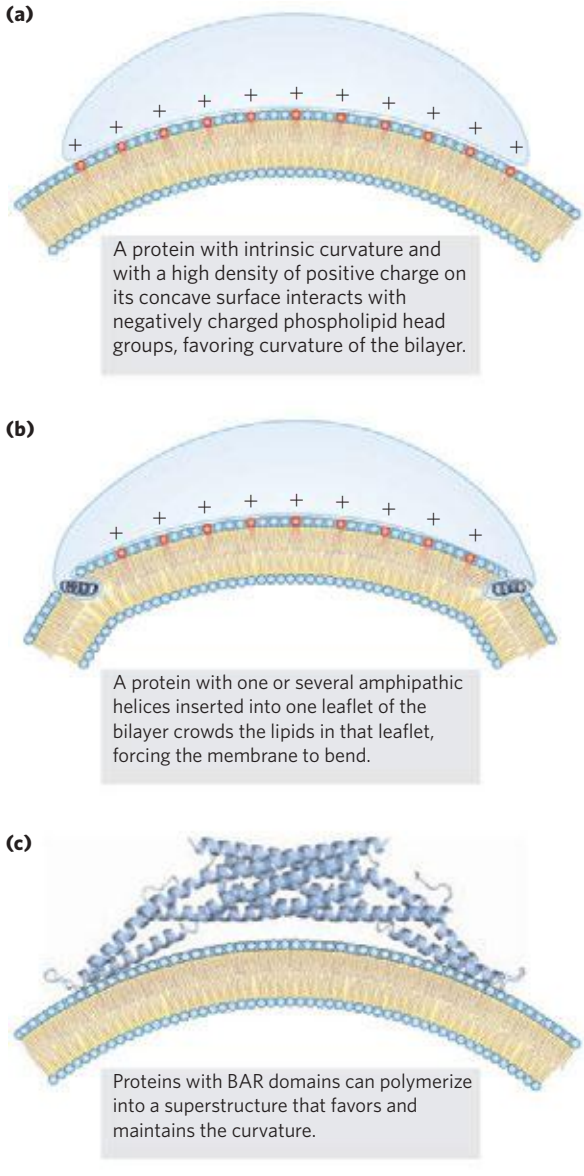


FIGURE 11–24 **Three models for protein-induced curvature of membranes.**

with neurotransmitter fuse with the plasma membrane. This process involves a family of proteins called SNARES **(Fig. 11–25)**. SNAREs in the cytoplasmic face of the intracellular vesicle are called **v-SNAREs**; those in the target membrane with which the vesicle fuses (the plasma membrane during exocytosis) are **t-SNAREs**. Two other proteins, SNAP25 and NSF, are also involved. During fusion, a v-SNARE and t-SNARE bind to each other and undergo a structural change that produces a bundle of long, thin rods made up of helices from both SNARES and two helices from SNAP25 (Fig. 11–25). The two SNAREs initially interact at their ends, then zip up into the bundle of helices. This structural change pulls the two membranes into contact and initiates the fusion of their lipid bilayers.

The complex of SNAREs and SNAP25 is the target of the powerful *Clostridium botulinum* toxin, a protease that cleaves specific bonds in these proteins,

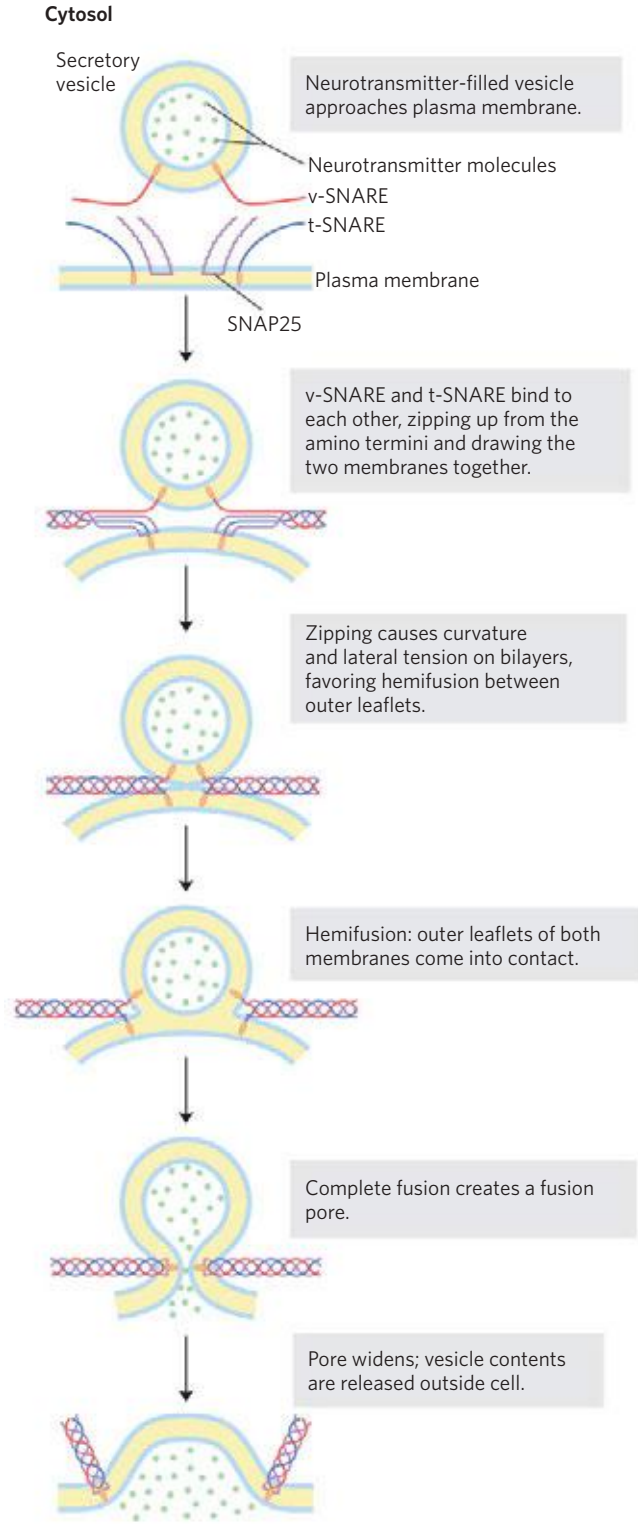


FIGURE 11–25 **Membrane fusion during neurotransmitter release at a synapse.** The secretory vesicle membrane contains the v-SNARE synaptobrevin (red). The target (plasma) membrane contains the t-SNAREs syntaxin (blue) and SNAP25 (violet). When a local increase in $[Ca^{2+}]$ signals release of neurotransmitter, the v-SNARE, SNAP25, and t-SNARE interact, forming a coiled bundle of four α helices, pulling the two membranes together and disrupting the bilayer locally. This leads first to hemifusion, joining the outer leaflets of the two membranes, then to complete membrane fusion and neurotransmitter release. NSF (*N*-ethylmaleimide-sensitive *f*usion factor) acts in disassembly of the SNARE complex when fusion is complete.

preventing neurotransmission and thereby causing the death of the organism. Because of its very high specificity for these proteins, purified botulinum toxin has served as a powerful tool for dissecting the mechanism of neurotransmitter release in vivo and in vitro.

Integral Proteins of the Plasma Membrane Are Involved in Surface Adhesion, Signaling, and Other Cellular Processes

Several families of integral proteins in the plasma membrane provide specific points of attachment between cells or between a cell and extracellular matrix proteins. **Integrins** are surface adhesion proteins that mediate a cell's interaction with the extracellular matrix and with other cells, including some pathogens. Integrins also carry signals in both directions across the plasma membrane, integrating information about the extracellular and intracellular environments. All integrins are heterodimeric proteins composed of two unlike subunits, α and β, each anchored to the plasma membrane by a single transmembrane helix. The large extracellular domains of the α and β subunits combine to form a specific binding site for extracellular proteins such as collagen and fibronectin, which contain a common determinant of integrin binding, the sequence Arg–Gly–Asp (RGD). We discuss the signaling functions of integrins in more detail in Chapter 12 (p. 470).

Other plasma membrane proteins involved in surface adhesion are the **cadherins**, which undergo homophilic ("with same kind") interactions with identical cadherins in an adjacent cell. **Selectins** have extracellular domains that, in the presence of Ca^{2+}, bind specific polysaccharides on the surface of an adjacent cell. Selectins are present primarily in the various types of blood cells and in the endothelial cells that line blood vessels (see Fig. 7–32). They are an essential part of the blood-clotting process.

Integral membrane proteins play roles in many other cellular processes. They serve as transporters and ion channels (discussed in Section 11.3) and as receptors for hormones, neurotransmitters, and growth factors (Chapter 12). They are central to oxidative phosphorylation and photophosphorylation (Chapter 19) and to cell-cell and cell-antigen recognition in the immune system (Chapter 5). Integral proteins are also important players in the membrane fusion that accompanies exocytosis, endocytosis, and the entry of many types of viruses into host cells.

SUMMARY 11.2 Membrane Dynamics

- Lipids in a biological membrane can exist in liquid-ordered or liquid-disordered states; in the latter state, thermal motion of acyl chains makes the interior of the bilayer fluid. Fluidity is affected by temperature, fatty acid composition, and sterol content.
- Flip-flop diffusion of lipids between the inner and outer leaflets of a membrane is very slow except when specifically catalyzed by flippases, floppases, or scramblases.
- Lipids and proteins can diffuse laterally within the plane of the membrane, but this mobility is limited by interactions of membrane proteins with internal cytoskeletal structures and interactions of lipids with lipid rafts. One class of lipid rafts consists of sphingolipids and cholesterol with a subset of membrane proteins that are GPI-linked or attached to several long-chain fatty acyl moieties.
- Caveolin is an integral membrane protein that associates with the inner leaflet of the plasma membrane, forcing it to curve inward to form caveolae, probably involved in membrane transport and signaling.
- Specific proteins containing BAR domains cause local membrane curvature and mediate the fusion of two membranes, which accompanies processes such as endocytosis, exocytosis, and viral invasion.
- Integrins are transmembrane proteins of the plasma membrane that act both to attach cells to each other and to carry messages between the extracellular matrix and the cytoplasm.

11.3 Solute Transport across Membranes

Every living cell must acquire from its surroundings the raw materials for biosynthesis and for energy production, and must release the byproducts of metabolism to its environment. A few nonpolar compounds can dissolve in the lipid bilayer and cross the membrane unassisted, but for transmembrane movement of any polar compound or ion, a membrane protein is essential. In some cases a membrane protein simply facilitates the diffusion of a solute down its concentration gradient, but transport can also occur against a gradient of concentration, electric charge, or both, in which case the process requires energy **(Fig. 11–26)**. The energy may come directly from ATP hydrolysis or may be supplied in the form of one solute moving down its electrochemical gradient, which provides sufficient energy to drive another solute up its gradient. Ions may also move across membranes via ion channels formed by proteins, or they may be carried across by ionophores, small molecules that mask the charge of ions and allow them to diffuse through the lipid bilayer. With very few exceptions, the traffic of small molecules across the plasma membrane is mediated by proteins such as transmembrane channels, carriers, or pumps. Within the eukaryotic cell, different compartments have different concentrations of ions and of metabolic intermediates and products, and these, too, must move across intracellular membranes in tightly regulated, protein-mediated processes.

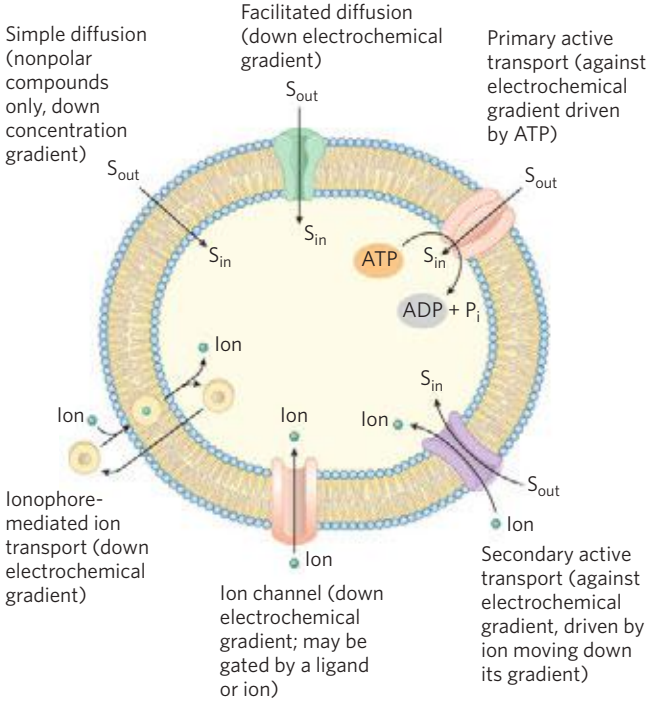


FIGURE 11–26 Summary of transporter types. Some types (ionophores, ion channels, and passive transporters) simply speed transmembrane movement of solutes down their electrochemical gradients, whereas others (active transporters) can pump solutes against a gradient, using ATP or a gradient of a second solute to provide the energy.

Passive Transport Is Facilitated by Membrane Proteins

When two aqueous compartments containing unequal concentrations of a soluble compound or ion are separated by a permeable divider (membrane), the solute moves by **simple diffusion** from the region of higher concentration, through the membrane, to the region of lower concentration, until the two compartments have equal solute concentrations **(Fig. 11–27a)**. When ions of opposite charge are separated by a permeable membrane, there is a transmembrane electrical gradient, a **membrane potential, V_m** (expressed in millivolts). This membrane potential produces a force opposing ion movements that increase V_m and driving ion movements that reduce V_m (Fig. 11–27b). Thus, the direction in which a charged solute tends to move spontaneously across a membrane depends on both the chemical gradient (the difference in solute concentration) and the electrical gradient (V_m) across the membrane. Together these two factors are referred to as the **electrochemical gradient** or **electrochemical potential**. This behavior of solutes is in accord with the second law of thermodynamics: molecules tend to spontaneously assume the distribution of greatest randomness and lowest energy.

To pass through a lipid bilayer, a polar or charged solute must first give up its interactions with the water molecules in its hydration shell, then diffuse about 3 nm (30 Å) through a substance (lipid) in which it is poorly soluble **(Fig. 11–28)**. The energy used to strip away the hydration shell and to move the polar compound from water into lipid, then through the lipid bilayer, is regained as the compound leaves the membrane on the other side and is rehydrated. However, the intermediate stage of transmembrane passage is a high-energy state comparable to the transition state in an enzyme-catalyzed chemical reaction. In both cases, an activation barrier must be overcome to reach the intermediate stage (Fig. 11–28; compare with Fig. 6–3). The energy of activation ($\Delta G^{\ddagger}$) for translocation of a polar solute across the bilayer is so large that pure lipid bilayers are virtually impermeable to polar and charged species over periods relevant to cell growth and division.

Membrane proteins lower the activation energy for transport of polar compounds and ions by providing an alternative path across the membrane for specific solutes. Proteins that bring about this **facilitated diffusion**, or

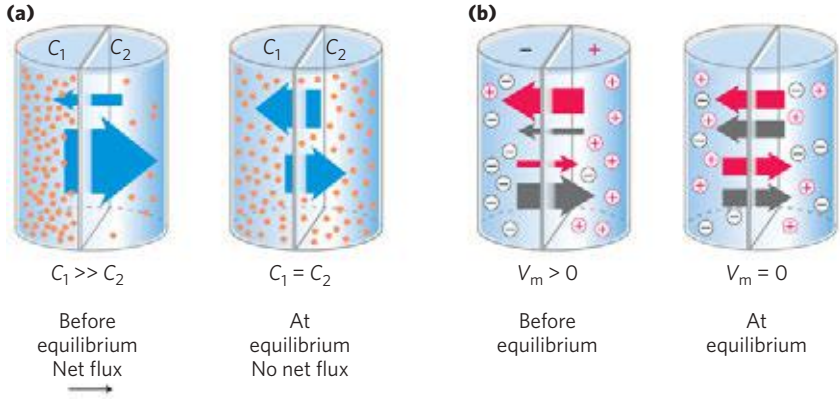


FIGURE 11–27 Movement of solutes across a permeable membrane. **(a)** Net movement of an electrically neutral solute is toward the side of lower solute concentration until equilibrium is achieved. The solute concentrations on the left and right sides of the membrane are designated C_1 and C_2. The rate of transmembrane solute movement (indicated by the arrows) is proportional to the concentration ratio. **(b)** Net movement of an electrically charged solute is dictated by a combination of the electrical potential (V_m) and the ratio of chemical concentrations (C_2/C_1) across the membrane; net ion movement continues until this electrochemical potential reaches zero.

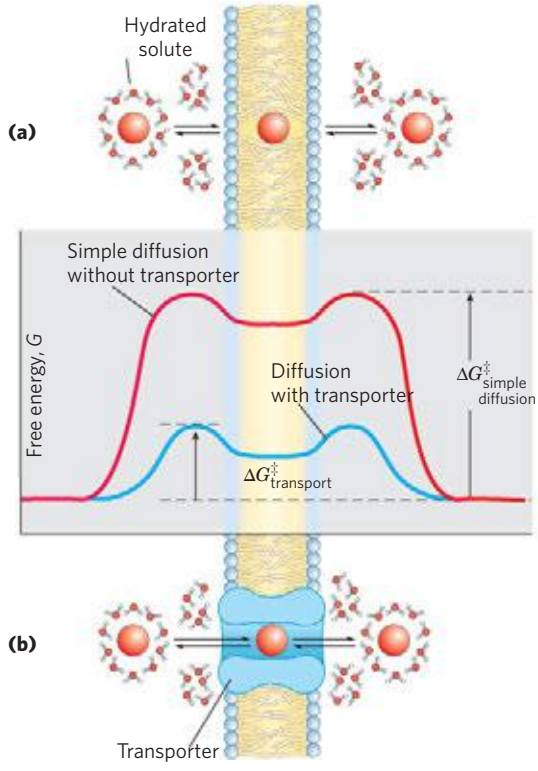


FIGURE 11–28 Energy changes accompanying passage of a hydrophilic solute through the lipid bilayer of a biological membrane. (a) In simple diffusion, removal of the hydration shell is highly endergonic, and the energy of activation ($\Delta G^{\ddagger}$) for diffusion through the bilayer is very high. **(b)** A transporter protein reduces the $\Delta G^{\ddagger}$ for transmembrane diffusion of the solute. It does this by forming noncovalent interactions with the dehydrated solute to replace the hydrogen bonding with water and by providing a hydrophilic transmembrane pathway.

passive transport, are not enzymes in the usual sense; their "substrates" are moved from one compartment to another but are not chemically altered. Membrane proteins that speed the movement of a solute across a membrane by facilitating diffusion are called **transporters** or **permeases**.

Like enzymes, transporters bind their substrates with stereochemical specificity through multiple weak, noncovalent interactions. The negative free-energy change associated with these weak interactions, $\Delta G_{\text{binding}}$, counterbalances the positive free-energy change that accompanies loss of the water of hydration from the substrate, $\Delta G_{\text{dehydration}}$, thereby lowering $\Delta G^{\ddagger}$ for transmembrane passage (Fig. 11–28). Transporters span the lipid bilayer several times, forming a transmembrane pathway lined with hydrophilic amino acid side chains. The pathway provides an alternative route for a specific substrate to move across the lipid bilayer without its having to dissolve in the bilayer, further lowering $\Delta G^{\ddagger}$ for transmembrane diffusion. The result is an increase of several to many orders of magnitude in the rate of transmembrane passage of the substrate.

Transporters and Ion Channels Are Fundamentally Different

We know from genomic studies that transporters constitute a significant fraction of all proteins encoded in the genomes of both simple and complex organisms. There are probably a thousand or more different genes in the human genome encoding proteins that allow molecules and ions to cross membranes. These proteins fall within two very broad categories: transporters and channels **(Fig. 11–29)**. **Transporters** for molecules and ions bind their substrates with high specificity, catalyze transport at rates well below the limits of free diffusion, and are saturable in the same sense as are enzymes: there is some substrate concentration above which further increases will not produce a greater rate of transport. **Channels** generally allow transmembrane movement of ions at rates that are orders of magnitude greater than those typical of transporters, approaching the limit of unhindered diffusion (tens of millions of ions per second per channel). Channels typically show some specificity for an ion, but are not saturable with the ion substrate, in contrast to the saturation kinetics seen with transporters. The direction of ion movement through an ion channel is dictated by the ion's charge and the electrochemical

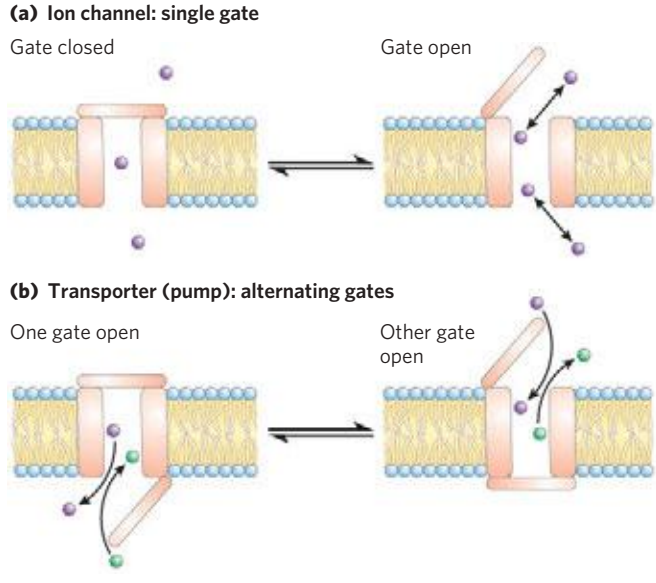


FIGURE 11–29 Differences between channels and transporters. (a) In an ion channel, a transmembrane pore is either open or closed, depending on the position of the single gate. When it is open, ions move through at a rate limited only by the maximum rate of diffusion. **(b)** Transporters (pumps) have two gates, and they are never both open. Movement of a substrate (an ion or a small molecule) through the membrane is therefore limited by the time needed for one gate to open and close (on one side of the membrane) and for the second gate to open. Rates of movement through ion channels can be orders of magnitude greater than rates through pumps, but channels simply allow the ion to flow down the electrochemical gradient, whereas pumps can move a substrate against a concentration gradient.

gradient across the membrane. Within each of these categories are families of various types, defined not only by their primary sequences but by their secondary structures. Among the transporters, some simply facilitate diffusion down a concentration gradient; they are the **passive transporters**. **Active transporters** can drive substrates across the membrane against a concentration gradient, some using energy provided directly by a chemical reaction (primary active transporters) and some coupling uphill transport of one substrate with downhill transport of another (secondary active transporters). We now consider some well-studied representatives of the main transporter and channel families. You will encounter some of these in Chapter 12 when we discuss transmembrane signaling and again in later chapters in the context of the metabolic pathways in which they participate.

The Glucose Transporter of Erythrocytes Mediates Passive Transport

Energy-yielding metabolism in erythrocytes depends on a constant supply of glucose from the blood plasma, where the glucose concentration is maintained at about 5 mM. Glucose enters the erythrocyte by facilitated diffusion via a specific glucose transporter, at a rate about 50,000 times greater than uncatalyzed transmembrane diffusion. The glucose transporter of erythrocytes (called GLUT1 to distinguish it from related glucose transporters in other tissues) is a type III integral protein (M_r ~45,000) with 12 hydrophobic segments, each of which is believed to form a membrane-spanning helix. The detailed structure of GLUT1 is not yet known, but one plausible model suggests that the side-by-side assembly of several helices produces a transmembrane channel lined with hydrophilic residues that can hydrogen-bond with glucose as it moves through the aqueous pore **(Fig. 11–30)**.

The process of glucose transport can be described by its analogy with an enzymatic reaction in which the "substrate" is glucose outside the cell (S_{out}), the "product" is glucose inside (S_{in}), and the "enzyme" is the transporter, T. When the initial rate of glucose uptake is measured as a function of external glucose concentration **(Fig. 11–31)**, the resulting plot is hyperbolic: at high external glucose concentrations the rate of uptake approaches V_{max}. Formally, such a transport process can be described by the equations

$$\begin{array}{ccc} S_{out} + T_1 & \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} & S_{out} \bullet T_1 \\ k_{-4}\Big\downarrow\Big\uparrow k_4 & & k_{-2}\Big\downarrow\Big\uparrow k_2 \\ S_{in} + T_2 & \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} & S_{in} \bullet T_2 \end{array}$$

in which k_1, k_{-1}, and so forth are the forward and reverse rate constants for each step; T_1 is the transporter conformation in which the glucose-binding site faces out, and T_2 is the conformation in which it faces in.

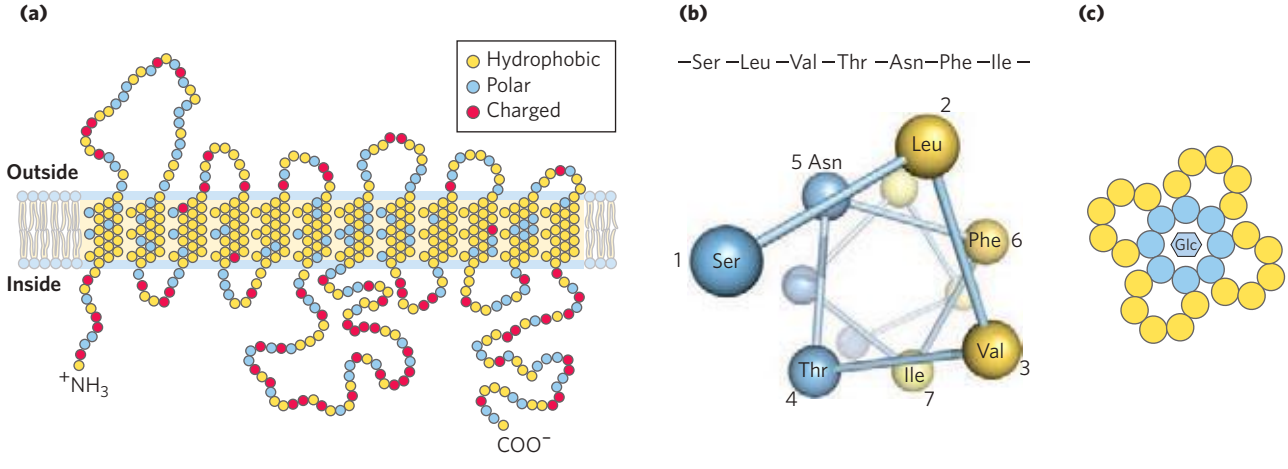


FIGURE 11–30 Membrane topology of the glucose transporter GLUT1. **(a)** Transmembrane helices are represented here as oblique (angled) rows of three or four amino acid residues, each row depicting one turn of the α helix. Nine of the 12 helices contain three or more polar or charged residues (blue or red), often separated by several hydrophobic residues (yellow). This representation of topology is not intended to depict three-dimensional structure. **(b)** A helical wheel diagram shows the distribution of polar and nonpolar residues on the surface of a helical segment. The helix is diagrammed as though observed along its axis from the amino terminus. Adjacent residues in the linear sequence are connected, and each residue is placed around the wheel in the position it occupies in the helix; recall that 3.6 residues are required to make one complete turn of the α helix. In this example, the polar residues (blue) are on one side of the helix and the hydrophobic residues (yellow) on the other. This is, by definition, an amphipathic helix. **(c)** Side-by-side association of four amphipathic helices, each with its polar face oriented toward the central cavity, can produce a transmembrane channel lined with polar (and charged) residues. This channel provides many opportunities for hydrogen bonding with glucose as it moves through.

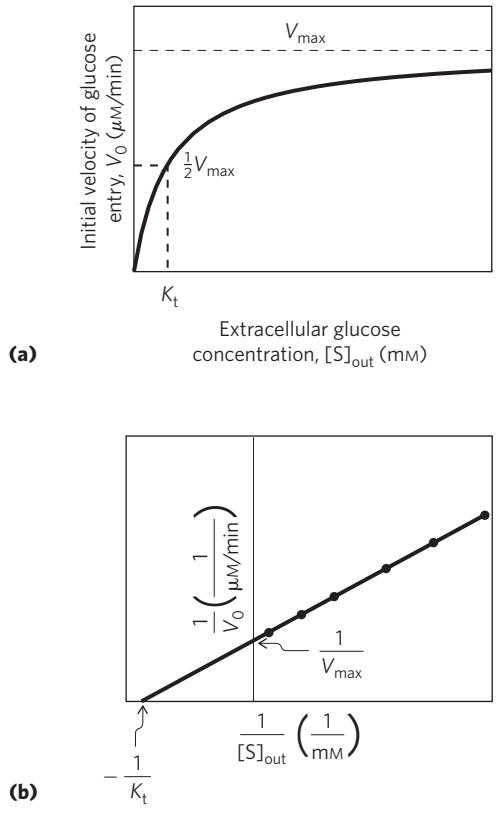


FIGURE 11–31 Kinetics of glucose transport into erythrocytes. (a) The initial rate of glucose entry into an erythrocyte, V_0, depends on the initial concentration of glucose on the outside, $[S]_{out}$. **(b)** Double-reciprocal plot of the data in (a). The kinetics of facilitated diffusion is analogous to the kinetics of an enzyme-catalyzed reaction. Compare these plots with Figure 6–11 and with Figure 1 in Box 6–1. Note that K_t is analogous to K_m, the Michaelis constant.

The steps are summarized in **Figure 11–32**. Given that every step in this sequence is reversible, the transporter is, in principle, equally able to move glucose into or out of the cell. However, with GLUT1, glucose

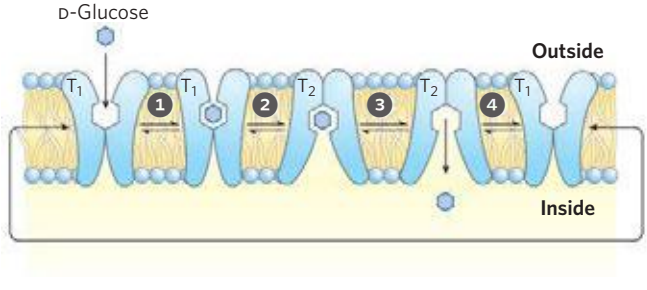


FIGURE 11–32 Model of glucose transport into erythrocytes by GLUT1. The transporter exists in two conformations: T_1, with the glucose-binding site exposed on the outer surface of the plasma membrane, and T_2, with the binding site exposed on the inner surface. Glucose transport occurs in four steps. ❶ Glucose in blood plasma binds to a stereospecific site on T_1; this lowers the activation energy for ❷ a conformational change from $glucose_{out} \cdot T_1$ to $glucose_{in} \cdot T_2$, effecting the transmembrane passage of the glucose. ❸ Glucose is released from T_2 into the cytoplasm, and ❹ the transporter returns to the T_1 conformation, ready to transport another glucose molecule.

always moves down its concentration gradient, which normally means *into* the cell. Glucose that enters a cell is generally metabolized immediately, and the intracellular glucose concentration is thereby kept low relative to its concentration in the blood.

The rate equations for glucose transport can be derived exactly as for enzyme-catalyzed reactions (Chapter 6), yielding an expression analogous to the Michaelis-Menten equation:

$$V_0 = \frac{V_{max}[S]_{out}}{K_t + [S]_{out}} \qquad (11–1)$$

in which V_0 is the initial velocity of accumulation of glucose inside the cell when its concentration in the surrounding medium is $[S]_{out}$ and $\boldsymbol{K_t}$ **(**$\boldsymbol{K_{transport}}$**)** is a constant analogous to the Michaelis constant, a combination of rate constants that is characteristic of each transport system. This equation describes the *initial* velocity, the rate observed when $[S]_{in} = 0$. As is the case for enzyme-catalyzed reactions, the slope-intercept form of the equation describes a linear plot of $1/V_0$ against $1/[S]_{out}$, from which we can obtain values of K_t and V_{max} (Fig. 11–31b). When $[S]_{out} = K_t$, the rate of uptake is $\frac{1}{2}V_{max}$; the transport process is half-saturated. The concentration of glucose in blood is 4.5 to 5 mM, close to the K_t, which ensures that GLUT1 is nearly saturated with substrate and operates near V_{max}.

Because no chemical bonds are made or broken in the conversion of S_{out} to S_{in}, neither "substrate" nor "product" is intrinsically more stable, and the process of entry is therefore fully reversible. As $[S]_{in}$ approaches $[S]_{out}$, the rates of entry and exit become equal. Such a system is therefore incapable of accumulating glucose within a cell at concentrations above that in the surrounding medium; it simply equilibrates glucose on the two sides of the membrane much faster than would occur in the absence of a specific transporter. GLUT1 is specific for D-glucose, with a measured K_t of about 6 mM. For the close analogs D-mannose and D-galactose, which differ only in the position of one hydroxyl group, the values of K_t are 20 and 30 mM, respectively, and for L-glucose, K_t exceeds 3,000 mM. Thus, GLUT1 shows the three hallmarks of passive transport: high rates of diffusion down a concentration gradient, saturability, and specificity.

Twelve passive glucose transporters are encoded in the human genome, each with its unique kinetic properties, patterns of tissue distribution, and function (Table 11–3). In the liver, GLUT2 transports glucose out of hepatocytes when liver glycogen is broken down to replenish blood glucose. GLUT2 has a large K_t (17 mM or greater) and can therefore respond to increased levels of intracellular glucose (produced by glycogen breakdown) by increasing outward transport. Skeletal and heart muscle and adipose tissue have yet another glucose transporter, GLUT4

TABLE 11–3 Glucose Transporters in Humans

Transporter	Tissue(s) where expressed	K_t (mM)*	Role†
GLUT1	Ubiquitous	3	Basal glucose uptake
GLUT2	Liver, pancreatic islets, intestine	17	In liver and kidney, removal of excess glucose from blood; in pancreas, regulation of insulin release
GLUT3	Brain (neuronal), testis (sperm)	1.4	Basal glucose uptake
GLUT4	Muscle, fat, heart	5	Activity increased by insulin
GLUT5	Intestine (primarily), testis, kidney	6‡	Primarily fructose transport
GLUT6	Spleen, leukocytes, brain	>5	Possibly no transporter function
GLUT7	Small intestine, colon	0.3	—
GLUT8	Testis	~2	—
GLUT9	Liver, kidney	0.6	—
GLUT10	Heart, lung, brain, liver, muscle, pancreas, kidney	0.3§	—
GLUT11	Heart, skeletal muscle, kidney	0.16	—
GLUT12	Skeletal muscle, heart, prostate, small intestine	—	—

*K_t for glucose, except as noted, from Augustin, R. (2010) The protein family of glucose transport facilitators: it's not only about glucose after all. *IUBMB Life* 62, 315–333.

† Dash indicates role uncertain.

‡ K_m for fructose.

§ K_m for 2-deoxyglucose.

(K_t = 5 mM), which is distinguished by its response to insulin: its activity increases when insulin signals a high blood glucose concentration, thus increasing the rate of glucose uptake into muscle and adipose tissue (Box 11–1 describes the effect of insulin on this transporter).

The Chloride-Bicarbonate Exchanger Catalyzes Electroneutral Cotransport of Anions across the Plasma Membrane

The erythrocyte contains another facilitated diffusion system, an anion exchanger that is essential in CO_2 transport to the lungs from tissues such as skeletal muscle and liver. Waste CO_2 released from respiring tissues into the blood plasma enters the erythrocyte, where it is converted to bicarbonate (HCO_3^-) by the enzyme carbonic anhydrase. (Recall that HCO_3^- is the primary buffer of blood pH; see Fig. 2–21.) The HCO_3^- reenters the blood plasma for transport to the lungs **(Fig. 11–33)**. Because HCO_3^- is much more soluble in blood plasma than is CO_2, this roundabout route increases the capacity of the blood to carry carbon dioxide from the tissues to the lungs. In the lungs, HCO_3^- reenters the erythrocyte and is converted to CO_2, which is eventually released into the lung space and exhaled. To be effective, this shuttle requires very rapid movement of HCO_3^- across the erythrocyte membrane.

The **chloride-bicarbonate exchanger**, also called the **anion exchange (AE) protein**, increases the rate of HCO_3^- transport across the erythrocyte membrane more than a millionfold. Like the glucose transporter, it is an integral protein that probably

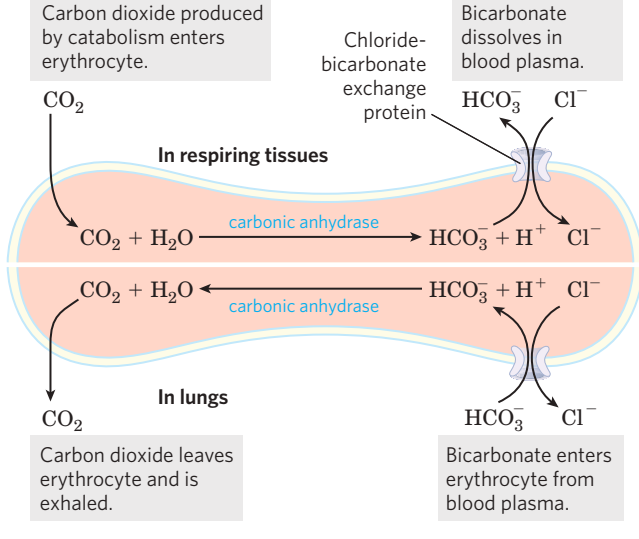


FIGURE 11–33 Chloride-bicarbonate exchanger of the erythrocyte membrane. This cotransport system allows the entry and exit of HCO_3^- without changing the membrane potential. Its role is to increase the CO_2-carrying capacity of the blood. The top half of the figure illustrates the events that take place in respiring tissues; the bottom half, the events in the lungs.

BOX 11–1 MEDICINE Defective Glucose and Water Transport in Two Forms of Diabetes

When ingestion of a carbohydrate-rich meal causes blood glucose to exceed the usual concentration between meals (about 5 mM), excess glucose is taken up by the myocytes of cardiac and skeletal muscle (which store it as glycogen) and by adipocytes (which convert it to triacylglycerols). Glucose uptake into myocytes and adipocytes is mediated by the glucose transporter GLUT4. Between meals, some GLUT4 is present in the plasma membrane, but most is sequestered in the membranes of small intracellular vesicles (Fig. 1). Insulin released from the pancreas in response to high blood glucose triggers the movement of these intracellular vesicles to the plasma membrane, with which they fuse, bringing GLUT4 molecules to the plasma membrane (see Fig. 12–16). With more GLUT4 molecules in action, the rate of glucose uptake increases 15-fold or more. When blood glucose levels return to normal, insulin release slows and most GLUT4 molecules are removed from the plasma membrane and stored in vesicles.

In type 1 (insulin-dependent) diabetes mellitus, the inability to release insulin (and thus to mobilize glucose transporters) results in low rates of glucose uptake into muscle and adipose tissue. One consequence is a prolonged period of high blood glucose after a carbohydrate-rich meal. This condition is the basis for the glucose tolerance test used to diagnose diabetes (Chapter 23).

The water permeability of epithelial cells lining the renal collecting duct in the kidney is due to the presence of an aquaporin (AQP2) in their apical plasma membranes (facing the lumen of the duct). Vasopressin (antidiuretic hormone, ADH) regulates the retention of water by mobilizing AQP2 molecules stored in vesicle membranes within the epithelial cells, much as insulin mobilizes GLUT4 in muscle and adipose tissue. When the vesicles fuse with the epithelial cell plasma membrane, water permeability greatly increases and more water is reabsorbed from the collecting duct and returned to the blood. When the vasopressin level drops, AQP2 is resequestered within vesicles, reducing water retention. In the relatively rare human disease diabetes insipidus, a genetic defect in AQP2 leads to impaired water reabsorption by the kidney. The result is excretion of copious volumes of very dilute urine. If the individual drinks enough water to replace that lost in the urine, there are no serious medical consequences, but insufficient water intake leads to dehydration and imbalances in blood electrolytes, which can lead to fatigue, headache, muscle pain, or even death.

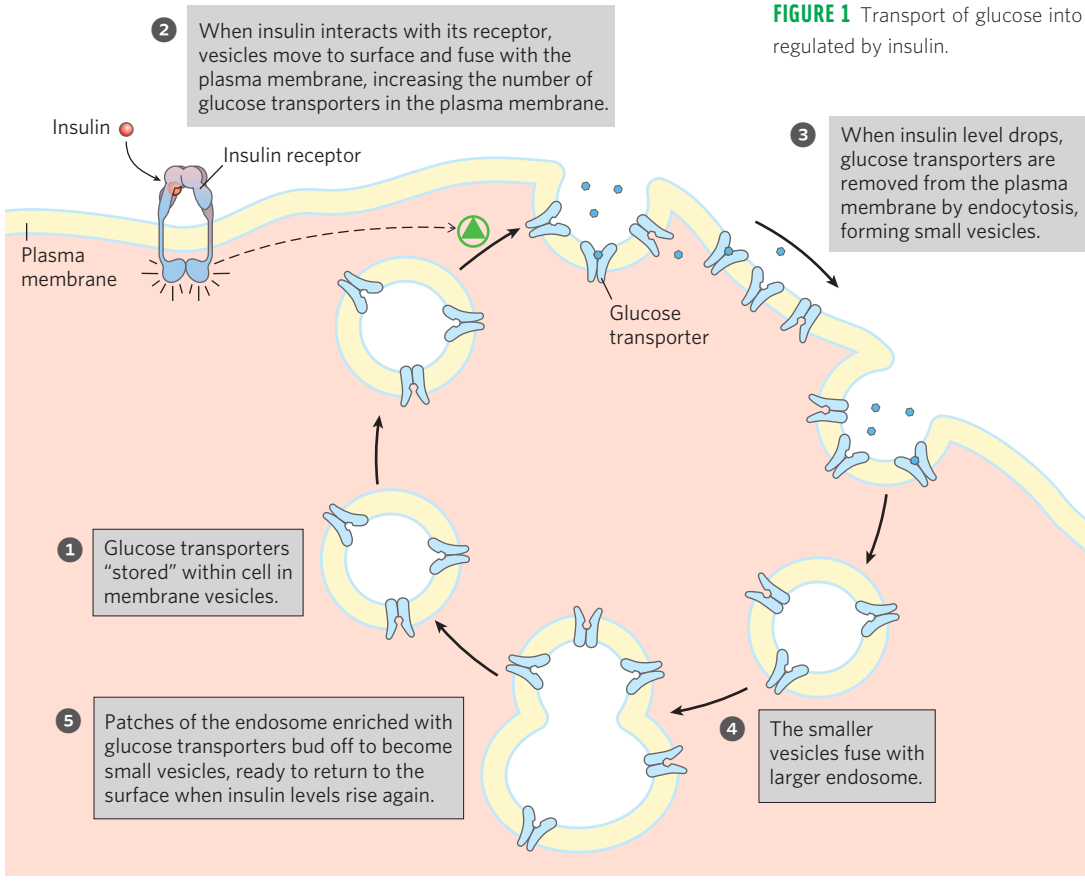


FIGURE 1 Transport of glucose into a myocyte by GLUT4 is regulated by insulin.

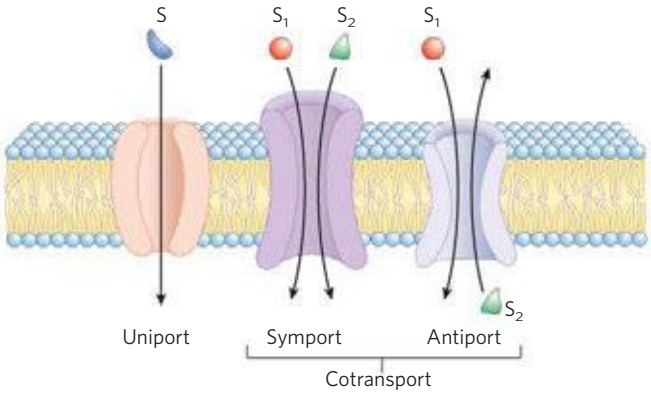


FIGURE 11–34 Three general classes of transport systems. Transporters differ in the number of solutes (substrates) transported and the direction in which each solute moves. Examples of all three types of transporter are discussed in the text. Note that this classification tells us nothing about whether these are energy-requiring (active transport) or energy-independent (passive transport) processes.

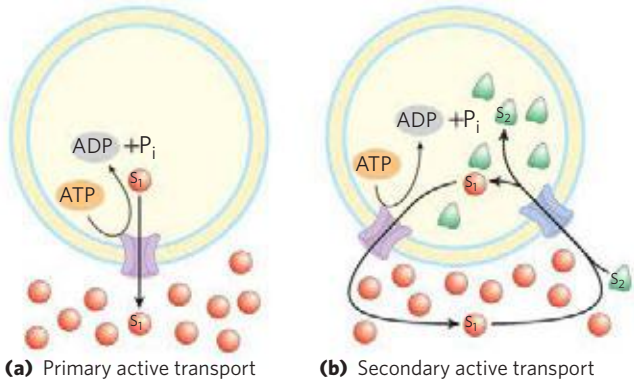


FIGURE 11–35 Two types of active transport. (a) In primary active transport, the energy released by ATP hydrolysis drives solute (S_1) movement against an electrochemical gradient. **(b)** In secondary active transport, a gradient of ion X (S_1) (often Na^+) has been established by primary active transport. Movement of X (S_1) down its electrochemical gradient now provides the energy to drive cotransport of a second solute (S_2) against its electrochemical gradient.

spans the membrane at least 12 times. This protein mediates the simultaneous movement of two anions: for each HCO_3^- ion that moves in one direction, one Cl^- ion moves in the opposite direction, with no net transfer of charge; the exchange is **electroneutral**. The coupling of Cl^- and HCO_3^- movements is obligatory; in the absence of chloride, bicarbonate transport stops. In this respect, the anion exchanger is typical of those systems, called **cotransport systems**, that simultaneously carry two solutes across a membrane **(Fig. 11–34)**. When, as in this case, the two substrates move in opposite directions, the process is **antiport**. In **symport**, two substrates are moved simultaneously in the same direction. Transporters that carry only one substrate, such as the erythrocyte glucose transporter, are known as **uniport** systems.

The human genome has genes for three closely related chloride-bicarbonate exchangers, all with the same predicted transmembrane topology. Erythrocytes contain the AE1 transporter, AE2 is prominent in the liver, and AE3 is present in plasma membranes of the brain, heart, and retina. Similar anion exchangers are also found in plants and microorganisms.

Active Transport Results in Solute Movement against a Concentration or Electrochemical Gradient

In passive transport, the transported species always moves down its electrochemical gradient and is not accumulated above the equilibrium concentration. **Active transport**, by contrast, results in the accumulation of a solute above the equilibrium point. Active transport is thermodynamically unfavorable (endergonic) and takes place only when coupled (directly or indirectly) to an exergonic process such as the absorption of sunlight, an oxidation reaction, the breakdown of ATP, or the concomitant flow of some other chemical species down its electrochemical gradient. In **primary active transport**, solute accumulation is coupled directly to an exergonic chemical reaction, such as conversion of ATP to ADP + P_i **(Fig. 11–35)**. **Secondary active transport** occurs when endergonic (uphill) transport of one solute is coupled to the exergonic (downhill) flow of a different solute that was originally pumped uphill by primary active transport.

The amount of energy needed for the transport of a solute against a gradient can be calculated from the initial concentration gradient. The general equation for the free-energy change in the chemical process that converts S to P is

$$\Delta G = \Delta G'^{\circ} + RT \ln ([P]/[S]) \qquad (11–2)$$

where $\Delta G'^{\circ}$ is the standard free-energy change, R is the gas constant, 8.315 J/mol · K, and T is the absolute temperature. When the "reaction" is simply transport of a solute from a region where its concentration is C_1 to a region where its concentration is C_2, no bonds are made or broken and $\Delta G'^{\circ}$ is zero. The free-energy change for transport, ΔG_t, is then

$$\Delta G_t = RT \ln (C_2/C_1) \qquad (11–3)$$

If there is a 10-fold difference in concentration between two compartments, the cost of moving 1 mol of an uncharged solute at 25 °C uphill across a membrane separating the compartments is

$$\Delta G_t = (8.315 \text{ J/mol·K})(298 \text{ K}) \ln (10/1) = 5{,}700 \text{ J/mol} = 5.7 \text{ kJ/mol}$$

Equation 11–3 holds for all uncharged solutes.

WORKED EXAMPLE 11–1 **Energy Cost of Pumping an Uncharged Solute**

Calculate the energy cost (free-energy change) of pumping an uncharged solute against a 1.0×10^4–fold concentration gradient at 25 °C.

Solution: Begin with Equation 11–3. Substitute 1.0×10^4 for (C_2/C_1), 8.315 J/mol · K for R, and 298 K for T:

$$\begin{aligned}\Delta G_t &= RT \ln(C_2/C_1) \\ &= (8.315 \text{ J/mol·K})(298 \text{ K}) \ln (1.0 \times 10^4) \\ &= 23 \text{ kJ/mol}\end{aligned}$$

When the solute is an *ion*, its movement without an accompanying counterion results in the endergonic separation of positive and negative charges, producing an electrical potential; such a transport process is said to be **electrogenic**. The energetic cost of moving an ion depends on the electrochemical potential (Fig 11–27), the sum of the chemical and electrical gradients:

$$\Delta G_t = RT \ln (C_2/C_1) + Z\mathcal{F}\Delta\psi \quad (11\text{–}4)$$

where Z is the charge on the ion, $\mathcal{F}$ is the Faraday constant (96,480 J/V · mol), and $\Delta\psi$ is the transmembrane electrical potential (in volts). Eukaryotic cells typically have plasma membrane potentials of about 0.05 V (with the inside negative relative to the outside), so the second term of Equation 11–4 can make a significant contribution to the total free-energy change for transporting an ion. Most cells maintain more than a 10-fold difference in ion concentrations across their plasma or intracellular membranes, and for many cells and tissues active transport is therefore a major energy-consuming process.

WORKED EXAMPLE 11–2 **Energy Cost of Pumping a Charged Solute**

Calculate the energy cost (free-energy change) of pumping Ca^{2+} from the cytosol, where its concentration is about 1.0×10^{-7} M, to the extracellular fluid, where its concentration is about 1.0 mM. Assume a temperature of 37 °C (body temperature in a mammal) and a standard transmembrane potential of 50 mV (inside negative) for the plasma membrane.

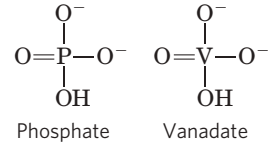
Solution: In this calculation, both the concentration gradient and the electrical potential must be taken into account. In Equation 11–4, substitute 8.315 J/mol · K for R, 310 K for T, 1.0×10^{-3} for C_2, 1.0×10^{-7} for C_1, 96,500 J/V · mol for $\mathcal{F}$, +2 (the charge on a Ca^{2+} ion) for Z, and 0.050 V for $\Delta\psi$. Note that the transmembrane potential is 50 mV (inside negative), so the change in potential when an ion moves from inside to outside is 50 mV.

$$\begin{aligned}\Delta G_t &= RT \ln(C_2/C_1) + Z\mathcal{F}\Delta\psi \\ &= (8.315 \text{ J/mol·K})(310 \text{ K}) \ln \frac{1.0 \times 10^{-3}}{1.0 \times 10^{-7}} + \\ &\quad 2(96{,}500 \text{ J/V·mol})(0.050 \text{ V}) \\ &= 33 \text{ kJ/mol}\end{aligned}$$

The mechanism of active transport is of fundamental importance in biology. As we shall see in Chapter 19, ATP is formed in mitochondria and chloroplasts by a mechanism that is essentially ATP-driven ion transport operating in reverse. The energy made available by the spontaneous flow of protons across a membrane is calculable from Equation 11–4; remember that ΔG for flow *down* an electrochemical gradient has a negative value and ΔG for transport of ions *against* an electrochemical gradient has a positive value.

P-Type ATPases Undergo Phosphorylation during Their Catalytic Cycles

The family of active transporters called **P-type ATPases** are cation transporters that are reversibly phosphorylated by ATP (thus the name P-type) as part of the transport cycle. Phosphorylation forces a conformational change that is central to movement of the cation across the membrane. The human genome encodes at least 70 P-type ATPases that share similarities in amino acid sequence and topology, especially near the Asp residue that undergoes phosphorylation. All are integral proteins with 8 or 10 predicted membrane-spanning regions in a single polypeptide (type III in Fig. 11–9), and all are sensitive to inhibition by the phosphate analog **vanadate**.

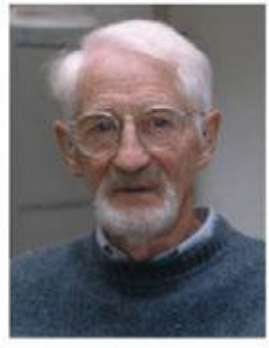


Phosphate Vanadate

The P-type ATPases are widespread in eukaryotes and bacteria. The Na^+K^+ ATPase of animal cells (an antiporter for Na^+ and K^+ ions) and the plasma membrane H^+ ATPase of plants and fungi set the transmembrane electrochemical potential in cells by establishing ion gradients across the plasma membrane. These gradients provide the driving force for secondary active transport and are also the basis for electrical signaling in neurons. In animal tissues, the ***sarcoplasmic/endoplasmic reticulum Ca^{2+} ATPase*** (**SERCA**) **pump** and the plasma membrane Ca^{2+} ATPase pump are uniporters for Ca^{2+} ions, which together maintain the cytosolic level of Ca^{2+} below 1 μM. Parietal cells in the lining of the mammalian stomach have a P-type ATPase that pumps H^+ and K^+ across the plasma membrane, thereby acidifying the stomach contents. Lipid flippases, as we noted earlier, are structurally and functionally

related to P-type transporters. Bacteria and eukaryotes use P-type ATPases to pump out toxic heavy metal ions such as Cd^{2+} and Cu^{2+}.

The P-type pumps have similar structures **(Fig. 11–36)** and similar mechanisms. The mechanism postulated for P-type ATPases takes into account the large conformational changes and the phosphorylation-dephosphorylation of the critical Asp residue in the P domain that is known to occur during a catalytic cycle. For the SERCA pump **(Fig. 11–37)**, each catalytic cycle moves two Ca^{2+} ions across the membrane and converts an ATP to ADP and P_i. ATP has two roles in this mechanism, one catalytic and one modulatory. The role of ATP binding and phosphoryl transfer to the enzyme is to bring about the interconversion of two conformations (E1 and E2) of the transporter. In the E1 conformation, the two Ca^{2+}-binding sites are exposed on the cytosolic side of the ER or sarcoplasmic reticulum and bind Ca^{2+} with high affinity. ATP binding and Asp phosphorylation drive a conformational change from E1 to E2 in which the Ca^{2+}-binding sites are now exposed on the lumenal side of the membrane and their affinity for Ca^{2+} is greatly reduced, causing Ca^{2+} release into the lumen. By this mechanism, the energy released by hydrolysis of ATP during one phosphorylation-dephosphorylation cycle drives Ca^{2+} across the membrane against a large electrochemical gradient.



Jens Skou

A variation on this basic mechanism is seen in the **Na^+K^+ ATPase** of the plasma membrane, discovered by Jens Skou in 1957. This cotransporter couples phosphorylation-dephosphorylation of the critical Asp residue to the simultaneous movement of both Na^+ and K^+ against their electrochemical gradients. The Na^+K^+ ATPase is responsible for maintaining low Na^+ and high K^+ concentrations in the cell relative to the extracellular fluid **(Fig. 11–38)**. For each molecule of ATP converted to ADP and P_i, the transporter moves two K^+ ions inward

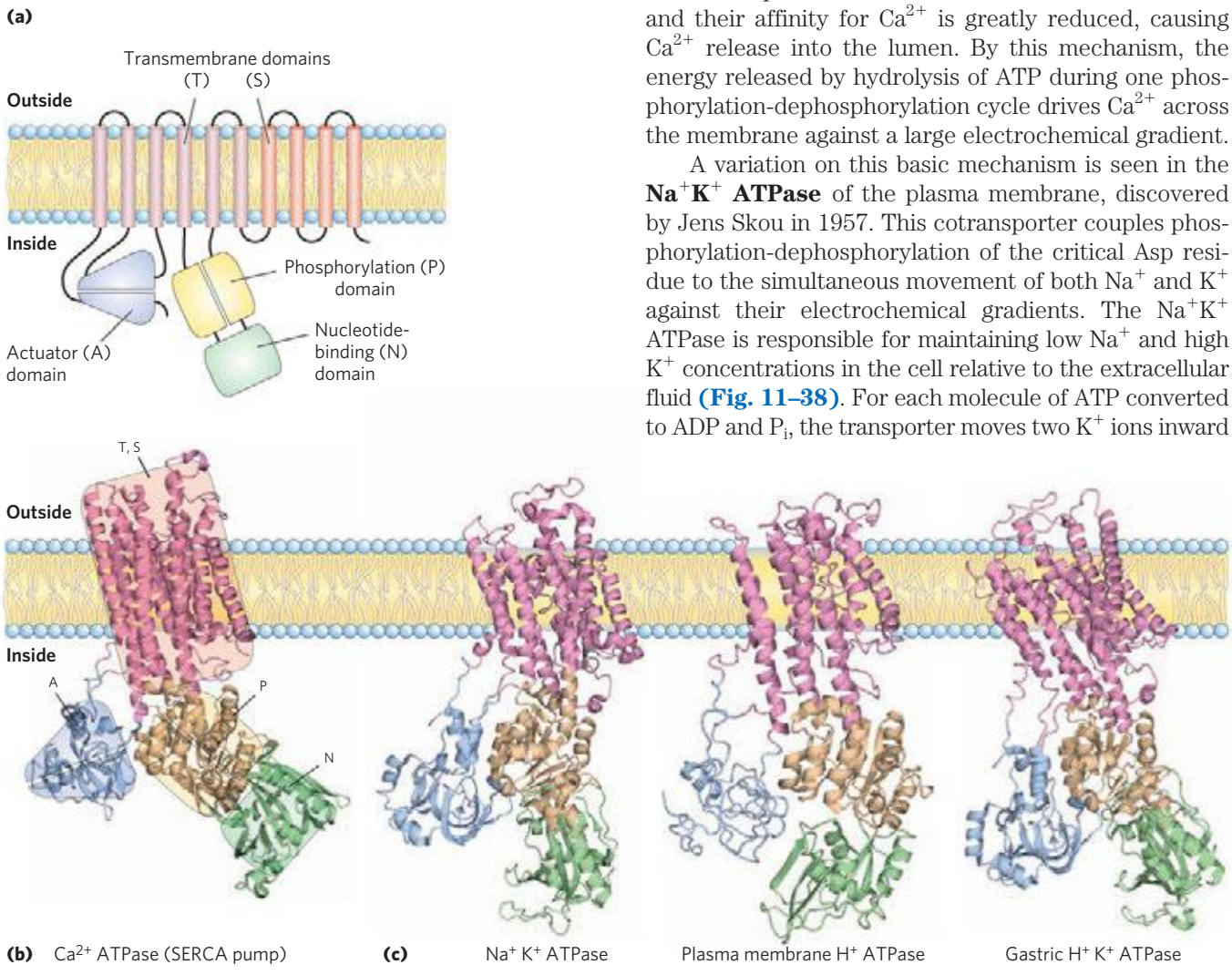


FIGURE 11–36 The general structure of the P-type ATPases. (a) P-type ATPases have three cytoplasmic domains (A, N, and P) and two transmembrane domains (T and S) consisting of multiple helices. The N (nucleotide) domain binds ATP and Mg^{2+}, and it has protein kinase activity that phosphorylates the specific Asp residue found in the P (phosphorylated) domain of all P-type ATPases. The A (actuator) domain has protein phosphatase activity and removes the phosphoryl group from the Asp residue with each catalytic cycle of the pump. A transport domain with six transmembrane helices (T) includes the ion-transporting structure, and four more transmembrane helices make up the support (S) domain, which provides physical support to the transport domain and may have other specialized function in certain P-type ATPases. The binding sites for the ions to be transported are near the middle of the membrane, 40 to 50 Å from the phosphorylated Asp residue—thus Asp phosphorylation-dephosphorylation does not *directly* affect ion binding. The A domain communicates movements of the N and P domains to the ion-binding sites. **(b)** A ribbon representation of the Ca^{2+} ATPase (SERCA pump) (PDB 1T5S). ATP binds to the N domain and the Ca^{2+} ions to be transported bind to the T domain. **(c)** Other P-type ATPases have domain structures, and presumably mechanisms, like the SERCA pump: Na^+K^+ ATPase (PDB ID 3KDP), the plasma membrane H^+ ATPase (PDB ID 3B8C), and the gastric H^+K^+ ATPase (derived from PDB ID 3IXZ).

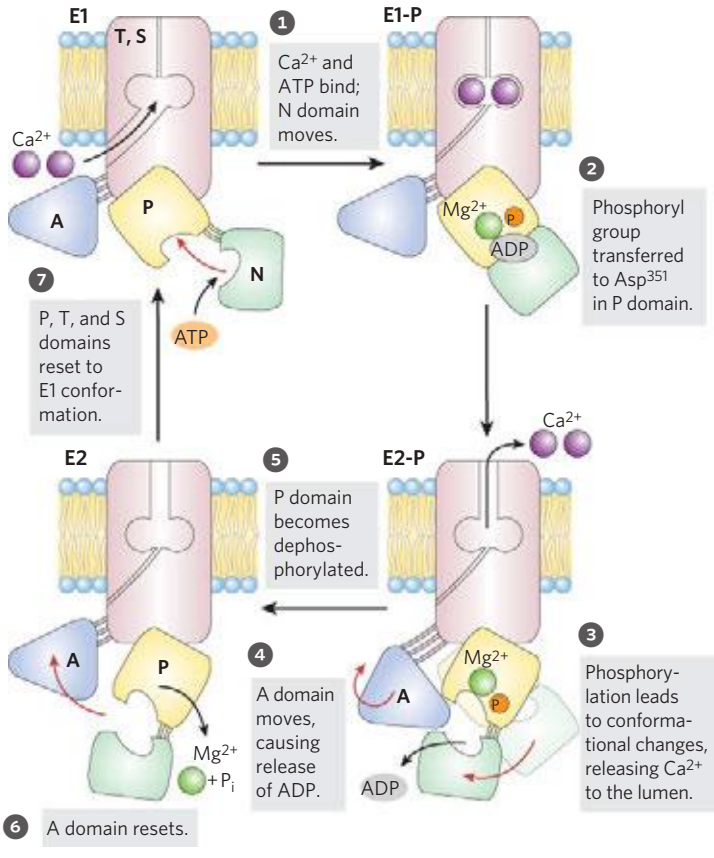


FIGURE 11–37 Postulated mechanism of the SERCA pump. The transport cycle begins with the protein in the E1 conformation, with the Ca^{2+}-binding sites facing the cytosol. Two Ca^{2+} ions bind, then ATP binds to the transporter and phosphorylates Asp^{351}, forming E1-P. Phosphorylation favors the second conformation, E2-P, in which the Ca^{2+}-binding sites, now with a reduced affinity for Ca^{2+}, are accessible on the other side of the membrane (the lumen or extracellular space), and the released Ca^{2+} diffuses away. Finally, E2-P is dephosphorylated, returning the protein to the E1 conformation for another round of transport.

and three Na^+ ions outward across the plasma membrane. Cotransport is therefore electrogenic—it creates a net separation of charge across the membrane; in animals, this produces the membrane potential of −50 to −70 mV (inside negative relative to outside) that is characteristic of most cells and is essential to the conduction of action potentials in neurons. The central role of the Na^+K^+ ATPase is reflected in the energy invested in this single reaction: about 25% of the total energy consumption of a human at rest!

V-Type and F-Type ATPases Are ATP-Driven Proton Pumps

V-type ATPases, a class of proton-transporting ATPases, are responsible for acidifying intracellular compartments in many organisms (thus *V* for *v*acuolar). Proton pumps of this type maintain the vacuoles of fungi and higher plants at a pH between 3 and 6, well below that of the surrounding cytosol (pH 7.5). V-type ATPases are also responsible for the acidification of lysosomes, endosomes, the Golgi complex, and secretory vesicles in animal cells. All V-type ATPases have a similar complex structure, with an integral (transmembrane) domain (V_o) that serves as a proton channel and a peripheral domain (V_1) that contains the ATP-binding site and the ATPase activity **(Fig 11–39a)**. The structure is similar to that of the well-characterized F-type ATPases.

F-type ATPase active transporters catalyze the uphill transmembrane passage of protons driven by ATP hydrolysis. The "F-type" designation derives from the identification of these ATPases as energy-coupling *f*actors. The F_o integral membrane protein complex (Fig. 11–39b; subscript *o* denotes its inhibition by the drug *o*ligomycin) provides a transmembrane pathway for protons, and the peripheral protein F_1 (subscript *1* indicating this was the first of several factors isolated from mitochondria) uses the energy of ATP to drive protons uphill (into a region of higher H^+ concentration). The

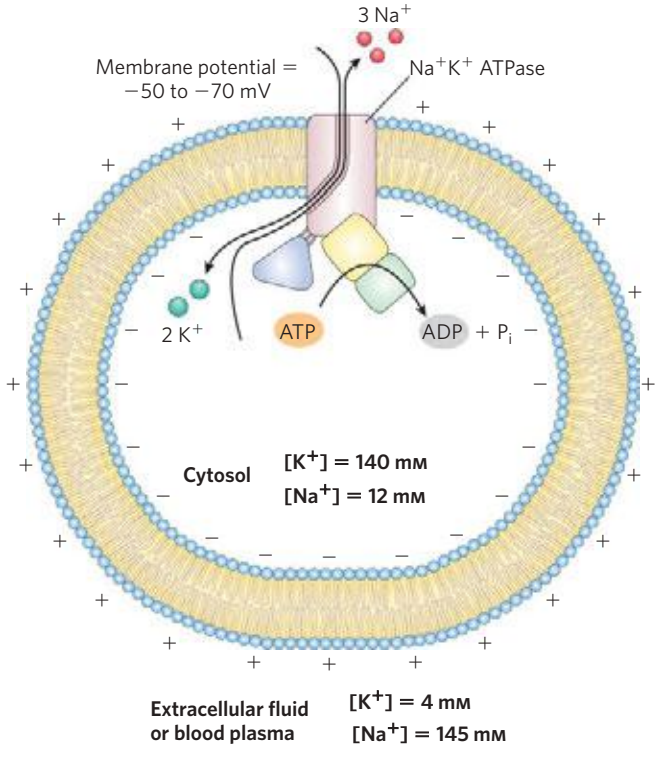


FIGURE 11–38 Role of the Na^+K^+ ATPase in animal cells. This active transport system is primarily responsible for setting and maintaining the intracellular concentrations of Na^+ and K^+ in animal cells and for generating the membrane potential. It does this by moving three Na^+ out of the cell for every two K^+ it moves in. The electrical potential across the plasma membrane is central to electrical signaling in neurons, and the gradient of Na^+ is used to drive the uphill cotransport of solutes in many cell types.

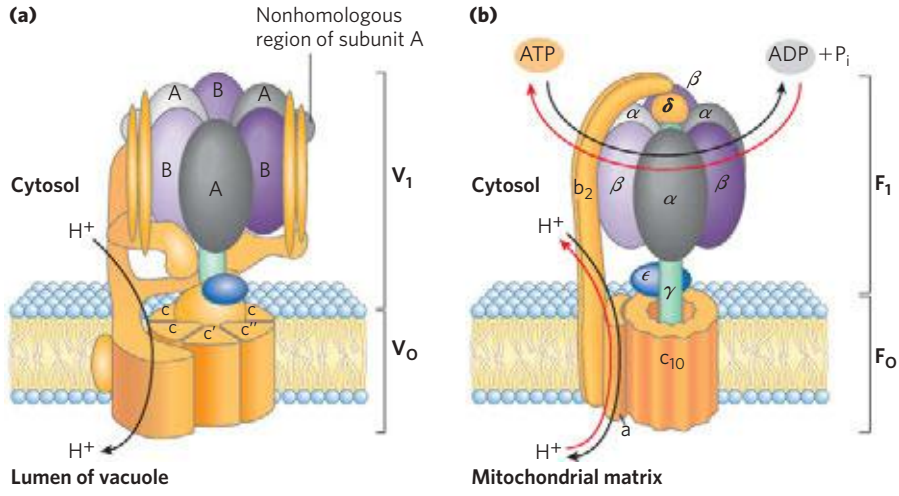


FIGURE 11–39 Two proton pumps with similar structures. (a) The V_oV_1 H^+ ATPase uses ATP to pump protons into vacuoles and lysosomes, creating their low internal pH. It has an integral (membrane-embedded) domain, V_o (orange), that includes multiple identical c subunits, and a peripheral domain that projects into the cytosol and contains the ATP hydrolysis sites, located on three identical B subunits (purple). **(b)** The F_oF_1 ATPase/ATP synthase of mitochondria has an integral domain, F_o (orange), with multiple copies of the c subunit, and a peripheral domain, F_1, consisting of three α subunits, three β subunits, and a central shaft joined to the integral domain. F_o, and presumably V_o, provides a transmembrane channel through which protons are pumped as ATP is hydrolyzed on the β subunits of F_1 (B subunits of V_1). The remarkable mechanism by which ATP hydrolysis is coupled to proton movement is described in detail in Chapter 19. It involves rotation of F_o in the plane of the membrane. The structures of the V_oV_1 ATPase and its analogs A_oA_1 ATPase (of archaea) and CF_oCF_1 ATPase (of chloroplasts) are essentially similar to that of F_oF_1, and the mechanisms are also conserved. An ATP-driven proton transporter also can catalyze ATP synthesis (red arrows) as protons flow *down* their electrochemical gradient. This is the central reaction in the processes of oxidative phosphorylation and photophosphorylation, both described in detail in Chapter 19.

F_oF_1 organization of proton-pumping transporters must have developed very early in evolution. Bacteria such as *E. coli* use an F_oF_1 ATPase complex in their plasma membrane to pump protons outward, and archaea have a closely homologous proton pump, the A_oA_1 ATPase.

Like all enzymes, F-type ATPases catalyze their reactions in both directions. Therefore, a sufficiently large proton gradient can supply the energy to drive the reverse reaction, ATP synthesis (Fig. 11–39b). When functioning in this direction, the F-type ATPases are more appropriately named **ATP synthases**. ATP synthases are central to ATP production in mitochondria during oxidative phosphorylation and in chloroplasts during photophosphorylation, as well as in bacteria and archaea. The proton gradient needed to drive ATP synthesis is produced by other types of proton pumps powered by substrate oxidation or sunlight. We provide a detailed description of these processes in Chapter 19 (p. 750).

ABC Transporters Use ATP to Drive the Active Transport of a Wide Variety of Substrates

ABC transporters (Fig. 11–40) constitute a large family of ATP-dependent transporters that pump amino acids, peptides, proteins, metal ions, various lipids, bile salts, and many hydrophobic compounds, including drugs, out of cells against a concentration gradient. One ABC transporter in humans, the **multidrug transporter** (**MDR1**; also called **P glycoprotein**), is responsible for the striking resistance of certain tumors to some generally effective antitumor drugs. MDR1 has broad substrate specificity for hydrophobic compounds, including, for example, the chemotherapeutic drugs adriamycin, doxorubicin, and vinblastine. By pumping these drugs out of the cell, the transporter prevents their accumulation within a tumor and thus blocks their therapeutic effects. MDR1 (Fig. 11–40a) is an integral membrane protein (M_r 170,000) with two homologous halves, each with six transmembrane helices and a cytoplasmic ATP-binding domain ("cassette"), which give the family its name: *A*TP-*b*inding *c*assette transporters. Overexpression of MDR1 is associated with treatment failure in cancers of the liver, kidney, and colon. A related ABC transporter, ABCC1, is overexpressed in drug-resistant prostate, lung, and breast cancer cells. Highly selective inhibitors of multidrug transporters, which would be expected to enhance the effectiveness of antitumor drugs that are otherwise pumped out of tumor cells, are the objects of current drug discovery and design. Why have multidrug transporters been conserved in evolution? MDR1 in the placental membrane and in the blood-brain barrier keep out toxic compounds that would damage the fetus or the brain.

ABC transporters are also present in simpler animals and in plants and microorganisms. Yeast has 31 genes that encode ABC transporters, *Drosophila* has 56, and *E. coli* has 80, representing 2% of its entire genome. ABC transporters, used by *E. coli* and other bacteria to import

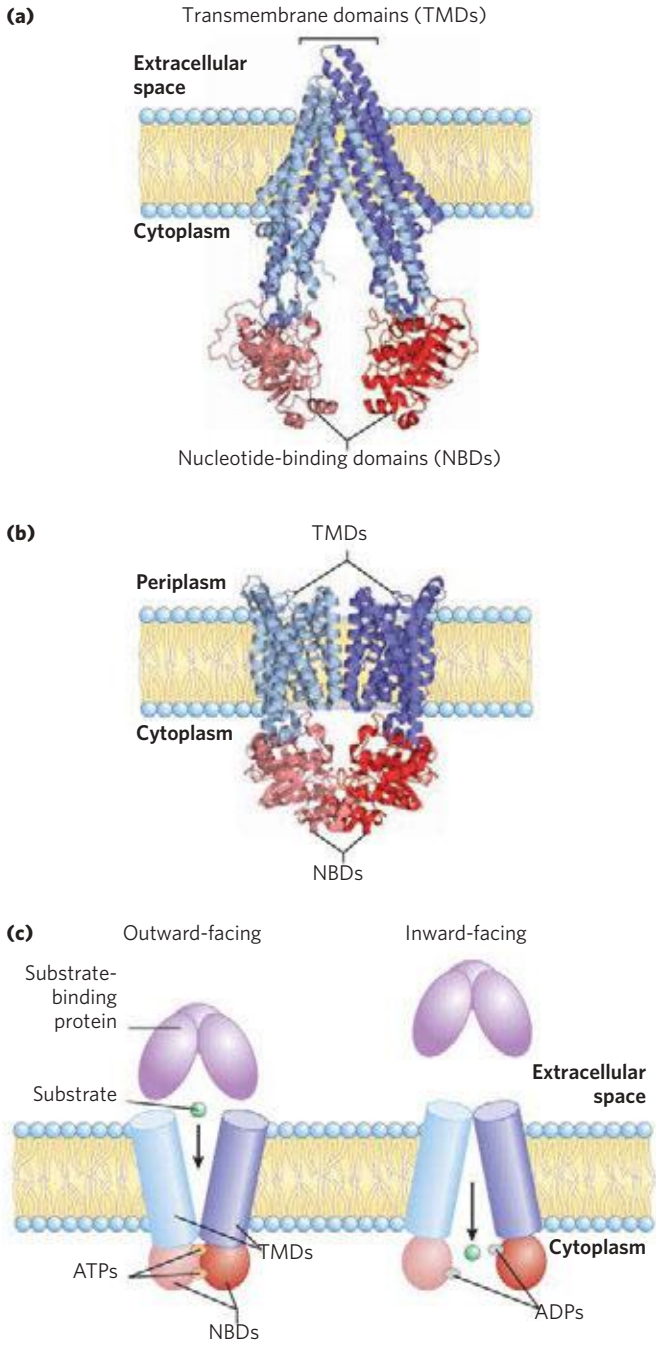


FIGURE 11–40 **Two ABC transporters. (a)** The multidrug transporter of animal cells (MDR1, also called P glycoprotein; PDB ID 3G60), responsible for pumping a variety of antitumor drugs out of human cells, has two homologous halves (blue and light blue), each with six transmembrane helices in two transmembrane domains (TMDs; blue), and a cytoplasmic nucleotide-binding domain (NBD; red). **(b)** The vitamin B_{12} importer BtuCD (PDB ID 1L7V) of *E. coli* is a homodimer with 10 transmembrane helices (blue, light blue) in each monomer and two NBDs (red) that extend into the cytoplasm. **(c)** Mechanisms proposed for the *E. coli* vitamin B_{12} ABC transporter coupling of ATP hydrolysis to transport. Substrate is brought to the transporter on the periplasmic side by a substrate-specific binding protein. With ATP bound to the NBD sites, the transporter is open to the outside (periplasm), but on substrate binding and ATP hydrolysis to ADP, a conformational change exposes the substrate to the inside surface, and it diffuses away from the transporter and into the cytosol.

essentials such as vitamin B_{12} (Fig. 11–40b), are the presumed evolutionary precursors of the MDRs of animal cells. The presence of ABC transporters that confer antibiotic resistance in pathogenic microbes (*Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Candida albicans*, *Neisseria gonorrhoeae*, and *Plasmodium falciparum*) is a serious public health concern and makes these transporters attractive targets for drug design. ■

All ABC transporters have two nucleotide-binding domains (NBDs) and two transmembrane domains containing multiple transmembrane helices. In some cases, all these domains are in a single long polypeptide; other ABC transporters have two subunits, each contributing an NBD and a domain with six transmembrane helices. Many ABC transporters are in the plasma membrane, but some types are also found in the endoplasmic reticulum and in the membranes of mitochondria and lysosomes. The CFTR protein (see Box 11–2) is an interesting case of an ion channel (for Cl^-), operated by ATP hydrolysis, that is apparently derived from an ABC transporter in which evolution has eliminated the pumping function but left a functional channel.

The NBDs of all ABC proteins are similar in sequence and presumably in three-dimensional structure; they are the conserved molecular motor that can be coupled to a wide variety of pumps and channels. When coupled with a pump, the ATP-driven motor moves solutes against a concentration gradient. The stoichiometry of ABC pumps is approximately one ATP hydrolyzed per molecule of substrate transported, but neither the mechanism of coupling nor the site of substrate binding is fully understood (Fig. 11–40c).

Some ABC transporters have very high specificity for a single substrate; others are more promiscuous. The human genome contains at least 48 genes that encode ABC transporters, many of which are involved in maintaining the lipid bilayer and in transporting sterols, sterol derivatives, and fatty acids throughout the body. The flippases that move membrane lipids from one leaflet of the bilayer to the other are ABC transporters, and the cellular machinery for exporting excess cholesterol includes an ABC transporter. Mutations in the genes that encode some of these proteins contribute to several genetic diseases, including cystic fibrosis (Box 11–2), Tangier disease (p. 874), retinal degeneration, anemia, and liver failure. ■

Ion Gradients Provide the Energy for Secondary Active Transport

The ion gradients formed by primary transport of Na^+ or H^+ can in turn provide the driving force for cotransport of other solutes. Many cell types contain transport systems that couple the spontaneous, downhill flow of these ions to the simultaneous uphill pumping of another ion, sugar, or amino acid (Table 11–4).

BOX 11–2 MEDICINE A Defective Ion Channel in Cystic Fibrosis

Cystic fibrosis (CF) is a serious and relatively common hereditary human disease. About 5% of white Americans are carriers, having one defective and one normal copy of the gene. Only individuals with two defective copies show the severe symptoms of the disease: obstruction of the gastrointestinal and respiratory tracts, commonly leading to bacterial infection of the airways and death due to respiratory insufficiency before the age of 30. In CF, the thin layer of mucus that normally coats the internal surfaces of the lungs is abnormally thick, obstructing air flow and providing a haven for pathogenic bacteria, particularly *Staphylococcus aureus* and *Pseudomonas aeruginosa*.

The defective gene in CF patients was discovered in 1989. It encodes a membrane protein called *c*ystic *f*ibrosis *t*ransmembrane conductance *r*egulator, or CFTR. This protein has two segments, each containing six transmembrane helices, two nucleotide-binding domains (NBDs), and a regulatory region (Fig. 1). CFTR is therefore very similar to other ABC transporter proteins except that it functions as an *ion channel* (for Cl^- ion), not as a pump. The channel conducts Cl^- across the plasma membrane when both NBDs have bound ATP, and it closes when the ATP on one of the NBDs is broken down to ADP and P_i. The Cl^- channel is further regulated by phosphorylation of several Ser residues in the regulatory domain, catalyzed by cAMP-dependent protein kinase (Chapter 12). When the regulatory domain is not phosphorylated, the Cl^- channel is closed. The mutation responsible for CF in 70% of cases results in deletion of a Phe residue at position 508. The mutant protein folds incorrectly, which interferes with its insertion in the plasma membrane, resulting in reduced Cl^- and H_2O movement across the plasma membranes of epithelial cells that line the airways (Fig. 2), the digestive tract, and exocrine glands (pancreas, sweat glands, bile ducts, and vas deferens). Liquid secretion is essential to keep the mucus on the surface of alveoli of the lung at just the right viscosity to trap and clear microorganisms that are inhaled.

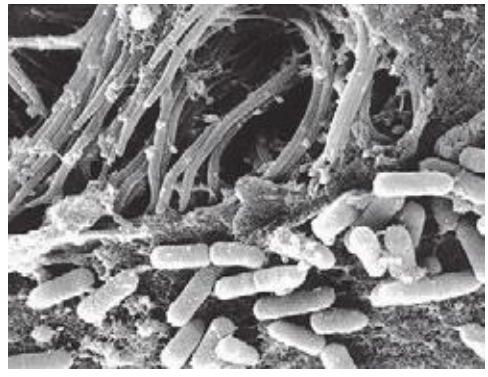


FIGURE 2 Mucus lining the surface of the lungs traps bacteria. In healthy lungs (shown here), these bacteria are killed and swept away by the action of cilia. In CF, this mechanism is impaired, resulting in recurring infections and progressive damage to the lungs.

Diminished export of Cl^- is accompanied by diminished export of water from cells, causing the mucus on the surfaces of the cells to become dehydrated, thick, and excessively sticky. In normal circumstances, cilia on the epithelial cells that line the inner surface of the lungs constantly sweep away bacteria that settle in this mucus, but the thick mucus in individuals with CF hinders this process. Frequent infections by bacteria such as *S. aureus* and *P. aeruginosa* result, causing progressive damage to the lungs and reduced respiratory efficiency. Respiratory failure is commonly the cause of death in people with CF.

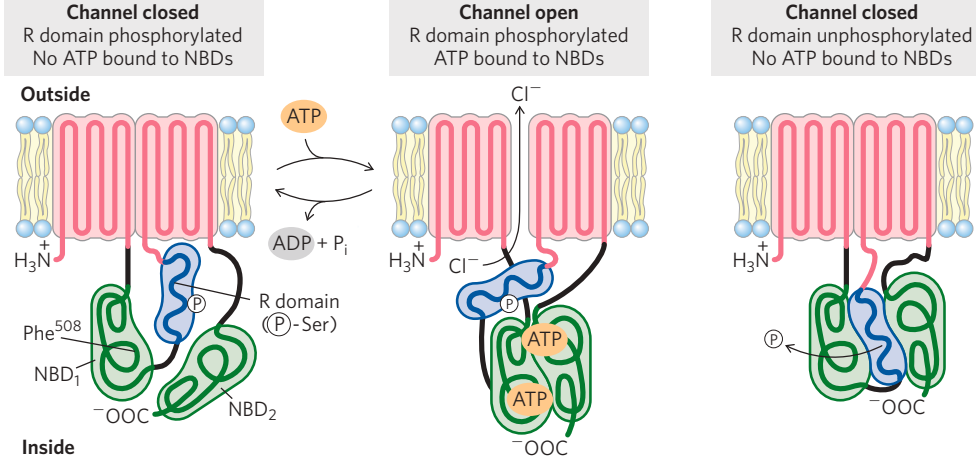


FIGURE 1 Three states of the cystic fibrosis transmembrane conductance regulator, CFTR. The protein has two segments, each with six transmembrane helices, and three functionally significant domains extend from the cytoplasmic surface: NBD_1 and NBD_2 (green) are nucleotide-binding domains that bind ATP, and a regulatory R domain (blue) is the site of phosphorylation by cAMP-dependent protein kinase. When this R domain is phosphorylated but no ATP is bound to the NBDs (left), the channel is closed. The binding of ATP opens the channel (middle) until the bound ATP is hydrolyzed. When the regulatory domain is unphosphorylated (right), it binds the NBD domains and prevents ATP binding and channel opening. The most commonly occurring mutation leading to CF is the deletion of Phe^{508} in the NBD_1 domain (left). CFTR is a typical ABC transporter in all but two respects: most ABC transporters lack the regulatory domain, and CFTR acts as an ion channel (for Cl^-), not as a typical transporter.

TABLE 11–4 Cotransport Systems Driven by Gradients of Na^+ or H^+

Organism/ tissue/cell type	Transported solute (moving against its gradient)	Cotransported solute (moving down its gradient)	Type of transport
E. coli	Lactose	H^+	Symport
	Proline	H^+	Symport
	Dicarboxylic acids	H^+	Symport
Intestine, kidney (vertebrates)	Glucose	Na^+	Symport
	Amino acids	Na^+	Symport
Vertebrate cells (many types)	Ca^{2+}	Na^+	Antiport
Higher plants	K^+	H^+	Antiport
Fungi (*Neurospora*)	K^+	H^+	Antiport

The **lactose transporter** (**lactose permease**, or **galactoside permease**) of *E. coli* is the well-studied prototype for proton-driven cotransporters. This protein consists of a single polypeptide chain (417 residues) that functions as a monomer to transport one proton and one lactose molecule into the cell, with the net accumulation of lactose **(Fig. 11–41)**. *E. coli* normally produces a gradient of protons and charge across its plasma membrane by oxidizing fuels and using the energy of oxidation to pump protons outward. (This mechanism is discussed in detail in Chapter 19.) The lipid bilayer is impermeable to protons, but the lactose transporter provides a route for proton reentry, and lactose is simultaneously carried into the cell by symport. The endergonic accumulation of lactose is thereby coupled to the exergonic flow of protons into the cell, with a negative overall free-energy change.

The lactose transporter is one member of the **major facilitator superfamily (MFS)** of transporters, which comprises 28 families. Almost all proteins in this superfamily have 12 transmembrane domains (the few exceptions have 14). The proteins share relatively little sequence homology, but the similarity of their secondary structures and topology suggests a common tertiary structure. The crystallographic solution of the *E. coli* lactose transporter provides a glimpse of this general structure **(Fig. 11–42a)**. The protein has 12 transmembrane helices, and connecting loops protrude into the cytoplasm or the periplasmic space (between the plasma membrane and outer membrane or cell wall). The six amino-terminal and six carboxyl-terminal helices form very similar domains to produce a structure with a rough twofold symmetry. In the crystallized form of the protein, a large aqueous cavity is exposed on the cytoplasmic side of the membrane. The substrate-binding site is in this cavity, more or less in the middle of the membrane. The side of the transporter facing outward (the periplasmic face) is closed tightly, with no channel big enough for lactose to enter. The proposed mechanism for transmembrane passage of the substrate (Fig. 11–42b) involves a rocking motion between the two domains, driven by substrate binding and proton movement, alternately exposing the substrate-binding domain to the cytoplasm and to the periplasm. This "rocking banana" model is similar to that shown in Figure 11–32 for GLUT1.

How is proton movement into the cell coupled with lactose uptake? Extensive genetic studies of the lactose transporter have established that of the 417 residues in the protein, only 6 are absolutely essential for cotransport of H^+ and lactose—some for lactose binding, others for

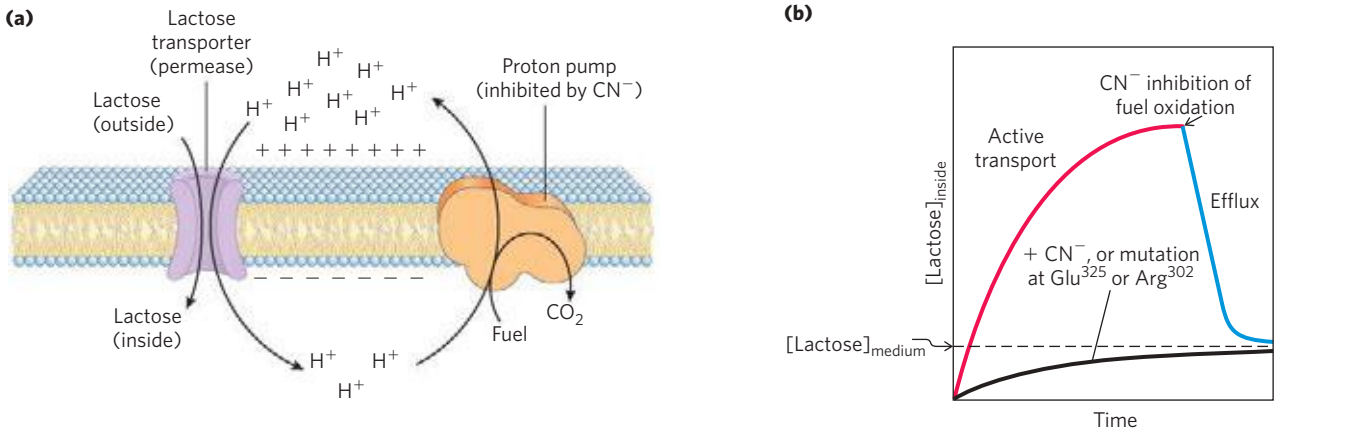


FIGURE 11–41 **Lactose uptake in *E. coli*.** **(a)** The primary transport of H^+ out of the cell, driven by the oxidation of a variety of fuels, establishes both a proton gradient and an electrical potential (inside negative) across the membrane. Secondary active transport of lactose into the cell involves symport of H^+ and lactose by the lactose transporter. The uptake of lactose against its concentration gradient is entirely dependent on this inflow of protons driven by the electrochemical gradient. **(b)** When the energy-yielding oxidation reactions of metabolism are blocked by cyanide (CN^-), the lactose transporter allows equilibration of lactose across the membrane via passive transport. Mutations that affect Glu^{325} or Arg^{302} have the same effect as cyanide. The dashed line represents the concentration of lactose in the surrounding medium.

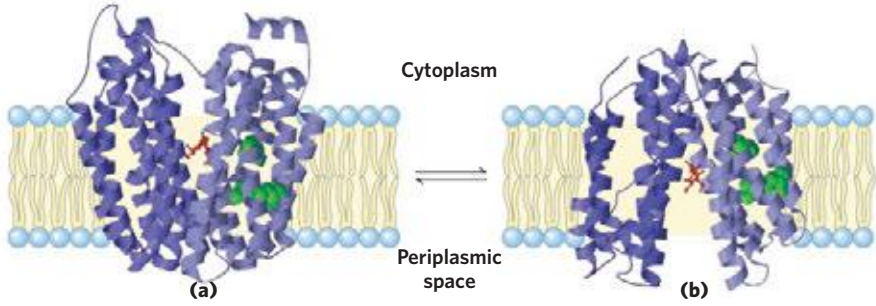


FIGURE 11–42 **The lactose transporter (lactose permease) of *E. coli*.** **(a)** Ribbon representation viewed parallel to the plane of the membrane shows the 12 transmembrane helices arranged in two nearly symmetric domains, shown in different shades of purple. In the form of the protein for which the crystal structure was determined, the substrate sugar (red) is bound near the middle of the membrane where the sugar is exposed to the cytoplasm (derived from PDB ID 1PV7). **(b)** The postulated second conformation of the transporter (PDB ID 2CFQ), related to the first by a large, reversible conformational change in which the substrate-binding site is exposed first to the periplasm, where lactose is picked up, then to the cytoplasm, where the lactose is released. The interconversion of the two forms is driven by changes in the pairing of charged (protonatable) side chains such as those of Glu^{325} and Arg^{302} (green), which is affected by the transmembrane proton gradient.

proton transport. Mutation in either of two residues (Glu^{325} and Arg^{302}; Fig. 11–42) results in a protein still able to catalyze facilitated diffusion of lactose but incapable of coupling H^+ flow to uphill lactose transport. A similar effect is seen in wild-type (unmutated) cells when their ability to generate a proton gradient is blocked with CN^-: the transporter carries out facilitated diffusion normally, but it cannot pump lactose against a concentration gradient (Fig. 11–41b). The balance between the two conformations of the lactose transporter is affected by changes in charge pairing between the side chains of Glu^{325} and Arg^{302}.

In intestinal epithelial cells, glucose and certain amino acids are accumulated by symport with Na^+, down the Na^+ gradient established by the Na^+K^+ ATPase of the plasma membrane **(Fig. 11–43)**. The apical surface of the intestinal epithelial cell is covered with microvilli, long, thin projections of the plasma membrane that greatly increase the surface area exposed to the intestinal contents. **Na^+-glucose symporters** in the apical plasma membrane take up glucose from the intestine in a process driven by the downhill flow of Na^+:

$$2Na^+_{out} + glucose_{out} \longrightarrow 2Na^+_{in} + glucose_{in}$$

The energy required for this process comes from two sources: the greater concentration of Na^+ outside than inside (the chemical potential) and the membrane (electrical) potential, which is inside negative and therefore draws Na^+ inward.

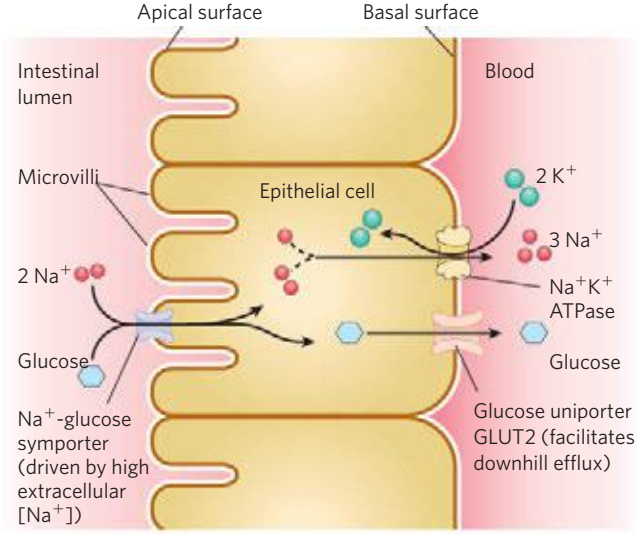


FIGURE 11–43 **Glucose transport in intestinal epithelial cells.** Glucose is cotransported with Na^+ across the apical plasma membrane into the epithelial cell. It moves through the cell to the basal surface, where it passes into the blood via GLUT2, a passive glucose uniporter. The Na^+K^+ ATPase continues to pump Na^+ outward to maintain the Na^+ gradient that drives glucose uptake.

WORKED EXAMPLE 11–3 Energetics of Pumping by Symport

Calculate the maximum $\frac{[glucose]_{in}}{[glucose]_{out}}$ ratio that can be achieved by the plasma membrane Na^+-glucose symporter of an epithelial cell when $[Na^+]_{in}$ is 12 mM, $[Na^+]_{out}$ is 145 mM, the membrane potential is −50 mV (inside negative), and the temperature is 37°C.

Solution: Using Equation 11–4 (p. 412), we can calculate the energy inherent in an electrochemical Na^+ gradient—that is, the cost of moving one Na^+ ion up this gradient:

$$\Delta G_t = RT \ln \frac{[Na^+]_{out}}{[Na^+]_{in}} + Z\mathcal{F}\Delta\psi$$

We then substitute standard values for R, T, and $\mathcal{F}$; the given values for $[Na^+]$ (expressed as molar concentrations); +1 for Z (because Na^+ has a positive charge); and 0.050 V for $\Delta\psi$. Note that the membrane potential

is −50 mV (inside negative), so the change in potential when an ion moves from inside to outside is 50 mV.

$$\Delta G_t = (8.315\ \text{J/mol·K})(310\ \text{K}) \ln \frac{1.45 \times 10^{-1}}{1.2 \times 10^{-2}} + 1(96{,}500\ \text{J/V·mol})(0.050\ \text{V})$$
$$= 11.2\ \text{kJ/mol}$$

When Na^+ reenters the cell, it releases the electrochemical potential created by pumping it out; ΔG for reentry is −11.2 kJ/mol of Na^+. This is the potential energy per mole of Na^+ that is available to pump glucose. Given that two Na^+ ions pass down their electrochemical gradient and into the cell for each glucose carried in by symport, the energy available to pump 1 mole of glucose is 2 × 11.2 kJ/mol = 22.4 kJ/mol. We can now calculate the maximum concentration ratio of glucose that can be achieved by this pump (from Equation 11–3, p. 411):

$$\Delta G_t = RT \ln \frac{[\text{glucose}]_{in}}{[\text{glucose}]_{out}}$$

Rearranging, then substituting the values of ΔG_t, R, and T, gives

$$\ln \frac{[\text{glucose}]_{in}}{[\text{glucose}]_{out}} = \frac{\Delta G_t}{RT} = \frac{22.4\ \text{kJ/mol}}{(8.315\ \text{J/mol·K})(310\ \text{K})} = 8.69$$
$$\frac{[\text{glucose}]_{in}}{[\text{glucose}]_{out}} = e^{8.69}$$
$$= 5.94 \times 10^3$$

Thus the cotransporter can pump glucose inward until its concentration inside the epithelial cell is about 6,000 times that outside (in the intestine). (This is the maximum theoretical ratio, assuming a perfectly efficient coupling of Na^+ reentry and glucose uptake.)

As glucose is pumped from the intestine into the epithelial cell at the apical surface, it is simultaneously moved from the cell into the blood by passive transport through a glucose transporter (GLUT2) in the basal surface (Fig. 11–43). The crucial role of Na^+ in symport and antiport systems such as this requires the continued outward pumping of Na^+ to maintain the transmembrane Na^+ gradient.

Because of the essential role of ion gradients in active transport and energy conservation, compounds that collapse ion gradients across cellular membranes are effective poisons, and those that are specific for infectious microorganisms can serve as antibiotics. One such substance is valinomycin, a small cyclic peptide that neutralizes the K^+ charge by surrounding it with six carbonyl oxygens **(Fig. 11–44)**. The hydrophobic peptide then acts as a shuttle, carrying K^+ across membranes down its concentration gradient and deflating that gradient. Compounds that shuttle ions across membranes in this way are called **ionophores** ("ion bearers"). Both valinomycin and monensin (a Na^+-carrying ionophore) are antibiotics; they kill microbial cells by disrupting secondary transport processes and energy-conserving reactions. Monensin is widely used as an antifungal and antiparasitic agent. ■

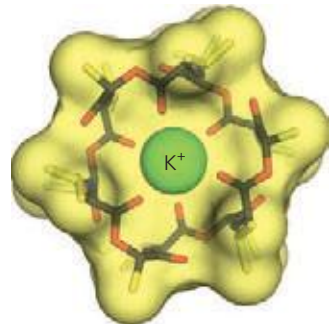


FIGURE 11–44 Valinomycin, a peptide ionophore that binds K^+. In this image, the surface contours are shown as a yellow envelope, through which a stick structure of the peptide and a K^+ ion (green) are visible. The oxygen atoms (red) that bind K^+ are part of a central hydrophilic cavity. Hydrophobic amino acid side chains (yellow) coat the outside of the molecule. Because the exterior of the K^+-valinomycin complex is hydrophobic, the complex readily diffuses through membranes, carrying K^+ down its concentration gradient. The resulting dissipation of the transmembrane ion gradient kills microbial cells, making valinomycin a potent antibiotic.

Aquaporins Form Hydrophilic Transmembrane Channels for the Passage of Water

Peter Agre

A family of integral membrane proteins discovered by Peter Agre, the **aquaporins (AQPs)**, provide channels for rapid movement of water molecules across all plasma membranes. Aquaporins are found in all organisms, and multiple aquaporin genes are generally present, encoding similar but not identical proteins. Eleven aquaporins are known in mammals, each with a specific localization and role (Table 11–5). Erythrocytes, which swell or shrink rapidly in response to abrupt changes in extracellular osmolarity as blood travels through the renal medulla, have a high density of aquaporin in their plasma membrane (2×10^5 copies of AQP1 per cell). Water secretion by the exocrine glands that produce sweat, saliva, and tears occurs through aquaporins. Seven different aquaporins play roles in urine production and water retention in the nephron (the functional unit of the kidney). Each renal AQP has a specific localization in the nephron, and each has specific properties and regulatory features. For example, AQP2 in the epithelial cells of the renal collecting duct is regulated by vasopressin (also called antidiuretic hormone): more water is reabsorbed in the kidney when the vasopressin level is high. Mutant mice with no AQP2 gene have increased urine output (polyuria) and decreased urine-concentrating ability, the

TABLE 11–5 Permeability Characteristics and Predominant Distribution of Known Mammalian Aquaporins

Aquaporin	Permeant (permeability)	Tissue distribution	Subcellular distribution*
AQP0	Water (low)	Lens	Plasma membrane
AQP1	Water (high)	Erythrocyte, kidney, lung, vascular endothelium, brain, eye	Plasma membrane
AQP2	Water (high)	Kidney, vas deferens	Apical plasma membrane, intracellular vesicles
AQP3	Water (high), glycerol (high), urea (moderate)	Kidney, skin, lung, eye, colon	Basolateral plasma membrane
AQP4	Water (high)	Brain, muscle, kidney, lung, stomach, small intestine	Basolateral plasma membrane
AQP5	Water (high)	Salivary gland, lacrimal gland, sweat gland, lung, cornea	Apical plasma membrane
AQP6	Water (low), anions ($NO_3^- > Cl^-$)	Kidney	Intracellular vesicles
AQP7	Water (high), glycerol (high), urea (high)	Adipose tissue, kidney, testis	Plasma membrane
AQP8†	Water (high)	Testis, kidney, liver, pancreas, small intestine, colon	Plasma membrane, intracellular vesicles
AQP9	Water (low), glycerol (high), urea (high), arsenite	Liver, leukocyte, brain, testis	Plasma membrane
AQP10	Water (low), glycerol (high), urea (high)	Small intestine	Intracellular vesicles

Source: Data from King, L.S., Kozono, D., & Agre, P. (2004) From structure to disease: the evolving tale of aquaporin biology. *Nat. Rev. Mol. Cell Biol.* 5, 688.

*Aquaporins that are present primarily in the apical or in the basolateral membrane are noted as localized in one of these membranes; those present in both membranes are described as localized in the plasma membrane.

†AQP8 might also be permeated by urea.

result of decreased water permeability of the proximal tubule. In humans, genetically defective AQPs are known to be responsible for a variety of diseases, including a relatively rare form of diabetes that is accompanied by polyuria (Box 11–1).

Water molecules flow through an AQP1 channel at a rate of about $10^9\ s^{-1}$. For comparison, the highest known turnover number for an enzyme is that for catalase, $4 \times 10^7\ s^{-1}$, and many enzymes have turnover numbers between $1\ s^{-1}$ and $10^4\ s^{-1}$ (see Table 6–7). The low activation energy for passage of water through aquaporin channels ($\Delta G^{\ddagger} < 15$ kJ/mol) suggests that water moves through the channels in a continuous stream, in the direction dictated by the osmotic gradient. (For a discussion of osmosis, see p. 56.) Aquaporins do not allow passage of protons (hydronium ions, H_3O^+), which would collapse membrane electrochemical gradients. What is the basis for this extraordinary selectivity?

We find an answer in the structure of AQP1, as determined by x-ray crystallography. AQP1 **(Fig. 11–45a)** consists of four identical monomers (each M_r 28,000), each of which forms a transmembrane pore with a diameter sufficient to allow passage of water molecules in single file. Each monomer has six transmembrane helical segments and two shorter helices, both of which contain the sequence Asn–Pro–Ala (NPA). The six transmembrane helices form the pore through the monomer, and the two short loops containing the NPA sequences extend toward the middle of the bilayer from opposite sides. Their NPA regions overlap in the middle of the membrane to form part of the specificity filter—the structure that allows only water to pass (Fig. 11–45b).

The water channel narrows to a diameter of 2.8 Å near the center of the membrane, severely restricting the size of molecules that can travel through. The positive charge of a highly conserved Arg residue at this bottleneck discourages the passage of cations such as H_3O^+. The residues that line the channel of each AQP1 monomer are generally nonpolar, but carbonyl oxygens in the peptide backbone, projecting into the narrow part of the channel at intervals, can hydrogen-bond with individual water molecules as they pass through; the two Asn residues (Asn^{76} and Asn^{192}) in the NPA loops also form hydrogen bonds with the water. The structure of the channel does not permit formation of a chain of water molecules close enough to allow proton hopping (see Fig. 2–14), which would effectively move protons across the membrane. Critical Arg and His residues and

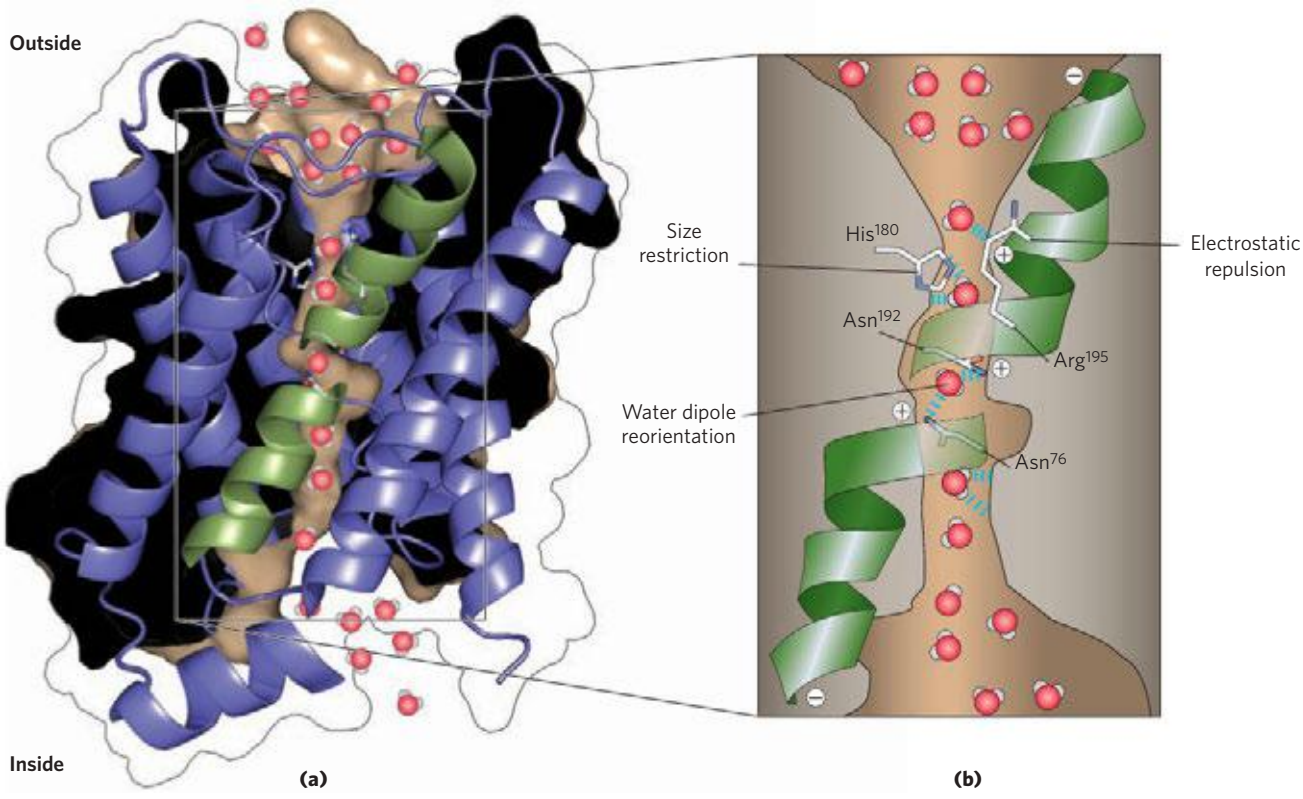


FIGURE 11–45 **Aquaporin.** The protein is a tetramer of identical subunits, each with a transmembrane pore. **(a)** A monomer of spinach aquaporin SoPIP2;1 (derived from PDB ID 2B5F), viewed in the plane of the membrane. The helices form a central pore, and two short helical segments (green) contain the Asn-Pro-Ala (NPA) sequences, found in all aquaporins, that form part of the water channel. **(b)** This cartoon of bovine aquaporin 1 (derived from PDB ID 1J4N) shows that the pore (brown; filled with water molecules shown in red and white) narrows at His^{180} to a diameter of 2.8 Å (about the size of a water molecule), limiting passage of molecules larger than H_2O. The positive charge of Arg^{195} repels cations, including H_3O^+, preventing their passage through the pore. The two short helices shown in green are oriented with their positively charged dipoles pointed at the pore in such a way as to force a water molecule to reorient as it passes through; this breaks up hydrogen-bonded chains of water molecules, preventing proton passage by "proton hopping."

electric dipoles formed by the short helices of the NPA loops provide positive charges in positions that repel any protons that might leak through the pore and prevent hydrogen bonding between adjacent water molecules.

An aquaporin isolated from spinach is known to be "gated"—open when two critical Ser residues near the intracellular end of the channel are phosphorylated, and closed when they are dephosphorylated. Both the open and closed structures have been determined by crystallography. Phosphorylation favors a conformation that presses two nearby Leu residues and a His residue into the channel, blocking the movement of water past that point and effectively closing the channel. Other aquaporins are regulated in other ways, allowing rapid changes in membrane permeability to water.

Although generally highly specific for water, some AQPs also allow glycerol or urea to pass at high rates (Table 11–5); these AQPs are believed to be important in the metabolism of glycerol. AQP7, for example, found in the plasma membranes of adipocytes (fat cells), transports glycerol efficiently. Mice with defective AQP7 develop obesity and non–insulin-dependent diabetes, presumably as a result of their inability to move glycerol into or out of adipocytes as triacylglycerols are converted to free fatty acids and glycerol, and as glycerol is acylated to triacylglycerol.

Ion-Selective Channels Allow Rapid Movement of Ions across Membranes

Ion-selective channels—first recognized in neurons and now known to be present in the plasma membranes of all cells, as well as in the intracellular membranes of eukaryotes—provide another mechanism for moving inorganic ions across membranes. Ion channels, together with ion pumps such as the Na^+K^+ ATPase, determine a plasma membrane's permeability to specific ions and regulate the cytosolic concentration of ions and the membrane potential. In neurons, very rapid changes in the activity of ion channels cause the changes in membrane potential (action potentials) that carry signals from one end of a neuron to the other. In myocytes, rapid opening of Ca^{2+} channels in the sarcoplasmic reticulum releases the Ca^{2+} that triggers muscle contraction. We discuss

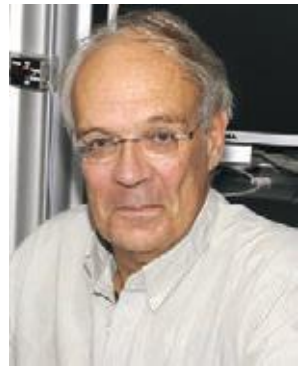
the signaling functions of ion channels in Chapter 12. Here we describe the structural basis for ion-channel function, using as examples a voltage-gated K^+ channel, the neuronal Na^+ channel, and the acetylcholine receptor ion channel.

Ion channels are distinct from ion transporters in at least three ways. First, the rate of flux through channels can be several orders of magnitude greater than the turnover number for a transporter—10^7 to 10^8 ions/s for an ion channel, approaching the theoretical maximum for unrestricted diffusion. By contrast, the turnover rate of the Na^+K^+ ATPase is about 100 s^{-1}. Second, ion channels are not saturable: rates do not approach a maximum at high substrate concentration. Third, they are gated in response to some cellular event. In **ligand-gated channels** (which are generally oligomeric), binding of an extracellular or intracellular small molecule forces an allosteric transition in the protein, which opens or closes the channel. In **voltage-gated ion channels**, a change in transmembrane electrical potential (V_m) causes a charged protein domain to move relative to the membrane, opening or closing the channel. Both types of gating can be very fast. A channel typically opens in a fraction of a millisecond and may remain open for only milliseconds, making these molecular devices effective for very fast signal transmission in the nervous system.

Ion-Channel Function Is Measured Electrically

Because a single ion channel typically remains open for only a few milliseconds, monitoring this process is beyond the limit of most biochemical measurements. Ion fluxes must therefore be measured electrically, either as changes in V_m (in the millivolt range) or as electric current I (in the microampere or picoampere range), using microelectrodes and appropriate amplifiers. In **patch-clamping**, a technique developed by Erwin Neher and Bert Sakmann in 1976, very small currents are measured through a tiny region of the membrane surface containing only one or a few ion-channel molecules **(Fig. 11–46)**. The researcher can measure the size and duration of the current that flows during one opening of an ion channel and can determine how often a channel opens and how that frequency is affected by membrane potential, regulatory ligands, toxins, and other agents. Patch-clamp studies have revealed that as many as 10^4 ions can move through a single ion channel in 1 ms. Such an ion flux represents a huge amplification of the initial signal; for example, only two acetylcholine molecules are needed to open an acetylcholine receptor channel (as described below).

Erwin Neher



Bert Sakmann

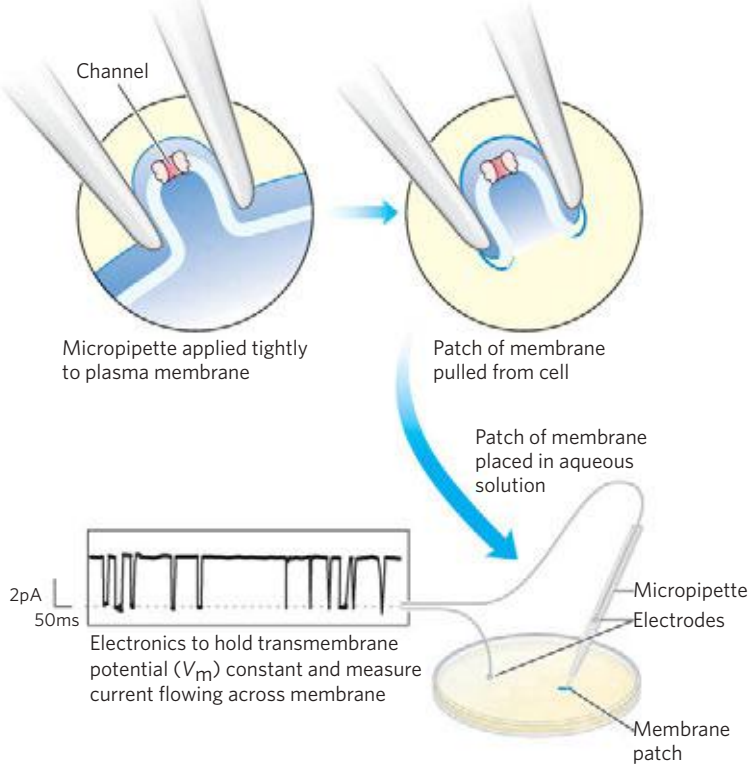


FIGURE 11–46 Electrical measurements of ion-channel function. The "activity" of an ion channel is estimated by measuring the flow of ions through it, using the patch-clamp technique. A finely drawn-out pipette (micropipette) is pressed against the cell surface, and negative pressure in the pipette forms a pressure seal between pipette and membrane. As the pipette is pulled away from the cell, it pulls off a tiny patch of membrane (which may contain one or a few ion channels). After placing the pipette and attached patch in an aqueous solution, the researcher can measure channel activity as the electric current that flows between the contents of the pipette and the aqueous solution. In practice, a circuit is set up that "clamps" the transmembrane potential at a given value and measures the current that must flow to maintain this voltage. With highly sensitive current detectors, researchers can measure the current flowing through a single ion channel, typically a few picoamperes. The trace shows the current through a single acetylcholine receptor channel as a function of time (in milliseconds), revealing how fast the channel opens and closes, how frequently it opens, and how long it stays open. Downward deflection represents channel opening. Clamping the V_m at different values permits determination of the effect of membrane potential on these parameters of channel function.

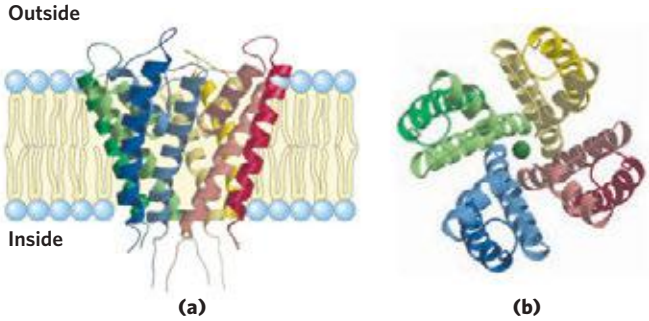
The Structure of a K^+ Channel Reveals the Basis for Its Specificity

Roderick MacKinnon

The structure of a potassium channel from the bacterium *Streptomyces lividans*, determined crystallographically by Roderick MacKinnon in 1998, provides important insight into the way ion channels work. This bacterial ion channel is related in sequence to all other known K^+ channels and serves as the prototype for such channels, including the voltage-gated K^+ channel of neurons. Among the members of this protein family, the similarities in sequence are greatest in the "pore region," which contains the ion selectivity filter that allows K^+ (radius 1.33 Å) to pass 10^4 times more readily than Na^+ (radius 0.95 Å)—at a rate (about 10^8 ions/s) approaching the theoretical limit for unrestricted diffusion.

The K^+ channel consists of four identical subunits that span the membrane and form a cone within a cone surrounding the ion channel, with the wide end of the double cone facing the extracellular space **(Fig. 11–47a)**. Each subunit has two transmembrane α helices as well as a third, shorter helix that contributes to the pore region. The outer cone is formed by one of the transmembrane helices of each subunit. The inner cone, formed by the other four transmembrane helices, surrounds the ion channel and cradles the ion selectivity filter. Viewed perpendicular to the plane of the membrane, the central channel is seen to be just wide enough to accommodate an unhydrated metal ion such as potassium (Fig. 11–47b).

Both the ion specificity and the high flux through the channel are understandable from what we know of the channel's structure (Fig. 11–47c). At the inner and outer plasma membrane surfaces, the entryways to the channel have several negatively charged amino acid residues, which presumably increase the local concentration of cations such as K^+ and Na^+. The ion path through the membrane begins (on the inner surface) as a wide, water-filled channel in which the ion can retain its hydration sphere. Further stabilization is provided by the short helices in the pore region of each subunit, with the partial negative charges of their electric dipoles pointed at K^+ in the channel. About two-thirds of the way through the membrane, this channel narrows in the region of the selectivity filter, forcing the ion to give up its hydrating water molecules. Carbonyl oxygen atoms in the backbone of the selectivity filter replace the water molecules in the hydration sphere, forming a series of perfect coordination shells through which the K^+ moves. This favorable interaction with the filter is not possible for Na^+, which is too small to make contact with all the potential oxygen ligands. The preferential stabilization of K^+ is the basis for the ion selectivity of the filter, and mutations that change residues in this part of the protein eliminate the channel's ion selectivity. The K^+-binding sites of the filter are flexible enough to collapse to fit any



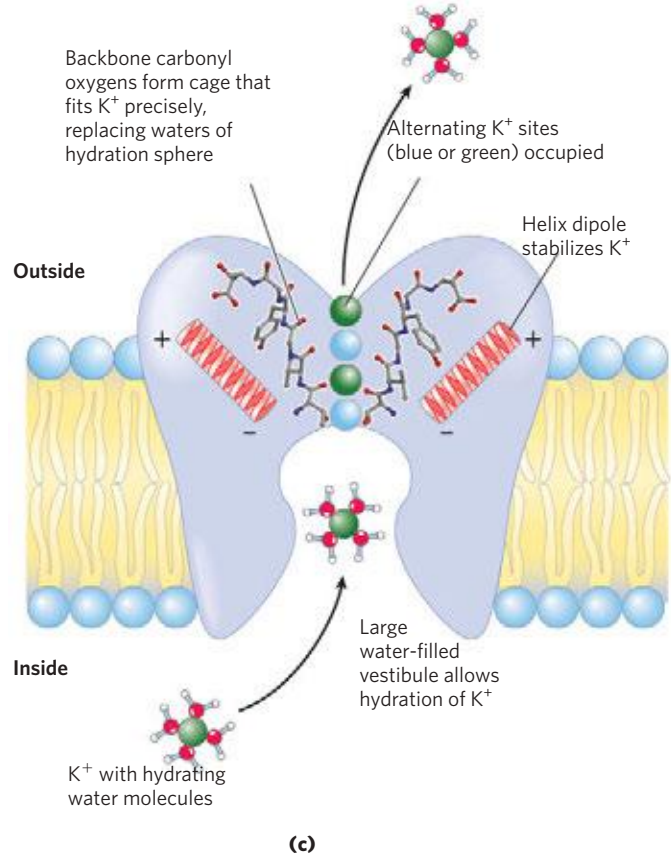


FIGURE 11–47 The K^+ channel of *Streptomyces lividans*. (PDB ID 1BL8) **(a)** Viewed in the plane of the membrane, the channel consists of eight transmembrane helices (two from each of four identical subunits), forming a cone with its wide end toward the extracellular space. The inner helices of the cone (lighter colored) line the transmembrane channel, and the outer helices interact with the lipid bilayer. Short segments of each subunit converge in the open end of the cone to make a selectivity filter. **(b)** This view, perpendicular to the plane of the membrane, shows the four subunits arranged around a central channel just wide enough for a single K^+ ion to pass. **(c)** Diagram of a K^+ channel in cross section, showing the structural features critical to function. Carbonyl oxygens (red) of the peptide backbone in the selectivity filter protrude into the channel, interacting with and stabilizing a K^+ ion passing through. These ligands are perfectly positioned to interact with each of four K^+ ions but not with the smaller Na^+ ions. This preferential interaction with K^+ is the basis for the ion selectivity.

Na^+ that enters the channel, and this conformational change closes the channel.

There are four potential K^+-binding sites along the selectivity filter, each composed of an oxygen "cage" that provides ligands for the K^+ ions (Fig. 11–47c). In the crystal structure, two K^+ ions are visible within the selectivity filter, about 7.5 Å apart, and two water molecules occupy the unfilled positions. K^+ ions pass through the filter in single file; their mutual electrostatic repulsion most likely just balances the interaction of each ion with the selectivity filter and keeps them moving. Movement of the two K^+ ions is concerted: first they occupy positions 1 and 3, then they hop to positions 2 and 4. The energetic difference between these two configurations (1, 3 and 2, 4) is very small; energetically, the selectivity pore is not a series of hills and valleys but a flat surface, which is ideal for rapid ion movement through the channel. The structure of the channel seems to have been optimized during evolution to give maximal flow rates and high specificity.

Voltage-gated K^+ channels are more complex structures than that illustrated in Figure 11–47, but they are variations on the same theme. For example, the mammalian voltage-gated K^+ channels in the *Shaker* family have an ion channel like that of the bacterial channel shown in Figure 11–47, but with additional protein domains that sense the membrane potential, move in response to a change in potential, and in moving trigger the opening or closing of the K^+ channel **(Fig. 11–48)**. The critical transmembrane helix in the voltage-sensing domain of

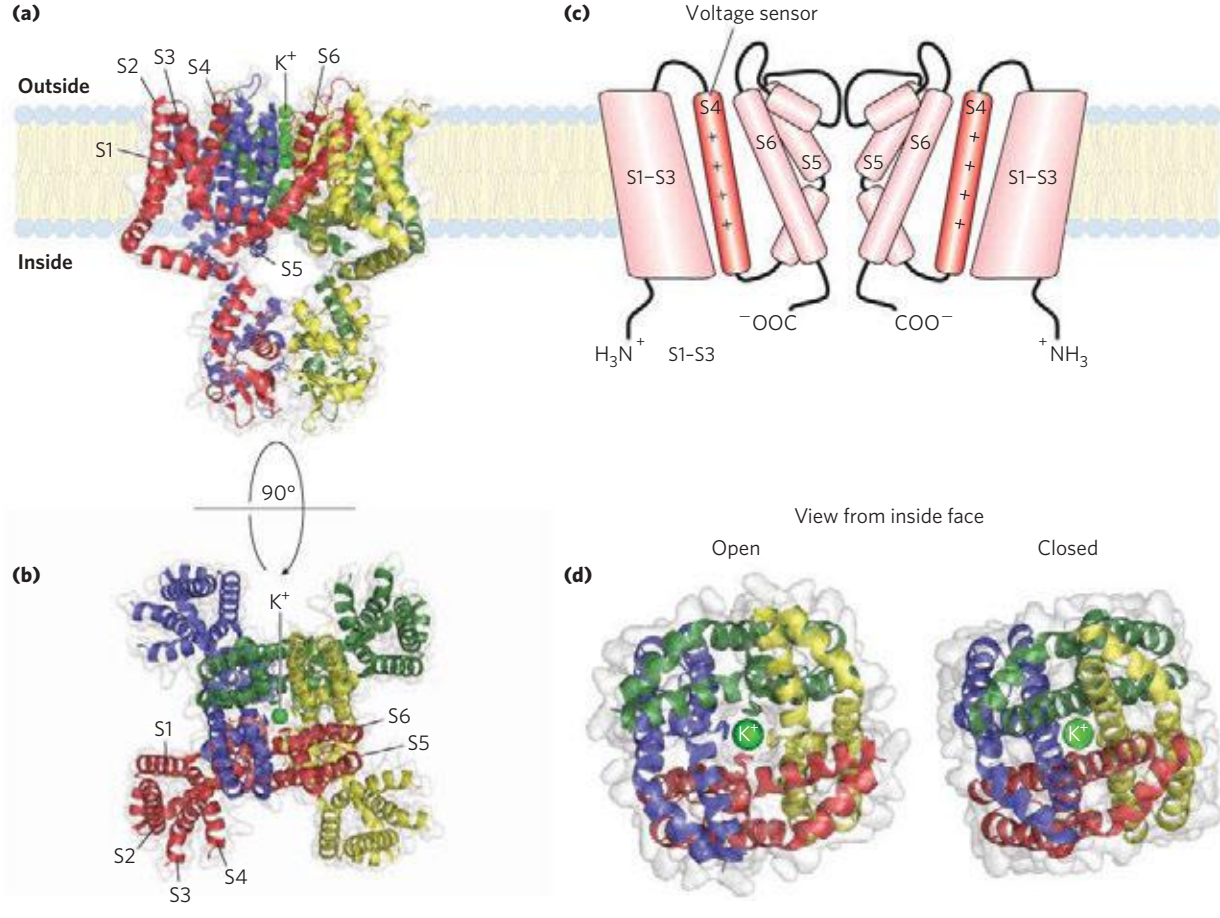


FIGURE 11–48 Structural basis for voltage gating in the K^+ channel. (PDB ID 2A79) This crystal structure of the Kv1.2-β2 subunit complex from rat brain shows the basic K^+ channel (corresponding to that shown in Fig. 11–47) with the extra machinery necessary to make the channel sensitive to gating by membrane potential: four transmembrane helical extensions of each subunit and four β subunits. The entire complex, viewed **(a)** in the plane of the membrane and **(b)** perpendicular to the plane (as viewed from outside the membrane), is represented as in Figure 11–47, with each subunit in a different color; each of the four β subunits is the same color as the subunit with which it associates. In (b), each transmembrane helix of one subunit (red) is numbered, S1 to S6. S5 and S6 from each of four subunits form the channel itself and are comparable to the two transmembrane helices of each subunit in Figure 11–47. S1 to S4 are four transmembrane helices. The S4 helix contains the highly conserved Arg residues and is believed to be the chief moving part of the voltage-sensing mechanism. **(c)** A schematic diagram of the voltage-gated channel, showing the basic pore structure (center) and the extra structures that make the channel voltage-sensitive; S4, the Arg-containing helix, is orange. For clarity, the β subunits are not shown in this view. In the resting membrane, the transmembrane electrical potential (inside negative) exerts a pull on positively charged Arg side chains in S4, toward the cytosolic side. When the membrane is depolarized, the pull is lessened, and with complete reversal of the membrane potential, S4 is drawn toward the extracellular side. **(d)** This movement of S4 is physically coupled to opening and closing of the K^+ channel, which is shown here in its open and closed conformations. Although K^+ is present in the closed channel, the pore closes on the bottom, near the cytosol, preventing K^+ passage.

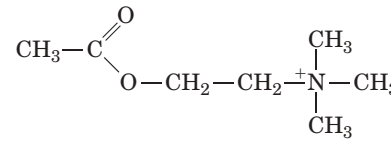
Shaker K^+ channels contains four Arg residues; the positive charges on these residues cause the helix to move relative to the membrane in response to changes in the transmembrane electric field (the membrane potential).

Cells also have channels that specifically conduct Na^+ or Ca^{2+} and exclude K^+. In each case, the ability to discriminate among cations requires both a cavity in the binding site of just the right size (neither too large nor too small) to accommodate the ion and the precise positioning within the cavity of carbonyl oxygens that can replace the ion's hydration shell. This fit can be achieved with molecules smaller than proteins; for example, valinomycin (Fig. 11–44) can provide the precise fit that gives high specificity for the binding of one ion rather than another. Chemists have designed small molecules with very high specificity for binding of Li^+ (radius 0.60 Å), Na^+ (radius 0.95 Å), K^+ (radius 1.33 Å), or Rb^+ (radius 1.48 Å). The biological versions, however—the channel proteins—not only *bind* specifically but *conduct* ions across membranes in a *gated* fashion.

Gated Ion Channels Are Central in Neuronal Function

Virtually all rapid signaling between neurons and their target tissues (such as muscle) is mediated by the rapid opening and closing of ion channels in plasma membranes. For example, Na^+ channels in neuronal plasma membranes sense the transmembrane electrical gradient and respond to changes by opening or closing. These voltage-gated ion channels are typically very selective for Na^+ over other monovalent or divalent cations (by factors of 100 or more) and have very high flux rates ($>10^7$ ions/s). Closed in the resting state, Na^+ channels are opened—activated—by a reduction in the membrane potential; they then undergo very rapid inactivation. Within milliseconds of opening, a channel closes and remains inactive for many milliseconds. Activation followed by inactivation of Na^+ channels is the basis for signaling by neurons (see Fig. 12–26).

Another very well-studied ion channel is the **nicotinic acetylcholine receptor**, which functions in the passage of an electric signal from a motor neuron to a muscle fiber at the neuromuscular junction (signaling the muscle to contract). Acetylcholine released by the motor neuron diffuses a few micrometers to the plasma membrane of a myocyte, where it binds to an acetylcholine receptor. This forces a conformational change in the receptor, causing its ion channel to open. The resulting inward movement of positively charged ions into the myocyte depolarizes its plasma membrane and triggers contraction. The acetylcholine receptor allows Na^+, Ca^{2+}, and K^+ to pass through its channel with equal ease, but other cations and all anions are unable to pass. Movement of Na^+ through an acetylcholine receptor ion channel is unsaturable (its rate is linear with respect to extracellular $[Na^+]$) and very fast—about 2×10^7 ions/s under physiological conditions.



Acetylcholine

The acetylcholine receptor channel is typical of many other ion channels that produce or respond to electric signals: it has a "gate" that opens in response to stimulation by a signal molecule (in this case acetylcholine) and an intrinsic timing mechanism that closes the gate after a split second. Thus the acetylcholine signal is transient—an essential feature of all electric signal conduction.

Based on similarities between the amino acid sequences of other ligand-gated ion channels and the acetylcholine receptor, neuronal receptor channels that respond to the extracellular signals γ-aminobutyric acid (GABA), glycine, and serotonin are grouped in the acetylcholine receptor superfamily and probably share three-dimensional structure and gating mechanisms. The $GABA_A$ and glycine receptors are anion channels specific for Cl^- or HCO_3^-, whereas the serotonin receptor, like the acetylcholine receptor, is cation-specific.

Another class of ligand-gated ion channels respond to *intracellular* ligands: 3′,5′-cyclic guanosine mononucleotide (cGMP) in the vertebrate eye, cGMP and cAMP in olfactory neurons, and ATP and inositol 1,4,5-trisphosphate (IP_3) in many cell types. These channels are composed of multiple subunits, each with six transmembrane helical domains. We discuss the signaling functions of these ion channels in Chapter 12.

Table 11–6 shows some transporters discussed in other chapters in the context of the pathways in which they act.

Defective Ion Channels Can Have Severe Physiological Consequences

The importance of ion channels to physiological processes is clear from the effects of mutations in specific ion-channel proteins (Table 11–7, Box 11–2). Genetic defects in the voltage-gated Na^+ channel of the myocyte plasma membrane result in diseases in which muscles are periodically either paralyzed (as in hyperkalemic periodic paralysis) or stiff (as in paramyotonia congenita). Cystic fibrosis is the result of a mutation that changes one amino acid in the protein CFTR, a Cl^- ion channel; the defective process here is not neurotransmission but secretion by various exocrine gland cells with activities tied to Cl^- ion fluxes.

Many naturally occurring toxins act on ion channels, and the potency of these toxins further illustrates the importance of normal ion-channel function. Tetrodotoxin (produced by the puffer fish, *Sphaeroides rubripes*) and saxitoxin (produced by the marine

TABLE 11–6 Transport Systems Described Elsewhere in This Text

Transport system and location	Figure	Role
Adenine nucleotide antiporter of mitochondrial inner membrane	19–30	Imports substrate ADP for oxidative phosphorylation and exports product ATP
Acetylcholine receptor/channel	12–28	Signals muscle contraction
Acyl-carnitine/carnitine transporter of mitochondrial inner membrane	17–6	Imports fatty acids into matrix for β oxidation
P_i-H^+ symporter of mitochondrial inner membrane	19–30	Supplies P_i for oxidative phosphorylation
Malate–α-ketoglutarate transporter of mitochondrial inner membrane	19–31	Shuttles reducing equivalents (as malate) from matrix to cytosol
Glutamate-aspartate transporter of mitochondrial inner membrane	19–31	Completes shuttling begun by malate–α-ketoglutarate shuttle
Citrate transporter of mitochondrial inner membrane	21–10	Provides cytosolic citrate as source of acetyl-CoA for lipid synthesis
Pyruvate transporter of mitochondrial inner membrane	21–10	Is part of mechanism for shuttling citrate from matrix to cytosol
Fatty acid transporter of myocyte plasma membrane	17–3	Imports fatty acids for fuel
Complex I, III, and IV proton transporters of mitochondrial inner membrane	19–16	Act as energy-conserving mechanism in oxidative phosphorylation, converting electron flow into proton gradient
Thermogenin (uncoupling protein 1), a proton pore of mitochondrial inner membrane	19–36, 23–34	Allows dissipation of proton gradient in mitochondria as means of thermogenesis and/or disposal of excess fuel
Cytochrome *bf* complex, a proton transporter of chloroplast thylakoid	19–61	Acts as proton pump, driven by electron flow through the Z scheme; source of proton gradient for photosynthetic ATP synthesis
Bacteriorhodopsin, a light-driven proton pump	19–69	Is light-driven source of proton gradient for ATP synthesis in halophilic bacterium
F_oF_1 ATPase/ATP synthase of mitochondrial inner membrane, chloroplast thylakoid, and bacterial plasma membrane	19–25, 19–62a, 19–66	Interconverts energy of proton gradient and ATP during oxidative phosphorylation and photophosphorylation
P_i–triose phosphate antiporter of chloroplast inner membrane	20–15, 20–16	Exports photosynthetic product from stroma; imports P_i for ATP synthesis
Bacterial protein transporter	27–44	Exports secreted proteins through plasma membrane
Protein translocase of ER	27–38	Transports into ER proteins destined for plasma membrane, secretion, or organelles
Nuclear pore protein translocase	27–42	Shuttles proteins between nucleus and cytoplasm
LDL receptor in animal cell plasma membrane	21–41	Imports, by receptor-mediated endocytosis, lipid-carrying particles
Glucose transporter of animal cell plasma membrane; regulated by insulin	12–16	Increases capacity of muscle and adipose tissue to take up excess glucose from blood
IP_3-gated Ca^{2+} channel of ER	12–10	Allows signaling via changes in cytosolic [Ca^{2+}]
cGMP-gated Ca^{2+} channel of retinal rod and cone cells	12–37	Allows signaling via rhodopsin linked to cAMP-dependent phosphodiesterase in vertebrate eye
Voltage-gated Na^+ channel of neuron	12–26	Creates action potentials in neuronal signal transmission

TABLE 11–7 Some Diseases Resulting from Ion Channel Defects

Ion channel	Affected gene	Disease
Na^+ (voltage-gated, skeletal muscle)	*SCN4A*	Hyperkalemic periodic paralysis (or paramyotonia congenita)
Na^+ (voltage-gated, neuronal)	*SCN1A*	Generalized epilepsy with febrile seizures
Na^+ (voltage-gated, cardiac muscle)	*SCN5A*	Long QT syndrome 3
Ca^{2+} (neuronal)	*CACNA1A*	Familial hemiplegic migraine
Ca^{2+} (voltage-gated, retina)	*CACNA1F*	Congenital stationary night blindness
Ca^{2+} (polycystin-1)	*PKD1*	Polycystic kidney disease
K^+ (neuronal)	*KCNQ4*	Dominant deafness
K^+ (voltage-gated, neuronal)	*KCNQ2*	Benign familial neonatal convulsions
Nonspecific cation (cGMP-gated, retinal)	*CNCG1*	Retinitis pigmentosa
Acetylcholine receptor (skeletal muscle)	*CHRNA1*	Congenital myasthenic syndrome
Cl^-	*CFTR*	Cystic fibrosis

dinoflagellate *Gonyaulax*, which causes "red tides") act by binding to the voltage-gated Na^+ channels of neurons and preventing normal action potentials. Puffer fish is an ingredient of the Japanese delicacy fugu, which may be prepared only by chefs specially trained to separate succulent morsel from deadly poison. Eating shellfish that have fed on *Gonyaulax* can also be fatal; shellfish are not sensitive to saxitoxin, but they concentrate it in their muscles, which become highly poisonous to organisms higher up the food chain. The venom of the black mamba snake contains dendrotoxin, which interferes with voltage-gated K^+ channels. Tubocurarine, the active component of curare (used as an arrow poison in the Amazon region), and two other toxins from snake venoms, cobrotoxin and bungarotoxin, block the acetylcholine receptor or prevent the opening of its ion channel. By blocking signals from nerves to muscles, all these toxins cause paralysis and possibly death. On the positive side, the extremely high affinity of bungarotoxin for the acetylcholine receptor ($K_d = 10^{-15}$ M) has proved useful experimentally: the radiolabeled toxin was used to quantify the receptor during its purification. ■

SUMMARY 11.3 Solute Transport across Membranes

- Movement of polar compounds and ions across biological membranes requires transporter proteins. Some transporters simply facilitate passive diffusion across the membrane from the side with higher concentration to the side with lower. Others transport solutes against an electrochemical gradient; this requires a source of metabolic energy.
- Carriers, like enzymes, show saturation and stereospecificity for their substrates. Transport via these systems may be passive or active. Primary active transport is driven by ATP or electron-transfer reactions; secondary active transport is driven by coupled flow of two solutes, one of which (often H^+ or Na^+) flows down its electrochemical gradient as the other is pulled up its gradient.
- The GLUT transporters, such as GLUT1 of erythrocytes, carry glucose into cells by facilitated diffusion. These transporters are uniporters, carrying only one substrate. Symporters permit simultaneous passage of two substances in the same direction; examples are the lactose transporter of *E. coli*, driven by the energy of a proton gradient (lactose-H^+ symport), and the glucose transporter of intestinal epithelial cells, driven by a Na^+ gradient (glucose-Na^+ symport). Antiporters mediate simultaneous passage of two substances in opposite directions; examples are the chloride-bicarbonate exchanger of erythrocytes and the ubiquitous Na^+K^+ ATPase.
- In animal cells, Na^+K^+ ATPase maintains the differences in cytosolic and extracellular concentrations of Na^+ and K^+, and the resulting Na^+ gradient is used as the energy source for a variety of secondary active transport processes.
- The Na^+K^+ ATPase of the plasma membrane and the Ca^{2+} transporters of the sarcoplasmic and endoplasmic reticulum (the SERCA pumps) are examples of P-type ATPases; they undergo reversible phosphorylation during their catalytic cycle. F-type ATPase proton pumps (ATP synthases) are central to energy-conserving mechanisms in mitochondria and chloroplasts. V-type ATPases produce gradients of protons across some intracellular membranes, including plant vacuolar membranes.
- ABC transporters carry a variety of substrates (including many drugs) out of cells, using ATP as energy source.

- Ionophores are lipid-soluble molecules that bind specific ions and carry them passively across membranes, dissipating the energy of electrochemical ion gradients.
- Water moves across membranes through aquaporins. Some aquaporins are regulated; some also transport glycerol or urea.
- Ion channels provide hydrophilic pores through which select ions can diffuse, moving down their electrical or chemical concentration gradients; they characteristically are unsaturable, have very high flux rates, and are highly specific for one ion. Most are voltage- or ligand-gated. The neuronal Na^+ channel is voltage-gated, and the acetylcholine receptor ion channel is gated by acetylcholine, which triggers conformational changes that open and close the transmembrane path.

Key Terms

Terms in bold are defined in the glossary.

fluid mosaic model 387
micelle 387
bilayer 387
vesicle 388
integral proteins 389
peripheral proteins 389
amphitropic proteins 390
annular lipid 391
hydropathy index 392
positive-inside rule 393
β barrel 393
porin 393
liquid-disordered state (l_d) 395
liquid-ordered state (l_o) 395
flippases 397
floppases 397
scramblases 397
FRAP 398
microdomains 398
rafts 399
GPI-anchored protein 399
caveolin 399
caveolae 399
BAR domain 400
fusion protein 400
v-SNAREs 401
t-SNAREs 401
selectins 402
simple diffusion 403
membrane potential (V_m) 403
electrochemical gradient 403
electrochemical potential 403
facilitated diffusion 403
passive transport 404
transporters 404
permeases 404
channels 404
K_t ($K_{transport}$) 406
electroneutral 409
cotransport 409
antiport 409
symport 409
uniport 409
active transport 409
electrogenic 410
P-type ATPases 410
SERCA pump 410
Na^+K^+ ATPase 411
V-type ATPases 412
F-type ATPases 412
ATP synthase 413
ABC transporters 413
multidrug transporters 413
lactose transporter 416
major facilitator superfamily (MFS) 416
Na^+-glucose symporters 417
ionophore 418
aquaporins (AQPs) 418
ion channel 420
ligand-gated channel 421
voltage-gated channel 421
patch-clamping 421
nicotinic acetylcholine receptor 424

Further Reading

Composition and Architecture of Membranes

Dowhan, W. (1997) Molecular basis for membrane phospholipids diversity: why are there so many lipids? *Annu. Rev. Biochem.* **66**, 199–232.

Ediden, M. (2002) Lipids on the frontier: a century of cell-membrane bilayers. *Nat. Rev. Mol. Cell Biol.* **4**, 414–418.
Short review of how the notion of a lipid bilayer membrane was developed and confirmed.

Leventis, P.A. & Grinstein, S. (2010) The distribution and function of phosphatidylserine in cellular membranes. *Annu. Rev. Biophys.* **39**, 407–427.

Maxfield, R.R. & van Meer, G. (2010) Cholesterol, the central lipid of mammalian cells *Curr. Opin. Cell Biol.* **22**, 422–429.

Von Heijne, G. (2006) Membrane protein topology. *Nat. Rev. Mol. Cell Biol.* **7**, 909–918.

White, S.H., Ladokhin, A.S., Jayasinghe, S., & Hristova, K. (2001) How membranes shape protein structure. *J. Biol. Chem.* **276**, 32,395–32,398.
Brief, intermediate-level review of the forces that shape transmembrane helices.

Wimley, W.C. (2003) The versatile β barrel membrane protein. *Curr. Opin. Struct. Biol.* **13**, 1–8.
Intermediate-level review.

Zeth, K. & Thein, M. (2010) Porins in prokaryotes and eukaryotes: common themes and variations. *Biochem. J.* **431**, 13–22.
Intermediate-level review of the β-barrel porins.

Membrane Dynamics

Daleke, D.L. (2007) Phospholipid flippases. *J. Biol. Chem.* **282**, 821–825.
Intermediate-level review.

Deveaux, P.F., Lopez-Montero, I., & Bryde, S. (2006) Proteins involved in lipid translocation in eukaryotic cells. *Chem. Phys. Lipids* **141**, 119–132.

Didier, M., Lenne, P.-F., Rigneault, H., & He, H.-T. (2006) Dynamics in the plasma membrane: how to combine fluidity and order. *EMBO J.* **25**, 3446–3457.
Intermediate-level review of studies of membrane dynamics, with fluorescent and other probes.

Frost, A., Unger, V.M., & De Camilli, P. (2009) The BAR domain superfamily: membrane-molding macromolecules. *Cell* **137**, 191–196.

Frye, L.D. & Ediden, M. (1970) The rapid intermixing of cell-surface antigens after formation of mouse-human heterokaryons. *J. Cell Sci.* **7**, 319–335.
The classic demonstration of membrane protein mobility.

Graham, T.R. (2004) Flippases and vesicle-mediated protein transport. *Trends Cell Biol.* **14**, 670–677.
Intermediate-level review of flippase function.

Graham, T.R. & Kozlov, M.M. (2010) Interplay of proteins and lipids in generating membrane curvature. *Curr. Opin. Cell Biol.* **22**, 430–436.

Hannich, J.T., Umebayashi, K., & Riezman, H. (2011) Distribution and functions of sterols and sphingolipids. In *The Biology of Lipids: Trafficking, Regulation, and Function* (Simons, K., ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; also in *CSH Perspect. Cell Biol.* doi:10.1101/cshperspect.a004762.

Jahn, R. & Scheller, R.H. (2006) SNAREs—engines for membrane fusion. *Nat. Rev. Cell Mol. Biol.* **7**, 631–643.
Excellent intermediate-level review of the role of SNAREs in membrane fusion and the fusion mechanism itself.

Janmey, P.A. & Kunnunen, P.K.J. (2006) Biophysical properties of lipids and dynamic membranes. *Trends Cell Biol.* **16**, 538–546.

Leventis, P.A. & Grinstein, S. (2010) The distribution and function of phosphatidylserine in cellular membranes. *Annu. Rev. Biophys.* **39**, 407–427.

Advanced review that includes discussion of flippases.

Lingwood, D. & Simons, K. (2010) Lipid rafts as a membrane-organizing principle. *Science* **327**, 46–50.

MacCallum, J.L. & Tieleman, D.P. (2011) Hydrophobicity scales: a thermodynamic looking glass into lipid-protein interactions. *Trends Biochem. Sci.* **36**, 653–662.

Intermediate-level review of several methods for determining an amino acid's hydrophobicity.

Marguet, D., Lenne, P.-F., Rigneault, H., & He, H.-T. (2006) Dynamics in the plasma membrane: how to combine fluidity and order. *EMBO J.* **25**, 3446–3457.

Intermediate-level review of the methods and results of studies on molecular motions in the membrane.

Martens, S. & McMahon, H.T. (2008) Mechanisms of membrane fusion: disparate players and common principles. *Nat. Rev. Mol. Cell Biol.* **9**, 543–566.

Niessen, C.M., Leckband, D., & Yap, A.S. (2011) Tissue organization by cadherin adhesion molecules: dynamic molecular and cellular mechanisms of morphogenetic regulation. *Physiol. Rev.* **91**, 691–731.

Palmgren, M.G. & Nissen, P. (2011) P-type ATPases. *Annu. Rev. Biophys.* **40**, 243–266.

Advanced review of transporters that include P4-type flippases.

Palsdottir, H. & Hunte, C. (2004) Lipids in membrane protein structures. *Biochim. Biophys. Acta* **1666**, 2–18.

Parton, R.G. & Simons, K. (2007) The multiple faces of caveolae. *Nat. Rev. Mol. Cell Biol.* **8**, 185–194.

Phillips, R., Ursell, T., Wiggins, P., & Sens, P. (2009) Emerging roles for lipids in shaping membrane-protein function. *Nature* **459**, 379–385.

Intermediate-level review.

Qualman, B., Koch, D., & Kessels, M.M. (2011) Let's go bananas: revisiting the endocytic BAR code. *EMBO J.* **30**, 3501–3515.

Intermediate-level review of the action of BAR domains in causing membrane curvature.

Quinn, P.J. & Wolf, C. (2009) The liquid ordered phase in membranes. *Biochim. Biophys. Acta* **1788**, 33–46.

Advanced review of the state of membrane lipids. This is one of 26 excellent reviews that appear in this issue of the journal on all aspects of the physical state of lipids in biological membranes.

Sanyal, S. & Menon, A.K. (2009) Flipping lipids: why an' what's the reason for? *ACS Chem. Biol.* **4**, 895–909.

Intermediate-level review of flippases in membrane biogenesis.

Sezgin, E. & Schwille, P. (2011) Fluorescence techniques to study lipid dynamics. In *The Biology of Lipids: Trafficking, Regulation, and Function* (Simons, K., ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; also in *CSH Perspect. Cell Biol.*, doi:10.1101/cshperspect.a009803.

Simons, K. & Sampaio, J.L. (2011) Membrane organization and lipid rafts. In *The Biology of Lipids: Trafficking, Regulation, and Function* (Simons, K., ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; also in *CSH Perspect. Cell Biol.*, doi:10.1101/cshperspect.a004697.

Tanaka, K., Fujimura-Kamada, K., & Yamamoto, T. (2011) Functions of phospholipid flippases. *J. Biochem.* **149**, 131–143.

Intermediate-level review of flippase structure and function.

van der Velden, L.M., van de Graaf, S.F.J., & Klomp, L.W.J. (2010) Biochemical and cellular functions of P_4 ATPases. *Biochem. J.* **431,** 1–11.

Intermediate-level review of the P_4 ATPase flippases.

van Deurs, B., Roepstorff, K., Hommelgaard, A.M., & Sandpit, K. (2003) Caveolae: anchored, multifunctional platforms in the lipid ocean. *Trends Cell Biol.* **13**, 92–100.

Van Meer, G. (2011) Dynamic transbilayer lipid asymmetry. In *The Biology of Lipids: Trafficking, Regulation, and Function* (Simons, K., ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; also in *CSH Perspect. Cell Biol.*, doi:10.1101/cshperspect.a004671.

Wickner, W. & Shekmana, R. (2010) Membrane fusion *Nat. Struct. Mol. Biol*. **15**, 658–664.

Brief and accessible review of the molecules involved in membrane fusion.

Zhang, Y.-M. & Rock, C.O. (2008) Membrane lipid homeostasis in bacteria. *Nat. Rev. Microbiol.* **6**, 222–233.

Zimmerberg, J. & Kozlov, M.M. (2006) How proteins produce cellular membrane curvature. *Nat. Rev. Mol. Cell Biol.* **7**, 9–19.

Transporters

Augustin, R. (2010) The protein family of glucose transport facilitators: it's not only about glucose after all. *IUBMB Life* **62**, 315–333.

Advanced review of the GLUT proteins' structure and function.

Brini, M. & Carafoli, E. (2009) Calcium pumps in health and disease *Physiol. Rev.* **89**, 1341–1378.

Bublitz, M., Poulson, H., Preben Morth, J., & Nissen, P. (2010) In and out of the cation pumps: P-type ATPase structure revisited. *Curr. Opin. Struct. Biol.* **20**, 431–439.

Bublitz, M., Preben Morth, J., & Nissen, P. (2011) P-type ATPases at a glance. *J. Cell Sci.* **124**, 2515–2519.

Fujiyoshi, Y., Mitsuoka, K., de Groot, B.L., Philippsen, A., Grubmüller, H., Agre, P., & Engel, A. (2002) Structure and function of water channels. *Curr. Opin. Struct. Biol.* **12**, 509–515.

Guan, L. & Kaback, H.R. (2006) Lessons from lactose permease. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 67–91.

Hoffman, N.J. & Elmendorf, J.S. (2011) Signaling, cytoskeletal and membrane mechanisms regulating GLUT4 exocytosis. *Trends Endocrinol. Metab.* **22**, 110–116.

Jones, P.M., O'Mara, M.L., & George, A.M. (2001) ABC transporters: a riddle wrapped in a mystery inside an enigma. *Trends Biochem. Sci.* **34**, 520–531.

Intermediate-level review of ATP transporter structure and function.

Kjellbom, P., Larsson, C., Johansson, I., Karlsson, M., & Johanson, U. (1999) Aquaporins and water homeostasis in plants. *Trends Plant Sci.* **4**, 308–314.

Intermediate-level review.

Krishnamurthy, H., Piscitelli, C.L., & Gouaus, E. (2009) Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347–355.

Intermediate-level review of the structure and function of the Na^+ cotransporters.

Preben Morth, J., Pedersen, B.P., Buch-Pedersen, M.J., Andersen, J.P., Vilsen, B., Palmgren, M.G., & Nissen, P. (2011) A structural overview of the plasma membrane Na^+,K^+-ATPase and H^+-ATPase ion pumps. *Nat. Rev. Mol. Cell Biol.* **12**, 60–70.

Rees, D.C., Johnson, E., & Lewinson, O. (2009) ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* **10**, 218–227.

Sui, H., Han, B.-G., Lee, J.K., Walian, P., & Jap, B.K. (2001) Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**, 872–878.

High-resolution solution of the aquaporin structure by x-ray crystallography.

Thorens, B. & Mueckler, M. (2010) Glucose transporters in the 21st century. *Am. J. Physiol. Endocrinol. Metab.* **298**, E141–E145.

Intermediate-level review of glucose facilitators.

Toyoshima, C. & Mizutani, T. (2004) Crystal structure of the calcium pump with a bound ATP analogue. *Nature* **430**, 529–535.

Toyoshima, C., Nomura, H., & Tsuda, T. (2004) Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogs. *Nature* **432**, 361–368.

The supplementary materials available with the online version of this article include an excellent movie of the putative gating mechanism.

Watson, R.T. & Pessin, J.E. (2006) Bridging the GAP between insulin signaling and GLUT4 translocation. *Trends Biochem. Sci.* **31**, 215–222.

Intermediate-level review of the regulation of glucose transport through GLUT4.

Wright, E.M., Loo, D.D.F., & Hirayama, B.A. (2011) Biology of human sodium glucose transporters. *Physiol. Rev.* **91**, 733–794.

Advanced review of the Na^+-glucose cotransporter's biochemistry and physiology.

Zeth, K. & Thein, M. (2010) Porins in prokaryotes and eukaryotes: common themes and variations. *Biochem. J.* **431**, 13–22.

Ion Channels

Ashcroft, F.M. (2006). From molecule to malady. *Nature* **440**, 440–447.

A short review of the many known cases in which genetic defects in ion channels lead to disease in humans.

Doyle, D.A., Cabral, K.M., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., & MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77.

The first crystal structure of an ion channel is described.

Gadsby, D.C., Vergani, P., & Csanady, L. (2006) The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* **440**, 477–483.

This is one of seven excellent reviews of ion channels published together in this issue of *Nature*.

Gouaux, E. & MacKinnon, R. (2005) Principles of selective ion transport in channels and pumps. *Science* **310**, 1461–1465.

Short review of the architectural features of channels and pumps that give each protein its ion specificity.

Guggino, W.B. & Stanton, B.A. (2006) New insights into cystic fibrosis: molecular switches that regulate CFTR. *Nat. Rev. Mol. Cell Biol.* **7**, 426–436.

Hille, B. (2001) *Ion Channels of Excitable Membranes*, 3rd edn, Sinauer Associates, Sunderland, MA.

Intermediate-level text emphasizing the function of ion channels.

Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B.T., & MacKinnon, R. (2003) X-ray structure of a voltage-dependent K^+ channel. *Nature* **423**, 33–41.

King, L.S., Kozono, D., & Agre, P. (2004) From structure to disease: the evolving tale of aquaporin biology. *Nat. Rev. Mol. Cell Biol.* **5**, 687–698.

Intermediate-level review of the localization of aquaporins in mammalian tissues and the effects of aquaporin defects on physiology.

Lee, A.G. & East, J.M. (2001) What the structure of a calcium pump tells us about its mechanism. *Biochem J.* **356**, 665–683.

Long, S.B., Campbell, E.B., & MacKinnon, R. (2005) Crystal structure of a mammalian voltage-dependent *Shaker* family K^+ channel. *Science* **309**, 897–902.

Neher, E. & Sakmann, B. (1992) The patch clamp technique. *Sci. Am.* (March) **266**, 44–51.

Clear description of the electrophysiological methods used to measure the activity of single ion channels, by the Nobel Prize–winning developers of this technique.

Tombola, F., Pathak, M.M., & Isacoff, E.Y. (2006) How does voltage open an ion channel? *Annu. Rev. Cell Dev. Biol.* **22**, 23–52.

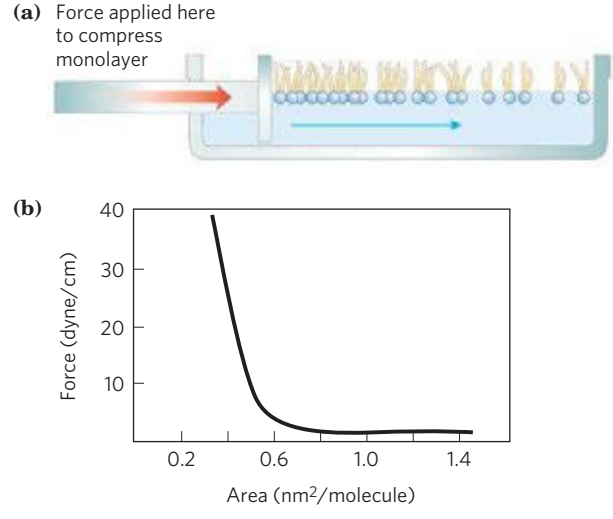
Advanced review of the mechanisms of voltage gating of ion channels.

Verkman, A.S. (2011) Aquaporins at a glance. *J. Cell Sci.* **124**, 2107–2112.

Intermediate-level, poster-style review of aquaporin structure and function.

Problems

1. Determining the Cross-Sectional Area of a Lipid Molecule When phospholipids are layered gently onto the surface of water, they orient at the air-water interface with their head groups in the water and their hydrophobic tails in the air. An experimental apparatus **(a)** has been devised that reduces the surface area available to a layer of lipids. By measuring the force necessary to push the lipids together, it is possible to determine when the molecules are packed tightly in a continuous monolayer; as that area is approached, the force needed to further reduce the surface area increases sharply **(b)**. How would you use this apparatus to determine the average area occupied by a single lipid molecule in the monolayer?



2. Evidence for a Lipid Bilayer In 1925, E. Gorter and F. Grendel used an apparatus like that described in Problem 1 to determine the surface area of a lipid monolayer formed by lipids extracted from erythrocytes of several animal species. They used a microscope to measure the dimensions of individual cells, from which they calculated the average surface area of one erythrocyte. They obtained the data shown in the table. Were these investigators justified in concluding that "chromocytes [erythrocytes] are covered by a layer of fatty substances that is two molecules thick" (i.e., a lipid bilayer)?

Animal	Volume of packed cells (mL)	Number of cells (per mm^3)	Total surface area of lipid monolayer from cells (m^2)	Total surface area of one cell (μm^2)
Dog	40	8,000,000	62	98
Sheep	10	9,900,000	6.0	29.8
Human	1	4,740,000	0.92	99.4

Source: Data from Gorter, E. & Grendel, F. (1925) On bimolecular layers of lipoids on the chromocytes of the blood. *J. Exp. Med.* 41, 439–443.

3. Number of Detergent Molecules per Micelle When a small amount of the detergent sodium dodecyl sulfate (SDS; $Na^+CH_3(CH_2)_{11}OSO_3^-$) is dissolved in water, the detergent ions enter the solution as monomeric species. As more detergent is added, a concentration is reached (the critical micelle concentration) at which the monomers associate to form micelles. The critical micelle concentration of SDS is 8.2 mM. The micelles have an average particle weight (the sum of the molecular weights of the constituent monomers) of 18,000. Calculate the number of detergent molecules in the average micelle.

4. Properties of Lipids and Lipid Bilayers Lipid bilayers formed between two aqueous phases have this important property: they form two-dimensional sheets, the edges of which close on each other and undergo self-sealing to form vesicles (liposomes).

(a) What properties of lipids are responsible for this property of bilayers? Explain.

(b) What are the consequences of this property for the structure of biological membranes?

5. Length of a Fatty Acid Molecule The carbon–carbon bond distance for single-bonded carbons such as those in a saturated fatty acyl chain is about 1.5 Å. Estimate the length of a single molecule of palmitate in its fully extended form. If two molecules of palmitate were placed end to end, how would their total length compare with the thickness of the lipid bilayer in a biological membrane?

6. Temperature Dependence of Lateral Diffusion The experiment described in Figure 11–18 was performed at 37 °C. If the experiment were carried out at 10 °C, what effect would you expect on the rate of diffusion? Why?

7. Synthesis of Gastric Juice: Energetics Gastric juice (pH 1.5) is produced by pumping HCl from blood plasma (pH 7.4) into the stomach. Calculate the amount of free energy required to concentrate the H^+ in 1 L of gastric juice at 37 °C. Under cellular conditions, how many moles of ATP must be hydrolyzed to provide this amount of free energy? The free-energy change for ATP hydrolysis under cellular conditions is about −58 kJ/mol (as explained in Chapter 13). Ignore the effects of the transmembrane electrical potential.

8. Energetics of the Na^+K^+ ATPase For a typical vertebrate cell with a membrane potential of −0.070 V (inside negative), what is the free-energy change for transporting 1 mol of Na^+ from the cell into the blood at 37 °C? Assume the concentration of Na^+ inside the cell is 12 mM and that in blood plasma is 145 mM.

9. Action of Ouabain on Kidney Tissue Ouabain specifically inhibits the Na^+K^+ ATPase activity of animal tissues but is not known to inhibit any other enzyme. When ouabain is added to thin slices of living kidney tissue, it inhibits oxygen consumption by 66%. Why? What does this observation tell us about the use of respiratory energy by kidney tissue?

10. Energetics of Symport Suppose you determined experimentally that a cellular transport system for glucose, driven by symport of Na^+, could accumulate glucose to concentrations 25 times greater than in the external medium, while the external $[Na^+]$ was only 10 times greater than the intracellular $[Na^+]$. Would this violate the laws of thermodynamics? If not, how could you explain this observation?

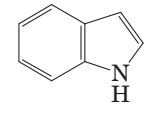
11. Location of a Membrane Protein The following observations are made on an unknown membrane protein, X. It can be extracted from disrupted erythrocyte membranes into a concentrated salt solution, and it can be cleaved into fragments by proteolytic enzymes. Treatment of erythrocytes with proteolytic enzymes followed by disruption and extraction of membrane components yields intact X. However, treatment of erythrocyte "ghosts" (which consist of just plasma membranes, produced by disrupting the cells and washing out the hemoglobin) with proteolytic enzymes followed by disruption and extraction yields extensively fragmented X. What do these observations indicate about the location of X in the plasma membrane? Do the properties of X resemble those of an integral or peripheral membrane protein?

12. Membrane Self-Sealing Cellular membranes are self-sealing—if they are punctured or disrupted mechanically, they quickly and automatically reseal. What properties of membranes are responsible for this important feature?

13. Lipid Melting Temperatures Membrane lipids in tissue samples obtained from different parts of a reindeer's leg have different fatty acid compositions. Membrane lipids from tissue near the hooves contain a larger proportion of unsaturated fatty acids than those from tissue in the upper leg. What is the significance of this observation?

14. Flip-Flop Diffusion The inner leaflet (monolayer) of the human erythrocyte membrane consists predominantly of phosphatidylethanolamine and phosphatidylserine. The outer leaflet consists predominantly of phosphatidylcholine and sphingomyelin. Although the phospholipid components of the membrane can diffuse in the fluid bilayer, this sidedness is preserved at all times. How?

15. Membrane Permeability At pH 7, tryptophan crosses a lipid bilayer at about one-thousandth the rate of indole, a closely related compound:



Suggest an explanation for this observation.

16. Water Flow through an Aquaporin A human erythrocyte has about 2×10^5 AQP1 monomers. If water molecules flow through the plasma membrane at a rate of 5×10^8 per AQP1 tetramer per second and the volume of an erythrocyte is 5×10^{-11} mL, how rapidly could an erythrocyte halve its volume as it encountered the high osmolarity (1 M) in the interstitial fluid of the renal medulla? Assume that the erythrocyte consists entirely of water.

17. Labeling the Lactose Transporter A bacterial lactose transporter, which is highly specific for lactose, contains a Cys residue that is essential to its transport activity. Covalent reaction of *N*-ethylmaleimide (NEM) with this Cys residue irreversibly inactivates the transporter. A high concentration of lactose in the medium prevents inactivation by NEM, presumably by sterically protecting the Cys residue, which is in or near the lactose-binding site. You know nothing else about the transporter protein. Suggest an experiment that might allow you to determine the M_r of the Cys-containing transporter polypeptide.

18. Predicting Membrane Protein Topology from Sequence You have cloned the gene for a human erythrocyte protein, which you suspect is a membrane protein. From the nucleotide sequence of the gene, you know the amino acid sequence. From this sequence alone, how would you evaluate the possibility that the protein is an integral protein? Suppose the protein proves to be an integral protein, either type I or type II. Suggest biochemical or chemical experiments that might allow you to determine which type it is.

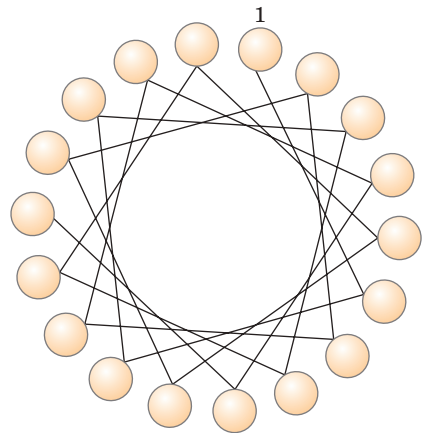
19. Intestinal Uptake of Leucine You are studying the uptake of L-leucine by epithelial cells of the mouse intestine. Measurements of the rate of uptake of L-leucine and several of its analogs, with and without Na^+ in the assay buffer, yield the results given in the table. What can you conclude about the properties and mechanism of the leucine transporter? Would you expect L-leucine uptake to be inhibited by ouabain?

	Uptake in presence of Na^+		Uptake in absence of Na^+	
Substrate	V_{max}	K_t (mM)	V_{max}	K_t (mM)
L-Leucine	420	0.24	23	0.2
D-Leucine	310	4.7	5	4.7
L-Valine	225	0.31	19	0.31

20. Effect of an Ionophore on Active Transport Consider the leucine transporter described in Problem 19. Would V_{max} and/or K_t change if you added a Na^+ ionophore to the assay solution containing Na^+? Explain.

21. Surface Density of a Membrane Protein *E. coli* can be induced to make about 10,000 copies of the lactose transporter (M_r 31,000) per cell. Assume that *E. coli* is a cylinder 1 μm in diameter and 2 μm long. What fraction of the plasma membrane surface is occupied by the lactose transporter molecules? Explain how you arrived at this conclusion.

22. Use of the Helical Wheel Diagram A helical wheel is a two-dimensional representation of a helix, a view along its central axis (see Fig. 11–30b; see also Fig. 4–4d). Use the helical wheel diagram shown here to determine the distribution of amino acid residues in a helical segment with the sequence –Val–Asp–Arg–Val–Phe–Ser–Asn–Val–Cys–Thr–His–Leu–Lys–Thr–Leu–Gln–Asp–Lys–



What can you say about the surface properties of this helix? How would you expect the helix to be oriented in the tertiary structure of an integral membrane protein?

23. Molecular Species in the *E. coli* Membrane The plasma membrane of *E. coli* is about 75% protein and 25% phospholipid by weight. How many molecules of membrane lipid are present for each molecule of membrane protein? Assume an average protein M_r of 50,000 and an average phospholipid M_r of 750. What more would you need to know to estimate the fraction of the membrane surface that is covered by lipids?

Using the Web

24. Membrane Protein Topology The receptor for the hormone epinephrine in animal cells is an integral membrane protein (M_r 64,000) that is believed to have seven membrane-spanning regions.

(a) Show that a protein of this size is capable of spanning the membrane seven times.

(b) Given the amino acid sequence of this protein, how would you predict which regions of the protein form the membrane-spanning helices?

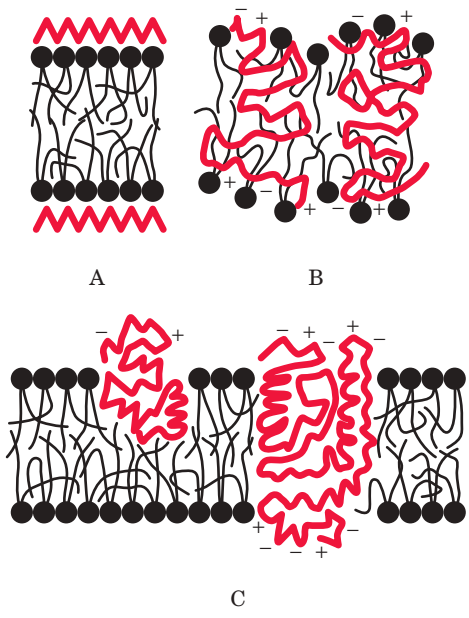
(c) Go to the Protein Data Bank (www.pdb.org). Use the PDB identifier 1DEP to retrieve the data page for a portion of the β-adrenergic receptor (one type of epinephrine receptor) isolated from a turkey. Using Jmol to explore the structure, predict whether this portion of the receptor is located within the membrane or at the membrane surface. Explain.

(d) Retrieve the data for a portion of another receptor, the acetylcholine receptor of neurons and myocytes, using the PDB identifier 1A11. As in (c), predict where this portion of the receptor is located and explain your answer.

If you have not used the PDB, see Box 4–4 (p. 132) for more information.

Data Analysis Problem

25. The Fluid Mosaic Model of Biological Membrane Structure Figure 11–3 shows the currently accepted fluid mosaic model of biological membrane structure. This model was presented in detail in a review article by S. J. Singer in 1971. In the article, Singer presented the three models of membrane structure that had been proposed by that time:



A. The Davson-Danielli-Robertson Model. This was the most widely accepted model in 1971, when Singer's review was published. In this model, the phospholipids are arranged as a bilayer. Proteins are found on both surfaces of the bilayer, attached to it by ionic interactions between the charged head groups of the phospholipids and charged groups in the proteins. Crucially, there is no protein in the interior of the bilayer.

B. The Benson Lipoprotein Subunit Model. Here the proteins are globular and the membrane is a protein-lipid mixture. The hydrophobic tails of the lipids are embedded in the hydrophobic parts of the proteins. The lipid head groups are exposed to the solvent. There is no lipid bilayer.

C. The Lipid-Globular Protein Mosaic Model. This is the model shown in Figure 11–3. The lipids form a bilayer and proteins are embedded in it, some extending through the bilayer and others not. Proteins are anchored in the bilayer by hydrophobic interactions between the hydrophobic tails of the lipids and hydrophobic portions of the protein.

For the data given below, consider how each piece of information aligns with each of the three models of membrane structure. Which model(s) are supported, which are not supported, and what reservations do you have about the data or their interpretation? Explain your reasoning.

(a) When cells were fixed, stained with osmium tetroxide, and examined in the electron microscope, the membranes showed a "railroad track" appearance, with two dark-staining lines separated by a light space.

(b) The thickness of membranes in cells fixed and stained in the same way was found to be 5 to 9 nm. The thickness of a "naked" phospholipid bilayer, without proteins, was 4 to 4.5 nm. The thickness of a single monolayer of proteins was about 1 nm.

(c) Singer wrote in his article: "The average amino acid composition of membrane proteins is not distinguishable from that of soluble proteins. In particular, a substantial fraction of the residues is hydrophobic" (p. 165).

(d) As described in Problems 1 and 2 of this chapter, researchers had extracted membranes from cells, extracted the lipids, and compared the area of the lipid monolayer with the area of the original cell membrane. The interpretation of the results was complicated by the issue illustrated in the graph of Problem 1: the area of the monolayer depended on how hard it was pushed. With very light pressures, the ratio of monolayer area to cell membrane area was about 2.0. At higher pressures—thought to be more like those found in cells—the ratio was substantially lower.

(e) Circular dichroism spectroscopy uses changes in polarization of UV light to make inferences about protein secondary structure (see Fig. 4–10). On average, this technique showed that membrane proteins have a large amount of α helix and little or no β sheet. This finding was consistent with most membrane proteins having a globular structure.

(f) Phospholipase C is an enzyme that removes the polar head group (including the phosphate) from phospholipids. In several studies, treatment of intact membranes with phospholipase C removed about 70% of the head groups without disrupting the "railroad track" structure of the membrane.

(g) Singer described in his article a study in which "a glycoprotein of molecular weight about 31,000 in human red blood cell membranes is cleaved by tryptic treatment of the membranes into soluble glycopeptides of about 10,000 molecular weight, while the remaining portions are quite hydrophobic" (p. 199). Trypsin treatment did not cause gross changes in the membranes, which remained intact.

Singer's review also included many more studies in this area. In the end, though, the data available in 1971 did not conclusively prove Model C was correct. As more data have accumulated, this model of membrane structure has been accepted by the scientific community.

Reference

Singer, S.J. (1971) The molecular organization of biological membranes. In *Structure and Function of Biological Membranes* (Rothfield, L.I., ed.), pp. 145–222, Academic Press, Inc., New York.

12

Biosignaling

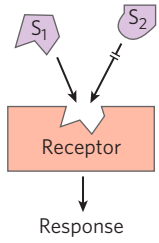
The ability of cells to receive and act on signals from beyond the plasma membrane is fundamental to life. Bacterial cells receive constant input from membrane proteins that act as information receptors, sampling the surrounding medium for pH, osmotic strength, the availability of food, oxygen, and light, and the presence of noxious chemicals, predators, or competitors for food. These signals elicit appropriate responses, such as motion toward food or away from toxic substances or the formation of dormant spores in a nutrient-depleted medium. In multicellular organisms, cells with different functions exchange a wide variety of signals. Plant cells respond to growth hormones and to variations in sunlight. Animal cells exchange information about the concentrations of ions and glucose in extracellular fluids, the interdependent metabolic activities taking place in different tissues, and, in an embryo, the correct placement of cells during development. In all these cases, the signal represents *information* that is detected by specific receptors and converted to a cellular response, which always involves a *chemical* process. This conversion of information into a chemical change, **signal transduction**, is a universal property of living cells.

12.1 General Features of Signal Transduction

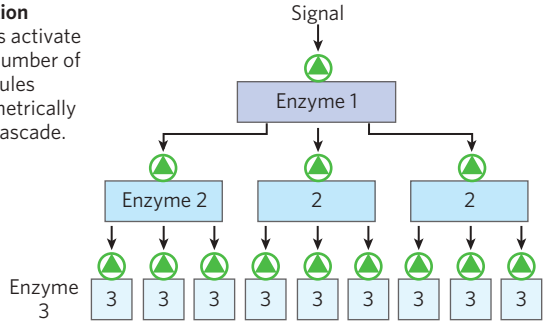
Signal transductions are remarkably specific and exquisitely sensitive. **Specificity** is achieved by precise molecular complementarity between the signal and receptor molecules **(Fig. 12–1a)**, mediated by the same kinds of weak (noncovalent) forces that mediate enzyme-substrate and antigen-antibody interactions. Multicellular organisms have an additional level of specificity, because the receptors for a given signal, or the intracellular targets of a given signal pathway, are present only in certain cell types. Thyrotropin-releasing hormone, for example, triggers responses in the cells of the anterior pituitary but not in hepatocytes, which lack receptors for this hormone. Epinephrine alters glycogen metabolism in hepatocytes but not in adipocytes; in this case, both cell types have receptors for the hormone, but whereas hepatocytes contain glycogen and the glycogen-metabolizing enzyme that is stimulated by epinephrine, adipocytes contain neither. Adipocytes respond to epinephrine by releasing fatty acids from triacylglycerols and exporting them to other tissues.

Three factors account for the extraordinary sensitivity of signal transduction: the high affinity of receptors for signal molecules, cooperativity (often but not always) in the ligand-receptor interaction, and amplification of the signal by enzyme cascades. The **affinity** between signal (ligand) and receptor can be expressed as the dissociation constant K_d, commonly 10^{-10} M or less—meaning that the receptor detects picomolar concentrations of a signal molecule. Receptor-ligand interactions are quantified by Scatchard analysis, which

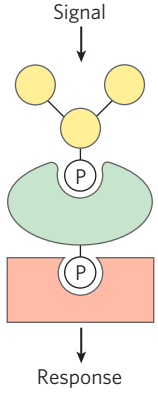
(a) Specificity
Signal molecule fits binding site on its complementary receptor; other signals do not fit.



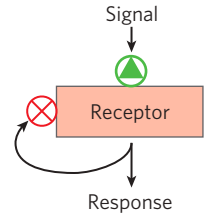
(b) Amplification
When enzymes activate enzymes, the number of affected molecules increases geometrically in an enzyme cascade.



(c) Modularity
Proteins with multivalent affinities form diverse signaling complexes from interchangeable parts. Phosphorylation provides reversible points of interaction.



(d) Desensitization/Adaptation
Receptor activation triggers a feedback circuit that shuts off the receptor or removes it from the cell surface.



(e) Integration
When two signals have opposite effects on a metabolic characteristic such as the concentration of a second messenger X, or the membrane potential V_m, the regulatory outcome results from the integrated input from both receptors.

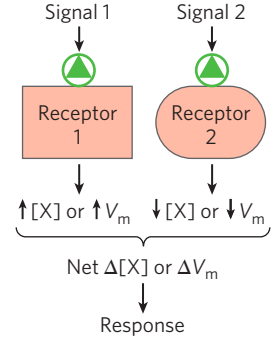


FIGURE 12–1 **Five features of signal-transducing systems.**

yields a quantitative measure of affinity (K_d) and the number of ligand-binding sites in a receptor sample (Box 12–1).

Cooperativity in receptor-ligand interactions results in large changes in receptor activation with small changes in ligand concentration (recall the effect of cooperativity on oxygen binding to hemoglobin; see Fig. 5–12). **Amplification** results when an enzyme associated with a signal receptor is activated and, in turn, catalyzes the activation of many molecules of a second enzyme, each of which activates many molecules of a third enzyme, and so on, in a so-called **enzyme cascade** (Fig. 12–1b). Such cascades can produce amplifications of several orders of magnitude within milliseconds. The response to a signal must also be terminated such that the downstream effects are in proportion to the strength of the original stimulus.

Modularity of interacting signaling proteins allows a cell to mix and match a set of signaling molecules to create complexes with different functions or cellular locations. Many signaling proteins have multiple domains that recognize specific features in other proteins, or in the cytoskeleton or plasma membrane, and the resulting multivalency of individual modules allows their assembly into a wide variety of multienzyme complexes. One common theme in such interactions is the binding of one modular signaling protein to phosphorylated residues in another protein; the resulting interaction can be regulated by phosphorylation or dephosphorylation of the protein partner (Fig. 12–1c). Nonenzymatic **scaffold proteins** with affinity for several enzymes that interact in cascades bring those proteins together, ensuring their interaction at specific cellular locations and at specific times.

The sensitivity of receptor systems is subject to modification. When a signal is present continuously, **desensitization** of the receptor system results (Fig. 12–1d); when the stimulus falls below a certain threshold, the system again becomes sensitive. Think of what happens to your visual transduction system when you walk from bright sunlight into a darkened room or from darkness into the light.

A final noteworthy feature of signal-transducing systems is **integration** (Fig. 12–1e), the ability of the system to receive multiple signals and produce a unified response appropriate to the needs of the cell or organism. Different signaling pathways converse with each other at several levels, generating complex cross talk that maintains homeostasis in the cell and the organism.

BOX 12–1 METHODS Scatchard Analysis Quantifies the Receptor-Ligand Interaction

The cellular actions of a hormone begin when the hormone (ligand, L) binds specifically and tightly to its protein receptor (R) on or in the target cell. Binding is mediated by noncovalent interactions (hydrogen-bonding, hydrophobic, and electrostatic) between the complementary surfaces of ligand and receptor. Receptor-ligand interaction brings about a conformational change that alters the biological activity of the receptor, which may be an enzyme, an enzyme regulator, an ion channel, or a regulator of gene expression.

Receptor-ligand binding is described by the equation

$$\underset{\text{Receptor}}{\text{R}} + \underset{\text{Ligand}}{\text{L}} \rightleftharpoons \underset{\text{Receptor-ligand complex}}{\text{RL}}$$

This binding, like that of an enzyme to its substrate, depends on the concentrations of the interacting components and can be described by an equilibrium constant:

$$\underset{\text{Receptor}}{\text{R}} + \underset{\text{Ligand}}{\text{L}} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} \underset{\text{Receptor-ligand complex}}{\text{RL}}$$

$$K_a = \frac{[\text{RL}]}{[\text{R}][\text{L}]} = \frac{k_{+1}}{k_{-1}} = 1/K_d$$

where K_a is the association constant and K_d is the dissociation constant.

Like enzyme-substrate binding, receptor-ligand binding is saturable. As more ligand is added to a fixed amount of receptor, an increasing fraction of receptor molecules is occupied by ligand (Fig. 1a). A rough measure of receptor-ligand affinity is given by the concentration of ligand needed to give half-saturation of the receptor. Using **Scatchard analysis** of receptor-ligand binding, we can estimate both the dissociation constant K_d and the number of receptor-binding sites in a given preparation. When binding has reached equilibrium, the total number of possible binding sites, B_{max}, equals the number of unoccupied sites, represented by [R], plus the number of occupied or ligand-bound sites, [RL]; that is, $B_{max} = [\text{R}] + [\text{RL}]$. The number of unbound sites can be expressed in terms of total sites minus occupied sites: $[\text{R}] = B_{max} - [\text{RL}]$. The equilibrium expression can now be written

$$K_a = \frac{[\text{RL}]}{[\text{L}](B_{max} - [\text{RL}])}$$

Rearranging to obtain the ratio of receptor-bound ligand to free (unbound) ligand, we get

$$\frac{[\text{Bound}]}{[\text{Free}]} = \frac{[\text{RL}]}{[\text{L}]} = K_a(B_{max} - [\text{RL}])$$
$$= \frac{1}{K_d}(B_{max} - [\text{RL}])$$

From this slope-intercept form of the equation, we can see that a plot of [bound ligand]/[free ligand] versus [bound ligand] should give a straight line with a slope of $-K_a$ ($-1/K_d$) and an intercept on the abscissa of B_{max}, the total number of binding sites (Fig. 1b). Hormone-ligand interactions typically have K_d values of 10^{-9} to 10^{-11} M, corresponding to very tight binding.

Scatchard analysis is reliable for the simplest cases, but as with Lineweaver-Burk plots for enzymes, when the receptor is an allosteric protein, the plots deviate from linearity.

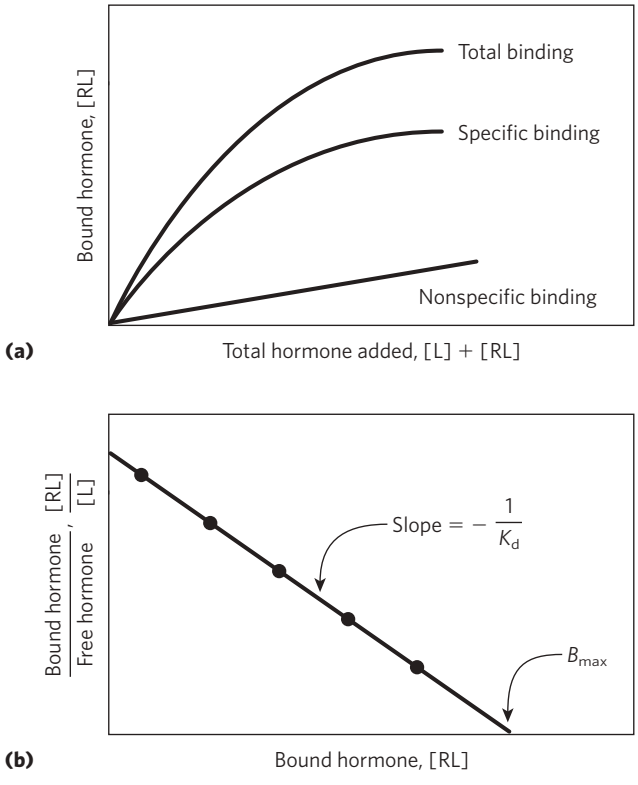


FIGURE 1 Scatchard analysis of a receptor-ligand interaction. A radiolabeled ligand (L)—a hormone, for example—is added at several concentrations to a fixed amount of receptor (R), and the fraction of the hormone bound to receptor is determined by separating the receptor-hormone complex (RL) from free hormone.

(a) A plot of [RL] versus [L] + [RL] (total hormone added) is hyperbolic, rising toward a maximum for [RL] as the receptor sites become saturated. To control for nonsaturable, nonspecific binding sites (eicosanoid hormones bind nonspecifically to the lipid bilayer, for example), a separate series of binding experiments is also necessary. A large excess of unlabeled hormone is added along with the dilute solution of labeled hormone. The unlabeled molecules compete with the labeled molecules for specific binding to the saturable site on the receptor, but not for the nonspecific binding. The true value for specific binding is obtained by subtracting nonspecific binding from total binding.

(b) A linear plot of [RL]/[L] versus [RL] gives K_d and B_{max} for the receptor-hormone complex. Compare these plots with those of V_0 versus [S] and $1/V_0$ versus 1/[S] for an enzyme-substrate complex (see Fig. 6–12, Box 6–1).

TABLE 12–1 Some Signals to Which Cells Respond

Antigens	Light
Cell surface glycoproteins/ oligosaccharides	Mechanical touch
Developmental signals	Microbial, insect pathogens
Extracellular matrix components	Neurotransmitters
Growth factors	Nutrients
Hormones	Odorants
Hypoxia	Pheromones
	Tastants

One of the revelations of research on signaling is the remarkable degree to which signaling mechanisms have been conserved during evolution. Although the number of different biological signals (Table 12–1) is probably in the thousands, and the kinds of response elicited by these signals are comparably numerous, the machinery for transducing all of these signals is built from about 10 basic types of protein components.

In this chapter we examine some examples of the major classes of signaling mechanisms, looking at how they are integrated in specific biological functions such as the transmission of nerve signals; responses to hormones and growth factors; the senses of sight, smell, and taste; and control of the cell cycle. Often, the end result of a signaling pathway is the phosphorylation of a few specific target-cell proteins, which changes their activities and thus the activities of the cell. Throughout our discussion we emphasize the conservation of fundamental mechanisms for the transduction of biological signals and the adaptation of these basic mechanisms to a wide range of signaling pathways.

We consider the molecular details of several representative signal-transduction systems, classified according to the type of receptor. The trigger for each system is different, but the general features of signal transduction are common to all: a signal interacts with a receptor; the activated receptor interacts with cellular machinery, producing a second signal or a change in the activity of a cellular protein; the metabolic activity of the target cell undergoes a change; and finally, the transduction event ends. To illustrate these general features of signaling systems, we will look at examples of six basic receptor types **(Fig. 12–2)**.

1. *G protein–coupled receptors* that *indirectly* activate (through GTP-binding proteins, or G proteins) enzymes that generate intracellular second messengers. This type of receptor is illustrated by the β-adrenergic receptor system that detects epinephrine (adrenaline) (Section 12.2).
2. *Receptor tyrosine kinases*, plasma membrane receptors that are also enzymes. When one of these receptors is activated by its extracellular

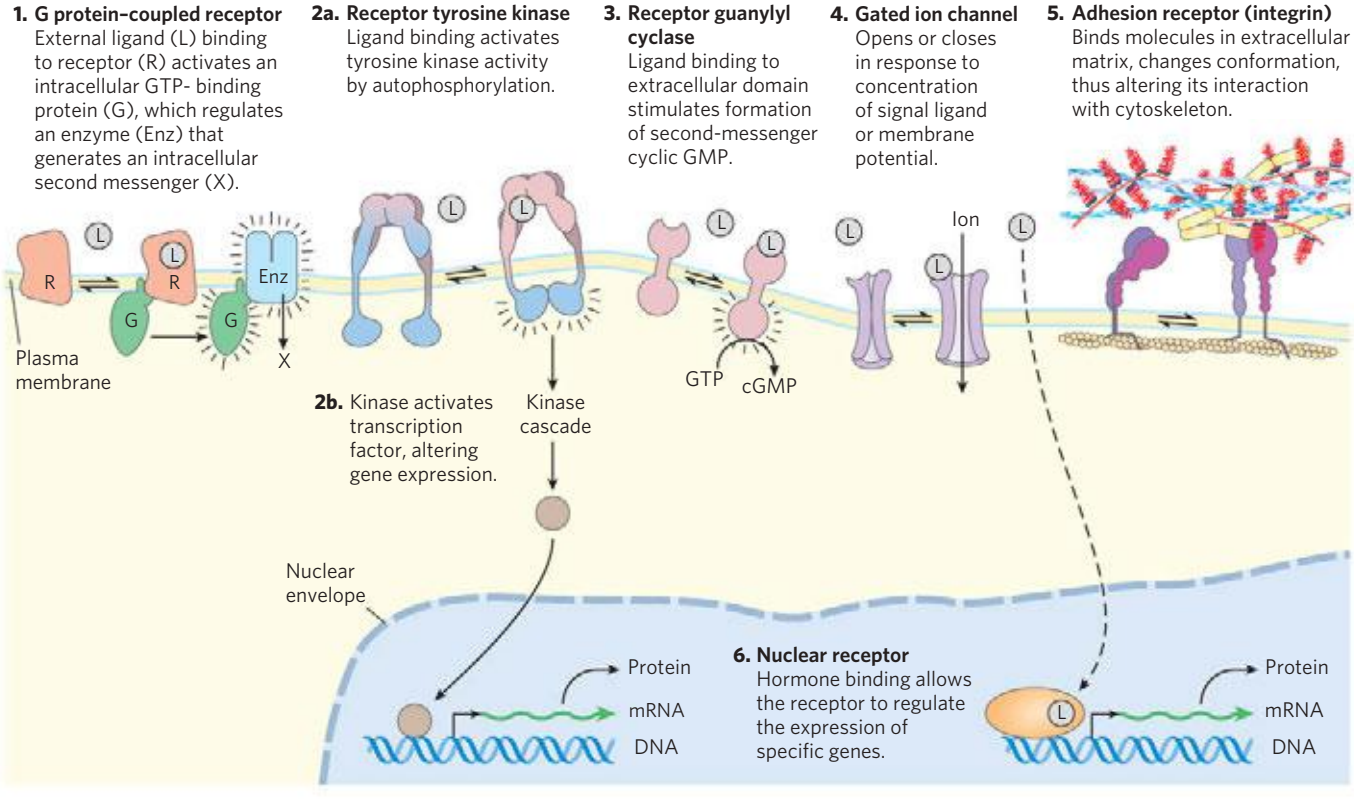


FIGURE 12–2 Six general types of signal transducers.

ligand, it catalyzes the phosphorylation of several cytosolic or plasma membrane proteins. The insulin receptor is one example (Section 12.3); the receptor for epidermal growth factor (EGFR) is another.

3. *Receptor guanylyl cyclases*, which are also plasma membrane receptors with an enzymatic cytoplasmic domain. The intracellular second messenger for these receptors, cyclic guanosine monophosphate (cGMP), activates a cytosolic protein kinase that phosphorylates cellular proteins and thereby changes their activities (Section 12.4).
4. *Gated ion channels* of the plasma membrane that open and close (hence the term "gated") in response to the binding of chemical ligands or changes in transmembrane potential. These are the simplest signal transducers. The acetylcholine receptor ion channel is an example of this mechanism (Section 12.6).
5. *Adhesion receptors* that interact with macromolecular components of the extracellular matrix (such as collagen) and convey instructions to the cytoskeletal system about cell migration or adherence to the matrix. Integrins illustrate this general type of transduction mechanism (Section 12.7).
6. *Nuclear receptors* that bind specific ligands (such as the hormone estrogen) and alter the rate at which specific genes are transcribed and translated into cellular proteins. Because steroid hormones function through mechanisms intimately related to the regulation of gene expression, we consider them here only briefly (Section 12.8) and defer a detailed discussion of their action until Chapter 28.

As we begin this discussion of biological signaling, a word about the nomenclature of signaling proteins is in order. These proteins are typically discovered in one context and named accordingly, then prove to be involved in a broader range of biological functions for which the original name is not helpful. For example, the retinoblastoma protein, pRb, was initially identified as the site of a mutation that contributes to cancer of the retina (retinoblastoma), but it is now known to function in many pathways essential to cell division in all cells, not just those of the retina. Some genes and proteins are given noncommittal names: the tumor suppressor protein p53, for example, is a *p*rotein of *53* kDa, but its name gives no clue to its great importance in the regulation of cell division and the development of cancer. In this chapter we generally define these protein names as we encounter them, introducing the names commonly used by researchers in the field. Don't be discouraged if you can't get them all straight the first time you encounter them!

SUMMARY 12.1 General Features of Signal Transduction

- All cells have specific and highly sensitive signal-transducing mechanisms, which have been conserved during evolution.
- A wide variety of stimuli act through specific protein receptors in the plasma membrane.
- The receptors bind the signal molecule and initiate a process that amplifies the signal, integrates it with input from other receptors, and transmits the information throughout the cell. If the signal persists, receptor desensitization reduces or ends the response.
- Multicellular organisms have six general types of signaling mechanisms: plasma membrane proteins that act through G proteins, receptor tyrosine kinases, receptor guanylyl cyclases that act through a protein kinase, gated ion channels, adhesion receptors that carry information between the extracellular matrix and the cytoskeleton, and nuclear receptors that bind steroids and alter gene expression.

12.2 G Protein–Coupled Receptors and Second Messengers

As their name implies, **G protein–coupled receptors (GPCRs)** are receptors that are closely associated with a member of the **guanosine nucleotide–binding protein (G protein)** family. Three essential components define signal transduction through GPCRs: a plasma membrane receptor with seven transmembrane helical segments, a G protein that cycles between active (GTP-bound) and inactive (GDP-bound) forms, and an effector enzyme (or ion channel) in the plasma membrane that is regulated by the activated G protein. The G protein, stimulated by the activated receptor, exchanges bound GDP for GTP, then dissociates from the occupied receptor and binds to the nearby effector enzyme, altering its activity. The activated enzyme then generates a **second messenger** that affects downstream targets. The human genome encodes about 350 GPCRs for detecting hormones, growth factors, and other endogenous ligands, and perhaps 500 that serve as olfactory (smell) and gustatory (taste) receptors.

GPCRs have been implicated in many common human diseases, including allergies, depression, blindness, diabetes, and various cardiovascular defects with serious health consequences. Close to half of *all* drugs on the market target one GPCR or another. For example, the β-adrenergic receptor, which mediates the effects of epinephrine, is the target of the "beta blockers," prescribed for such diverse conditions as hypertension, cardiac arrhythmia, glaucoma, anxiety, and migraine headache. At least 150 of the GPCRs found in the human genome are still "orphan receptors": their natural ligands

are not yet identified, and so we know nothing about their biology. The β-adrenergic receptor, with well-understood biology and pharmacology, is the prototype for all GPCRs, and our discussion of signal-transducing systems begins there. ■

The β-Adrenergic Receptor System Acts through the Second Messenger cAMP

Epinephrine sounds the alarm when some threat requires the organism to mobilize its energy-generating machinery; it signals the need to fight or flee. Epinephrine action begins when the hormone binds to a protein receptor in the plasma membrane of an epinephrine-sensitive cell. **Adrenergic receptors** ("adrenergic" reflects the alternative name for epinephrine, adrenaline) are of four general types, α_1, α_2, β_1, and β_2, defined by differences in their affinities and responses to a group of agonists and antagonists. **Agonists** are structural analogs that bind to a receptor and mimic the effects of its natural ligand; **antagonists** are analogs that bind the receptor without triggering the normal effect and thereby block the effects of agonists, including the biological ligand. In some cases, the affinity of the synthetic agonist or antagonist for the receptor is greater than that of the natural agonist **(Fig. 12–3)**. The four types of adrenergic receptors are found in different target tissues and mediate different responses to epinephrine. Here we focus on the ***β*-adrenergic receptors** of muscle, liver, and adipose tissue. These receptors mediate changes in fuel metabolism, as described in Chapter 23, including

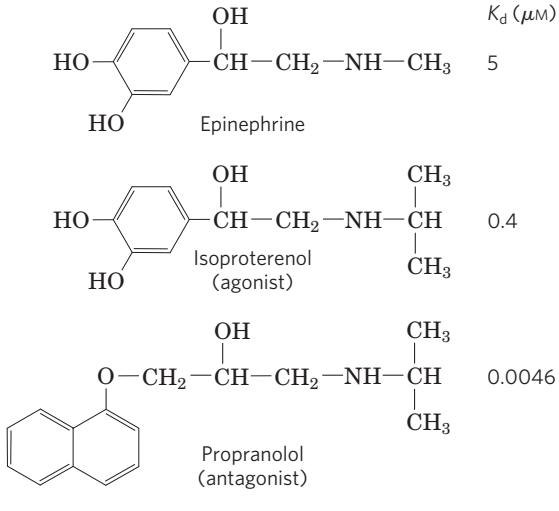


FIGURE 12–3 Epinephrine and its synthetic analogs. Epinephrine, also called adrenaline, is released from the adrenal gland and regulates energy-yielding metabolism in muscle, liver, and adipose tissue. It also serves as a neurotransmitter in adrenergic neurons. Its affinity for its receptor is expressed as a dissociation constant for the receptor-ligand complex. Isoproterenol and propranolol are synthetic analogs, one an agonist with an affinity for the receptor that is higher than that of epinephrine, and the other an antagonist with extremely high affinity.

the increased breakdown of glycogen and fat. Adrenergic receptors of the β_1 and β_2 subtypes act through the same mechanism, so in our discussion, "β-adrenergic" applies to both types.

Like all GPCRs, the β-adrenergic receptor is an integral protein with seven hydrophobic, helical regions of 20 to 28 amino acid residues that span the plasma membrane seven times, thus the alternative name for GPCRs: **heptahelical receptors**. The binding of epinephrine to a site on the receptor deep within the plasma membrane (**Fig. 12–4a**, step ❶) promotes a conformational change in the receptor's intracellular domain that affects its interaction with an associated G protein, promoting the dissociation of GDP and the binding of GTP (step ❷). For all GPCRs, the G protein is heterotrimeric, composed of three different subunits: α, β, and γ. Such G proteins are therefore known as **trimeric G proteins**. In this case, it is the α subunit that binds GDP or GTP and transmits the signal from the activated receptor to the effector protein. Because this G protein activates its effector, it is referred to as a **stimulatory G protein**, or $\mathbf{G_s}$. Like other G proteins (Box 12–2), G_s functions as a biological "switch": when the nucleotide-binding site of G_s (on the α subunit) is occupied by GTP, G_s is turned on and can activate its effector protein (adenylyl cyclase in the present case); with GDP bound to the site, G_s is switched off. In the active form, the β and γ subunits of G_s dissociate from the α subunit as a βγ dimer, and $G_{s\alpha}$, with its bound GTP, moves in the plane of the membrane from the receptor to a nearby molecule of adenylyl cyclase (step ❸). $G_{s\alpha}$ is held to the membrane by a covalently attached palmitoyl group (see Fig. 11–15).

Adenylyl cyclase is an integral protein of the plasma membrane, with its active site on the cytoplasmic face. The association of active $G_{s\alpha}$ with adenylyl cyclase stimulates the cyclase to catalyze cAMP synthesis from ATP (Fig. 12–4a, step ❹, and Fig. 12–4b), raising the cytosolic [cAMP]. The interaction between $G_{s\alpha}$ and adenylyl cyclase is possible only when $G_{s\alpha}$ is bound to GTP.

The stimulation by $G_{s\alpha}$ is self-limiting; *$G_{s\alpha}$ has intrinsic GTPase activity that inactivates $G_{s\alpha}$* by converting its bound GTP to GDP **(Fig. 12–5)**. The now inactive $G_{s\alpha}$ dissociates from adenylyl cyclase, rendering the cyclase inactive. $G_{s\alpha}$ reassociates with the βγ dimer ($G_{s\beta\gamma}$), and inactive G_s is again available to interact with a hormone-bound receptor.

The role of $G_{s\alpha}$ in serving as a biological "switch" protein is not unique. A variety of G proteins act as binary switches in signaling systems with GPCRs and in many processes that involve membrane fusion or fission (Box 12–2). **Trimeric G Proteins: Molecular On/Off Switches**

Epinephrine exerts its downstream effects through the increase in [cAMP] that results from the activation of adenylyl cyclase. Cyclic AMP, in turn, allosterically activates **cAMP-dependent protein kinase**, also called **protein kinase A** or **PKA** (Fig. 12–4a, step ❺),

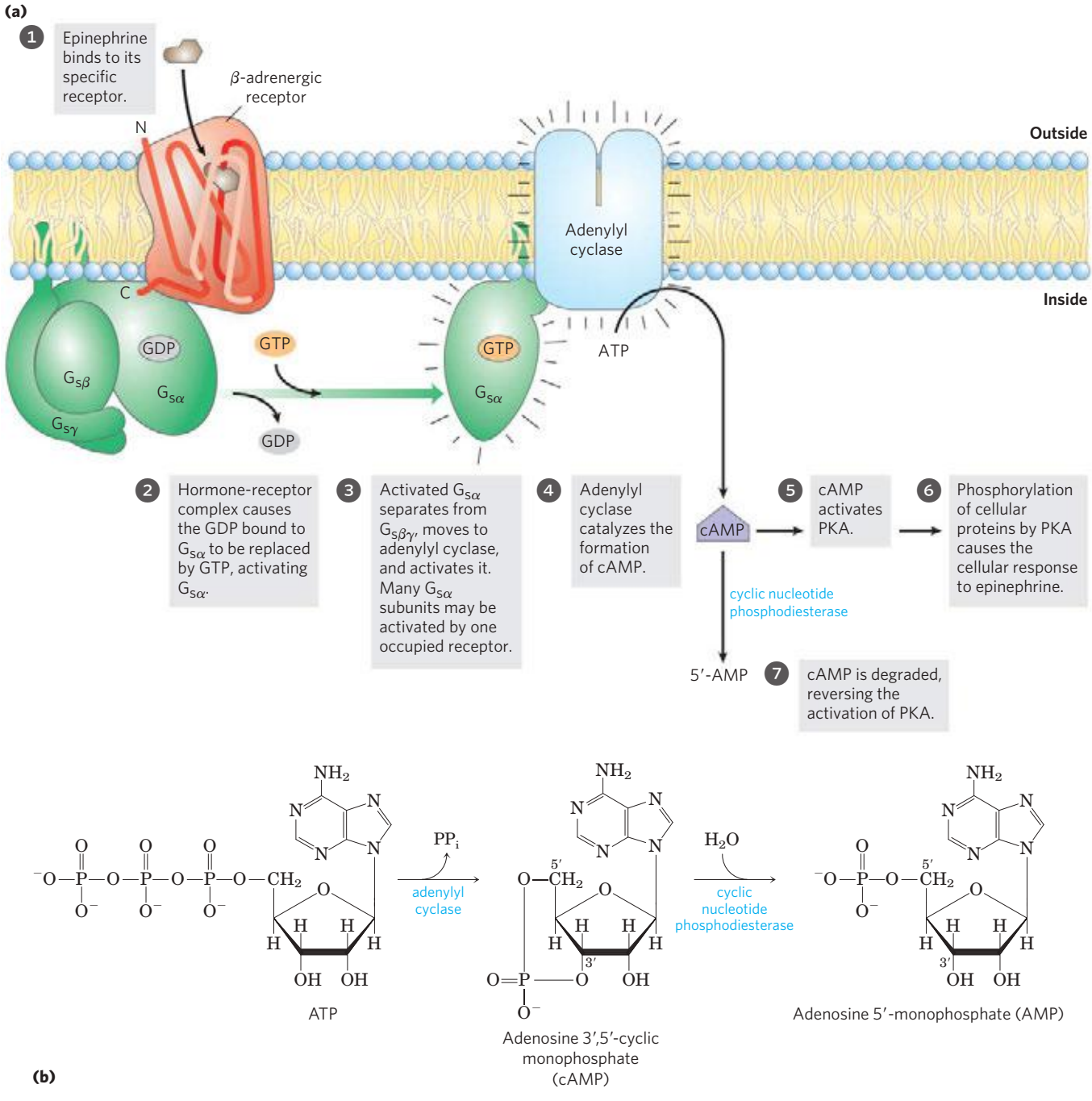


FIGURE 12–4 Transduction of the epinephrine signal: the β-adrenergic pathway. (a) The mechanism that couples binding of epinephrine to its receptor with activation of adenylyl cyclase; the seven steps are discussed further in the text. The same adenylyl cyclase molecule in the plasma membrane may be regulated by a stimulatory G protein (G_S), as shown, or an inhibitory G protein (G_i, not shown). G_S and G_i are under the influence of different hormones. Hormones that induce GTP binding to G_i cause *inhibition* of adenylyl cyclase, resulting in lower cellular [cAMP]. **(b)** The combined action of the enzymes that catalyze steps ❹ and ❼, synthesis and hydrolysis of cAMP by adenylyl cyclase and cAMP phosphodiesterase, respectively.

which catalyzes the phosphorylation of specific Ser or Thr residues of targeted proteins, including glycogen phosphorylase *b* kinase. This enzyme is active when phosphorylated and can begin the process of mobilizing glycogen stores in muscle and liver in anticipation of the need for energy, as signaled by epinephrine.

The inactive form of PKA contains two identical catalytic subunits (C) and two identical regulatory subunits (R) **(Fig. 12–6a)**. The tetrameric R_2C_2 complex is catalytically inactive, because an autoinhibitory domain of each R subunit occupies the substrate-binding cleft of each C subunit. When cAMP binds to the R subunits,

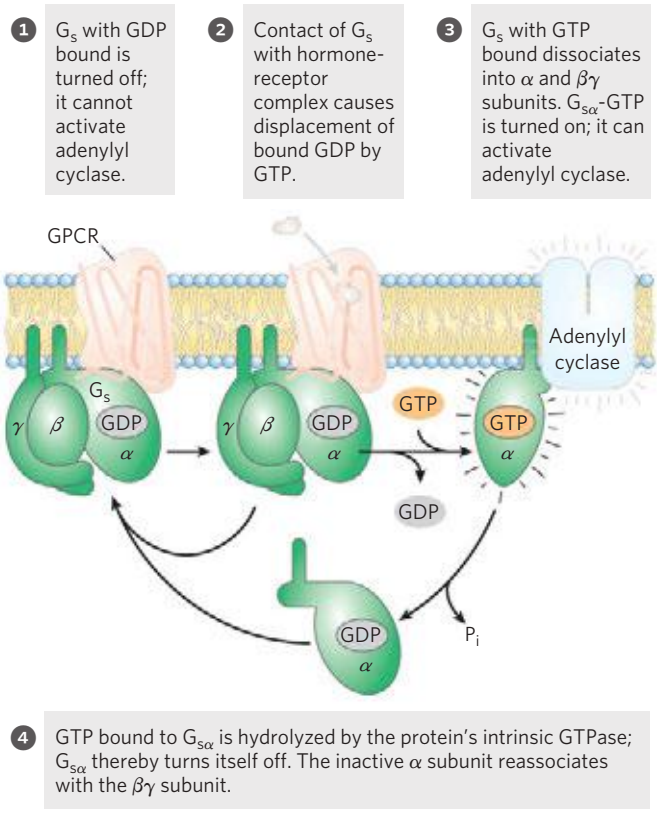


FIGURE 12–5 The GTPase switch. G proteins cycle between GDP-bound (off) and GTP-bound (on). The protein's intrinsic GTPase activity, in many cases stimulated by RGS proteins (regulators of *G*-protein signaling; see Box 12-2), determines how quickly bound GTP is hydrolyzed to GDP and thus how long the G protein remains active.

they undergo a conformational change that moves the autoinhibitory domain of R out of the catalytic domain of C, and the R_2C_2 complex dissociates to yield two free, catalytically active C subunits. This same basic mechanism—displacement of an autoinhibitory domain—mediates the allosteric activation of many types of protein kinases by their second messengers (as in Figs 12–14 and 12–22, for example). The structure of the substrate-binding cleft in PKA is the prototype for all known protein kinases (Fig. 12–6b); certain residues in this cleft region have identical counterparts in all of the more than 1,000 known protein kinases. The ATP-binding site of each catalytic subunit positions ATP perfectly for the transfer of its terminal (γ) phosphoryl group to the —OH in the side chain of a Ser or Thr residue.

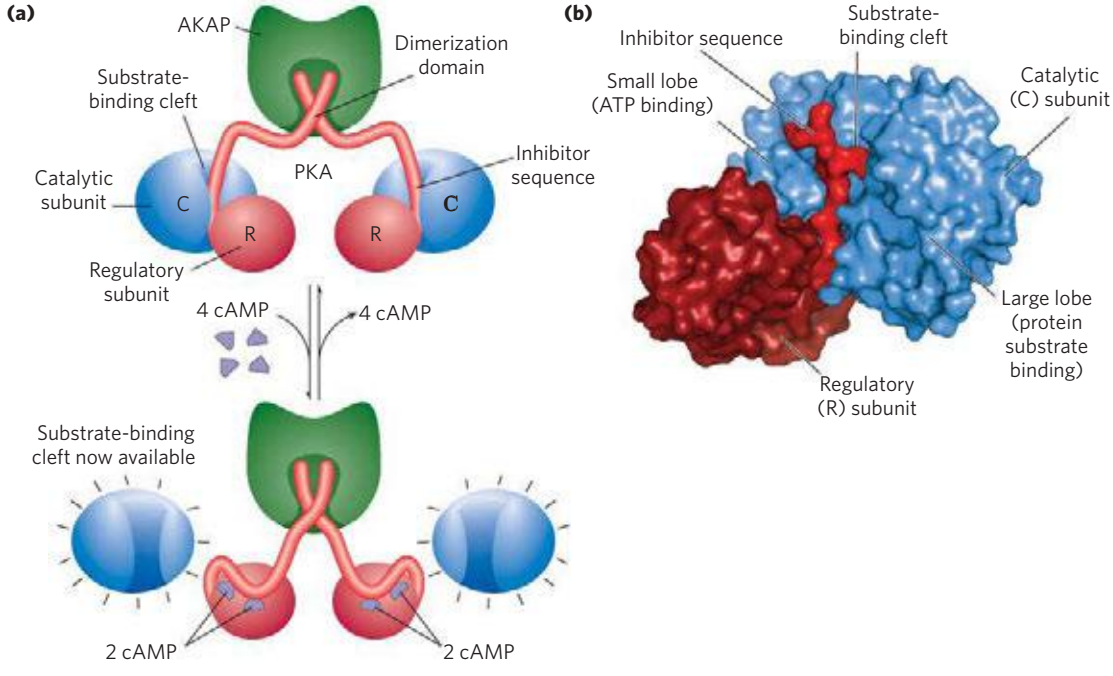


FIGURE 12–6 Activation of cAMP-dependent protein kinase (PKA). **(a)** When [cAMP] is low, the two identical regulatory subunits (R; red) associate with the two identical catalytic subunits (C). In this R_2C_2 complex, the inhibitor sequences of the R subunits lie in the substrate-binding cleft of the C subunits and prevent binding of protein substrates; the complex is therefore catalytically inactive. The amino-terminal sequences of the R subunits interact to form an R_2 dimer, the site of binding to an A kinase anchoring protein (AKAP), described later in the text. When [cAMP] rises in response to a hormonal signal, each R subunit binds two cAMP molecules and undergoes a dramatic reorganization that pulls its inhibitory sequence away from the C subunit, opening up the substrate-binding cleft and releasing each C subunit in its catalytically active form. **(b)** A crystal structure showing part of the R_2C_2 complex (PDB ID 1U7E)—one C subunit and part of one R subunit. The amino-terminal dimerization region of the R subunit is omitted for simplicity. The small lobe of C contains the ATP-binding site, and the large lobe surrounds and defines the cleft where the protein substrate binds and undergoes phosphorylation at a Ser or Thr residue, with the phosphoryl group transferred from ATP. In this inactive form, the inhibitor sequence of R blocks the substrate-binding cleft of C, inactivating it.

BOX 12–2 MEDICINE G Proteins: Binary Switches in Health and Disease

Alfred G. Gilman and Martin Rodbell (Fig. 1) discovered the critical roles of guanosine nucleotide–binding proteins (G proteins) in a wide variety of cellular processes, including sensory perception, signaling for cell division, growth and differentiation, intracellular movements of proteins and membrane vesicles, and protein synthesis. The human genome encodes nearly 200 of these proteins, which differ in size and subunit structure, intracellular location, and function. But all G proteins share a common feature: they can become activated and then, after a brief period, can inactivate themselves, thereby serving as molecular binary switches with built-in timers. This superfamily of proteins includes the trimeric G proteins involved in adrenergic signaling (G_s and G_i) and vision (transducin); small G proteins such as that involved in insulin signaling (Ras) and others that function in vesicle trafficking (ARF and Rab), transport into and out of the nucleus (Ran; see Fig. 27–42), and timing of the cell cycle (Rho); and several proteins involved in protein synthesis (initiation factor IF2 and elongation factors EF-Tu and EF-G; see Chapter 26). Many G proteins have covalently bound lipids, which give them an affinity for membranes and dictate their locations in the cell.

All G proteins have the same core structure and use the same mechanism for switching between an inactive conformation, favored when GDP is bound, and an active conformation, favored when GTP is bound. We can use the Ras protein (~20 kDa), a minimal signaling unit, as a prototype for all members of this superfamily (Fig. 2).

In the GTP-bound conformation, the G protein exposes previously buried regions (called **switch I** and **switch II**) that interact with proteins downstream in the signaling pathway, until the G protein inactivates itself by hydrolyzing its bound GTP to GDP. The critical determinant of G-protein conformation is the γ phosphate of GTP, which interacts with a region called the **P loop** (*p*hosphate-binding; Fig. 3). In Ras, the γ phosphate of GTP binds to a Lys residue in the P loop and to two critical residues, Thr^{35} in switch I and Gly^{60} in switch II, that hydrogen-bond with the oxygens of the γ phosphate of GTP. These

(*Continued on next page*)

FIGURE 1 Alfred G. Gilman (left) and Martin Rodbell (1925–1998). Their Nobel lectures on the discovery and exploration of G proteins are available at www.nobelprize.org.

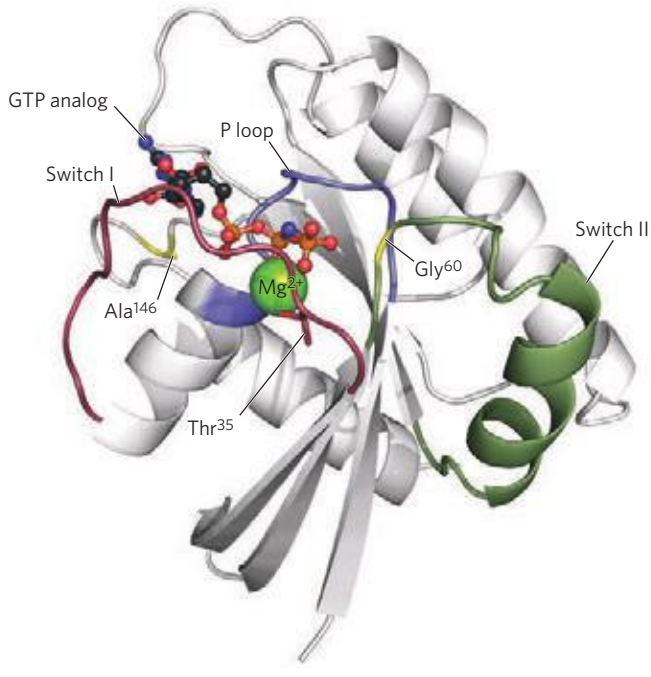


FIGURE 2 The Ras protein, the prototype for all G proteins (PDB ID 5P21). Mg^{2+}-GTP is held by critical residues in the phosphate-binding P loop (blue), and by Thr^{35} in the switch I (red) and Gly^{60} in the switch II (green) regions. Ala^{146} gives specificity for GTP over ATP. In this structure, the nonhydrolyzable GTP analog Gpp(NH)p is in the GTP-binding site.

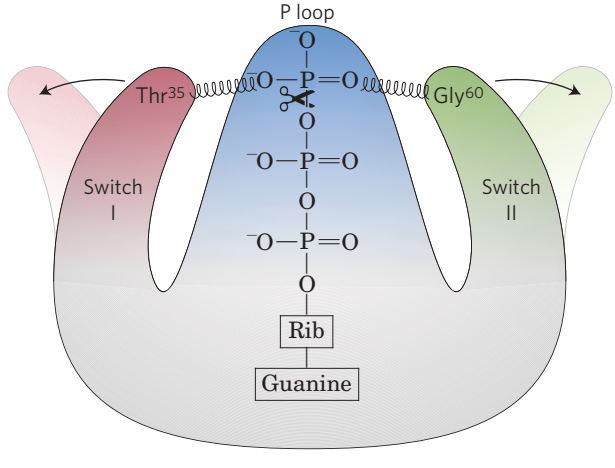


FIGURE 3 When bound GTP is hydrolyzed by the GTPase activities of Ras and its GAP, loss of hydrogen bonds to Thr^{35} and Gly^{60} allows the switch I and switch II regions to relax into a conformation in which they are no longer available to interact with downstream targets such as Raf.

BOX 12–2 MEDICINE G Proteins: Binary Switches in Health and Disease (*Continued*)

hydrogen bonds act like a pair of springs holding the protein in its active conformation. When GTP is cleaved to GDP and P_i is released, these hydrogen bonds are lost; the protein relaxes into its inactive conformation, burying the sites that interact with other partners in its active state. Ala^{146} hydrogen-bonds to the guanine oxygen, allowing GTP, but not ATP, to bind.

The intrinsic GTPase activity of G proteins is increased up to 10^5-fold by **GTPase activator proteins (GAPs)**, also called, in the case of heterotrimeric G proteins, **regulators of G protein signaling (RGSs**; Fig. 4). GAPs (and RGSs) thus determine how long the switch remains "on." They contribute a critical Arg residue that reaches into the G-protein GTPase active site and assists in catalysis. The intrinsically slow process of replacing bound GDP with GTP, switching the protein on, is catalyzed by **guanosine nucleotide–exchange factors (GEFs)** associated with the G protein (Fig. 4).

Because G proteins play crucial roles in so many signaling processes, it is not surprising that defects in G proteins lead to a variety of diseases. In about 25% of all human cancers (and in a much higher proportion of certain types of cancer), there is a mutation in a Ras protein—typically in one of the critical residues around the GTP-binding site or in the P loop—that virtually eliminates its GTPase activity. Once activated by GTP binding, this Ras protein remains constitutively active, promoting cell division in cells that should not divide. The tumor suppressor gene *NF1* encodes a GAP that enhances the GTPase activity of normal Ras. Mutations in *NF1* that result in a nonfunctioning GAP leave Ras with only its intrinsic GTPase activity, which is very weak (has a very low turnover number); once activated by GTP binding, Ras stays active for an extended period, continuing to send the signal: divide.

Defective heterotrimeric G proteins can also lead to disease. Mutations in the gene that encodes the α subunit of G_s (which mediates changes in [cAMP] in response to hormonal stimuli) may result in a G_α that is permanently active or permanently inactive. "Activating" mutations generally occur in residues crucial to GTPase activity; they lead to a continuously elevated [cAMP], with significant downstream consequences, including undesirable cell proliferation. For example,

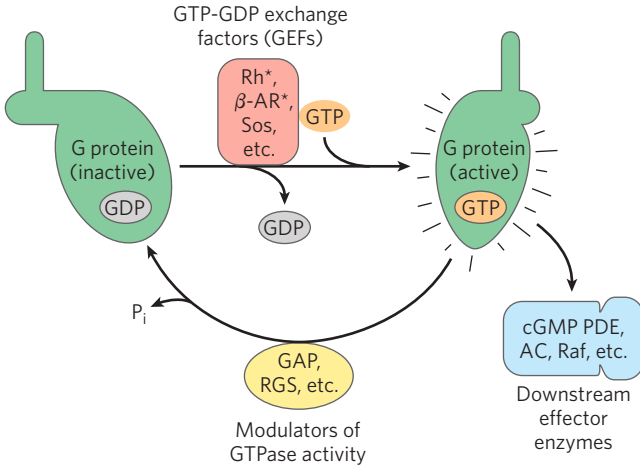


FIGURE 4 Factors that regulate the activity of G proteins (green). Inactive G proteins (both small G proteins such as Ras and heterotrimeric G proteins such as G_s) interact with upstream GTP-GDP exchange factors, GEFs (red; often these are activated (*) receptors such as rhodopsin, β-adrenergic receptors, and Sos), and are activated by GTP binding. In this form, the G proteins activate downstream effector enzymes (blue; enzymes such as cGMP phosphodiesterase, adenylyl cyclase, and Raf). GTPase activator proteins (GAPs, in the case of small G proteins) and regulators of G protein signaling (RGSs) (yellow), by modulating the GTPase activity of G proteins, determine how long the G protein will remain active.

such mutations are found in about 40% of pituitary tumors (adenomas). Individuals with "inactivating" mutations in G_α are unresponsive to hormones (such as thyroid hormone) that act through cAMP. Mutation in the gene for the transducin α subunit (T_α), which is involved in visual signaling, leads to a type of night blindness, apparently due to defective interaction between the activated T_α subunit and the phosphodiesterase of the rod outer segment (see Fig. 12–39). A sequence variation in the gene encoding the β subunit of a heterotrimeric G protein is commonly found in individuals with hypertension (high blood pressure), and this variant gene is suspected of involvement in obesity and atherosclerosis.

The pathogenic bacteria that cause cholera and pertussis (whooping cough) produce toxins that target G proteins, interfering with normal signaling in host cells. **Cholera toxin**, secreted by *Vibrio cholerae* in the intestine of an infected person, is a heterodimeric protein. Subunit B recognizes and binds to specific

As indicated in Figure 12–4a (step 6), PKA regulates several enzymes downstream in the signaling pathway (Table 12–2). Although these downstream targets have diverse functions, they share a region of sequence similarity around the Ser or Thr residue that undergoes phosphorylation, a sequence that marks them for regulation by PKA. The substrate-binding cleft of PKA recognizes these sequences and phosphorylates their Thr or Ser residue. Comparison of the sequences of various protein substrates for PKA has

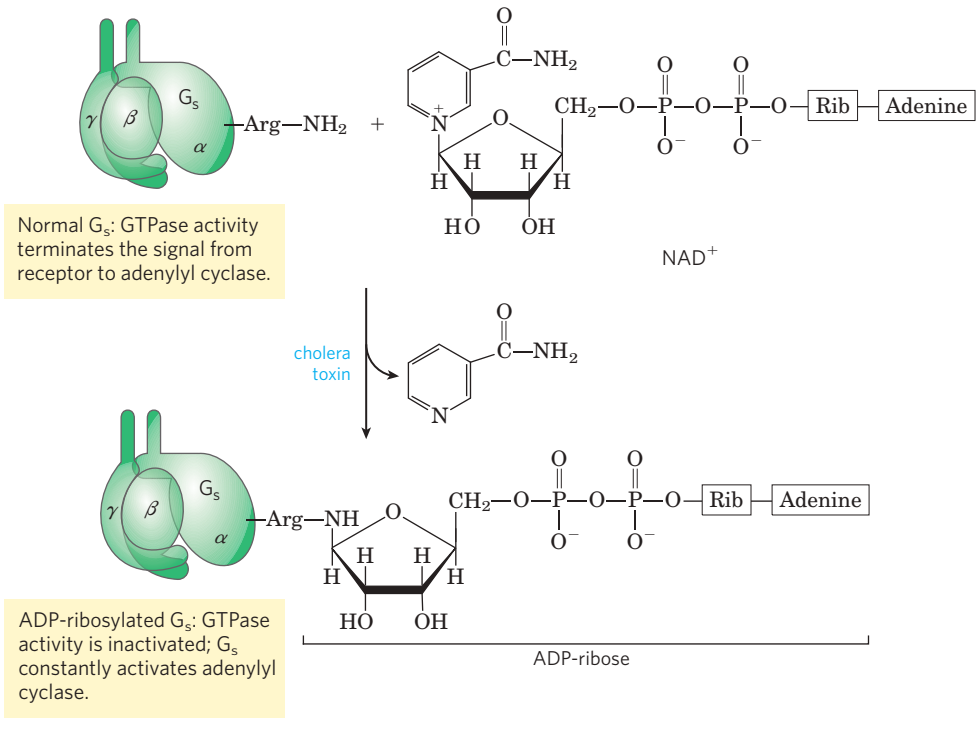


FIGURE 5 The bacterial toxins that cause cholera and whooping cough (pertussis) are enzymes that catalyze transfer of the ADP-ribose moiety of NAD^+ to an Arg residue of G_s (in the case of cholera toxin, as shown here) or a Cys residue of G_i (pertussis toxin). The G proteins thus modified fail to respond to normal hormonal stimuli. The pathology of both diseases results from defective regulation of adenylyl cyclase and overproduction of cAMP.

gangliosides on the surface of intestinal epithelial cells and provides a route for subunit A to enter these cells. After entry, subunit A is broken into two pieces: the A1 fragment and the A2 fragment. A1 then associates with the ADP-ribosylation factor ARF6, a small G protein in host cells, through residues in its switch I and switch II regions—which are accessible only when ARF6 is in its active (GTP-bound) form. This association with ARF6 activates A1, which catalyzes the transfer of ADP-ribose from NAD^+ to the critical Arg residue in the P loop of the α subunit of G_s (Fig. 5). ADP-ribosylation blocks the GTPase activity of G_s and thereby renders G_s permanently active. This results in continuous activation of the adenylyl cyclase of intestinal epithelial cells, chronically high [cAMP], and chronically active PKA. PKA phosphorylates the CFTR Cl^- channel (see Box 11–2) and a Na^+-H^+ exchanger in the intestinal epithelial cells. The resultant efflux of NaCl triggers massive water loss through the intestine as cells respond to the ensuing osmotic imbalance. Severe dehydration and electrolyte loss are the major pathologies in cholera. These can be fatal in the absence of prompt rehydration therapy.

The **pertussis toxin**, produced by *Bordetella pertussis*, catalyzes ADP-ribosylation of the α subunit of G_i, in this case preventing GDP-GTP exchange and blocking inhibition of adenylyl cyclase by G_i. The bacterium infects the respiratory tract, where it destroys the ciliated epithelial cells that normally sweep away mucus. Without this ciliary action, vigorous coughing is needed to clear the tract; this is the gasping cough that gives the disease its name (and spreads the bacterium to others). How the defect in G-protein signaling kills ciliated epithelial cells is not yet clear.

Given the large number of G protein–coupled receptors in the human genome, it seems likely that future studies will reveal many more examples of how defective G-protein signaling affects human health.

yielded the **consensus sequence**—the neighboring residues needed to mark a Ser or Thr residue for phosphorylation (see Table 12–2).

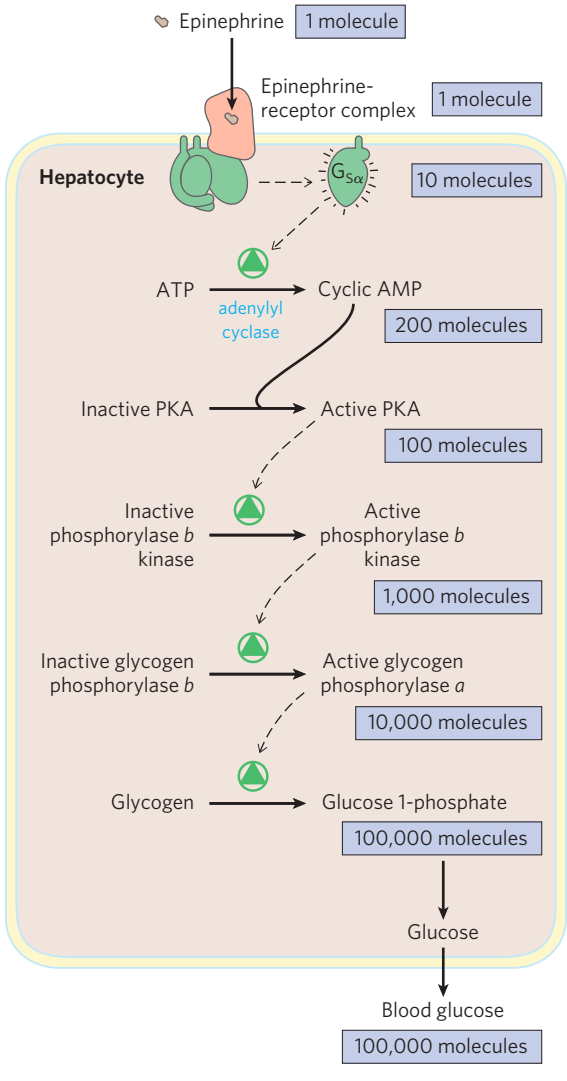
As in many signaling pathways, signal transduction by adenylyl cyclase entails several steps that *amplify* the original hormone signal **(Fig. 12–7)**. First, the binding of one hormone molecule to one receptor molecule catalytically activates many G_s molecules that associate with the activated receptor, one after the other. Next, by activating one molecule of adenylyl

TABLE 12–2 Some Enzymes and Other Proteins Regulated by cAMP-Dependent Phosphorylation (by PKA)

Enzyme/protein	Sequence phosphorylated*	Pathway/process regulated
Glycogen synthase	RA**S**CTSSS	Glycogen synthesis
Phosphorylase *b* kinase α subunit β subunit	 VEFRRL**S**I RTKR**S**GSV	Glycogen breakdown
Pyruvate kinase (rat liver)	GVLRRA**S**VAZL	Glycolysis
Pyruvate dehydrogenase complex (type L)	GYLRRA**S**V	Pyruvate to acetyl-CoA
Hormone-sensitive lipase	PMRR**S**V	Triacylglycerol mobilization and fatty acid oxidation
Phosphofructokinase-2/fructose 2,6-bisphosphatase	LQRRRG**S**SIPQ	Glycolysis/gluconeogenesis
Tyrosine hydroxylase	FIGRRQ**S**L	Synthesis of L-dopa, dopamine, norepinephrine, and epinephrine
Histone H1	AKRKA**S**GPPVS	DNA condensation
Histone H2B	KKAKA**S**RKESYSVYVYK	DNA condensation
Cardiac phospholamban (cardiac pump regulator)	AIRRA**S**T	Intracellular [Ca^{2+}]
Protein phosphatase-1 inhibitor-1	IRRRRP**T**P	Protein dephosphorylation
PKA consensus sequence†	xR[RK]x[**S**T]B	Many

*The phosphorylated S or T residue is shown in red. All residues are given as their one-letter abbreviations (see Table 3–1).

†x is any amino acid; B is any hydrophobic amino acid. See Box 3–2 for conventions used in displaying consensus sequences.



cyclase, each active $G_{s\alpha}$ molecule stimulates the catalytic synthesis of *many* molecules of cAMP. The second messenger cAMP now activates PKA, each molecule of which catalyzes the phosphorylation of *many* molecules of the target protein—phosphorylase *b* kinase in Figure 12–7. This kinase activates glycogen phosphorylase *b*, which leads to the rapid mobilization of glucose from glycogen. The net effect of the cascade is amplification of the hormonal signal by several orders of magnitude, which accounts for the very low concentration of epinephrine (or any other hormone) required for hormone activity.

Several Mechanisms Cause Termination of the β-Adrenergic Response

To be useful, a signal-transducing system has to *turn off* after the hormonal or other stimulus has ended, and mechanisms for shutting off the signal are intrinsic to

FIGURE 12–7 Epinephrine cascade. Epinephrine triggers a series of reactions in hepatocytes in which catalysts activate catalysts, resulting in great amplification of the original hormone signal. The numbers of molecules shown are simply to illustrate amplification and are almost certainly gross underestimates. Binding of one molecule of epinephrine to one β-adrenergic receptor on the cell surface activates a number (possibly hundreds) of G proteins, one after another, each of which goes on to activate a molecule of the enzyme adenylyl cyclase. Adenylyl cyclase acts catalytically, producing many molecules of cAMP for each activated adenylyl cyclase. (Because two molecules of cAMP are required to activate one PKA catalytic subunit, this step does not amplify the signal.)

all signaling systems. Most systems also adapt to the continued presence of the signal by becoming less sensitive to it, in the process of desensitization. The β-adrenergic system illustrates both. When the concentration of epinephrine in the blood drops below the K_d for its receptor, the hormone dissociates from the receptor and the latter reassumes the inactive conformation, in which it can no longer activate G_s.

A second means of ending the response to β-adrenergic stimulation is the hydrolysis of GTP bound to the G_α subunit, catalyzed by the intrinsic GTPase activity of the G protein. Conversion of bound GTP to GDP favors the return of G_α to the conformation in which it binds the $G_{\beta\gamma}$ subunits—the conformation in which the G protein is unable to interact with or stimulate adenylyl cyclase. This ends the production of cAMP. The rate of inactivation of G_s depends on the GTPase activity, which for G_α alone is very feeble. However, GTPase activator proteins (GAPs) strongly stimulate this GTPase activity, causing more rapid inactivation of the G protein (see Box 12–2). GAPs can themselves be regulated by other factors, providing a fine-tuning of the response to β-adrenergic stimulation. A third mechanism for terminating the response is to remove the second messenger: hydrolysis of cAMP to 5′-AMP (not active as a second messenger) by **cyclic nucleotide phosphodiesterase** (Fig. 12–4a, step ❼; 12–4b).

Finally, at the end of the signaling pathway, the metabolic effects that result from enzyme phosphorylation are reversed by the action of phosphoprotein phosphatases, which hydrolyze phosphorylated Ser, Thr, or Tyr residues, releasing inorganic phosphate (P_i). About 150 genes in the human genome encode phosphoprotein phosphatases, fewer than the number encoding protein kinases (~500). Some of these phosphatases are known to be regulated; others may act constitutively. When [cAMP] drops and PKA returns to its inactive form (step ❼ in Fig. 12–4a), the balance between phosphorylation and dephosphorylation is tipped toward dephosphorylation by these phosphatases.

The β-Adrenergic Receptor Is Desensitized by Phosphorylation and by Association with Arrestin

The mechanisms for signal termination described above take effect when the stimulus ends. A different mechanism, desensitization, damps the response *even while the signal persists*. Desensitization of the β-adrenergic receptor is mediated by a protein kinase that phosphorylates the receptor on the intracellular domain that normally interacts with G_s **(Fig. 12–8)**. When the receptor is occupied by epinephrine, ***β*-adrenergic receptor kinase**, or ***β*ARK** (also commonly called **GRK2**; see below), phosphorylates several Ser residues

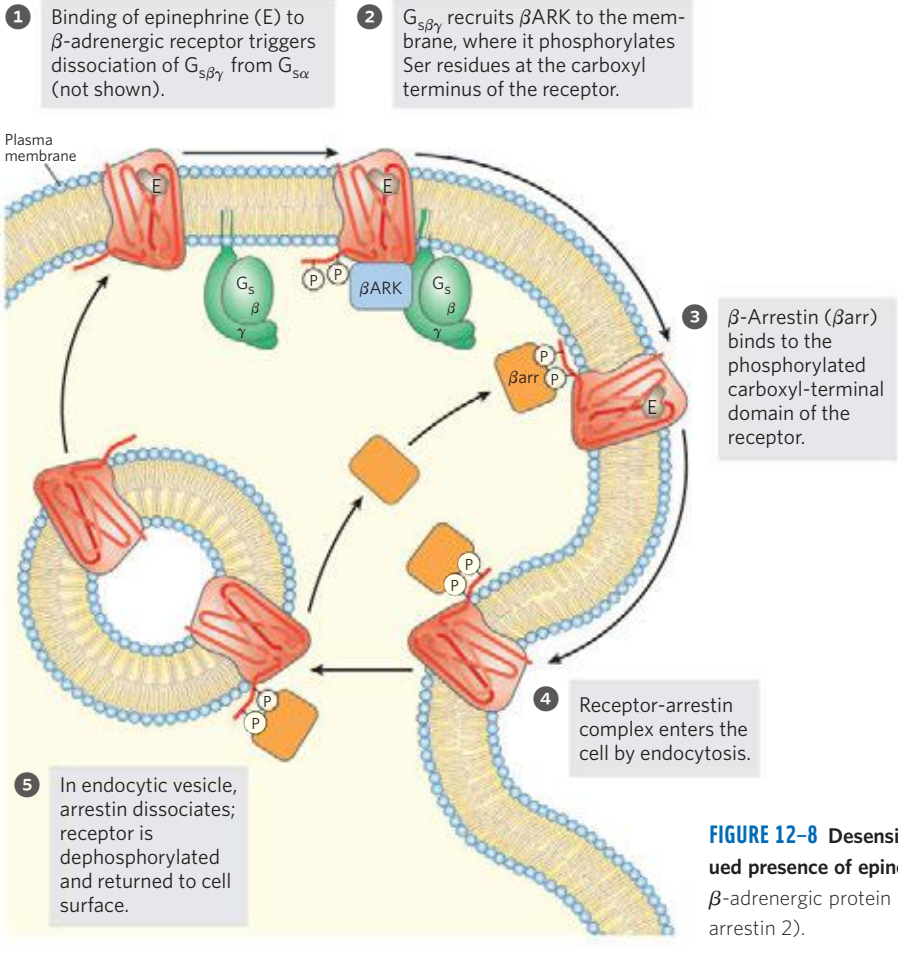


FIGURE 12–8 Desensitization of the β-adrenergic receptor in the continued presence of epinephrine. This process is mediated by two proteins: β-adrenergic protein kinase (βARK) and β-arrestin (βarr; also known as arrestin 2).

near the carboxyl terminus of the receptor, which is on the cytoplasmic side of the plasma membrane. Usually located in the cytosol, βARK is drawn to the plasma membrane by its association with the $G_{s\beta\gamma}$ subunits and is thus positioned to phosphorylate the receptor. Receptor phosphorylation creates a binding site for the protein ***β*-arrestin**, or ***β*arr** (also called arrestin 2), and binding of β-arrestin effectively prevents further interaction between the receptor and the G protein. The binding of β-arrestin also facilitates receptor sequestration, the removal of receptor molecules from the plasma membrane by endocytosis into small intracellular vesicles. The arrestin-receptor complex recruits two proteins involved in vesicle formation (see Fig. 27–45), the AP-2 complex and clathrin, which initiate membrane invagination, leading to the formation of endosomes containing the adrenergic receptor. In this state the receptors are inaccessible to epinephrine and therefore inactive. Receptors in the endocytic vesicles are eventually dephosphorylated and returned to the plasma membrane, completing the circuit and resensitizing the system to epinephrine. β-Adrenergic receptor kinase is a member of a family of **G protein–coupled receptor kinases (GRKs)**, all of which phosphorylate GPCRs on their carboxyl-terminal cytoplasmic domains and play roles similar to that of βARK in desensitization and resensitization of their receptors. At least five different GRKs and four different arrestins are encoded in the human genome; each GRK is capable of desensitizing a particular subset of GPCRs, and each arrestin can interact with many different types of phosphorylated receptors.

Cyclic AMP Acts as a Second Messenger for Many Regulatory Molecules

Epinephrine is just one of many hormones, growth factors, and other regulatory molecules that act by changing the intracellular [cAMP] and thus the activity of PKA (Table 12–3). For example, glucagon binds to its receptors in the plasma membrane of adipocytes, activating (via a G_s protein) adenylyl cyclase. PKA, stimulated by the resulting rise in [cAMP], phosphorylates and activates two proteins critical to the mobilization of the fatty acids of stored fats (see Fig. 17–3). Similarly, the peptide hormone ACTH (adrenocorticotropic hormone, also called corticotropin), produced by the anterior pituitary, binds to specific receptors in the adrenal cortex, activating adenylyl cyclase and raising the intracellular [cAMP]. PKA then phosphorylates and activates several of the enzymes required for the synthesis of cortisol and other steroid hormones. In many cell types, the catalytic subunit of PKA can also move into the nucleus, where it phosphorylates the **cAMP response element binding protein (CREB)**, which alters the expression of specific genes regulated by cAMP.

Some hormones act by *inhibiting* adenylyl cyclase, thus *lowering* [cAMP] and *suppressing* protein phosphorylation. For example, the binding of somatostatin to its receptor leads to activation of an **inhibitory G protein**, or $\mathbf{G_i}$, structurally homologous to G_s, that inhibits adenylyl cyclase and lowers [cAMP]. Somatostatin therefore counterbalances the effects of glucagon. In adipose tissue, prostaglandin E_1 (PGE_1; see Fig. 10–18) inhibits adenylyl cyclase, thus lowering [cAMP] and slowing the mobilization of lipid reserves triggered by epinephrine and glucagon. In certain other tissues PGE_1 stimulates cAMP synthesis: its receptors are coupled to adenylyl cyclase through a stimulatory G protein, G_s. In tissues with α_2-adrenergic receptors, epinephrine lowers [cAMP]; in this case, the receptors are coupled to adenylyl cyclase through an inhibitory G protein, G_i. In short, an extracellular signal such as epinephrine or PGE_1 can have quite different effects on different tissues or cell types, depending on three factors: the type of receptor in the tissue, the type of G protein (G_s or G_i) with which the receptor is coupled, and the set of PKA target enzymes in the cells. By summing the influences that tend to increase and decrease [cAMP], a cell achieves the integration of signals that we noted as a general feature of signal-transducing mechanisms (Fig. 12–1e).

Another factor that explains how so many types of signals can be mediated by a single second messenger (cAMP) is the confinement of the signaling process to a specific region of the cell by **adaptor proteins**—noncatalytic proteins that hold together other protein molecules that function in concert (further described

TABLE 12–3 Some Signals That Use cAMP as Second Messenger

Corticotropin (ACTH)
Corticotropin-releasing hormone (CRH)
Dopamine [D_1, D_2]
Epinephrine (β-adrenergic)
Follicle-stimulating hormone (FSH)
Glucagon
Histamine [H_2]
Luteinizing hormone (LH)
Melanocyte-stimulating hormone (MSH)
Odorants (many)
Parathyroid hormone
Prostaglandins E_1, E_2 (PGE_1, PGE_2)
Serotonin [5-HT-1a, 5-HT-2]
Somatostatin
Tastants (sweet, bitter)
Thyroid-stimulating hormone (TSH)

Note: Receptor subtypes in square brackets. Subtypes may have different transduction mechanisms. For example, serotonin is detected in some tissues by receptor subtypes 5-HT-1a and 5-HT-1b, which act through adenylyl cyclase and cAMP, and in other tissues by receptor subtype 5-HT-1c, acting through the phospholipase C–IP_3 mechanism (see Table 12–4).

below). **AKAPs (A kinase anchoring proteins)** are multivalent adaptor proteins; one part binds to the R subunits of PKA (see Fig. 12–6a) and another to a specific structure in the cell, confining the PKA to the vicinity of that structure. For example, specific AKAPs bind PKA to microtubules, actin filaments, ion channels, mitochondria, or the nucleus. Different types of cells have different complements of AKAPs, so cAMP might stimulate phosphorylation of mitochondrial proteins in one cell and phosphorylation of actin filaments in another. In some cases, an AKAP connects PKA with the enzyme that triggers PKA activation (adenylyl cyclase) or terminates PKA action (cAMP phosphodiesterase or phosphoprotein phosphatase) **(Fig. 12–9)**. The very close proximity of these activating and inactivating enzymes presumably achieves a highly localized, and very brief, response.

As is now clear, to fully understand cellular signaling, researchers need tools precise enough to detect and study the spatiotemporal aspects of signaling processes at the subcellular level and in real time. In studies of the intracellular localization of biochemical changes, biochemistry meets cell biology, and techniques that cross this boundary have become essential in understanding signaling pathways. Fluorescent probes have found wide application in signaling studies. Labeling of functional proteins with a fluorescent tag such as the green fluorescent protein (GFP) reveals their subcellular localizations (see Fig. 9–16c). Changes in the state of association of two proteins (such as the R and C subunits of PKA) can be seen by measuring the nonradiative transfer of energy between fluorescent probes attached to each protein, a technique called fluorescence resonance energy transfer (FRET; Box 12–3).

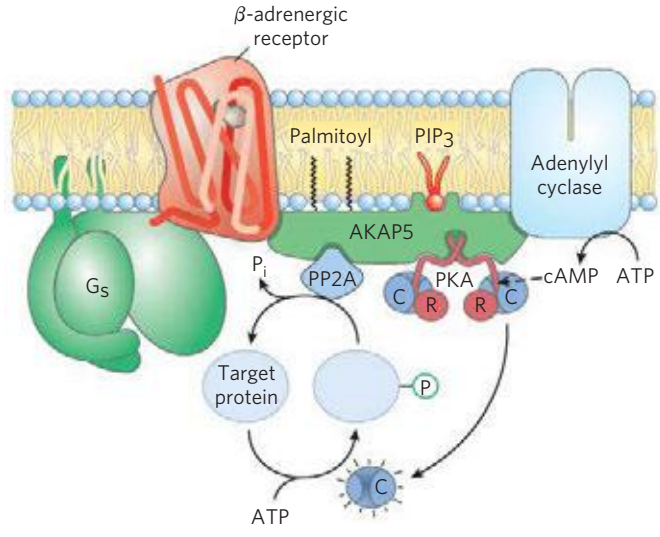


FIGURE 12–9 Nucleation of supramolecular complexes by A kinase anchoring proteins (AKAPs). AKAP5 is one of a family of proteins that act as multivalent scaffolds, holding PKA catalytic subunits—through the AKAP's interaction with the PKA regulatory subunits—in proximity to a particular region or structure in the cell. AKAP5 is targeted to rafts in the cytoplasmic surface of the plasma membrane by two covalently attached palmitoyl groups and a site that binds phosphatidylinositol 3,4,5-trisphosphate (PIP_3) in the membrane. AKAP5 also has binding sites for the β-adrenergic receptor, adenylyl cyclase, PKA, and a phosphoprotein phosphatase (PP2A), bringing them all together in the plane of the membrane. When epinephrine binds to the β-adrenergic receptor, $G_{s\alpha}$ triggers adenylyl cyclase produces cAMP, which reaches the nearby PKA quickly and with very little dilution. PKA phosphorylates its target protein, altering its activity, until the phosphoprotein phosphatase removes the phosphoryl group and returns the target protein to its prestimulus state. The AKAPs in this and other cases bring about a high local concentration of enzymes and second messengers, so that the signaling circuit remains highly localized, and the duration of the signal is limited.

Diacylglycerol, Inositol Trisphosphate, and Ca^{2+} Have Related Roles as Second Messengers

A second broad class of GPCRs are coupled through a G protein to a plasma membrane **phospholipase C (PLC)** that is specific for the membrane phospholipids phosphatidylinositol 4,5-bisphosphate, or PIP_2 (see Fig. 10–16). When one of the hormones that acts by this mechanism (Table 12–4) binds its specific receptor in the plasma membrane (**Fig. 12–10**, step ❶), the receptor-hormone complex catalyzes GTP-GDP exchange on an associated G protein, $\mathbf{G_q}$ (step ❷), activating it in much the same way that the β-adrenergic receptor activates G_s (Fig. 12–4). The activated G_q activates the PIP_2-specific PLC (Fig. 12–10, step ❸), which catalyzes (step ❹) the production of two potent second messengers, **diacylglycerol** and **inositol 1,4,5-trisphosphate**, or $\mathbf{IP_3}$ (not to be confused with PIP_3, p. 456).

TABLE 12–4 Some Signals That Act through Phospholipase C, IP_3, and Ca^{2+}

Acetylcholine [muscarinic M_1]	Gastrin-releasing peptide	Platelet-derived growth factor (PDGF)
α_1-Adrenergic agonists	Glutamate	Serotonin [5-HT-1c]
Angiogenin	Gonadotropin-releasing hormone (GRH)	Thyrotropin-releasing hormone (TRH)
Angiotensin II	Histamine [H_1]	Vasopressin
ATP [P_{2x}, P_{2y}]	Light (*Drosophila*)	
Auxin	Oxytocin	

Note: Receptor subtypes are in square brackets; see footnote to Table 12–3.

BOX 12–3 METHODS FRET: Biochemistry Visualized in a Living Cell

Fluorescent probes are commonly used to detect rapid biochemical changes in single living cells. They can be designed to give an essentially instantaneous report (within nanoseconds) on the changes in intracellular concentration of a second messenger or in the activity of a protein kinase. Furthermore, fluorescence microscopy has sufficient resolution to reveal where in the cell such changes are occurring. In one widely used procedure, the fluorescent probes are derived from a naturally occurring fluorescent protein, the **green fluorescent protein (GFP)** of the jellyfish *Aequorea victoria* (Fig. 1).

When excited by absorption of a photon of light, GFP emits a photon (that is, it fluoresces) in the green region of the spectrum. The light-absorbing/emitting center of GFP (its chromophore) comprises an oxidized form of the tripeptide –Ser^{65}–Tyr^{66}–Gly^{67}– (Fig. 2). Oxidation of the tripeptide is catalyzed by the GFP protein itself (Fig. 3), so it is possible to clone the protein into virtually any cell, where it can serve as a fluorescent marker for any protein to which it is fused (see Fig. 9–18). Variants of GFP, with different fluorescence spectra, are produced by genetic engineering. For example, in the yellow fluorescent protein (YFP), Ala^{206} in GFP is replaced by a Lys residue, changing the wavelength of light absorption and fluorescence. Other variants of GFP fluoresce blue (BFP) or cyan (CFP) light, and a related protein (mRFP1) fluoresces red light (Fig. 4). GFP and its variants are compact structures that retain their ability to fold into their native β-barrel conformation even when fused with another protein. These fluorescent hybrid proteins act as spectroscopic rulers for measuring distances between interacting proteins within a cell and, indirectly, to measure local concentrations of compounds that change the distance between two proteins.

FIGURE 1 *Aequorea victoria*, a jellyfish abundant in Puget Sound, Washington State.

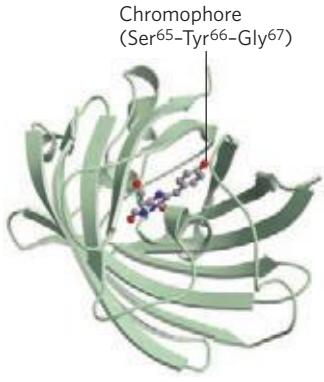


FIGURE 2 Green fluorescent protein (GFP), with the fluorescent chromophore shown in ball-and-stick form (derived from PDB ID 1GFL).

An excited fluorescent molecule such as GFP or YFP can dispose of the energy from the absorbed photon in either of two ways: (1) by fluorescence, emitting a photon of slightly longer wavelength (lower energy) than the exciting light, or (2) by nonradiative **fluorescence resonance energy transfer (FRET)**, in which the energy of the excited molecule (the donor) passes directly to a nearby molecule (the acceptor) *without emission of a photon*, exciting the acceptor (Fig. 5). The acceptor can now decay to its ground state by fluorescence; the emitted photon has a longer wavelength (lower energy) than both the original exciting light and the fluorescence emission of the donor. This second mode of decay (FRET) is possible only when donor and acceptor are close to each other (within 1 to 50 Å); the efficiency of FRET is inversely proportional to the *sixth power* of the distance between donor and acceptor. Thus very small changes in the distance between donor and acceptor register as very large changes in FRET, measured as the fluorescence of the acceptor molecule when the donor is excited. With sufficiently sensitive light detectors, this fluorescence signal can be located to specific regions of a single, living cell.

FRET has been used to measure [cAMP] in living cells. The gene for GFP is fused with that for the regulatory subunit (R) of cAMP-dependent protein kinase (PKA), and the gene for BFP is fused with that for the catalytic subunit (C) (Fig. 6). When these two hybrid

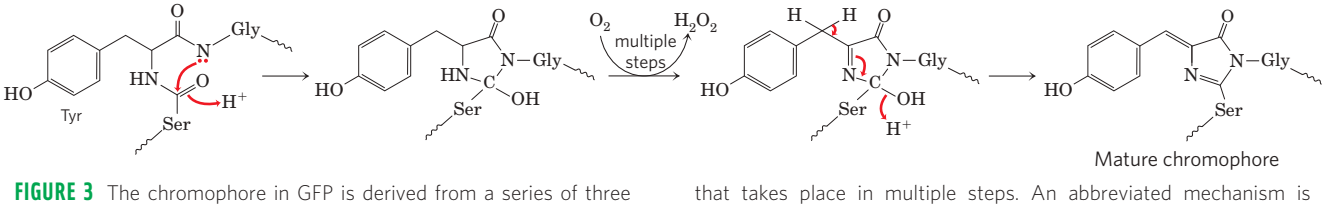


FIGURE 3 The chromophore in GFP is derived from a series of three amino acids: -Ser^{65}-Tyr^{66}-Gly^{67}-. Maturation of the chromophore involves an internal rearrangement, coupled to an oxidation reaction that takes place in multiple steps. An abbreviated mechanism is shown here.

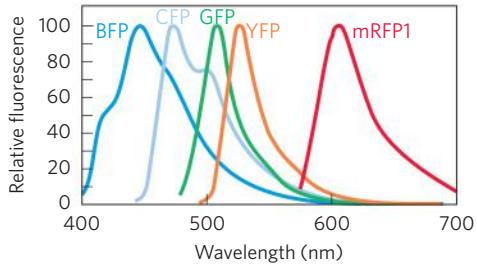


FIGURE 4 Emission spectra of GFP variants.

proteins are expressed in a cell, BFP (donor; excitation at 380 nm, emission at 460 nm) and GFP (acceptor; excitation at 475 nm, emission at 545 nm) in the inactive PKA (R_2C_2 tetramer) are close enough to undergo FRET. Wherever in the cell [cAMP] increases, the R_2C_2 complex dissociates into R_2 and 2C and the FRET signal is lost, because donor and acceptor are now too far apart for efficient FRET. Viewed in the fluorescence microscope, the region of higher [cAMP] has a minimal GFP signal and higher BFP signal. Measuring the ratio of emission at 460 nm and 545 nm gives a sensitive measure of the change in [cAMP]. By determining this ratio for all regions of the cell, the investigator can generate a false color image of the cell in which the ratio, or relative [cAMP], is represented by the intensity of the color. Images recorded at timed intervals reveal changes in [cAMP] over time.

A variation of this technology has been used to measure the activity of PKA in a living cell (Fig. 7). Researchers create a phosphorylation target for PKA by producing a hybrid protein containing four elements: YFP (acceptor); a short peptide with a Ser residue surrounded by the consensus sequence for PKA; a Ⓟ–Ser-binding domain (called 14-3-3); and

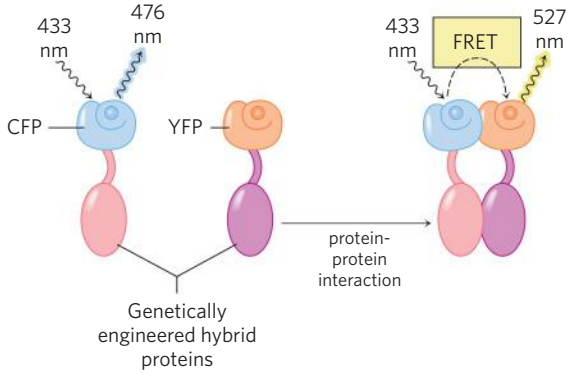


FIGURE 5 When the donor protein (CFP) is excited with monochromatic light of wavelength 433 nm, it emits fluorescent light at 476 nm (left). When the (red) protein fused with CFP interacts with the (purple) protein fused with YFP, that interaction brings CFP and YFP close enough to allow fluorescence resonance energy transfer (FRET) between them. Now, when CFP absorbs light of 433 nm, instead of fluorescing at 476 nm, it transfers energy directly to YFP, which then fluoresces at its characteristic emission wavelength, 527 nm. The ratio of light emission at 527 and 476 nm is therefore a measure of the interaction of the red and purple proteins.

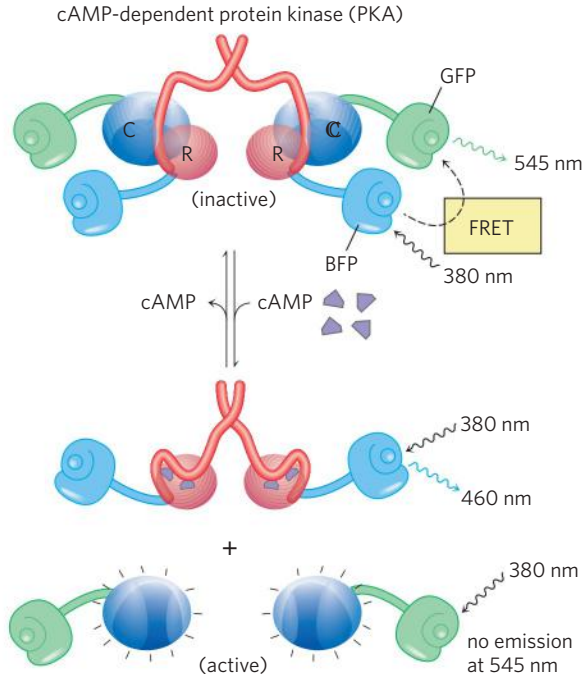


FIGURE 6 Measuring [cAMP] with FRET. Gene fusion creates hybrid proteins that exhibit FRET when the PKA regulatory (R) and catalytic (C) subunits are associated (low [cAMP]). When [cAMP] rises, the subunits dissociate and FRET ceases. The ratio of emission at 460 nm (dissociated) and 545 nm (complexed) thus offers a sensitive measure of [cAMP].

CFP (donor). When the Ser residue is not phosphorylated, 14-3-3 has no affinity for the Ser residue and the hybrid protein exists in an extended form, with the donor and acceptor too far apart to generate a FRET signal. Wherever PKA is active in the cell, it phosphorylates the Ser residue of the hybrid protein, and 14-3-3 binds to the Ⓟ–Ser. In doing so, it draws YFP and CFP together and a FRET signal is detected with the fluorescence microscope, revealing the presence of active PKA.

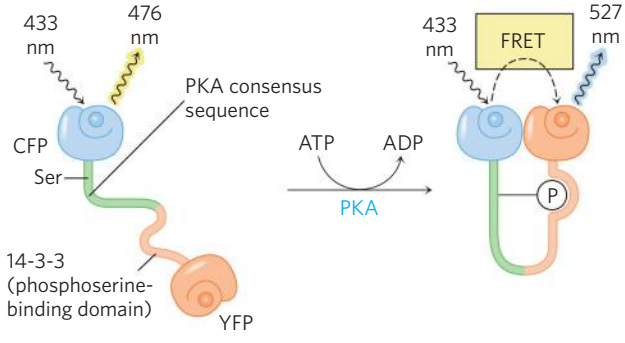


FIGURE 7 Measuring the activity of PKA with FRET. An engineered protein links YFP and CFP via a peptide that contains a Ser residue surrounded by the consensus sequence for phosphorylation by PKA, and the 14-3-3 Ⓟ-Ser-binding domain. Active PKA phosphorylates the Ser residue, which docks with the 14-3-3 binding domain, bringing the fluorescence proteins close enough to allow FRET to occur, revealing the presence of active PKA.

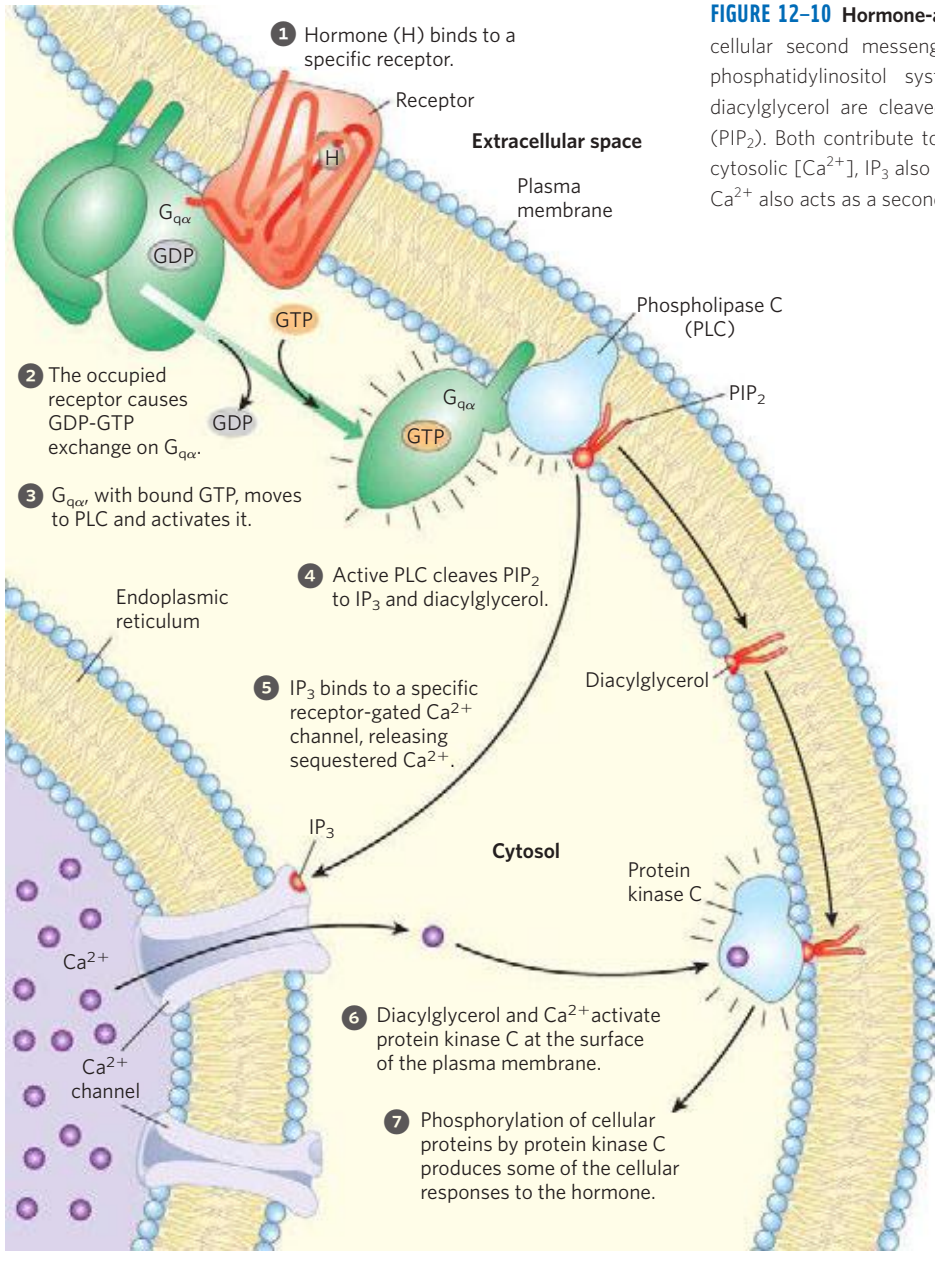
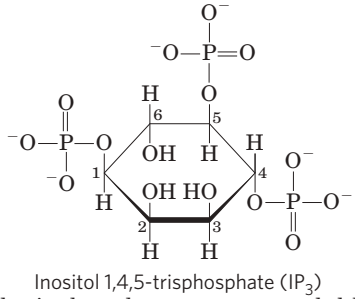


FIGURE 12–10 Hormone-activated phospholipase C and IP_3. Two intracellular second messengers are produced in the hormone-sensitive phosphatidylinositol system: inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol are cleaved from phosphatidylinositol 4,5-bisphosphate (PIP_2). Both contribute to the activation of protein kinase C. By raising cytosolic [Ca^{2+}], IP_3 also activates other Ca^{2+}-dependent enzymes; thus Ca^{2+} also acts as a second messenger.



Inositol 1,4,5-trisphosphate (IP_3)

Inositol trisphosphate, a water-soluble compound, diffuses from the plasma membrane to the endoplasmic reticulum (ER), where it binds to specific IP_3-gated Ca^{2+} channels, causing them to open. The action of the SERCA pump (see p. 410) ensures that [Ca^{2+}] in the ER is orders of magnitude higher than that in the cytosol, so when these gated Ca^{2+} channels open, Ca^{2+} rushes into the cytosol (Fig. 12–10, step 5), and the cytosolic [Ca^{2+}] rises sharply to about 10^{-6} M. One effect of elevated [Ca^{2+}] is the activation of **protein kinase C (PKC)**. Diacylglycerol cooperates with Ca^{2+} in activating PKC, thus also acting as a second messenger (step 6). Activation involves the movement of a PKC domain (the pseudosubstrate domain) away from its location in the substrate-binding region of the enzyme, allowing the enzyme to bind and phosphorylate proteins that contain a PKC consensus sequence—Ser or Thr residues embedded in an amino acid sequence recognized by PKC (step 7). There are several isozymes of PKC, each with a characteristic tissue distribution, target protein specificity, and role. Their targets include cytoskeletal proteins, enzymes, and nuclear proteins that regulate gene expression. Taken together, this family of enzymes has a wide range of cellular

actions, affecting neuronal and immune function and the regulation of cell division, for example.

Calcium Is a Second Messenger That May Be Localized in Space and Time

There are many variations on this basic scheme for Ca^{2+} signaling. In many cell types that respond to extracellular signals, Ca^{2+} serves as a second messenger that triggers intracellular responses, such as exocytosis in neurons and endocrine cells, contraction in muscle, and cytoskeletal rearrangements during amoeboid movement. In unstimulated cells, cytosolic [Ca^{2+}] is kept very low ($<10^{-7}$ M) by the action of Ca^{2+} pumps in the ER, mitochondria, and plasma membrane (as further discussed below). Hormonal, neural, or other stimuli cause either an influx of Ca^{2+} into the cell through specific Ca^{2+} channels in the plasma membrane or the release of sequestered Ca^{2+} from the ER or mitochondria, in either case raising the cytosolic [Ca^{2+}] and triggering a cellular response.

Changes in intracellular [Ca^{2+}] are detected by Ca^{2+}-binding proteins that regulate a variety of Ca^{2+}-dependent enzymes. **Calmodulin** (**CaM**; M_r 17,000) is an acidic protein with four high-affinity Ca^{2+}-binding sites. When intracellular [Ca^{2+}] rises to about 10^{-6} M (1 μM), the binding of Ca^{2+} to calmodulin drives a conformational change in the protein **(Fig. 12–11a)**. Calmodulin associates with a variety of proteins and, in its Ca^{2+}-bound state, modulates their activities (Fig. 12–11b). It is a member of a family of Ca^{2+}-binding proteins that also includes troponin (see Fig. 5–32), which triggers skeletal muscle contraction in response to increased [Ca^{2+}]. This family shares a characteristic Ca^{2+}-binding structure, the EF hand (Fig. 12–11c).

Calmodulin is an integral subunit of the **Ca^{2+}/calmodulin-dependent protein kinases** (**CaM kinases**, types I through IV). When intracellular [Ca^{2+}] increases in response to a stimulus, calmodulin binds Ca^{2+}, undergoes a change in conformation, and activates the CaM kinase. The kinase then phosphorylates target enzymes, regulating their activities. Calmodulin is also a regulatory subunit of phosphorylase *b* kinase of muscle, which is activated by Ca^{2+}. Thus Ca^{2+} triggers ATP-requiring muscle contractions while also activating glycogen breakdown, providing fuel for ATP synthesis. Many other enzymes are also known to be modulated by Ca^{2+} through calmodulin (Table 12–5). The activity of the second messenger Ca^{2+}, like that of cAMP, can be spatially restricted; after its release triggers a local response, Ca^{2+} is generally removed before it can diffuse to distant parts of the cell.

Very commonly, Ca^{2+} level does not simply rise and then decrease, but rather oscillates with a period of a few seconds **(Fig. 12–12)**—even when the extracellular concentration of the triggering hormone remains constant. The mechanism underlying [Ca^{2+}] oscillations presumably entails feedback regulation by Ca^{2+} on

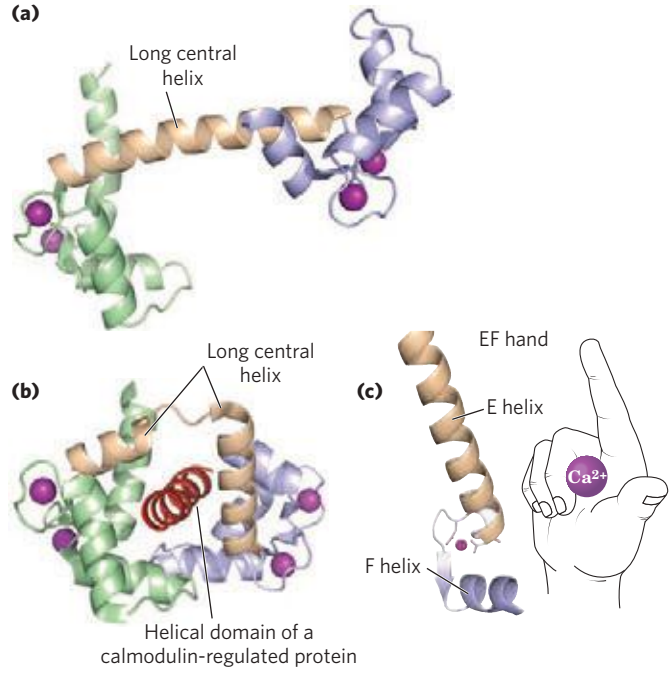


FIGURE 12–11 Calmodulin. This is the protein mediator of many Ca^{2+}-stimulated enzymatic reactions. Calmodulin has four high-affinity Ca^{2+}-binding sites ($K_d \approx 0.1$ to 1 μM). **(a)** A ribbon model of the crystal structure of calmodulin (PDB ID 1CLL). The four Ca^{2+}-binding sites are occupied by Ca^{2+} (purple). The amino-terminal domain is on the left; the carboxyl-terminal domain on the right. **(b)** Calmodulin associated with a helical domain (red) of one of the many enzymes it regulates, calmodulin-dependent protein kinase II (PDB ID 1CDL). Notice that the long central α helix of calmodulin visible in (a) has bent back on itself in binding to the helical substrate domain. The central helix of calmodulin is clearly more flexible in solution than in the crystal. **(c)** Each of the four Ca^{2+}-binding sites occurs in a helix-loop-helix motif called the EF hand, also found in many other Ca^{2+}-binding proteins.

TABLE 12–5 Some Proteins Regulated by Ca^{2+} and Calmodulin

Adenylyl cyclase (brain)
Ca^{2+}/calmodulin-dependent protein kinases (CaM kinases I to IV)
Ca^{2+}-dependent Na^+ channel (*Paramecium*)
Ca^{2+}-release channel of sarcoplasmic reticulum
Calcineurin (phosphoprotein phosphatase 2B)
cAMP phosphodiesterase
cAMP-gated olfactory channel
cGMP-gated Na^+, Ca^{2+} channels (rod and cone cells)
Glutamate decarboxylase
Myosin light-chain kinases
NAD^+ kinase
Nitric oxide synthase
Phosphatidylinositol 3-kinase
Plasma membrane Ca^{2+} ATPase (Ca^{2+} pump)
RNA helicase (p68)

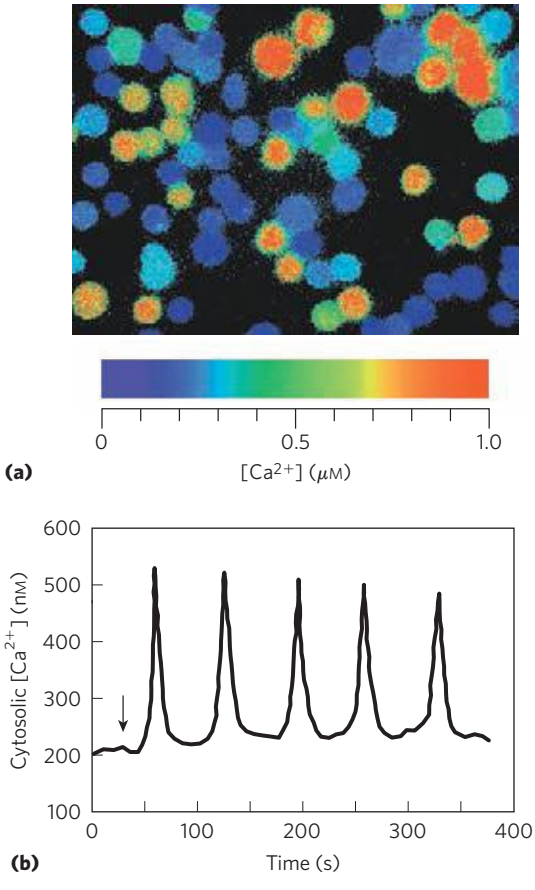


FIGURE 12–12 Triggering of oscillations in intracellular [Ca^{2+}] by extracellular signals. (a) A dye (fura) that undergoes fluorescence changes when it binds Ca^{2+} is allowed to diffuse into cells, and its instantaneous light output is measured by fluorescence microscopy. Fluorescence intensity is represented by color; the color scale relates intensity of color to [Ca^{2+}], allowing determination of the absolute [Ca^{2+}]. In this case, thymocytes (cells of the thymus) have been stimulated with extracellular ATP, which raises their internal [Ca^{2+}]. The cells are heterogeneous in their responses; some have high intracellular [Ca^{2+}] (red), others much lower (blue). **(b)** When such a probe is used in a single hepatocyte, the agonist norepinephrine (added at the arrow) causes oscillations of [Ca^{2+}] from 200 to 500 nM. Similar oscillations are induced in other cell types by other extracellular signals.

some part of the Ca^{2+}-release process. Whatever the mechanism, the effect is that one kind of signal (hormone concentration, for example) is converted into another (frequency and amplitude of intracellular [Ca^{2+}] "spikes"). The Ca^{2+} signal diminishes as Ca^{2+} diffuses away from the initial source (the Ca^{2+} channel), is sequestered in the ER, or is pumped out of the cell.

There is significant cross talk between the Ca^{2+} and cAMP signaling systems. In some tissues, both the enzyme that produces cAMP (adenylyl cyclase) and the enzyme that degrades cAMP (phosphodiesterase) are stimulated by Ca^{2+}. Temporal and spatial changes in [Ca^{2+}] can therefore produce transient, localized changes in [cAMP]. We have noted already that PKA, the enzyme that responds to cAMP, is often part of a highly localized supramolecular complex assembled on scaffold proteins such as AKAPs. This subcellular localization of target enzymes, combined with temporal and spatial gradients in [Ca^{2+}] and [cAMP], allows a cell to respond to one or several signals with subtly nuanced metabolic changes, localized in space and time.

GPCRs Mediate the Actions of a Wide Variety of Signals

The human genome encodes about 1,000 G protein–coupled receptors, recognizable by their seven transmembrane helical segments and certain highly conserved residues. Each is expressed selectively, in certain cell types or under certain conditions. Together, they allow cells and tissues to respond to a wide array of different stimuli, including various low molecular weight amines, peptides, proteins, eicosanoids and other lipids, as well as light and compounds detected by olfaction and gustation. The determination of several GCPR structures by crystallography **(Fig. 12–13)**, including the β-adrenergic receptor with its G protein, and the histamine receptor, has stimulated great interest in both the transduction mechanism(s) and the possibilities of altering receptor activity with drugs. These two receptors are the targets of a variety of widely used beta-blocker and antihistamine medications, respectively. The similarities in GPCR structures go beyond the common seven–transmembrane helix pattern; the structures of five different GPCRs are almost superimposable (Fig. 12–13c). Clearly, something about this three-dimensional structure makes it effective as a transducer of many disparate signals.

SUMMARY 12.2 G Protein–Coupled Receptors and Second Messengers

- G protein–coupled receptors (GPCRs) share a common structural arrangement of seven transmembrane helices and act through heterotrimeric G proteins. On ligand binding, GPCRs catalyze the exchange of GTP for GDP on the G protein, causing dissociation of the G_α subunit; G_α then stimulates or inhibits the activity of an effector enzyme, changing the level of its second-messenger product.
- The β-adrenergic receptor activates a stimulatory G protein, G_s, thereby activating adenylyl cyclase and raising the concentration of the second messenger cAMP. Cyclic AMP stimulates cAMP-dependent protein kinase to phosphorylate key target enzymes, changing their activities.
- Enzyme cascades, in which a single molecule of hormone activates a catalyst to activate another catalyst, and so on, result in the large signal amplification that is characteristic of hormone receptor systems.

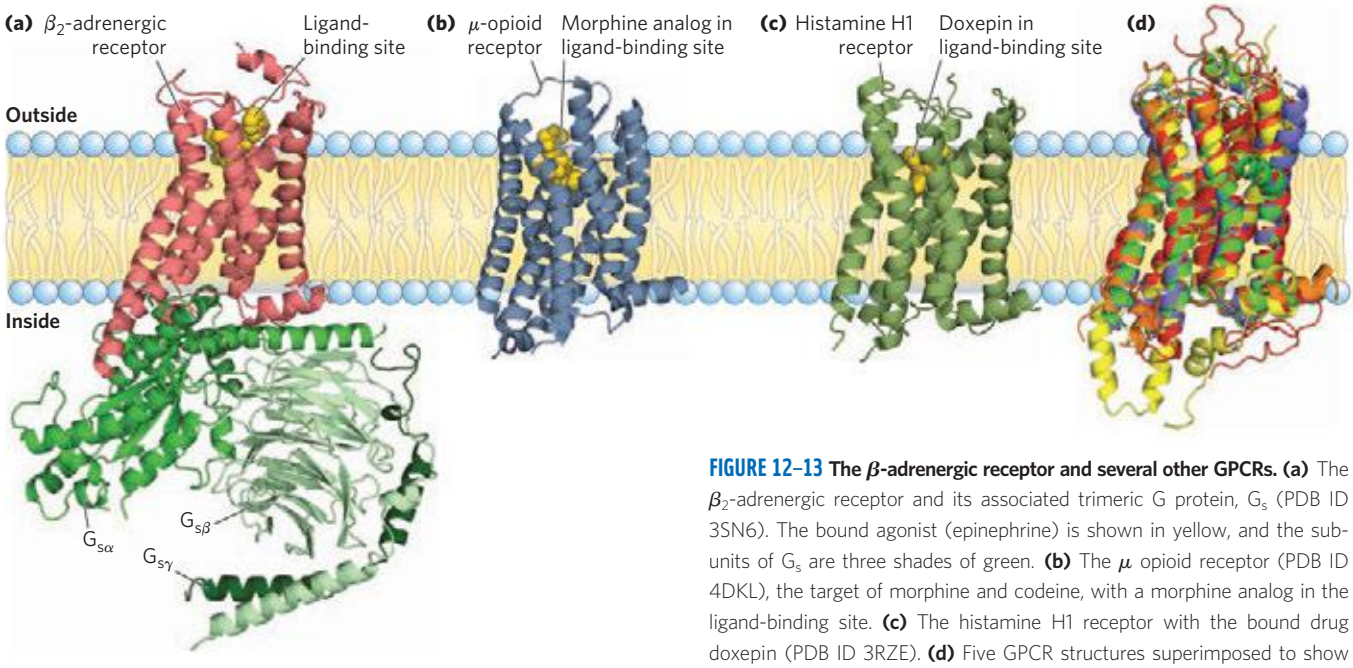


FIGURE 12–13 The β-adrenergic receptor and several other GPCRs. (a) The β_2-adrenergic receptor and its associated trimeric G protein, G_s (PDB ID 3SN6). The bound agonist (epinephrine) is shown in yellow, and the subunits of G_s are three shades of green. **(b)** The μ opioid receptor (PDB ID 4DKL), the target of morphine and codeine, with a morphine analog in the ligand-binding site. **(c)** The histamine H1 receptor with the bound drug doxepin (PDB ID 3RZE). **(d)** Five GPCR structures superimposed to show the remarkable conservation of structure. Shown are the human A2A adenosine receptor (orange; PDB ID 3EML); turkey β_1-adrenergic receptor (blue; PDB ID 2VT4), human β_2-adrenergic receptor (green; PDB ID 2RH1), rhodopsin from squid (yellow; PDB ID 2Z73); and bovine rhodopsin, (red; PDB ID 1U19).

- Cyclic AMP concentration is eventually reduced by cAMP phosphodiesterase, and G_s turns itself off by hydrolysis of its bound GTP to GDP, acting as a self-limiting binary switch.
- When the epinephrine signal persists, β-adrenergic receptor–specific protein kinase and β-arrestin temporarily desensitize the receptor and cause it to move into intracellular vesicles.
- Some receptors stimulate adenylyl cyclase through G_s; others inhibit it through G_i. Thus cellular [cAMP] reflects the integrated input of two (or more) signals.
- Noncatalytic adaptor proteins such as AKAPs hold together proteins involved in a signaling process, increasing the efficiency of their interactions and in some cases confining the process to a specific subcellular location.
- Some GPCRs act via a plasma membrane phospholipase C that cleaves PIP_2 to diacylglycerol and IP_3. By opening Ca^{2+} channels in the endoplasmic reticulum, IP_3 raises cytosolic [Ca^{2+}]. Diacylglycerol and Ca^{2+} act together to activate protein kinase C, which phosphorylates and changes the activity of specific cellular proteins. Cellular [Ca^{2+}] also regulates (often through calmodulin) many other enzymes and proteins involved in secretion, cytoskeletal rearrangements, or contraction.

12.3 Receptor Tyrosine Kinases

The **receptor tyrosine kinases (RTKs)**, a large family of plasma membrane receptors with intrinsic protein kinase activity, transduce extracellular signals by a mechanism fundamentally different from that of GPCRs. RTKs have a ligand-binding domain on the extracellular face of the plasma membrane and an enzyme active site on the cytoplasmic face, connected by a single transmembrane segment. The cytoplasmic domain is a protein kinase that phosphorylates Tyr residues in specific target proteins—a Tyr kinase. The receptors for insulin and epidermal growth factor are prototypes for this group.

Stimulation of the Insulin Receptor Initiates a Cascade of Protein Phosphorylation Reactions

Insulin regulates both metabolic enzymes and gene expression. Insulin does not enter cells, but initiates a signal that travels a branched pathway from the plasma membrane receptor to insulin-sensitive enzymes in the cytosol and to the nucleus, where it stimulates the transcription of specific genes. The active insulin receptor protein (INSR) consists of two identical α subunits protruding from the outer face of the plasma membrane and two transmembrane β subunits with their carboxyl termini protruding into the cytosol—a dimer of $\alpha\beta$ monomers **(Fig. 12–14)**. The α subunits contain the insulin-binding domain, and the intracellular domains of the β subunits contain the protein kinase activity that transfers a phosphoryl group from ATP to the hydroxyl group of Tyr residues in specific target proteins. Signaling through INSR begins when the binding of one insulin molecule between the two subunits of the dimer activates the Tyr kinase activity, and each β subunit phosphorylates three critical Tyr residues near the carboxyl terminus of the other β subunit. This **autophosphorylation** opens up the active site so that the enzyme can phosphorylate Tyr residues of other target proteins.

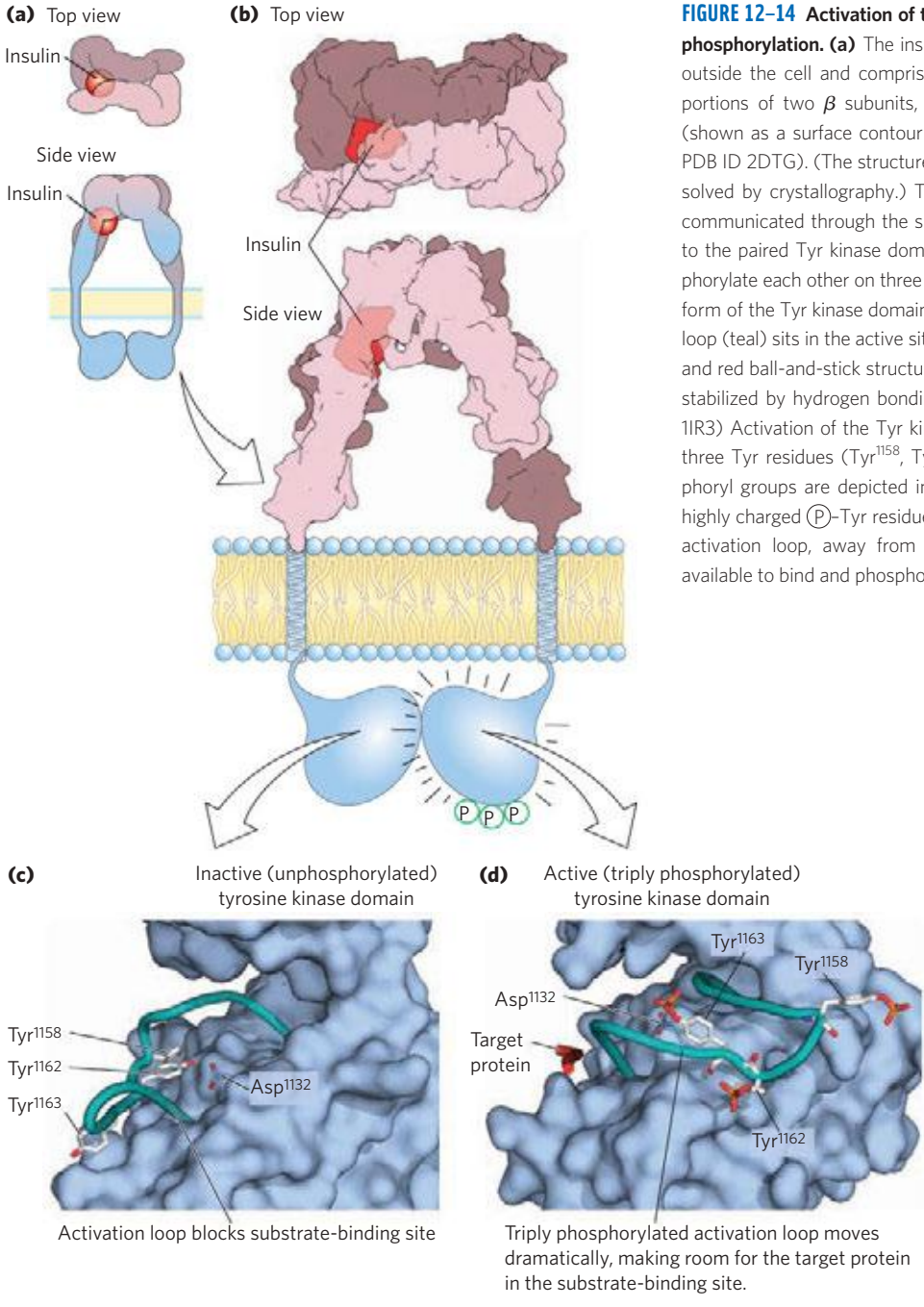


FIGURE 12–14 Activation of the insulin-receptor tyrosine kinase by autophosphorylation. (a) The insulin-binding region of the insulin receptor lies outside the cell and comprises **(b)** two α subunits and the extracellular portions of two β subunits, intertwined to form the insulin-binding site (shown as a surface contour model of the crystal structure, derived from PDB ID 2DTG). (The structure of the transmembrane domain has not been solved by crystallography.) The binding of insulin (red; PDB ID 2CEU) is communicated through the single transmembrane helix of each β subunit to the paired Tyr kinase domains inside the cell, activating them to phosphorylate each other on three Tyr residues. **(c)** (PDB ID 1IRK) In the inactive form of the Tyr kinase domain (before Tyr phosphorylations), the activation loop (teal) sits in the active site, and none of the critical Tyr residues (white and red ball-and-stick structures) are phosphorylated. This conformation is stabilized by hydrogen bonding between Tyr^{1162} and Asp^{1132}. **(d)** (PDB ID 1IR3) Activation of the Tyr kinase allows each β subunit to phosphorylate three Tyr residues (Tyr^{1158}, Tyr^{1162}, Tyr^{1163}) on the other β subunit. (Phosphoryl groups are depicted in red and orange.) The introduction of three highly charged Ⓟ–Tyr residues forces a 30 Å change in the position of the activation loop, away from the substrate-binding site, which becomes available to bind and phosphorylate a target protein.

The mechanism of activation of the INSR protein kinase is similar to that described for PKA and PKC: a region of the cytoplasmic domain (an autoinhibitory sequence) that usually occludes the active site moves out of the active site after being phosphorylated, opening up the site for the binding of target proteins (Fig. 12–14).

When INSR is autophosphorylated (**Fig. 12–15**, step ❶), one of its targets is insulin receptor substrate-1 (IRS-1; step ❷). Once phosphorylated on several of its Tyr residues, IRS-1 becomes the point of nucleation for a complex of proteins (step ❸) that carry the message from the insulin receptor to end targets in the cytosol and nucleus, through a long series of intermediate proteins. First, a Ⓟ–Tyr residue of IRS-1 binds to the **SH2 domain** of the protein Grb2. (SH2 is an abbreviation of *S*rc *h*omology *2*, so named because the sequence of an SH2 domain is similar to that of a domain in Src (pronounced *sark*), another protein Tyr kinase.) Several signaling proteins contain SH2 domains, all of which bind Ⓟ–Tyr residues in a protein partner. Grb2 is an adaptor protein, with no intrinsic enzymatic activity. Its function is to bring together two proteins (in this case, IRS-1 and the protein Sos) that must interact to enable signal transduction. In addition to its SH2 (Ⓟ–Tyr-binding) domain, Grb2 also contains a second protein-binding domain, SH3, that binds to a proline-rich region of Sos, recruiting Sos to the growing receptor complex. When bound to Grb2, Sos acts as a guanosine nucleotide–exchange factor

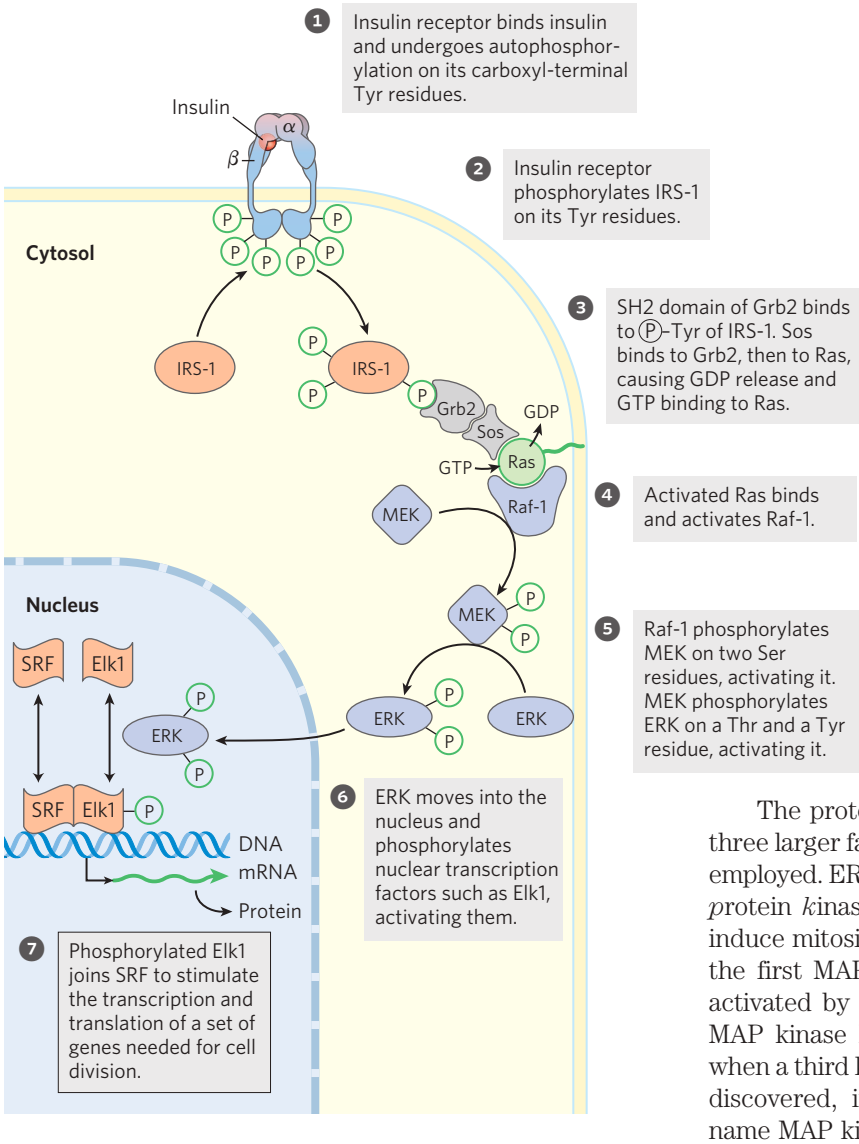


FIGURE 12–15 Regulation of gene expression by insulin through a MAP kinase cascade. The insulin receptor (INSR) consists of two α subunits on the outer face of the plasma membrane and two β subunits that traverse the membrane and protrude from the cytosolic face. Binding of insulin to the α subunits triggers a conformational change that allows the autophosphorylation of Tyr residues in the carboxyl-terminal domain of the β subunits. Autophosphorylation further activates the Tyr kinase domain, which then catalyzes phosphorylation of other target proteins. The signaling pathway by which insulin regulates the expression of specific genes consists of a cascade of protein kinases, each of which activates the next. INSR is a Tyr-specific kinase; the other kinases (all shown in blue) phosphorylate Ser or Thr residues. MEK is a dual-specificity kinase, which phosphorylates both a Thr and a Tyr residue in ERK (extracellular *r*egulated *k*inase); MEK is mitogen-activated, ERK-activating kinase; SRF is serum response factor.

(GEF), catalyzing the replacement of bound GDP with GTP on Ras, a G protein.

Ras is the prototype of a family of **small G proteins** that mediate a wide variety of signal transductions (see Box 12–2). Like the trimeric G protein that functions with the β-adrenergic system (Fig. 12–5), Ras can exist in either the GTP-bound (active) or GDP-bound (inactive) conformation, but Ras (~20 kDa) acts as a monomer. When GTP binds, Ras can activate a protein kinase, Raf-1 (Fig. 12–15, step 4), the first of three protein kinases—Raf-1, MEK, and ERK—that form a cascade in which each kinase activates the next by phosphorylation (step 5). The protein kinases MEK and ERK are activated by phosphorylation of both a Thr and a Tyr residue. When activated, ERK mediates some of the biological effects of insulin by entering the nucleus and phosphorylating transcription factors, such as Elk1 (step 6), that modulate the transcription of about 100 insulin-regulated genes (step 7), some of which encode proteins essential for cell division. Thus, insulin acts as a growth factor.

The proteins Raf-1, MEK, and ERK are members of three larger families, for which several nomenclatures are employed. ERK is in the **MAPK** family (*m*itogen-*a*ctivated *p*rotein *k*inases; mitogens are extracellular signals that induce mitosis and cell division). Soon after discovery of the first MAPK enzyme, that enzyme was found to be activated by another protein kinase, which was named MAP kinase kinase (MEK belongs to this family), and when a third kinase that activated MAP kinase kinase was discovered, it was given the slightly ludicrous family name MAP kinase kinase kinase (Raf-1 is in this family). Somewhat less cumbersome are the abbreviations for these three families: MAPK, MAPKK, and MAPKKK. Kinases in the MAPK and MAPKKK families are specific for Ser or Thr residues, and MAPKKs (here, MEK) phosphorylate both a Ser and a Tyr residue in their substrate, a MAPK (here, ERK).

Biochemists now recognize this insulin pathway as but one instance of a more general scheme in which hormone signals, via pathways similar to that shown in Figure 12–15, result in phosphorylation of target enzymes by protein kinases. The target of phosphorylation is often another protein kinase, which then phosphorylates a third protein kinase, and so on. The result is a cascade of reactions that amplifies the initial signal by many orders of magnitude (see Fig. 12–1b). **MAPK cascades** (Fig. 12–15) mediate signaling initiated by a variety of growth factors, such as platelet-derived growth factor (PDGF) and epidermal growth factor (EGF). Another general scheme exemplified by the insulin receptor pathway is the use of nonenzymatic adaptor proteins to bring together the components of a branched signaling pathway, to which we now turn.

The Membrane Phospholipid PIP_3 Functions at a Branch in Insulin Signaling

The signaling pathway from insulin branches at IRS-1 (Fig. 12–15, step ❷). Grb2 is not the only protein that associates with phosphorylated IRS-1. The enzyme phosphoinositide 3-kinase (PI3K) binds IRS-1 through PI3K's SH2 domain **(Fig. 12–16)**. Thus activated, PI3K converts the membrane lipid phosphatidylinositol 4,5-bisphosphate (PIP_2) to phosphatidylinositol 3,4,5-trisphosphate (PIP_3). The multiply charged head group of PIP_3, protruding on the cytoplasmic side of the plasma membrane, is the starting point for a second signaling branch involving another cascade of protein kinases. When bound to PIP_3, protein kinase B (PKB; also called Akt) is phosphorylated and activated by yet another protein kinase, PDK1. The activated PKB then phosphorylates Ser or Thr residues in its target proteins, one of which is glycogen synthase kinase 3 (GSK3). In its active, nonphosphorylated form, GSK3 phosphorylates glycogen synthase, inactivating it and thereby contributing to the slowing of glycogen synthesis. (This mechanism is only part of the explanation for the effects of insulin on glycogen metabolism.) When phosphorylated by PKB, GSK3 is inactivated. By thus preventing inactivation of glycogen synthase in liver and muscle, the cascade of protein phosphorylations initiated by insulin stimulates glycogen synthesis (Fig. 12–16). In a third signaling branch in muscle and fat tissue, PKB triggers the clathrin-aided movement of glucose transporters (GLUT4) from internal vesicles to the plasma membrane, stimulating glucose uptake from the blood (Fig. 12–16, step ❺; see also Box 11–1).

As in all signaling pathways, there is a mechanism for terminating the activity of the PI3K–PKB pathway. A PIP_3-specific phosphatase (PTEN in humans) removes the phosphoryl group at the 3 position of PIP_3 to produce PIP_2, which no longer serves as a binding site for PKB, and the signaling chain is broken. In various types of cancer, it is often found that the PTEN gene has undergone mutation, resulting in a defective regulatory circuit and abnormally high levels of PIP_3 and of PKB activity. The result seems to be a continuing signal for cell division and thus tumor growth. ■

The insulin receptor is the prototype for several receptor enzymes with a similar structure and RTK activity **(Fig. 12–17)**. The receptors for EGF and PDGF, for example, have structural and sequence similarities to INSR, and both have a protein Tyr kinase activity that phosphorylates IRS-1. Many of these receptors dimerize after binding ligand; INSR is the exception, as it is already an $(\alpha\beta)_2$ dimer before insulin binds.

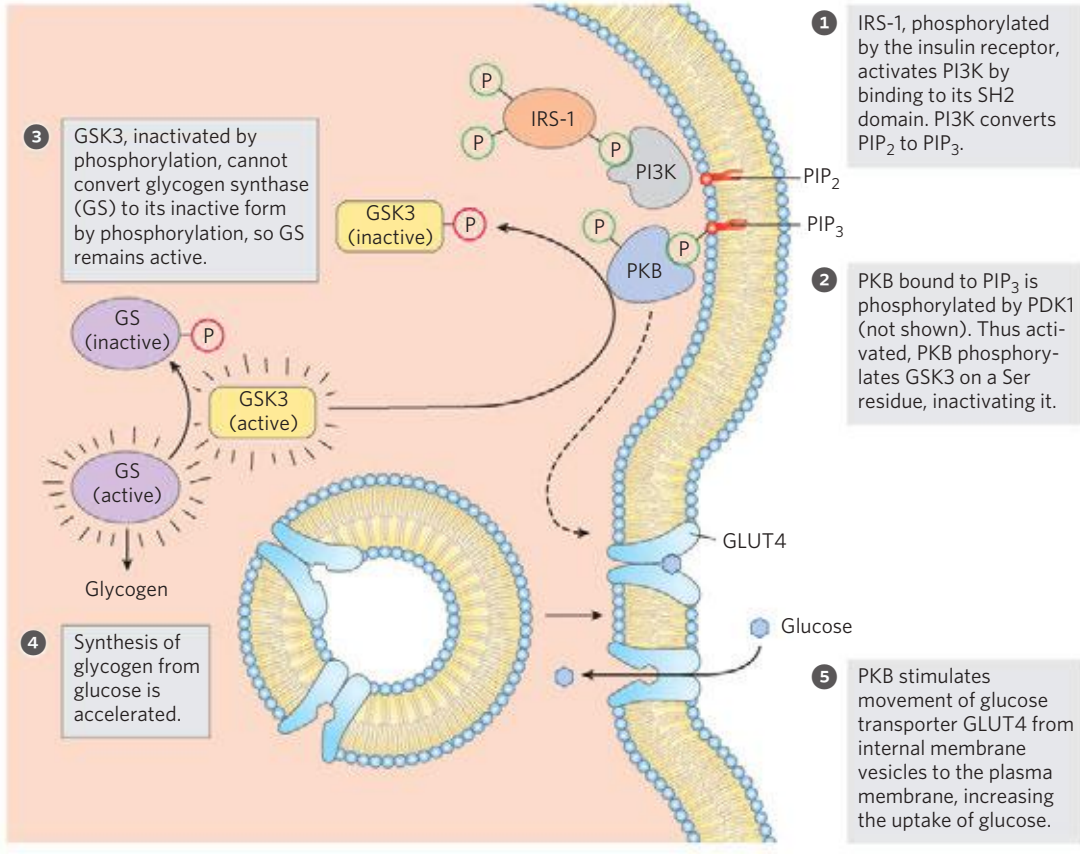


FIGURE 12–16 Insulin action on glycogen synthesis and GLUT4 movement to the plasma membrane. The activation of PI3 kinase (PI3K) by phosphorylated IRS-1 signals (through protein kinase B, PKB) movement of the glucose transporter GLUT4 to the plasma membrane, and the activation of glycogen synthase.

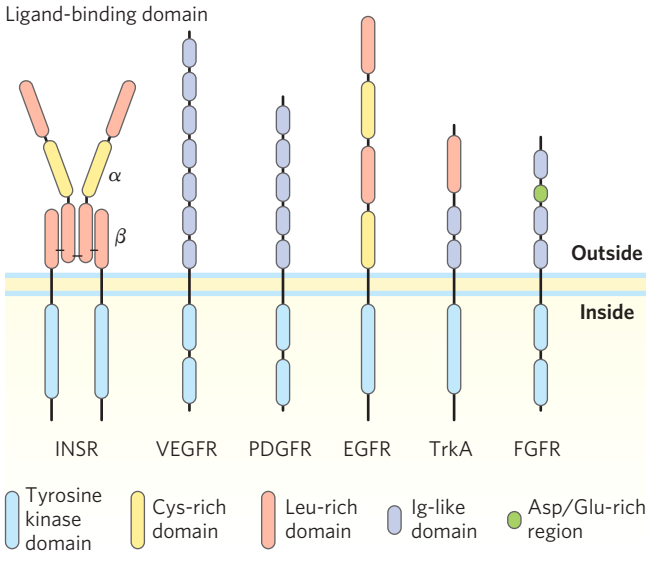


FIGURE 12–17 Receptor tyrosine kinases. Growth factor receptors that signal through Tyr kinase activity include those for insulin (INSR), vascular epidermal growth factor (VEGFR), platelet-derived growth factor (PDGFR), epidermal growth factor (EGFR), high-affinity nerve growth factor (TrkA), and fibroblast growth factor (FGFR). All these receptors have a Tyr kinase domain on the cytoplasmic side of the plasma membrane (blue). The extracellular domain is unique to each type of receptor, reflecting the different growth-factor specificities. These extracellular domains are typically combinations of structural motifs such as cysteine- or leucine-rich segments and segments containing one of several motifs common to immunoglobulins (Ig-like domains; see Fig. 4–22). Many other receptors of this type are encoded in the human genome, each with a different extracellular domain and ligand specificity.

(The protomer of the insulin receptor is one $\alpha\beta$ unit.) The binding of adaptor proteins such as Grb2 to Ⓟ–Tyr residues is a common mechanism for promoting protein-protein interactions initiated by RTKs, a subject to which we return in Section 12.5.

In addition to the many receptors that act as protein Tyr kinases (the RTKs), several receptorlike plasma membrane proteins have protein Tyr phosphatase activity. Based on the structures of these proteins, we can surmise that their ligands are components of the extracellular matrix or are surface molecules on other cells. Although their signaling roles are not yet as well understood as those of the RTKs, they clearly have the potential to reverse the actions of signals that stimulate RTKs.

What spurred the evolution of such complicated regulatory machinery? This system allows one activated receptor to activate several IRS-1 molecules, amplifying the insulin signal, and it provides for the integration of signals from different receptors such as EGFR and PDGFR, each of which can phosphorylate IRS-1. Furthermore, because IRS-1 can activate any of several proteins that contain SH2 domains, a single receptor acting through IRS-1 can trigger two or more signaling pathways; insulin affects gene expression through the Grb2-Sos-Ras-MAPK pathway and affects glycogen metabolism and glucose transport through the PI3K–PKB pathway. Finally, there are several closely related IRS proteins (IRS2, IRS3), each with its own characteristic tissue distribution and function, further enriching the signaling possibilities in pathways initiated by RTKs.

The JAK-STAT Signaling System Also Involves Tyrosine Kinase Activity

A variation on the basic theme of receptor Tyr kinases is receptors that have no intrinsic protein kinase activity but, when occupied by their ligand, bind a *cytosolic* Tyr kinase. One example is the system that regulates the formation of erythrocytes in mammals. The developmental signal, or **cytokine**, for this system is erythropoietin (EPO), a 165 amino acid protein produced in the kidneys. When EPO binds to its plasma membrane receptor **(Fig. 12–18)**, the receptor dimerizes, and the dimer can bind and activate the soluble protein kinase JAK (*Ja*nus *k*inase). The activated JAK phosphorylates

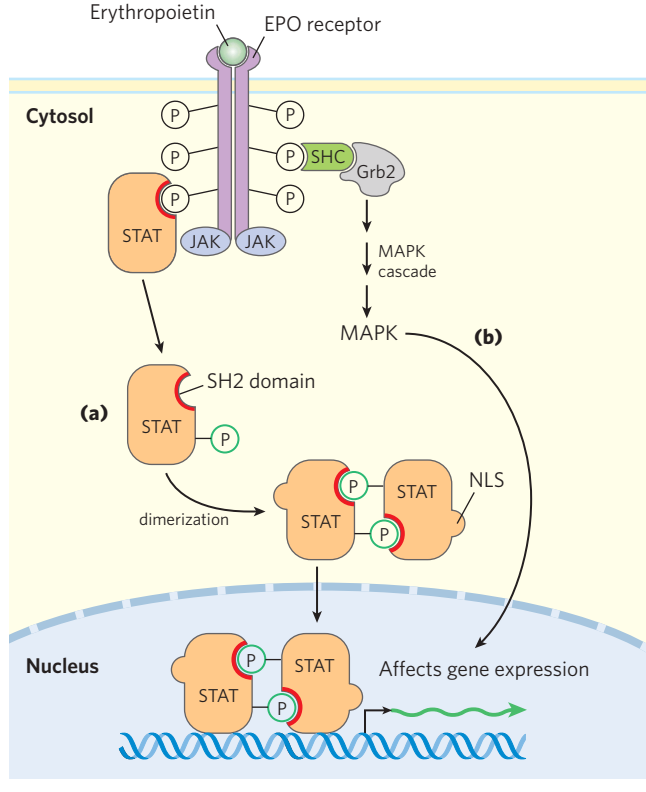


FIGURE 12–18 The JAK-STAT transduction mechanism for the erythropoietin receptor. Binding of erythropoietin (EPO) causes dimerization of the EPO receptor, which allows JAK, a soluble Tyr kinase, to bind to the internal domain of the receptor and phosphorylate it on several Tyr residues. **(a)** In one signaling pathway, the SH2 domain of the STAT protein STAT5 binds to Ⓟ-Tyr residues on the receptor, bringing it into proximity with JAK. Following phosphorylation of STAT5 by JAK, two STAT5 molecules dimerize, each binding the other's Ⓟ-Tyr residue, thus exposing a nuclear localization sequence (NLS) that targets the dimer for transport into the nucleus. In the nucleus, STAT5 turns on the expression of EPO-controlled genes. **(b)** In a second signaling pathway, following EPO binding and autophosphorylation of JAK, the adaptor protein SHC binds the Ⓟ-Tyr of the receptor, then Grb2 binds SHC and triggers the MAPK cascade, as in the insulin system (see Fig. 12–15).

several Tyr residues in the cytoplasmic domain of the EPO receptor. A family of transcription factors, collectively called STATs (*s*ignal *t*ransducers and *a*ctivators of *t*ranscription), are also targets of JAK. An SH2 domain in STAT5 binds Ⓟ–Tyr residues in the EPO receptor, positioning the STAT for phosphorylation by JAK in response to EPO. The phosphorylated STAT5 forms dimers, exposing a signal that causes it to be transported into the nucleus. There, STAT5 induces the expression (transcription) of specific genes essential for erythrocyte maturation. This JAK-STAT system also operates in other signaling pathways, including that for the hormone leptin, described in detail in Chapter 23 (see Fig. 23–36). Activated JAK can also trigger, through Grb2, the MAPK cascade (Fig. 12–18b), which leads to altered expression of specific genes.

Src is another soluble protein Tyr kinase that associates with certain receptors when they bind their ligands. Src was the first protein found to have the characteristic Ⓟ–Tyr-binding domain that was subsequently named the Src homology (SH2) domain.

Cross Talk among Signaling Systems Is Common and Complex

Although, for simplicity, we have treated individual signaling pathways as separate sequences of events leading to separate metabolic consequences, there is in fact extensive cross talk among signaling systems. The regulatory circuitry that governs metabolism is richly interwoven and multilayered. We have discussed the signaling pathways for insulin and epinephrine separately, but they do not operate independently. Insulin opposes the metabolic effects of epinephrine in most tissues, and activation of the insulin signaling pathway directly attenuates signaling through the β-adrenergic signaling system. For example, the INSR kinase directly phosphorylates two Tyr residues in the cytoplasmic tail of a β_2-adrenergic receptor, and PKB, activated by insulin **(Fig. 12–19)**, phosphorylates two Ser residues in the same region. Phosphorylation of these four residues triggers clathrin-aided internalization of the β_2-adrenergic receptor, taking it out of service and lowering the cell's sensitivity to epinephrine. A second type of cross talk between these receptors occurs when Ⓟ–Tyr residues on the β_2-adrenergic receptor, phosphorylated by INSR, serve as nucleation points for SH2 domain–containing proteins such as Grb2 (Fig. 12–19, left side). Activation of the MAPK ERK by insulin (see Fig. 12–15) is 5- to 10-fold greater in the presence of the β_2-adrenergic receptor, presumably because of this cross talk. Signaling systems that use cAMP and Ca^{2+} also show extensive interaction; each second messenger affects the generation and concentration of the other. One of the major challenges of systems biology is to sort out the effects of such interactions on the overall metabolic patterns in each tissue—a daunting task.

SUMMARY 12.3 Receptor Tyrosine Kinases

- The insulin receptor, INSR, is the prototype of receptor enzymes with Tyr kinase activity. When insulin binds, each $\alpha\beta$ unit of INSR phosphorylates the β subunit of its partner, activating the receptor's Tyr kinase activity. The kinase catalyzes the phosphorylation of Tyr residues on other proteins, such as IRS-1.
- Phosphotyrosine residues in IRS-1 serve as binding sites for proteins with SH2 domains. Some of these proteins, such as Grb2, have two or more protein-binding domains and can serve as adaptors that bring two proteins into proximity.
- Sos bound to Grb2 catalyzes GDP-GTP exchange on Ras (a small G protein), which in turn activates a MAPK cascade that ends with the phosphorylation of target proteins in the cytosol

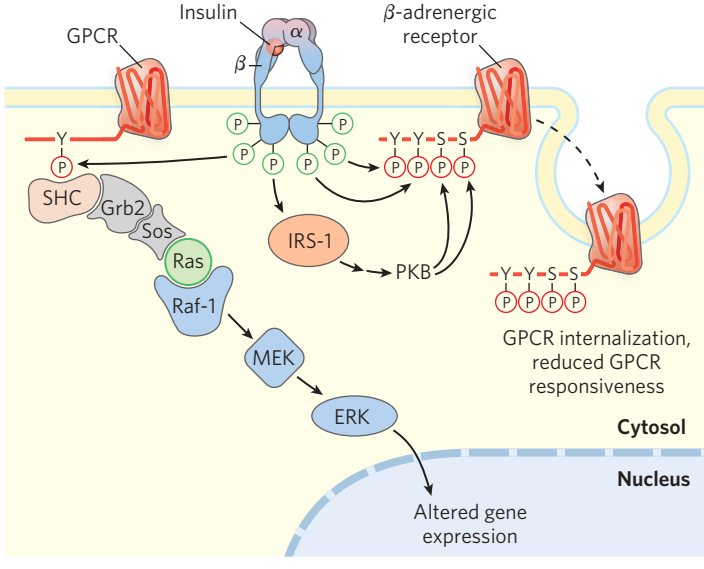


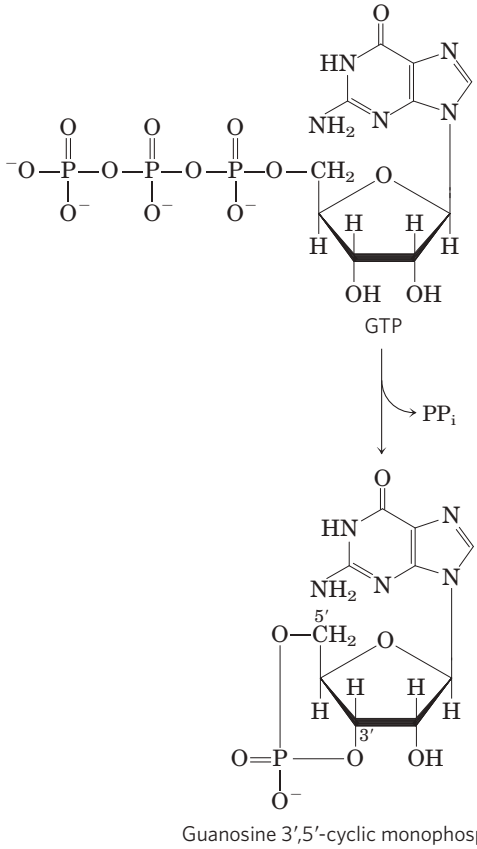
FIGURE 12–19 Cross talk between the insulin receptor and the β_2-adrenergic receptor (or other GPCR). When INSR is activated by insulin binding, its Tyr kinase directly phosphorylates the β_2-adrenergic receptor (right side) on two Tyr residues (Tyr^{350} and Tyr^{364}) near its carboxyl terminus, and indirectly (through activation of protein kinase B (PKB); see Fig. 12–16) causes phosphorylation of two Ser residues in the same region. The effect of these phosphorylations is internalization of the adrenergic receptor, reducing the response to the adrenergic stimulus. Alternatively (left side), INSR-catalyzed phosphorylation of a GPCR (an adrenergic or other receptor) on a carboxyl-terminal Tyr creates the point of nucleation for activating the MAPK cascade (see Fig. 12–15), with Grb2 serving as the adaptor protein. In this case, INSR has used the GPCR to enhance its own signaling.

and nucleus. The result is specific metabolic changes and altered gene expression.

- The enzyme PI3K, activated by interaction with IRS-1, converts the membrane lipid PIP_2 to PIP_3, which becomes the point of nucleation for proteins in a second and third branch of insulin signaling.
- In the JAK-STAT signaling system, a soluble protein Tyr kinase (JAK) is activated by association with a receptor, and then phosphorylates the transcription factor STAT, which enters the nucleus and alters the expression of a set of genes.
- There are extensive interconnections among signaling pathways, allowing integration and fine-tuning of multiple hormonal effects.

12.4 Receptor Guanylyl Cyclases, cGMP, and Protein Kinase G

Guanylyl cyclases **(Fig. 12–20)** are receptor enzymes that, when activated, convert GTP to the second messenger **guanosine 3′,5′-cyclic monophosphate (cyclic GMP, cGMP)**:



Many of the actions of cGMP in animals are mediated by **cGMP-dependent protein kinase**, also called **protein kinase G (PKG)**. On activation by cGMP, PKG phosphorylates Ser and Thr residues in target proteins.

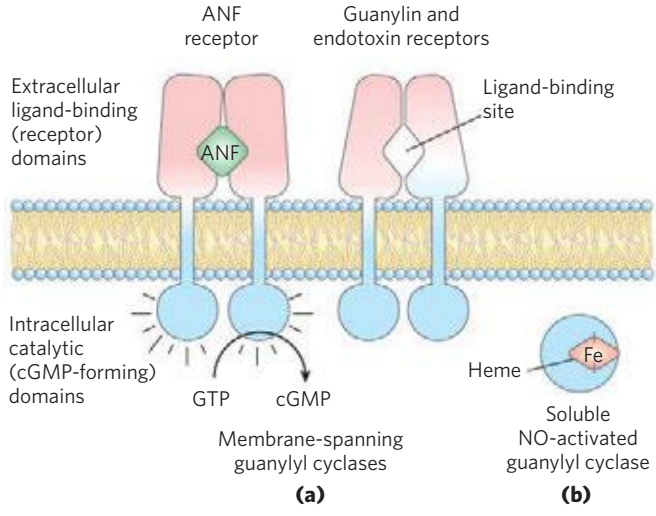


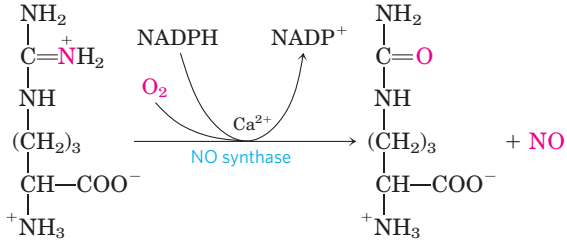
FIGURE 12–20 Two types of guanylyl cyclase that participate in signal transduction. (a) One type is a homodimer with a single membrane-spanning segment in each monomer, connecting the extracellular ligand-binding domain and the intracellular guanylyl cyclase domain. Receptors of this type are used to detect two extracellular ligands: atrial natriuretic factor (ANF; receptors in cells of the renal collecting ducts and vascular smooth muscle) and guanylin (peptide hormone produced in the intestine, with receptors in intestinal epithelial cells). The guanylin receptor is also the target of a bacterial endotoxin that triggers severe diarrhea. **(b)** The other type is a soluble heme-containing enzyme that is activated by intracellular nitric oxide (NO); this form is present in many tissues, including smooth muscle of the heart and blood vessels.

The catalytic and regulatory domains of this enzyme are in a single polypeptide (M_r ~80,000). Part of the regulatory domain fits snugly in the substrate-binding cleft. Binding of cGMP forces this pseudosubstrate out of the binding site, opening the site to target proteins containing the PKG consensus sequence.

Cyclic GMP carries different messages in different tissues. In the kidney and intestine it triggers changes in ion transport and water retention; in cardiac muscle (a type of smooth muscle) it signals relaxation; in the brain it may be involved both in development and in adult brain function. Guanylyl cyclase in the kidney is activated by the peptide hormone **atrial natriuretic factor (ANF)**, which is released by cells in the cardiac atrium when the heart is stretched by increased blood volume. Carried in the blood to the kidney, ANF activates guanylyl cyclase in cells of the collecting ducts (Fig. 12–20a). The resulting rise in [cGMP] triggers increased renal excretion of Na^+ and consequently of water, driven by the change in osmotic pressure. Water loss reduces the blood volume, countering the stimulus that initially led to ANF secretion. Vascular smooth muscle also has an ANF receptor–guanylyl cyclase; on binding to this receptor, ANF causes relaxation (vasodilation) of the blood vessels, which increases blood flow while decreasing blood pressure.

A similar receptor guanylyl cyclase in the plasma membrane of epithelial cells lining the intestine is activated by the peptide **guanylin** (Fig. 12–20a), which regulates Cl^- secretion in the intestine. This receptor is also the target of a heat-stable peptide endotoxin produced by *Escherichia coli* and other gram-negative bacteria. The elevation in [cGMP] caused by the endotoxin increases Cl^- secretion and consequently decreases reabsorption of water by the intestinal epithelium, producing diarrhea.

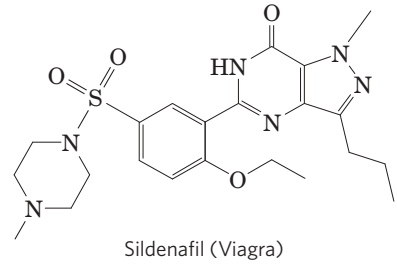
A distinctly different type of guanylyl cyclase is a cytosolic protein with a tightly associated heme group (Fig. 12–20b), an enzyme activated by nitric oxide (NO). Nitric oxide is produced from arginine by Ca^{2+}-dependent **NO synthase**, present in many mammalian tissues, and diffuses from its cell of origin into nearby cells.



NO is sufficiently nonpolar to cross plasma membranes without a carrier. In the target cell, it binds to the heme group of guanylyl cyclase and activates cGMP production. In the heart, cGMP-dependent protein kinase reduces the forcefulness of contractions by stimulating the ion pump(s) that remove Ca^{2+} from the cytosol.

NO-induced relaxation of cardiac muscle is the same response brought about by nitroglycerin and other nitrovasodilators taken to relieve **angina pectoris**, the pain caused by contraction of a heart deprived of O_2 because of blocked coronary arteries. Nitric oxide is unstable and its action is brief; within seconds of its formation, it undergoes oxidation to nitrite or nitrate. Nitrovasodilators produce long-lasting relaxation of cardiac muscle because they break down over several hours, yielding a steady stream of NO. The value of nitroglycerin as a treatment for angina was discovered serendipitously in factories producing nitroglycerin as an explosive in the 1860s. Workers with angina reported that their condition was much improved during the workweek but worsened on weekends. The physicians treating these workers heard this story so often that they made the connection, and a drug was born.

The effects of increased cGMP synthesis diminish after the stimulus ceases, because a specific phosphodiesterase (cGMP PDE) converts cGMP to the inactive 5′-GMP. Humans have several isoforms of cGMP PDE, with different tissue distributions. The isoform in the blood vessels of the penis is inhibited by the drug sildenafil (Viagra), which therefore causes [cGMP] to remain elevated once raised by an appropriate stimulus, accounting for the usefulness of this drug in the treatment of erectile dysfunction.



Sildenafil (Viagra)

Cyclic GMP has another mode of action in the vertebrate eye: it causes ion-specific channels to open in the retinal rod and cone cells. We return to this role of cGMP in the discussion of vision in Section 12.10.

SUMMARY 12.4 Receptor Guanylyl Cyclases, cGMP, and Protein Kinase G

- Several signals, including atrial natriuretic factor and guanylin, act through receptor enzymes with guanylyl cyclase activity. The cGMP so produced is a second messenger that activates cGMP-dependent protein kinase (PKG). This enzyme alters metabolism by phosphorylating specific enzyme targets.
- Nitric oxide is a short-lived messenger that stimulates a soluble guanylyl cyclase, raising [cGMP] and stimulating PKG.

12.5 Multivalent Adaptor Proteins and Membrane Rafts

Two generalizations have emerged from studies of signaling systems such as those we have discussed so far: (1) protein kinases that phosphorylate Tyr, Ser, and Thr residues are central to signaling, *directly* affecting the activities of a large number of protein substrates by phosphorylation, and (2) protein-protein interactions brought about by the reversible phosphorylation of Tyr, Ser, and Thr residues in signaling proteins create *docking sites* for other proteins that bring about *indirect* effects on proteins downstream in the signaling pathway. In fact, many signaling proteins are *multivalent*—they can interact with several different proteins simultaneously to form multiprotein signaling complexes. In this section we present a few examples to illustrate the general principles of phosphorylation-dependent protein interactions in signaling pathways.

Protein Modules Bind Phosphorylated Tyr, Ser, or Thr Residues in Partner Proteins

The protein Grb2 in the insulin signaling pathway (Figs 12–15 and 12–19) binds through its SH2 domain to other proteins that have exposed Ⓟ–Tyr residues. The

human genome encodes at least 87 SH2-containing proteins, many already known to participate in signaling. The Ⓟ–Tyr residue is bound in a deep pocket in an SH2 domain, with each of its phosphate oxygens participating in hydrogen bonding or electrostatic interactions; the positive charges on two Arg residues figure prominently in the binding. Subtle differences in the structure of SH2 domains account for the specificities of the interactions of SH2-containing proteins with various Ⓟ–Tyr-containing proteins. The SH2 domain typically interacts with a Ⓟ–Tyr (which is assigned the index position 0) and the next three residues toward the carboxyl terminus (designated +1, +2, +3). Some SH2 domains (Src, Fyn, Hck, Nck) favor negatively charged residues in the +1 and +2 positions; others (PLCγ1, SHP2) have a long hydrophobic groove that binds to aliphatic residues in positions +1 to +5. These differences define subclasses of SH2 domains that have different partner specificities.

Phosphotyrosine-binding domains **(PTB domains**; **Fig. 12–21)** are another binding partner for Ⓟ–Tyr proteins, but their critical sequences and three-dimensional structure distinguish them from SH2 domains. The human genome encodes 24 proteins that contain PTB domains, including IRS-1, which we have already encountered in its role as an adaptor protein in insulin-signal transduction (Fig. 12–15). The Ⓟ–Tyr binding sites for SH2 and PTB domains on partner proteins are created by Tyr kinases and eliminated by protein tyrosine phosphatases (PTPases).

Other signaling protein kinases, including PKA, PKC, PKG, and members of the MAPK cascade, phosphorylate Ser or Thr residues in their target proteins, which in some cases acquire the ability to interact with partner proteins through the phosphorylated residue, triggering a downstream process. An alphabet soup of domains that bind Ⓟ–Ser or Ⓟ–Thr residues has been identified, and more are sure to be found. Each domain favors a certain sequence around the phosphorylated residue, so the domains represent families of highly specific recognition sites, able to bind to a specific subset of phosphorylated proteins.

In some cases, the region on a protein that binds Ⓟ–Tyr of a substrate protein is masked by its interaction with a Ⓟ–Tyr in the same protein. For example, the soluble protein Tyr kinase Src, when phosphorylated on a critical Tyr residue, is rendered inactive; an SH2 domain needed to bind to the substrate protein instead binds to the internal Ⓟ–Tyr. When this Ⓟ–Tyr residue is hydrolyzed by a phosphoprotein phosphatase, the Tyr kinase activity of Src is activated **(Fig. 12–22a)**. Similarly, glycogen synthase kinase 3 (GSK3) is inactive

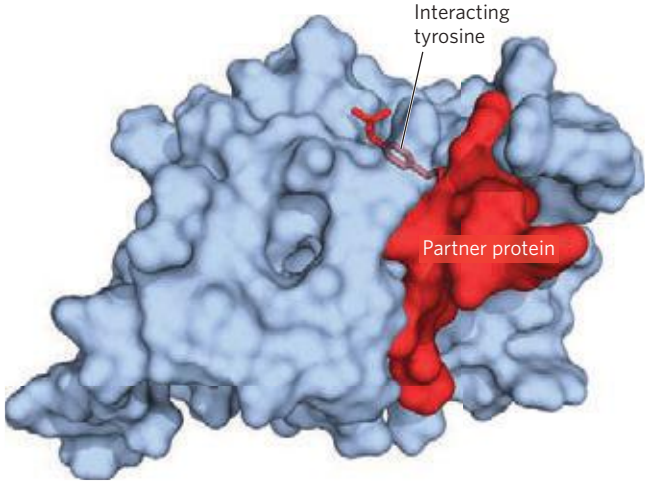


FIGURE 12–21 Interaction of a PTB domain with a Ⓟ–Tyr residue in a partner protein. (PDB ID 1SHC) The PTB domain is represented as a blue surface contour. The Ⓟ–Tyr residue of the partner protein (red) projects into a binding pocket in the PTB domain and is held firmly by multiple noncovalent interactions.

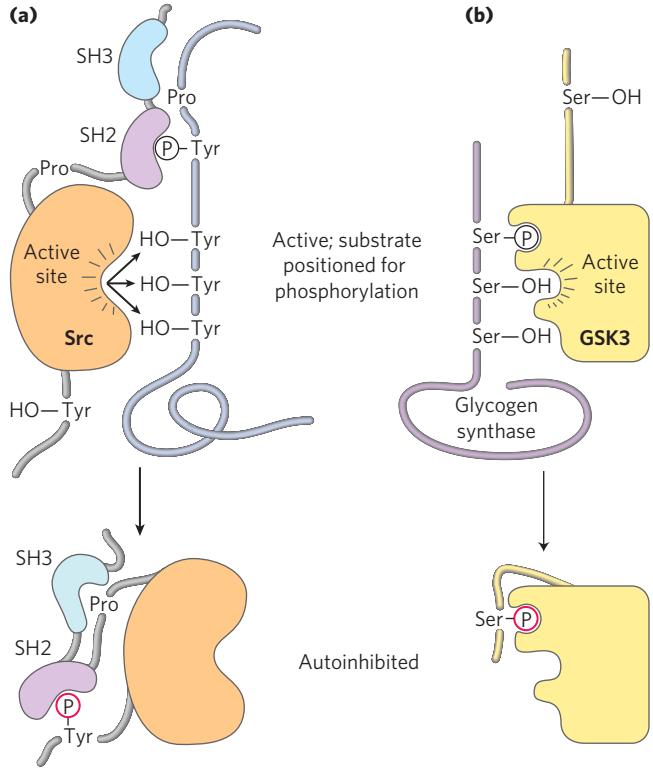


FIGURE 12–22 Mechanism of autoinhibition of Src and GSK3. (a) In the active form of the Tyr kinase Src, an SH2 domain binds a Ⓟ–Tyr in the protein substrate, and an SH3 domain binds a proline-rich region of the substrate, lining up the active site of the kinase with several target Tyr residues in the substrate (top). When Src is phosphorylated on a specific Tyr residue (bottom), the SH2 domain binds the internal Ⓟ–Tyr instead of the Ⓟ–Tyr of the substrate, and the SH3 domain binds an internal proline-rich region, preventing productive enzyme-substrate binding; the enzyme is thus autoinhibited. **(b)** In the active form of glycogen synthase kinase 3 (GSK3), an internal Ⓟ–Ser-binding domain is available to bind Ⓟ–Ser in its substrate (glycogen synthase), and thus to position the kinase to phosphorylate neighboring Ser residues (top). Phosphorylation of an internal Ser residue allows this internal kinase segment to occupy the Ⓟ–Ser-binding site, blocking substrate binding (bottom).

when phosphorylated on a Ser residue in its autoinhibitory domain (Fig. 12–22b). Dephosphorylation of that domain frees the enzyme to bind (and then phosphorylate) its target proteins.

In addition to the three commonly phosphorylated residues in proteins, there is a fourth structure that nucleates the formation of supramolecular complexes of signaling proteins: the phosphorylated head group of the membrane phosphatidylinositols. Many signaling proteins contain domains such as SH3 and PH (plextrin homology domain) that bind tightly to PIP_3 protruding from the inner leaflet of the plasma membrane. Wherever the enzyme PI3K creates this head group (as it does in response to the insulin signal), proteins that bind it will congregate at the membrane surface.

Most of the proteins involved in signaling at the plasma membrane have one or more protein- or phospholipid-binding domains; many have three or more, and thus are multivalent in their interactions with other signaling proteins. **Figure 12–23** shows just a few of the multivalent proteins known to participate in signaling. Many of the complexes include components with membrane-binding domains. Given the location of so many signaling processes at the inner surface of the plasma membrane, the molecules that must collide to produce the signaling response are effectively confined to two-dimensional space—the membrane surface; collisions here are far more likely than in the three-dimensional space of the cytosol.

In summary, a remarkable picture of signaling pathways has emerged from studies of many signaling proteins and their multiple binding domains. An initial signal results in phosphorylation of the receptor or a target protein, triggering the assembly of large multiprotein complexes, held together on scaffolds with multivalent binding capacities. Some of these complexes contain several protein kinases that activate each other in turn, producing a cascade of phosphorylation and a great amplification of the initial signal. The interactions between cascade kinases are not left to the vagaries of random collisions in three-dimensional space. In the MAPK cascade, for example, a scaffold protein, KSR,

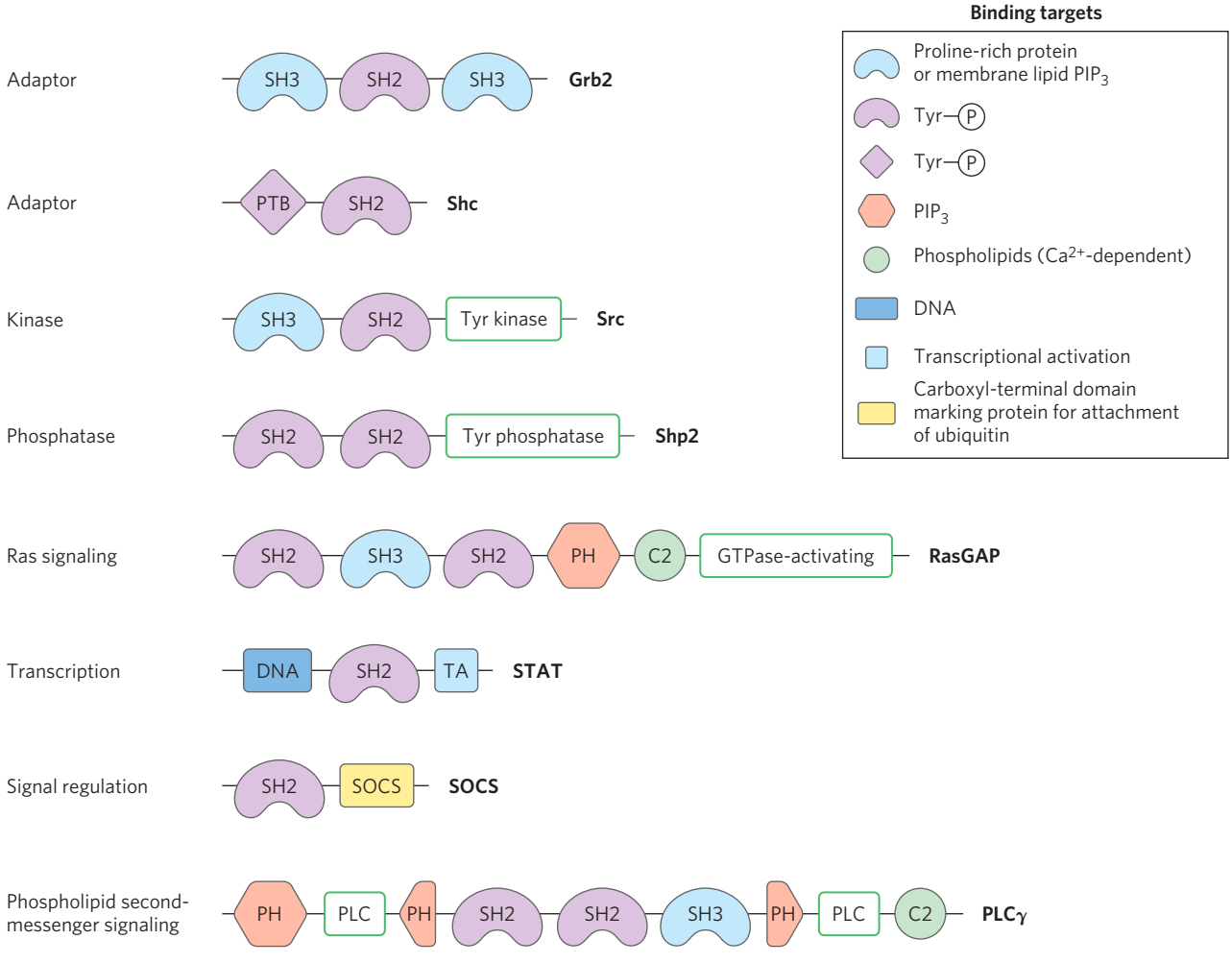


FIGURE 12–23 Some binding modules of signaling proteins. Each protein is represented by a line (with the amino terminus to the left); symbols indicate the location of conserved binding domains (with specificities as listed in the key; abbreviations are explained in the text); green boxes indicate catalytic activities. The name of each protein is given at its carboxyl-terminal end. These signaling proteins interact with phosphorylated proteins or phospholipids in many permutations and combinations to form integrated signaling complexes.

binds all three kinases (MAPK, MAPKK, and MAPKKK), assuring their proximity and correct orientation and even conferring allosteric properties on the interactions among the kinases, which makes their serial phosphorylation sensitive to very small stimuli **(Fig. 12–24)**.

Phosphotyrosine phosphatases remove the phosphate from Ⓟ–Tyr residues, reversing the effect of phosphorylation. Some of these are receptorlike membrane proteins, presumably controlled by extracellular factors not yet identified; other PTPases are soluble and contain SH2 domains. In addition, animal cells have protein Ⓟ–Ser and Ⓟ–Thr phosphatases, which reverse the effects of Ser- and Thr-specific protein kinases. We can see, then, that signaling occurs in *protein circuits*, which are effectively hardwired from signal receptor to response effector and can be switched off instantly by the hydrolysis of a single upstream phosphate ester bond.

The multivalency of signaling proteins allows for the assembly of many different combinations of signaling modules, each combination suited to particular signals, cell types, and metabolic circumstances, yielding diverse signaling circuits of extraordinary complexity.

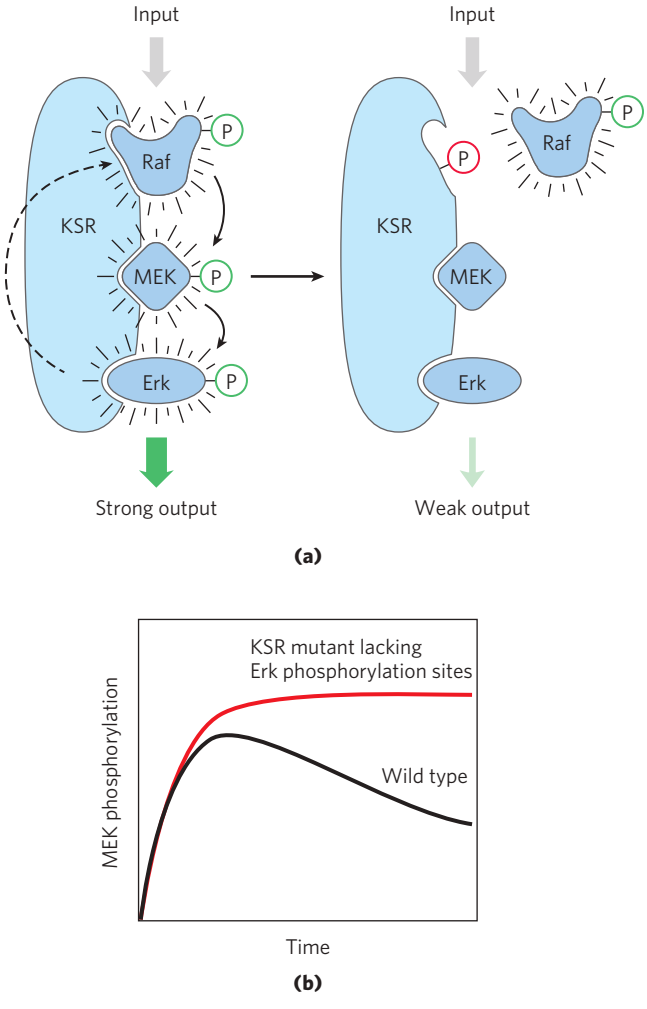


FIGURE 12–24 A scaffold protein from yeast that organizes and regulates a protein kinase cascade. (a) The scaffold protein KSR has binding sites for all three of the kinases in the Raf/MEK/Erk cascade. By binding all three in appropriate orientations, the scaffold makes interactions among the proteins rapid and efficient. When Erk has been activated (left), it phosphorylates the binding site for Raf (right), forcing a conformational change that displaces Raf and thereby prevents the phosphorylation of MEK. The result of this feedback regulation is that MEK phosphorylation is temporary. **(b)** In yeast cells with mutant KSR lacking the phosphorylation sites (red curve), no feedback occurs, producing a different time course of signaling.

Membrane Rafts and Caveolae May Segregate Signaling Proteins

Membrane rafts (Chapter 11) are regions of the membrane bilayer enriched in sphingolipids, sterols, and certain proteins, including many attached to the bilayer by GPI anchors. The β-adrenergic receptor is segregated in rafts that contain G proteins, adenylyl cyclase, PKA, and a specific protein phosphatase, PP2, which together provide a highly integrated signaling unit. By segregating in a small region of the plasma membrane all of the elements required for responding to and ending the signal, the cell is able to produce a highly localized and brief "puff" of second messenger.

Some RTKs (EGFR and PDGFR) seem to be localized in rafts, and this sequestration is very probably functionally significant. When cholesterol is removed from rafts by treatment of the membrane with cyclodextrin (which binds and removes cholesterol), the rafts are disrupted and the RTK signaling pathways become defective.

If an RTK in a raft is phosphorylated, and the only locally available PTPase that reverses this phosphorylation is in another raft, then dephosphorylation of the RTK is slowed or prevented. Interactions between adaptor proteins might be strong enough to recruit into a raft a signaling protein not usually located there, or might even be strong enough to pull receptors out of a raft. For example, the EGFR in isolated fibroblasts is usually concentrated in specialized rafts called caveolae (see Fig. 11–22), but treatment with EGF causes the receptor to leave the raft. This migration depends on the receptor's protein kinase activity; mutant receptors lacking this activity remain in the raft during treatment with EGF. Caveolin, an integral membrane protein localized in caveolae, is phosphorylated on Tyr residues in response to insulin, and the now-activated EGFR may be able to draw its binding partners into the raft. Spatial segregation of signaling proteins in rafts adds yet another dimension to the already complex processes initiated by extracellular signals.

SUMMARY 12.5 Multivalent Adaptor Proteins and Membrane Rafts

- Many signaling proteins have domains that bind phosphorylated Tyr, Ser, or Thr residues in other proteins; the binding specificity for each domain is

determined by sequences that adjoin the phosphorylated residue in the substrate.

- SH2 and PTB domains bind to proteins containing Ⓟ–Tyr residues; other domains bind Ⓟ–Ser and Ⓟ–Thr residues in various contexts.
- SH3 and PH domains bind the membrane phospholipid PIP_3.
- Many signaling proteins are multivalent, with several different binding modules. By combining the substrate specificities of various protein kinases with the specificities of domains that bind phosphorylated Ser, Thr, or Tyr residues, and with phosphatases that can rapidly inactivate a signaling pathway, cells create a large number of multiprotein signaling complexes.
- Membrane rafts and caveolae sequester groups of signaling proteins in small regions of the plasma membrane, enhancing their interactions and making signaling more efficient.

12.6 Gated Ion Channels

Ion Channels Underlie Electrical Signaling in Excitable Cells

Certain cells in multicellular organisms are "excitable": they can detect an external signal, convert it into an electrical signal (specifically, a change in membrane potential), and pass it on. Excitable cells play central roles in nerve conduction, muscle contraction, hormone secretion, sensory processes, and learning and memory. The excitability of sensory cells, neurons, and myocytes depends on ion channels, signal transducers that provide a regulated path for the movement of inorganic ions such as Na^+, K^+, Ca^{2+}, and Cl^- across the plasma membrane in response to various stimuli. Recall from Chapter 11 that these ion channels are "gated": they may be open or closed, depending on whether the associated receptor has been activated by the binding of its specific ligand (a neurotransmitter, for example) or by a change in the transmembrane electrical potential, V_m. The Na^+K^+ ATPase is electrogenic; it creates a charge imbalance across the plasma membrane by carrying 3 Na^+ out of the cell for every 2 K^+ carried in **(Fig. 12–25a)**, making the inside negative relative to the outside. The membrane is said to be polarized.

KEY CONVENTION: V_m is negative when the inside of the cell is negative relative to the outside. For a typical animal cell, $V_m = -50$ to -70 mV. ■

Because ion channels generally allow passage of either anions or cations but not both, ion flux through a channel causes a redistribution of charge on the two sides of the membrane, changing V_m. Influx of a positively charged ion such as Na^+, or efflux of a negatively charged ion such as Cl^-, *depolarizes* the membrane and brings V_m closer to zero. Conversely, efflux of K^+ *hyperpolarizes* the membrane and V_m becomes more negative. These ion fluxes through channels are passive, in contrast to active transport by the Na^+K^+ ATPase.

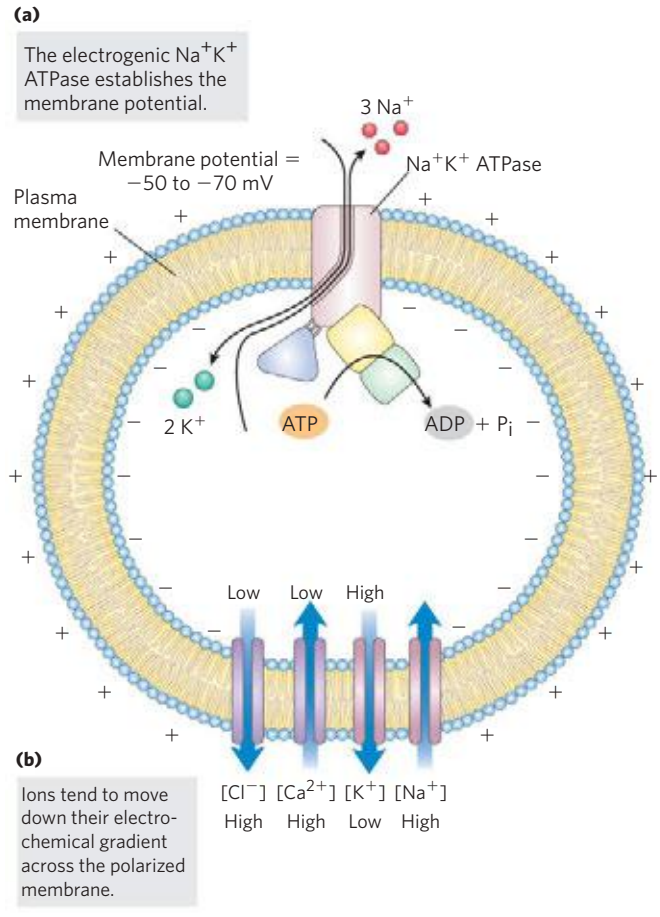


FIGURE 12–25 Transmembrane electrical potential. (a) The electrogenic Na^+K^+ ATPase produces a transmembrane electrical potential of about −60 mV (inside negative). **(b)** Blue arrows show the direction in which ions tend to move spontaneously across the plasma membrane in an animal cell, driven by the combination of chemical and electrical gradients. The chemical gradient drives Na^+ and Ca^{2+} inward (producing depolarization) and K^+ outward (producing hyperpolarization). The electrical gradient drives Cl^- outward, against its concentration gradient (producing depolarization).

The direction of spontaneous ion flow across a polarized membrane is dictated by the electrochemical potential of that ion across the membrane, which has two components: the difference in concentration (C) of the ion on the two sides of the membrane, and the difference in electrical potential, typically expressed in millivolts. The force (ΔG) that causes a cation (say, Na^+) to pass spontaneously inward through an ion channel is a function of the ratio of its concentrations on the two sides of the membrane (C_{in}/C_{out}) and of the difference in electrical potential (V_m or $\Delta\psi$):

$$\Delta G = RT \ln (C_{in}/C_{out}) + Z\mathcal{F}V_m \qquad (12\text{–}1)$$

where R is the gas constant, T the absolute temperature, Z the charge on the ion, and $\mathcal{F}$ the Faraday constant. (Note that the sign of the charge on the ion determines the sign of the second term in Eqn 12–1.) In a typical neuron or myocyte, the concentrations of Na^+, K^+, Ca^{2+}, and Cl^- in the cytosol are very different from those in the extracellular fluid (Table 12–6). Given these concentration differences, the resting V_m of about −60 mV, and the relationship shown in Equation 12–1, the opening of a Na^+ or Ca^{2+} channel will result in a spontaneous inward flow of Na^+ or Ca^{2+} (and depolarization), whereas opening of a K^+ channel will result in a spontaneous outward flux of K^+ (and hyperpolarization) (Fig. 12–25b). In this case, K^+ moves outward, against the electrical gradient, because the large concentration difference inside and outside the cell produces a more powerful, outward chemical force on the ion.

A given ionic species continues to flow through a channel only as long as the combination of concentration gradient and electrical potential provides a driving force. For example, as Na^+ flows down its concentration gradient, it depolarizes the membrane. When the membrane potential reaches +70 mV, the effect of this membrane potential (resistance to further entry of Na^+) exactly equals the effect of the $[Na^+]$ gradient (promotion of Na^+ flow inward). At this equilibrium potential (E), the driving force (ΔG) tending to move a Na^+ ion is zero. The equilibrium potential is different for each ionic species, because the concentration gradients differ.

The number of ions that must flow to produce a physiologically significant change in the membrane potential is negligible relative to the concentrations of Na^+, K^+, and Cl^- in cells and extracellular fluid, so the ion fluxes that occur during signaling in excitable cells have essentially no effect on the concentrations of these ions. With Ca^{2+}, the situation is different; because the intracellular $[Ca^{2+}]$ is generally very low ($\sim 10^{-7}$ M), inward flow of Ca^{2+} can significantly alter the cytosolic $[Ca^{2+}]$.

The membrane potential of a cell at a given time is the result of the types and numbers of ion channels open at that instant. In most cells at rest, more K^+ channels than Na^+, Cl^-, or Ca^{2+} channels are open and thus the resting potential is closer to the E for K^+ (–98 mV) than that for any other ion. When channels for Na^+, Ca^{2+}, or Cl^- open, the membrane potential moves toward the E for that ion. The precisely timed opening and closing of ion channels and the resulting transient changes in membrane potential underlie the electrical signaling by which the nervous system stimulates the skeletal muscles to contract, the heart to beat, or secretory cells to release their contents. Moreover, many hormones exert their effects by altering the membrane potential of their target cells. These mechanisms are not limited to animals; ion channels play important roles in the responses of bacteria, protists, and plants to environmental signals.

To illustrate the action of ion channels in cell-to-cell signaling, we describe the mechanisms by which a neuron passes a signal along its length and across a synapse to the next neuron (or to a myocyte) in a cellular circuit, using acetylcholine as the neurotransmitter.

Voltage-Gated Ion Channels Produce Neuronal Action Potentials

Signaling in the nervous system is accomplished by networks of neurons, specialized cells that carry an electrical impulse (action potential) from one end of the cell (the cell body) through an elongated cytoplasmic extension (the axon). The electrical signal triggers release of neurotransmitter molecules at the synapse, carrying the signal to the next cell in the circuit. Three types of **voltage-gated ion channels** are essential to this signaling mechanism. Along the entire length of the axon are **voltage-gated Na^+ channels (Fig. 12–26)**, which are closed when the membrane is at rest (V_m = −60 mV) but open briefly when the membrane is depolarized locally in response to acetylcholine (or some other neurotransmitter). Also distributed along the axon are **voltage-gated K^+ channels**, which open, a split second later, in response to the depolarization when nearby Na^+ channels open. The depolarizing flow of Na^+ into the axon (influx) is thus rapidly countered by a repolarizing flow of K^+ out (efflux). At the distal end of the axon are **voltage-gated Ca^{2+} channels**, which open when the wave of depolarization (step ❶) and repolarization (step ❷) caused by the activity of Na^+ and K^+ channels arrives, triggering release of the neurotransmitter acetylcholine—which carries the signal to another neuron (fire an action potential!) or to a muscle fiber (contract!).

TABLE 12–6 Ion Concentrations in Cells and Extracellular Fluids (mM)

	K^+		Na^+		Ca^{2+}		Cl^-	
Cell type	In	Out	In	Out	In	Out	In	Out
Squid axon	400	20	50	440	≤0.4	10	40–150	560
Frog muscle	124	2.3	10.4	109	<0.1	2.1	1.5	78

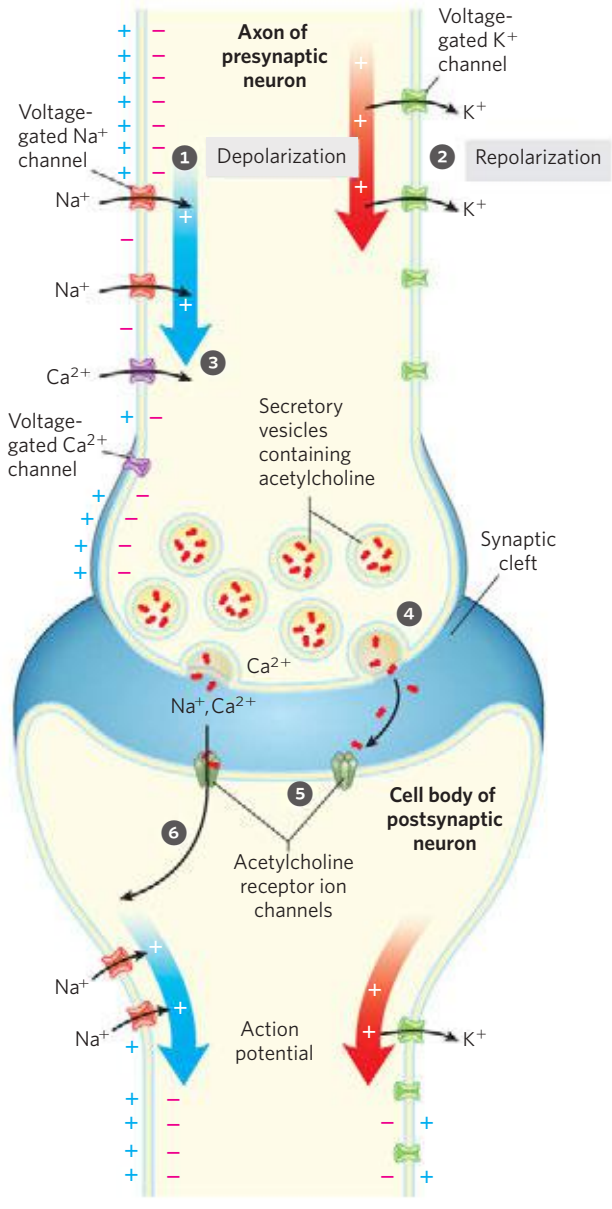


FIGURE 12–26 Role of voltage-gated and ligand-gated ion channels in neural transmission. Initially, the plasma membrane of the presynaptic neuron is polarized (inside negative) through the action of the electrogenic Na^+K^+ ATPase, which pumps out 3 Na^+ for every 2 K^+ pumped in (see Fig. 12–25). ❶ A stimulus to this neuron (not shown) causes an action potential to move along the axon (blue arrow), away from the cell body. The opening of a voltage-gated Na^+ channel allows Na^+ entry, and the resulting local depolarization causes the adjacent Na^+ channel to open, and so on. The directionality of movement of the action potential is ensured by the brief refractory period that follows the opening of each voltage-gated Na^+ channel. ❷ A split second after the action potential passes a point in the axon, voltage-operated K^+ channels open, allowing K^+ exit that brings about repolarization of the membrane (red arrow), to make it ready for the next action potential. (For clarity, Na^+ channels and K^+ channels are drawn on opposite sides of the axon; both types of channels are uniformly distributed in the axonal membrane.) ❸ When the wave of depolarization reaches the axon tip, voltage-gated Ca^{2+} channels open, allowing Ca^{2+} entry. ❹ The resulting increase in internal $[Ca^{2+}]$ triggers exocytic release of the neurotransmitter acetylcholine into the synaptic cleft. ❺ Acetylcholine binds to a receptor on the postsynaptic neuron (or myocyte), causing its ligand-gated ion channel to open. ❻ Extracellular Na^+ and Ca^{2+} enter through this channel, depolarizing the postsynaptic cell. The electrical signal has thus passed to the cell body of the postsynaptic neuron (or myocyte) and will move along its axon to a third neuron (or a myocyte) by this same sequence of events.

The voltage-gated Na^+ channels are very selective for Na^+ over other cations (by a factor of 100 or more) and have a very high flux rate ($>10^7$ ions/s). After being opened—activated—by a reduction in transmembrane electrical potential, a Na^+ channel undergoes very rapid inactivation—within milliseconds, the channel closes and remains inactive for many milliseconds. As voltage-gated K^+ channels open in response to the depolarization induced by the opening of Na^+ channels (step ❶ in Fig. 12–26), the resulting efflux of K^+ repolarizes the membrane locally (it reestablishes the inside-negative membrane potential; step ❷). A brief pulse of depolarization thus traverses the axon as local depolarization triggers the brief opening of neighboring Na^+ channels, then K^+ channels. The short refractory period that follows the opening of each Na^+ channel, during which it cannot open again, ensures that a unidirectional wave of depolarization—the action potential—sweeps from the nerve cell body toward the end of the axon.

When the wave of depolarization reaches the voltage-gated Ca^{2+} channels, they open (step ❸), and Ca^{2+} enters from the extracellular space. The rise in cytoplasmic $[Ca^{2+}]$ then triggers release of acetylcholine by exocytosis into the synaptic cleft (step ❹). Acetylcholine diffuses to the postsynaptic cell (another neuron or a myocyte), where it binds to acetylcholine receptors and triggers depolarization. Thus the message is passed to the next cell in the circuit. We see, then, that gated ion channels convey signals in either of two ways: by changing the cytoplasmic concentration of an ion (such as Ca^{2+}), which then serves as an intracellular second messenger, or by changing V_m and affecting other membrane proteins that are sensitive to V_m. The passage of an electrical signal through one neuron and on to the next illustrates both types of mechanism.

We discussed the structure and mechanism of voltage-gated K^+ channels in some detail in Section 11.3 (see Figs 11–47 and 11–48). Here we take a closer look at Na^+ channels. The essential component of a Na^+ channel is a single, large polypeptide (1,840 amino acid residues) organized into four domains clustered around a central channel **(Fig. 12–27a, b)**, providing a path for Na^+ through the membrane. The path is made Na^+-specific by a "pore region" composed of the segments between transmembrane helices 5 and 6 of each

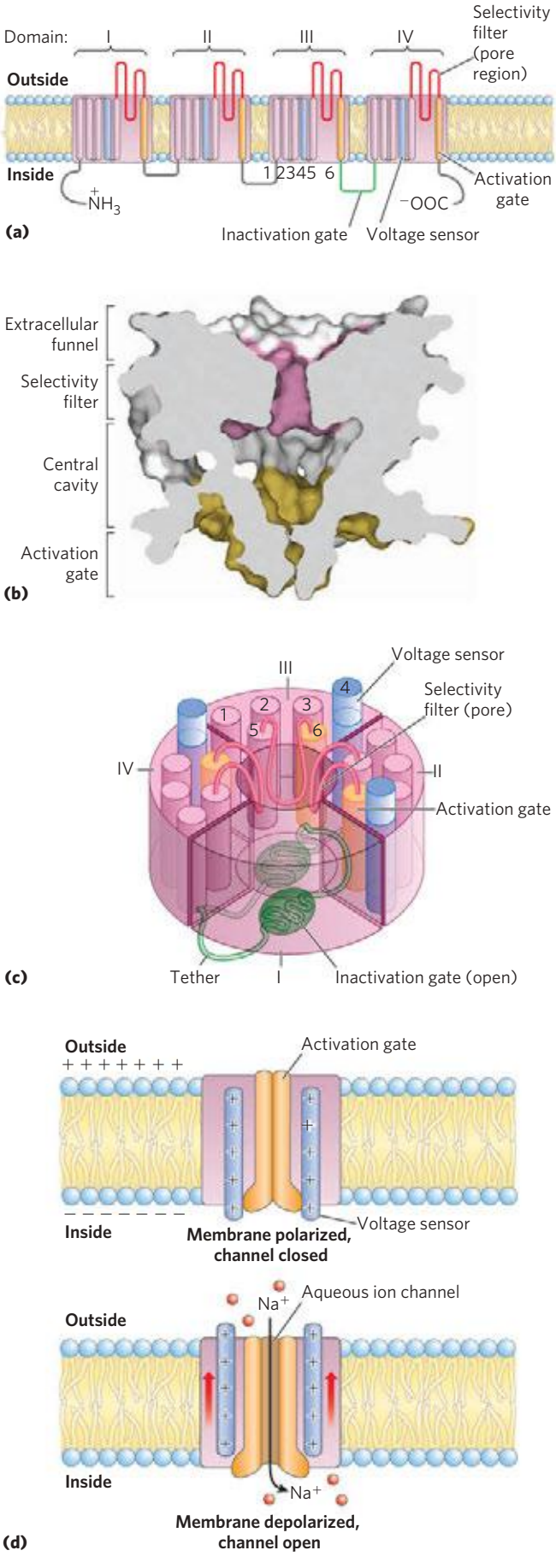


FIGURE 12–27 Voltage-gated Na^+ channels of neurons. Sodium channels of different tissues and organisms have a variety of subunits, but only the principal subunit (α) is essential. **(a)** The α subunit is a large protein with four homologous domains (I to IV, shown spread out here to illustrate the parts), each containing six transmembrane helices (1 to 6). Helix 4 in each domain (blue) is the voltage sensor; helix 6 (orange) is thought to be the activation gate. The segments between helices 5 and 6, the pore region (red), form the selectivity filter, and the segment connecting domains III and IV (green) is the inactivation gate. **(b)** (PDB ID 3RW0) Structure of a voltage-gated Na^+ channel (from the bacterium *Arcobacter butzleri*, but probably similar to channels of vertebrate neurons), with a funnel-shaped opening on the extracellular side, an ion selectivity filter, a central aqueous cavity, and the activation domain on the cytoplasmic side. **(c)** A schematic view of the Na^+ channel. The four domains are wrapped about a central transmembrane channel lined with polar amino acid residues. The four pore regions (red) come together near the extracellular surface to form the selectivity filter, which is conserved in all Na^+ channels. The filter gives the channel its ability to discriminate between Na^+ and other ions of similar size. The inactivation gate (green) closes (dotted lines) soon after the activation gate opens. **(d)** The voltage-sensing mechanism involves movement of helix 4 (blue) perpendicular to the plane of the membrane in response to a change in transmembrane potential. As shown at the top, the strong positive charge on helix 4 allows it to be pulled inward in response to the inside-negative membrane potential (V_m). Depolarization lessens this pull, and helix 4 relaxes by moving outward (bottom). This movement is communicated to the activation gate (orange), inducing conformational changes that open the channel.

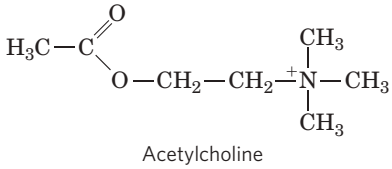
domain, which fold into the channel. Helix 4 of each domain has a high density of positively charged Arg residues; this segment is believed to move within the membrane in response to changes in the transmembrane voltage, from the resting potential of about −60 mV to about +30 mV. The movement of helix 4 triggers opening of the channel, and this is the basis for the voltage gating (Fig. 12–27c).

Inactivation of the channel (during the refractory period) is thought to occur by a ball-and-chain mechanism. A protein domain on the cytoplasmic surface of the Na^+ channel, the inactivation gate (the ball), is tethered to the channel by a short segment of the polypeptide (the chain; Fig. 12–27b). This domain is free to move about when the channel is closed, but when it opens, a site on the inner face of the channel becomes available for the tethered ball to bind, blocking the channel. The length of the tether seems to determine how long an ion channel stays open: the longer the tether, the longer the open period. Other gated ion channels may be inactivated by a similar mechanism.

The Acetylcholine Receptor Is a Ligand-Gated Ion Channel

The **nicotinic acetylcholine receptor** mediates the passage of the signal from an electrically excited neuron

at some types of synapses and at neuromuscular junctions (between motor neuron and muscle fiber), triggering muscle contraction. (Nicotinic acetylcholine receptors were originally distinguished from muscarinic acetylcholine receptors by the sensitivity of the former to nicotine, the latter to the mushroom alkaloid muscarine. They are structurally and functionally different.) Acetylcholine released by the presynaptic neuron or motor neuron diffuses a few micrometers to the plasma membrane of the postsynaptic neuron or myocyte, where it binds to the acetylcholine receptor. This forces a conformational change in the receptor, causing its ion channel to open. The resulting inward movement of cations depolarizes the plasma membrane. In a muscle fiber, this triggers contraction. The acetylcholine receptor allows ready passage of Na^+, Ca^{2+}, and K^+ ions, but other cations and all anions are unable to pass. Movement of Na^+ through an acetylcholine receptor ion channel is unsaturable (its rate is linear with respect to extracellular $[Na^+]$) and very fast—about 2×10^7 ions/s under physiological conditions.



Acetylcholine

Like other gated ion channels, the acetylcholine receptor opens in response to stimulation by its signal molecule and has an intrinsic timing mechanism that closes the gate milliseconds later. Thus the acetylcholine signal is transient—as we have seen, an essential feature of electrical signal conduction. We understand the structural changes underlying gating in the acetylcholine receptor, but not the exact mechanism of "desensitization," in which the gate remains closed even in the continued presence of acetylcholine.

The nicotinic acetylcholine receptor has five subunits ($\alpha_2\beta\gamma\delta$), each having four transmembrane helical segments (M1 to M4) **(Fig. 12–28)**. The five subunits, which are related in sequence and tertiary structure, surround a central pore, which is lined with their M2 helices. The pore is about 20 Å wide in the parts of the channel that protrude on the cytoplasmic and extracellular surfaces, but narrows as it passes through the lipid bilayer. Near the center of the bilayer is a ring of bulky hydrophobic side chains of Leu residues in the M2 helices, positioned so close together that they prevent ions from passing through the channel (Fig. 12–28d). Binding of acetylcholine to sites on each α subunit forces all M2 helices to rotate slightly, moving the bulky Leu residues aside and replacing them with smaller, polar residues. The widening of the pore allows passage of ions (Na^+ and Ca^{2+}).

Neurons Have Receptor Channels That Respond to Different Neurotransmitters

Animal cells, especially those of the nervous system, contain a variety of ion channels gated by ligands, voltage, or both. Receptors that are themselves ion channels are classified as **ionotropic** to distinguish them from receptors that generate a second messenger (metabotropic receptors). We have so far focused on acetylcholine as neurotransmitter, but there are many others. 5-Hydroxytryptamine (serotonin), glutamate, and glycine all can act through receptor channels that are structurally related to the acetylcholine receptor. Serotonin and glutamate trigger the opening of cation (K^+, Na^+, Ca^{2+}) channels, whereas glycine opens Cl^--specific channels. Cation and anion channels are distinguished by subtle differences in the amino acid residues that line the hydrophilic channel. Cation channels have negatively charged Glu and Asp side chains at crucial positions. When a few of these acidic residues are experimentally replaced with basic residues, the cation channel is converted to an anion channel.

Depending on which ion passes through a channel, binding of the ligand (neurotransmitter) for that channel results in either depolarization or hyperpolarization of the target cell. A single neuron normally receives input from many other neurons, each releasing its own characteristic neurotransmitter with its characteristic depolarizing or hyperpolarizing effect. The target cell's V_m therefore reflects the *integrated* input (Fig. 12–1e) from multiple neurons. The cell responds with an action potential only if the integrated input adds up to a net depolarization of sufficient size.

The receptor channels for acetylcholine, glycine, glutamate, and γ-aminobutyric acid (GABA) are gated by *extracellular* ligands. *Intracellular* second messengers—such as cAMP, cGMP, IP_3, Ca^{2+}, and ATP—regulate ion channels of another class, which, as we shall see in Section 12.10, participate in the sensory transductions of vision, olfaction, and gustation.

Toxins Target Ion Channels

Many of the most potent toxins found in nature act on ion channels. For example, dendrotoxin (from the black mamba snake) blocks the action of voltage-gated K^+ channels, tetrodotoxin (produced by puffer fish) acts on voltage-gated Na^+ channels, and cobrotoxin disables acetylcholine receptor ion channels. Why, in the course of evolution, have ion channels become the preferred target of toxins, rather than some critical metabolic target such as an enzyme essential in energy metabolism?

Ion channels are extraordinary amplifiers; opening of a single channel can allow the flow of 10 million ions per second. Consequently, relatively few molecules of

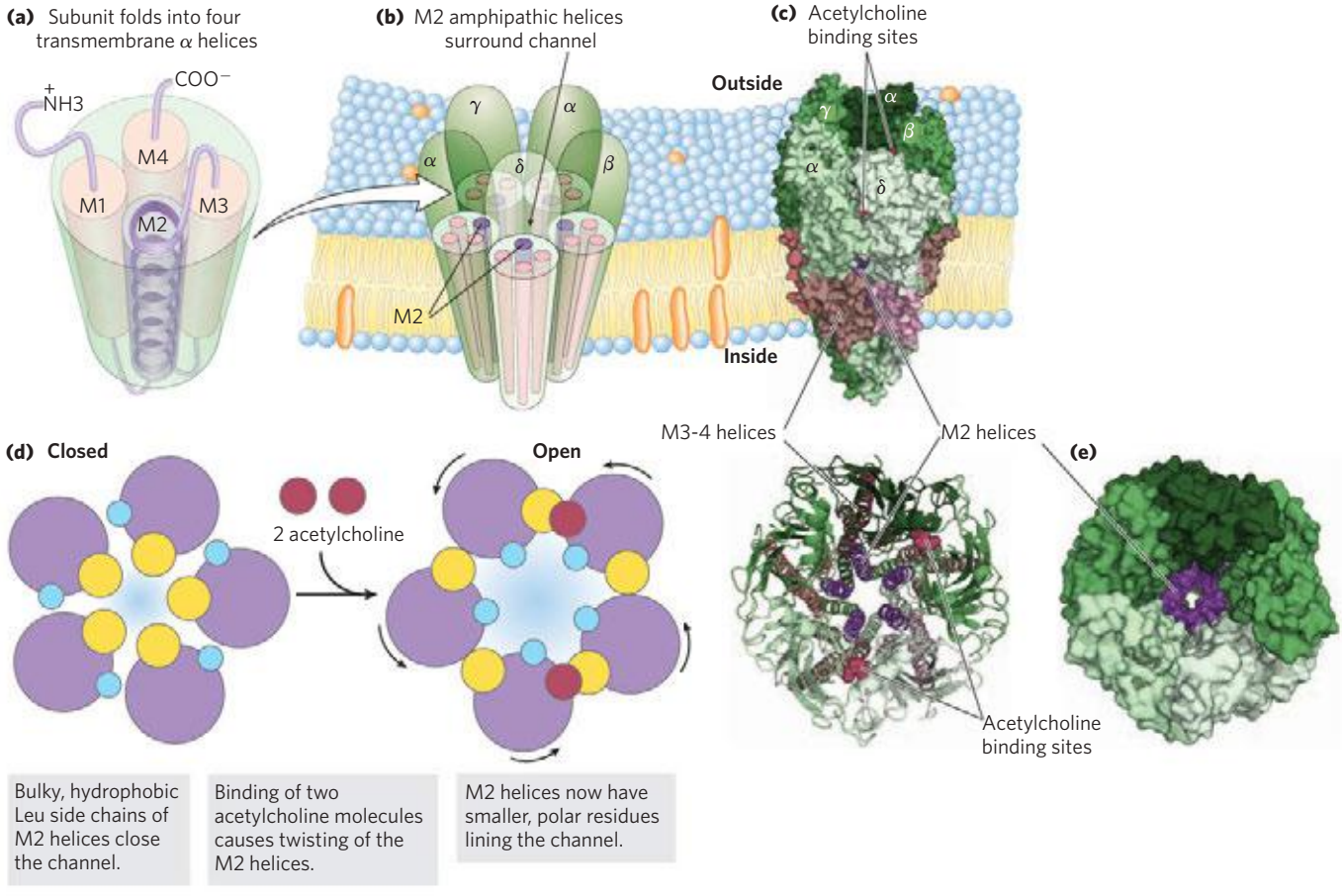


FIGURE 12–28 The acetylcholine receptor ion channel. (a) Each of the five homologous subunits ($\alpha_2\beta\gamma\delta$) has four transmembrane helices, M1 to M4. The M2 helices are amphipathic; the others have mainly hydrophobic residues. **(b)** The five subunits are arranged around a central transmembrane channel, which is lined with the polar sides of the M2 helices. At the top and bottom of the channel are rings of negatively charged amino acid residues. **(c)** A molecular model of the acetylcholine receptor, based on x-ray structure determination of a related protein (the acetylcholine-binding protein from a mollusk; PDB ID 1UV6). **(d)** This cartoon view of a cross section through the center of the M2 helices shows five Leu side chains (yellow), one from each M2 helix, protruding into the channel and constricting it to a diameter too small to allow passage of Ca^{2+}, Na^+, or K^+. When both acetylcholine receptor sites (one on each α subunit) are occupied, a conformational change occurs. As the M2 helices twist slightly, the five Leu residues rotate away from the channel and are replaced by smaller, polar residues (blue). This gating mechanism opens the channel, allowing the passage of Ca^{2+}, Na^+, or K^+. **(e)** Molecular model of the acetylcholine receptor viewed perpendicular to the membrane, showing the small central pore that allows ion passage.

an ion channel protein are needed per neuron for signaling functions. This means that a relatively small number of toxin molecules with high affinity for ion channels, acting from outside the cell, can have a very pronounced effect on neurosignaling throughout the body. A comparable effect by way of a metabolic enzyme, typically present in cells at much higher concentrations than ion channels, would require far more copies of the toxin molecule.

SUMMARY 12.6 Gated Ion Channels

- Ion channels gated by membrane potential or ligands are central to signaling in neurons and other cells.
- The voltage-gated Na^+ and K^+ channels of neuronal membranes carry the action potential along the axon as a wave of depolarization (Na^+ influx) followed by repolarization (K^+ efflux).
- The gating mechanism for voltage-sensitive channels involves the movement, perpendicular to the plane of the membrane, of a transmembrane peptide with a high charge density, due to the presence of Arg or other charged residues.
- Arrival of an action potential at the distal end of a presynaptic neuron triggers neurotransmitter release. The neurotransmitter (acetylcholine, for example) diffuses to the postsynaptic neuron (or the myocyte, at a neuromuscular junction), binds to specific receptors in the plasma membrane, and triggers a change in V_m.

- The acetylcholine receptor of neurons and myocytes is a ligand-gated ion channel; acetylcholine binding triggers a conformational change that opens the channel to Na^+ and Ca^{2+} ions.
- Neurotoxins produced by many organisms attack neuronal ion channels, and are therefore fast-acting and deadly.

12.7 Integrins: Bidirectional Cell Adhesion Receptors

Integrins are proteins of the plasma membrane that mediate the adhesion of cells to each other and to the extracellular matrix, and carry signals in both directions across the membrane **(Fig. 12–29)**. The mammalian genome encodes 18 different α subunits and 8 different β subunits, which are found in a range of combinations with various ligand-binding specificities in various tissues. Each of the 24 different integrins found thus far seems to have a unique function. Because they can inform cells about the extracellular neighborhood, integrins play crucial roles in processes that require selective cell-cell interactions, such as embryonic development, blood clotting, immune cell function, normal differentiation, and tumor growth and metastasis.

The extracellular ligands that interact with integrins include collagen, fibrinogen, fibronectin, and many other proteins that have the sequence recognized by integrins: —Arg–Gly–Asp– (RGD, using single-letter amino acid abbreviations). The short, cytoplasmic extensions of the α and β subunits interact with cytoskeletal proteins just beneath the plasma membrane—talin, α-actinin, vinculin, paxillin, and others—modulating the assembly of actin-based cytoskeletal structures. The dual association of integrins with the extracellular matrix and the cytoskeleton allows the cell to integrate information about its extracellular and intracellular environments and to coordinate cytoskeletal positioning with extracellular adhesion sites. In this capacity, integrins govern the shape, motility, polarity, and differentiation of many cell types. In "outside-in" signaling, the extracellular domains of an integrin undergo dramatic, global conformational changes when ligand binds at a site many angstroms from the transmembrane helices. These changes somehow alter the dispositions of the cytoplasmic tails of the α and β subunits, changing their interactions with intracellular proteins and thereby conducting the signal inward.

The conformation and adhesiveness of integrin extracellular domains are also dramatically altered by "inside-out" signaling initiated by signals from *inside* the cell. In one conformation, the extracellular domains have no affinity for the proteins of the extracellular

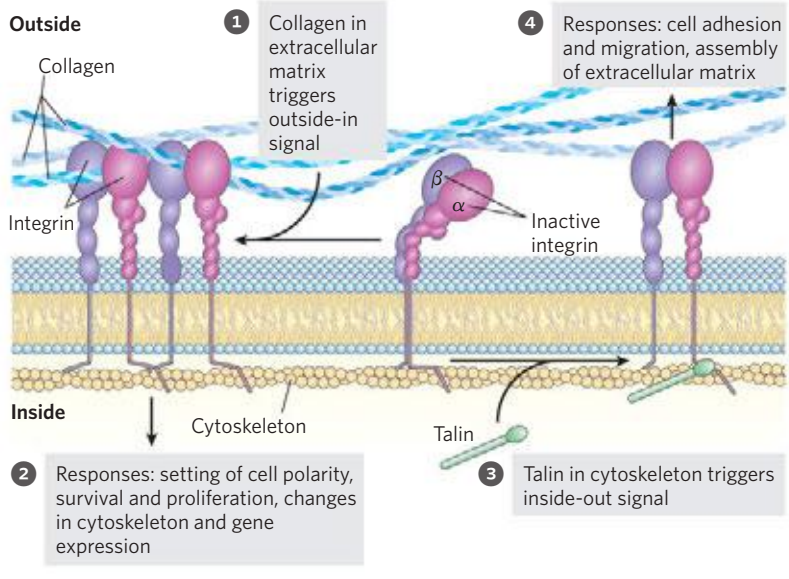


FIGURE 12–29 Two-way signaling by integrins. All integrins have one α and one β subunit, each with a short cytoplasmic extension, a single transmembrane helix, and a large extracellular domain with the ligand-binding site. The β subunit (purple) is rich in Cys residues and has extensive intrachain disulfide bonding. The α subunit (pink) in many integrins has several binding sites for divalent cations such as Ca^{2+}, which are intrinsic to the ligand-binding activity. In its inactive state, integrin's extracellular domain is folded upon itself (center). Contact with an extracellular ligand (collagen or heparan sulfate, for example) straightens the extracellular domain and moves the cytosolic tails of the α and β subunits apart (left), altering their interactions with intracellular proteins such as talin, which in turn connect the integrin to actin filaments in the cytoskeleton. In inside-out signaling, contact of the cytosolic domain with talin produces a dramatic unbending of the extracellular domain (right) and an increase in its affinity for extracellular binding partners, allowing interactions with extracellular proteins or proteoglycans and changing the cell's adhesion to the extracellular matrix. Protein ligands in the extracellular matrix have the RGD sequence recognized by integrins.

matrix, but signals from the cell can favor another conformation in which integrins adhere tightly to extracellular proteins (Fig. 12–29).

Regulation of adhesiveness is central to leukocyte homing to the site of an infection (see Fig. 7–32), interactions between immune cells, and phagocytosis by macrophages. During an immune response, for example, leukocyte integrins are activated (exposing their extracellular ligand-binding sites) from inside the cell via a signaling pathway *triggered* by cytokines (extracellular developmental signals). Thus activated, the integrins can mediate the attachment of leukocytes to other immune cells or can target cells for phagocytosis. Mutation in an integrin gene encoding the β subunit known as CD18 is the cause of leukocyte adhesion deficiency, a rare human genetic disease in which leukocytes fail to pass out of blood vessels to reach sites of infection. Infants with a severe defect in CD18 commonly die of infections before the age of two.

An integrin specific to platelets ($\alpha_{IIb}\beta_3$) is involved in both normal and pathological blood clotting. Local damage to blood vessels at a site of injury exposes high-affinity binding sites (RGD sequences in thrombin and collagen, for example) for the integrins of platelets, which attach themselves to the lesion, to other platelets, and to the clotting protein fibrinogen, leading to clot formation that prevents further bleeding. Mutations in the α or β subunit of platelet integrin $\alpha_{IIb}\beta_3$ lead to a bleeding disorder known as Glanzmann thrombasthenia, in which individuals bleed excessively after a relatively minor injury. Overly effective blood coagulation is also undesirable. Dysregulation of platelet adhesion can lead to pathological blood clot formation, resulting in blockage of the arteries that supply blood to the heart and brain and increasing the risk of heart attack and stroke. Drugs such as tirofiban and eptifibatide that block the external ligand-binding sites of platelet integrin reduce clot formation and are useful in treating and preventing heart attacks and strokes.

When tumors metastasize, tumor cells lose their adhesion to the originating tissue and invade new locations. Both the changes in tumor cell adhesion and the development of new blood vessels (angiogenesis) to support the tumor at a new location are modulated by specific integrins. These proteins are therefore potential targets for drugs that suppress the migration and relocation of tumor cells. ■

SUMMARY 12.7 Integrins: Bidirectional Cell Adhesion Receptors

- Integrins are a family of dimeric ($\alpha\beta$) plasma membrane receptors that interact with extracellular macromolecules and the cytoskeleton, carrying signals in and out of the cell.
- The active and inactive forms of an integrin differ in the conformation of their extracellular domains. Intracellular events and signals can interconvert the active and inactive forms.
- Integrins mediate various aspects of the immune response, blood clotting, and angiogenesis, and they play a role in tumor metastasis.

12.8 Regulation of Transcription by Nuclear Hormone Receptors

The steroid, retinoic acid (retinoid), and thyroid hormones form a large group of hormones (receptor ligands) that exert at least part of their effects by a mechanism fundamentally different from that of other hormones: they act in the nucleus to alter gene expression. We discuss their mode of action in detail in Chapter 28, along with other mechanisms for regulating gene expression. Here we give a brief overview.

Steroid hormones (estrogen, progesterone, and cortisol, for example), too hydrophobic to dissolve readily in the blood, are transported on specific carrier proteins from their point of release to their target tissues. In target cells, these hormones pass through the plasma membrane by simple diffusion and bind to specific receptor proteins in the nucleus **(Fig. 12–30)**. Steroid hormone receptors with no bound ligand (aporeceptors) often act to suppress the transcription of target genes. Hormone binding triggers changes in the conformation of a receptor protein so that it becomes capable of interacting with specific regulatory sequences in DNA called **hormone response elements (HREs)**, thus altering gene expression (see Fig. 28–33). The bound receptor-hormone complex enhances the expression of specific genes adjacent to HREs, with the help of several other proteins essential for transcription. Hours or days are required for these regulators to have their full effect—the time required for the changes in RNA synthesis and subsequent protein synthesis to become evident in altered metabolism.

The specificity of the steroid-receptor interaction is exploited in the use of the drug **tamoxifen** to treat breast cancer. In some types of breast cancer, division of the cancerous cells depends on the continued presence of estrogen. Tamoxifen is an estrogen antagonist; it competes with estrogen for binding to the estrogen receptor, but the tamoxifen-receptor complex has little or no effect on gene expression. Consequently, tamoxifen administered after surgery or during chemotherapy for hormone-dependent breast cancer slows or stops the growth of remaining cancerous cells. Another steroid analog, the drug **mifepristone (RU486)**, binds to the progesterone receptor and blocks hormone actions essential to implantation of the fertilized ovum in the uterus, and thus functions as a contraceptive.

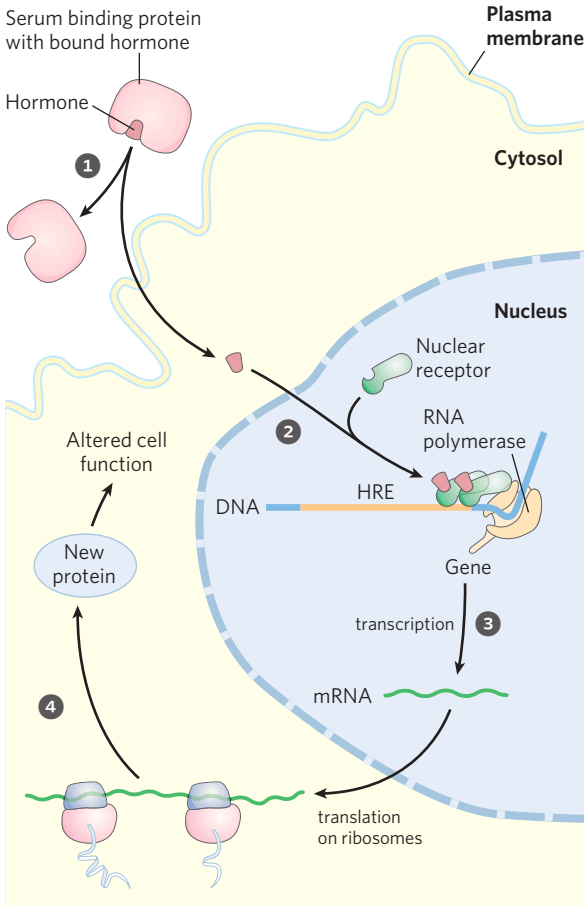
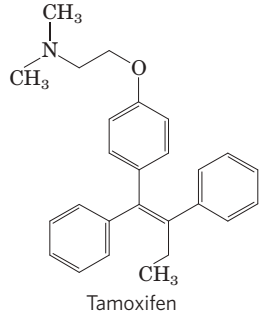
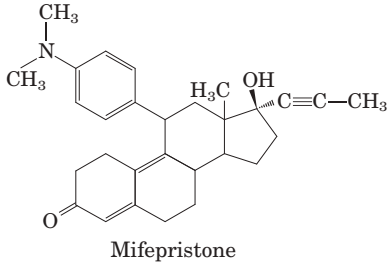


FIGURE 12–30 General mechanism by which steroid and thyroid hormones, retinoids, and vitamin D regulate gene expression. The details of transcription and protein synthesis are discussed in Chapters 26 and 27. Some steroids also act through plasma membrane receptors by a completely different mechanism.



Tamoxifen



Mifepristone (RU486)

Certain effects of steroids seem to occur too fast to be the result of altered protein synthesis via the classic mechanism of steroid hormone action through nuclear receptors. For example, the estrogen-mediated dilation of blood vessels is known to be independent of gene transcription or protein synthesis, as is the steroid-induced decrease in cellular [cAMP]. Another transduction mechanism involving plasma membrane receptors is believed to be responsible for some of these effects.

SUMMARY 12.8 Regulation of Transcription by Nuclear Hormone Receptors

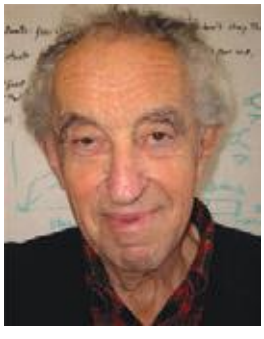
- Steroid hormones enter cells and bind to specific receptor proteins.
- The hormone-receptor complex binds specific regions of DNA, the hormone response elements, and interacts with other proteins to regulate the expression of nearby genes.
- Certain effects of steroid hormones may occur through a different, faster signaling pathway.

12.9 Signaling in Microorganisms and Plants

Much of what we have said here about signaling relates to mammalian tissues or cultured cells from such tissues. Bacteria, archaea, eukaryotic microorganisms, and vascular plants must also respond to a variety of external signals—O_2, nutrients, light, noxious chemicals, and so on. We turn here to a brief consideration of the kinds of signaling machinery used by microorganisms and plants.

Bacterial Signaling Entails Phosphorylation in a Two-Component System

In pioneering studies of chemotaxis in bacteria, Julius Adler showed that *Escherichia coli* responds to nutrients in its environment, including sugars and amino acids, by swimming toward them, propelled by one or a few flagella. A family of membrane proteins have binding domains on the outside of the plasma membrane to which specific **attractants** (sugars or amino acids) bind **(Fig. 12–31)**. Ligand binding causes an intrinsic kinase activity of the receptor to phosphorylate a His residue in its cytosolic domain. This first component of the **two-component system**, the **receptor histidine kinase**, then catalyzes transfer of the phosphoryl group from the His residue to an Asp residue on a second, soluble protein, the **response regulator**. This phosphoprotein moves to the base of the flagellum, carrying the signal from the membrane receptor. The flagellum is driven by a rotary motor that can propel the cell through its medium or cause it to stall, depending on the direction of motor rotation. The change in attractant concentration over time, signaled through the receptor, allows the cell to determine whether it is moving toward or away from the source of the attractant. If its motion is toward the attractant, the response regulator signals the cell to continue in a straight line (a run); if away from it, the cell tumbles momentarily, acquiring a new direction. Repetition of this behavior results in a random path, biased toward movement in the direction of increasing attractant concentration.



Julius Adler

E. coli detects not only sugars and amino acids but also O_2, extremes of temperature, and other environmental factors, using this basic two-component system. Two-component systems have been detected in many other bacteria, both gram-positive and gram-negative, and in archaea, as well as in protists and fungi. Clearly, this signaling mechanism developed early in the course of cellular evolution and has been conserved.

Various signaling systems used by animal cells also have analogs in bacteria. As the full genomic sequences of more, and more diverse, bacteria become known, researchers have discovered genes that encode proteins similar to protein Ser or Thr kinases, Ras-like proteins regulated by GTP binding, and proteins with SH3 domains. Receptor Tyr kinases have not been detected in bacteria, but Ⓟ–Tyr residues do occur in some bacteria.

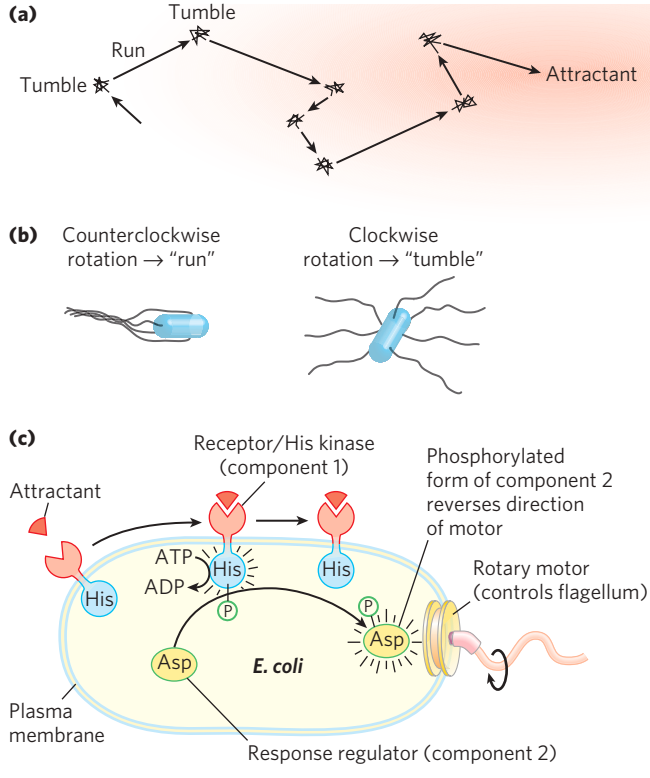


FIGURE 12–31 The two-component signaling mechanism in bacterial chemotaxis. (a) When placed near a source of an attractant solute, *E. coli* performs a random walk, biased toward the attractant. **(b)** Flagella have intrinsic helical structure, and when all flagella rotate counterclockwise, the flagellar helices twist together and move in concert to propel the cell forward in a "run." When the flagella rotate clockwise, the flagellar bundles fly apart, and the cell tumbles briefly until counterclockwise rotation resumes and the cell begins to swim forward again in a new, random direction. When moving toward the attractant, the cell has fewer tumbles, and therefore longer runs; when moving away, the frequent tumbles eventually result in movement toward the attractant. **(c)** Flagellar rotation is controlled by a two-component system consisting of a receptor-histidine kinase and an effector protein. When an attractant ligand binds to the receptor domain of the membrane-bound receptor, a protein kinase in the cytosolic domain (component 1) is activated and autophosphorylates a His residue. This phosphoryl group is then transferred to an Asp residue on component 2. After phosphorylation, component 2 moves to the base of the flagellum, where it causes counterclockwise rotation of the flagella, producing a run.

Signaling Systems of Plants Have Some of the Same Components Used by Microbes and Mammals

Like animals, vascular plants must have a means of communication between tissues to coordinate and direct growth and development; to adapt to conditions of O_2,

nutrients, light, temperature, and water availability; and to warn of the presence of noxious chemicals and damaging pathogens **(Fig. 12–32)**. At least a billion years of evolution have passed since the plant and animal branches of the eukaryotes diverged, which is reflected in the differences in signaling mechanisms: some plant mechanisms are conserved—that is, are similar to those in animals (protein kinases, adaptor proteins, cyclic nucleotides, electrogenic ion pumps, and gated ion channels); some are similar to bacterial two-component systems; and some are unique to plants (light-sensing mechanisms that reflect seasonal changes in the angle, and hence color, of sunlight, for example) (Table 12–7). The genome of the plant *Arabidopsis thaliana* encodes about 1,000 protein Ser/Thr kinases, including about 60 MAPKs and nearly 400 membrane-associated receptor kinases that phosphorylate Ser or Thr residues; a variety of protein phosphatases; adaptor proteins that form scaffolds on which proteins assemble in signaling complexes; enzymes for the synthesis and degradation of cyclic nucleotides; and 100 or more ion channels, including about 20 gated by cyclic nucleotides. Inositol phospholipids are present, as are kinases that interconvert them by phosphorylation of inositol head groups. Even given the fact that *Arabidopsis* has multiple copies of many genes, the presence of this many genes certainly reflects a wide array of signaling potential.

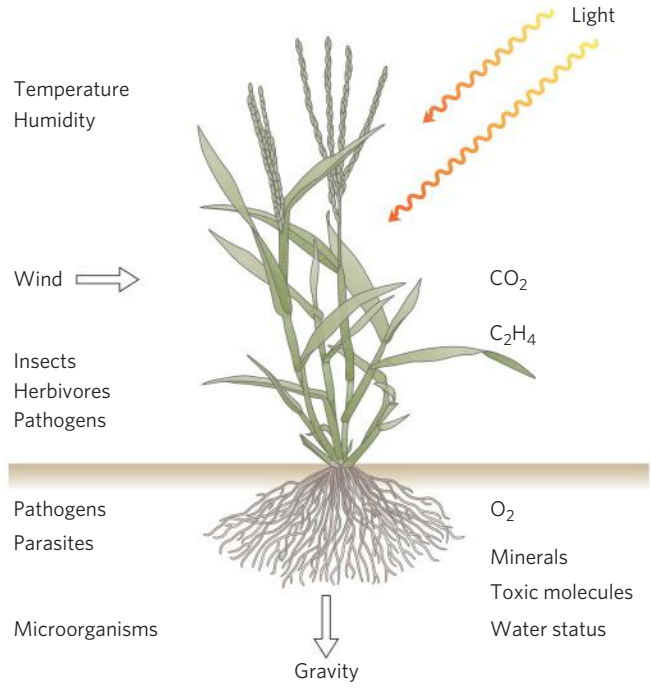


FIGURE 12–32 **Some stimuli that produce responses in plants.**

TABLE 12–7 **Signaling Components Present in Mammals, Plants, or Bacteria**

Signaling component	Mammals	Plants	Bacteria
Ion channels	+	+	+
Electrogenic ion pumps	+	+	+
Two-component His kinases	+	+	+
Adenylyl cyclase	+	+	+
Guanylyl cyclase	+	+	?
Receptor protein kinases (Ser/Thr)	+	+	?
Ca^{2+} as second messenger	+	+	?
Ca^{2+} channels	+	+	?
Calmodulin, CaM-binding protein	+	+	–
MAPK cascade	+	+	–
Cyclic nucleotide–gated channels	+	+	–
IP_3-gated Ca^{2+} channels	+	+	–
Phosphatidylinositol kinases	+	+	–
GPCRs	+	+/–	+
Trimeric G proteins	+	+/–	–
PI-specific phospholipase C	+	?	–
Tyrosine kinase receptors	+	?	–
SH2 domains	+	?	?
Nuclear steroid receptors	+	–	–
Protein kinase A	+	–	–
Protein kinase G	+	–	–

However, some types of signaling proteins common in animal tissues are not present in plants, or are represented by only a few genes. Cyclic nucleotide–dependent protein kinases (PKA and PKG) seem to be absent, for example. Heterotrimeric G proteins and protein Tyr kinase genes are much less prominent in the plant genome, and genes for GPCRs, the largest family of proteins in the human genome (~1,000 genes), are very sparsely represented in the plant genome. DNA-binding nuclear steroid receptors are certainly not prominent, and may be absent from plants. Although plants lack the most widely conserved light-sensing mechanism present in animals (rhodopsin, with retinal as pigment), they have a rich collection of other light-detecting mechanisms not found in animal tissues—phytochromes and cryptochromes, for example (Chapter 19).

The kinds of compounds that elicit signals in plants are similar to certain signaling molecules in animals **(Fig. 12–33)**. Instead of prostaglandins, plants have jasmonate; instead of steroid hormones, brassinosteroids. About 100 different small peptides serve as plant signals, and both plants and animals use compounds derived from aromatic amino acids as signals.

Plants Detect Ethylene through a Two-Component System and a MAPK Cascade

The gaseous plant hormone ethylene ($CH_2{=}CH_2$), which stimulates the ripening of fruits (among other functions), acts through receptors that are related in primary sequence to the receptor His kinases of the bacterial two-component systems and probably evolved from them. In *Arabidopsis*, the two-component signaling system is contained within a single integral membrane protein of the endoplasmic reticulum (*not* the plasma membrane). Ethylene diffuses into the cell through the plasma membrane and into the ER. The first downstream component affected by ethylene signaling is a protein Ser/Thr kinase (CTR1; **Fig. 12–34**) with sequence homology to Raf, the protein kinase that begins the MAPK cascade in the mammalian response to insulin (see Fig. 12–15). In plants, in the absence of ethylene, the CTR1 kinase is active and *inhibits* the MAPK cascade, preventing transcription of ethylene-responsive genes. Exposure to ethylene *inactivates* the CTR1 kinase, thereby activating the MAPK cascade that leads to activation of the transcription factor EIN3. Active EIN3 stimulates the synthesis of a second transcription factor (ERF1), which in turn activates transcription of

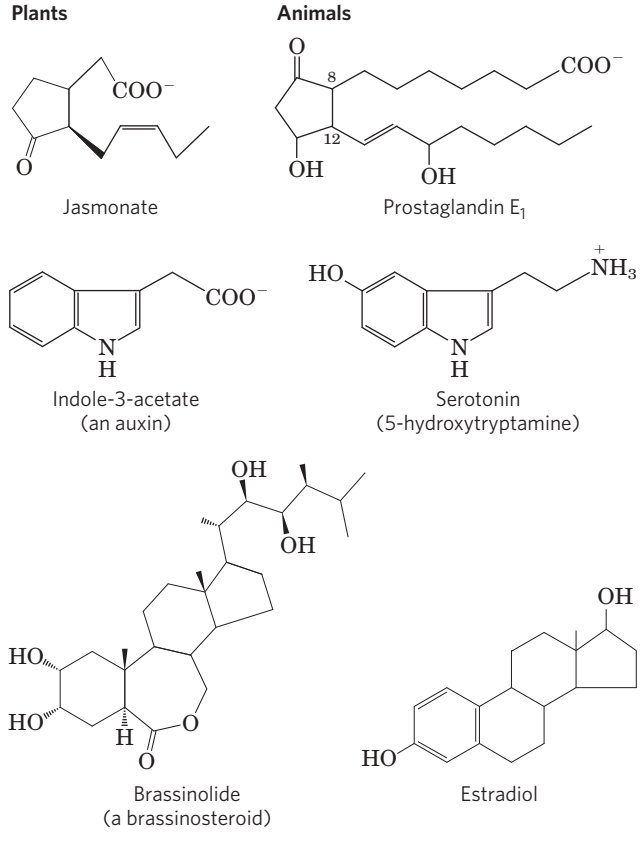


FIGURE 12–33 Structural similarities between plant and animal signals.

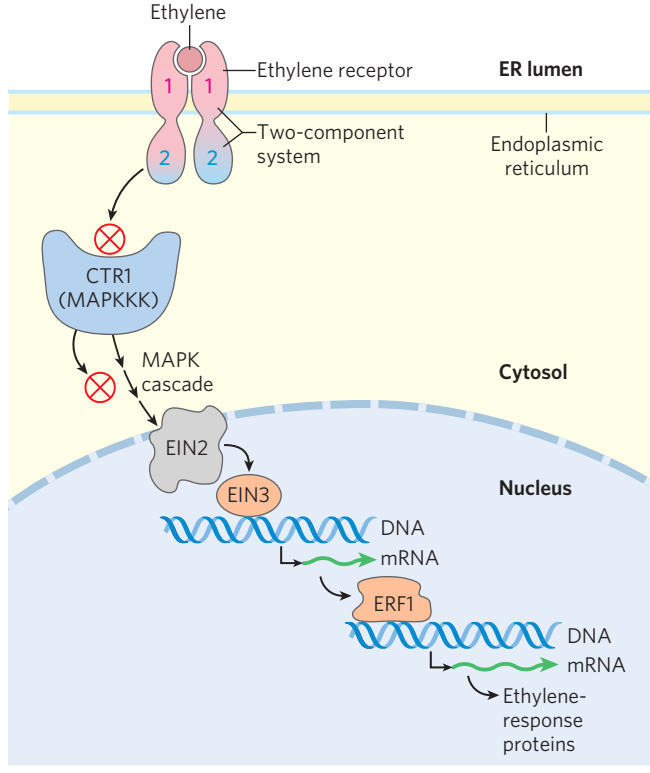


FIGURE 12–34 Transduction mechanism for detection of ethylene by plants. The ethylene receptor (pink) in the endoplasmic reticulum is a two-component system contained in a single protein, with a receptor domain (component 1) and a response regulator domain (component 2). The receptor controls (in ways we do not yet understand) the activity of CTR1, a protein kinase similar to MAPKKKs and therefore presumed to be part of a MAPK cascade. CTR1 is a negative regulator of the ethylene response; when CTR1 is *inactive*, the ethylene signal is transmitted through the gene product EIN2 (thought to be a nuclear envelope protein), which causes increased synthesis of ERF1, a transcription factor. ERF1 stimulates expression of proteins specific to the ethylene response.

ethylene-responsive genes; the gene products affect processes ranging from seedling development to fruit ripening. Although apparently derived from the bacterial two-component signaling system, the ethylene system in *Arabidopsis* is different in that the His kinase activity that defines component 1 in bacteria is not essential to signal transduction in *Arabidopsis*.

Receptorlike Protein Kinases Transduce Signals from Peptides

One common motif in plant signaling involves **receptor-like kinases (RLKs)**, which have a single helical segment in the plasma membrane that connects a receptor domain on the outside with a protein Ser/Thr kinase on the cytoplasmic side. This type of receptor participates in the defense mechanism triggered by infection with a bacterial pathogen **(Fig. 12–35a)**. The signal to turn on the genes needed for defense against infection is a peptide (flg22) released by breakdown of flagellin, the major protein of the bacterial flagellum. Binding of flg22 to the FLS2 receptor of *Arabidopsis* induces receptor dimerization and autophosphorylation on Ser and Thr residues, and the downstream effect is activation of a MAPK cascade like that described above for insulin action. The final kinase in this cascade activates a specific transcription factor, triggering synthesis of the proteins that defend against the bacterial infection. The steps between receptor phosphorylation and the MAPK cascade are not yet known. A phosphoprotein phosphatase (KAPP) associates with the active receptor protein and inactivates it by dephosphorylation to end the response.

The MAPK cascade in the plant's defense against bacterial pathogens is remarkably similar to the innate immune response in mammals (Fig. 12–35b) that is triggered by bacterial lipopolysaccharide and mediated by the Toll-like receptors (TLRs, a name derived from a *Drosophila* mutant originally called Toll (German for "mad"); TLRs were subsequently found in many other organisms and were shown to function in embryonic development). Other membrane receptors use similar mechanisms to activate a MAPK cascade, ultimately activating transcription factors and turning on the genes essential to the defense response.

Most of the several hundred RLKs in plants are presumed to act in similar ways: ligand binding induces dimerization and autophosphorylation, and the activated receptor kinase triggers downstream responses by phosphorylating key proteins at Ser or Thr residues.

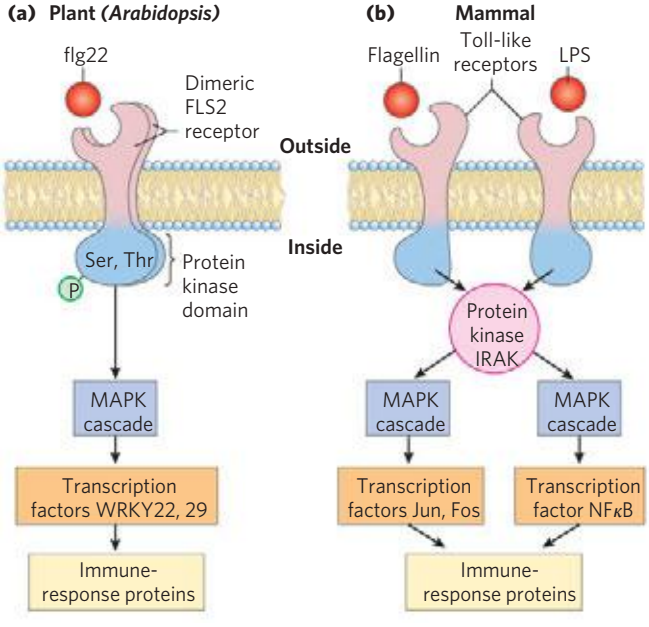


FIGURE 12–35 Similarities between the signaling pathways that trigger immune responses in plants and animals. (a) In *Arabidopsis thaliana*, the peptide flg22, derived from the flagella of a bacterial pathogen, binds to its receptor (FLS) in the plasma membrane, causing the receptor to form dimers and triggering autophosphorylation of the cytosolic protein kinase domain on a Ser or Thr residue (*not* a Tyr). Thus activated, the protein kinase phosphorylates downstream proteins (not shown). The activated receptor also activates (by means unknown) a MAPK cascade, which leads to phosphorylation of a nuclear protein that normally inhibits the transcription factors WRKY22 and 29; this phosphorylation triggers proteolytic degradation of the inhibitor and frees the transcription factors to stimulate gene expression related to the immune response. **(b)** In mammals, a toxic bacterial lipopolysaccharide (LPS; see Fig. 7–31) is detected by plasma membrane receptors, which then associate with and activate a soluble protein kinase (IRAK). The major flagellar protein of pathogenic bacteria acts through a similar receptor, also activating IRAK. The activated IRAK initiates two distinct MAPK cascades that end in the nucleus, causing the synthesis of proteins needed in the immune response. Jun, Fos, and NFκB are transcription factors.

SUMMARY 12.9 Signaling in Microorganisms and Plants

- Bacteria and eukaryotic microorganisms have a variety of sensory systems that allow them to sample and respond to their environment. In the two-component system, a receptor His kinase senses the signal and autophosphorylates a His residue, then phosphorylates an Asp residue of the response regulator.
- Plants respond to many environmental stimuli and employ hormones and growth factors to coordinate the development and metabolic activities of their tissues. Plant genomes encode hundreds of signaling proteins, including some very similar to those of mammals.
- Two-component signaling mechanisms common in bacteria are found in modified forms in plants, used in the detection of chemical signals and light.
- Plant receptorlike kinases (RLKs) participate in detecting a wide variety of stimuli, including brassinosteroids, peptides that originate from pathogens, and developmental signals. RLKs autophosphorylate Ser/Thr residues, then activate downstream proteins, which in some cases are MAPK cascades. The end result is increased transcription of specific genes.

12.10 Sensory Transduction in Vision, Olfaction, and Gustation

The detection of light, odors, and tastes (vision, olfaction, and gustation, respectively) in animals is accomplished by specialized sensory neurons that use signal-transduction mechanisms fundamentally similar to those that detect hormones, neurotransmitters, and growth factors. An initial sensory signal is amplified greatly by mechanisms that include gated ion channels and intracellular second messengers; the system adapts to continued stimulation by changing its sensitivity to the stimulus (desensitization); and sensory input from several receptors is integrated before the final signal goes to the brain.

The Visual System Uses Classic GPCR Mechanisms

In the vertebrate eye, light entering through the pupil is focused on a highly organized collection of light-sensitive neurons **(Fig. 12–36)**. The light-sensing neurons are of two types: rods (about 10^9 per retina), which sense low levels of light but cannot discriminate colors, and cones (about 3×10^6 per retina), which are less sensitive to light but can discriminate colors. Both cell types are long, narrow, specialized sensory neurons with two distinct cellular compartments: the outer segment contains dozens of membranous disks loaded with receptor proteins and their photosensitive chromophore **retinal**; the inner segment contains the nucleus and many mitochondria, which produce the ATP essential to phototransduction.

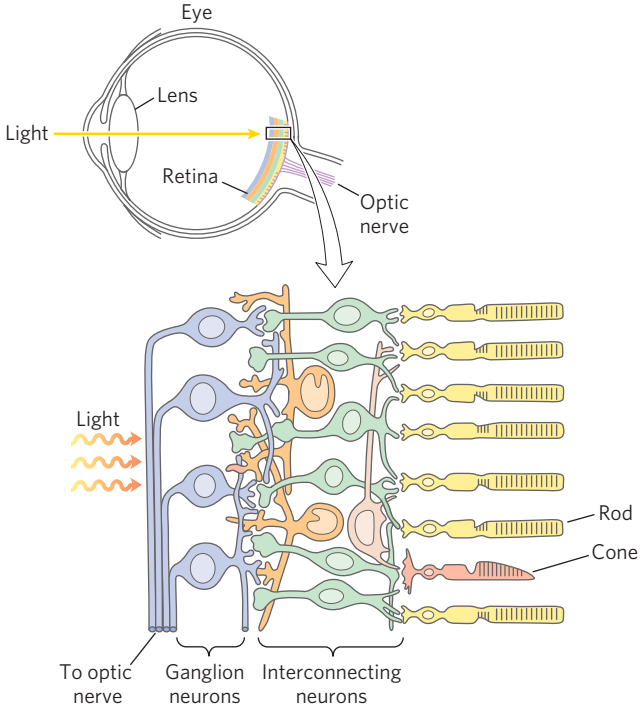


FIGURE 12–36 Light reception in the vertebrate eye. The lens focuses light on the retina, which is composed of layers of neurons. The primary photosensory neurons are rod cells (yellow), which are responsible for high-resolution and night vision, and cone cells of three subtypes (pink), which initiate color vision. The rods and cones form synapses with several ranks of interconnecting neurons that convey and integrate the electrical signals. The signals eventually pass from ganglion neurons through the optic nerve to the brain. Note that light must pass through the layers of ganglion neurons and interconnecting neurons before reaching the rod and cone cells.

Like other neurons, rods and cones have a transmembrane electrical potential (V_m), produced by the electrogenic pumping of the Na^+K^+ ATPase in the plasma membrane of the inner segment **(Fig. 12–37)**. Also contributing to the membrane potential is an ion channel in the outer segment that permits passage of either Na^+ or Ca^{2+} and is gated (opened) by cGMP. In the dark, rod cells contain enough cGMP to keep this channel open. The membrane potential is therefore determined by the difference between the amount of Na^+ and K^+ pumped by the inner segment (which polarizes the membrane) and the influx of Na^+ through the ion channels of the outer segment (which tends to depolarize the membrane).

The essence of signaling in the rod or cone cell is a light-induced decrease in [cGMP], which causes the cGMP-gated ion channel to close. The plasma membrane then becomes hyperpolarized by the Na^+K^+ ATPase. Rod and cone cells synapse with interconnecting neurons (Fig. 12–36) that carry information about the electrical activity to ganglion neurons near the inner surface of the retina. The ganglion neurons integrate the output from many rod or cone cells and send the resulting signal through the optic nerve to the visual cortex of the brain.

Visual transduction begins when light falls on rhodopsin, many thousands of molecules of which are present in each disk of the outer segments of rod and cone cells. **Rhodopsin** (M_r 40,000) is an integral protein with seven membrane-spanning α helices **(Fig. 12–38)**, the characteristic GPCR architecture. The light-absorbing pigment (chromophore) 11-*cis*-retinal is covalently attached to **opsin**, the protein component of rhodopsin, through a Schiff base to a Lys residue. The retinal molecule lies near the middle of the bilayer (Fig. 12–38), oriented with its long axis approximately in the plane of the membrane. When a photon is absorbed by the retinal component of rhodopsin, the energy causes a photochemical change; 11-*cis*-retinal is converted to all-*trans*-retinal (see Figs 1–19b and 10–21). This change in the structure of the chromophore forces conformational changes in the rhodopsin molecule—the first stage in visual transduction.

Retinal is derived from vitamin A_1 (retinol), which is produced from β-carotene (see Fig. 10–21). Dietary deficiency of vitamin A leads to night blindness (the inability to adapt to low light levels), which is relatively common in some developing countries. Vitamin A supplements or vegetables rich in carotene (such as carrots) supply the vitamin and reverse the night blindness. ■

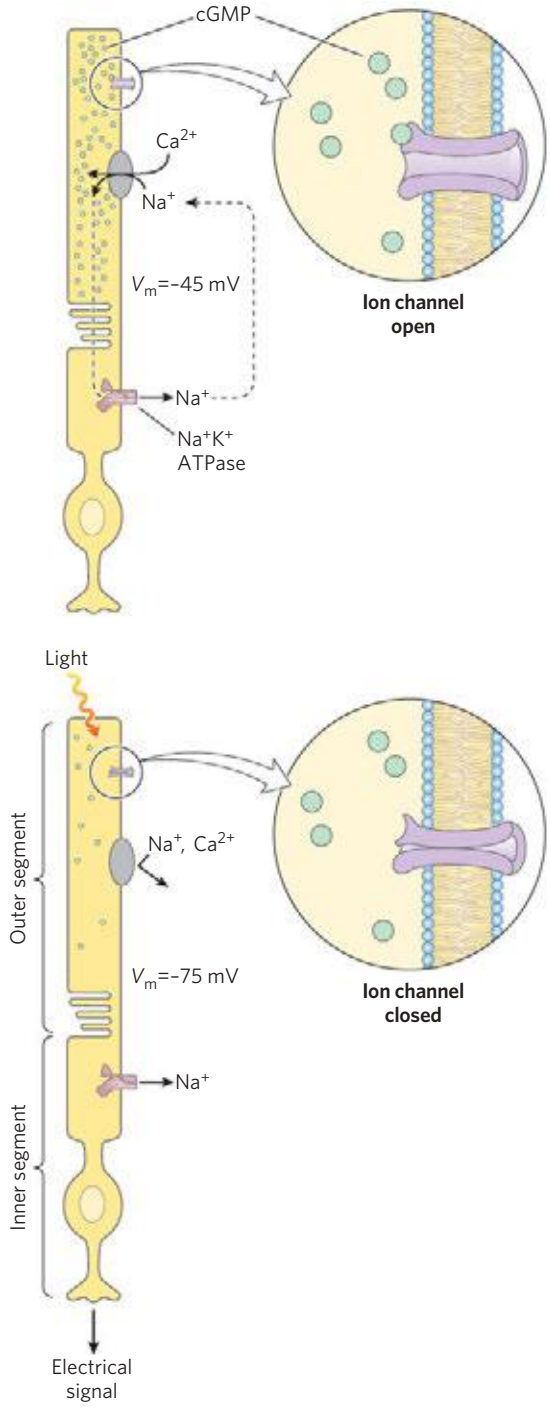


FIGURE 12–37 Light-induced hyperpolarization of rod cells. The rod cell consists of an outer segment, filled with stacks of membranous disks (not shown) containing the photoreceptor rhodopsin, and an inner segment that contains the nucleus and other organelles (not shown). The inner segment forms a synapse with interconnecting neurons (Fig. 12–36). Cones have a similar structure. ATP in the inner segment powers the Na^+K^+ ATPase, which creates a transmembrane electrical potential by pumping 3 Na^+ out for every 2 K^+ pumped in. The membrane potential is reduced by the inflow of Na^+ and Ca^{2+} through cGMP-gated cation channels in the outer-segment plasma membrane. When rhodopsin absorbs light, it triggers degradation of cGMP (green dots) in the outer segment, causing closure of the ion channel. Without cation influx through this channel, the cell becomes hyperpolarized. This electrical signal is passed to the brain through the ranks of neurons shown in Figure 12–36.

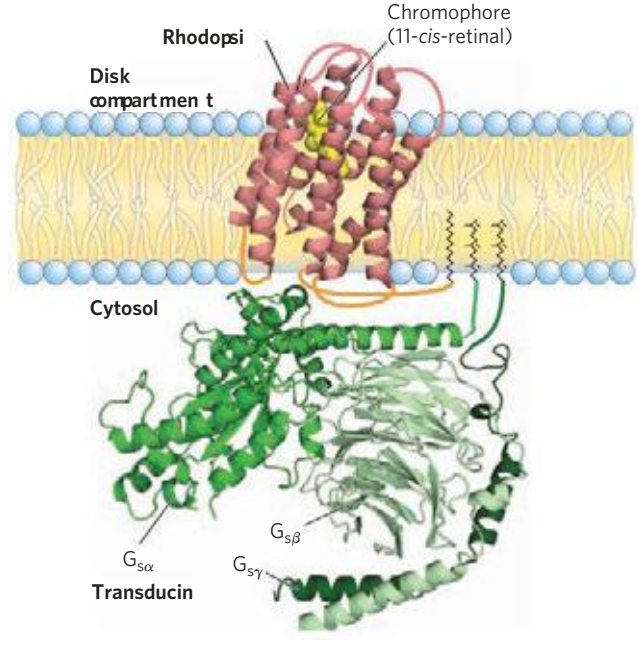


FIGURE 12–38 Complex of rhodopsin with the G protein transducin. (PDB ID 1BAC) Rhodopsin (red) has seven transmembrane helices embedded in the disk membranes of rod outer segments and is oriented with its carboxyl terminus on the cytosolic side and its amino terminus inside the disk. The chromophore 11-*cis*-retinal (yellow space-filling structure), attached through a Schiff base linkage to Lys^{256} of the seventh helix, lies near the center of the bilayer. (This location is similar to that of the epinephrine-binding site in the β-adrenergic receptor.) Several Ser and Thr residues near the carboxyl terminus are substrates for phosphorylations that are part of the desensitization mechanism for rhodopsin. Cytosolic loops that interact with the G protein transducin are shown in orange; their exact positions are not yet known. The three subunits of transducin (green) are shown in their likely arrangement. Rhodopsin is palmitoylated at its carboxyl terminus, and both the α and γ subunits of transducin have attached lipids (yellow) that assist in anchoring them to the membrane.

Excited Rhodopsin Acts through the G Protein Transducin to Reduce the cGMP Concentration

In its excited conformation, rhodopsin interacts with a second protein, **transducin**, which hovers nearby on the cytoplasmic face of the disk membrane (Fig. 12–38). Transducin (T) belongs to the same family of heterotrimeric GTP-binding proteins as G_s and G_i. Although specialized for visual transduction, transducin shares many functional features with G_s and G_i. It can bind either GDP or GTP. In the dark, GDP is bound, all three subunits of the protein (T_α, T_β, and T_γ) remain together,

and no signal is sent. When rhodopsin is excited by light, it interacts with transducin, catalyzing the replacement of bound GDP by GTP from the cytosol (**Fig. 12–39**, steps ❶ and ❷). Transducin then dissociates into T_α and $T_{\beta\gamma}$, and the T_α-GTP carries the signal from the excited receptor to the next element in the transduction pathway, a cGMP phosphodiesterase; this enzyme converts cGMP to 5′-GMP (steps ❸ and ❹). Note that this is not the same cyclic nucleotide phosphodiesterase that hydrolyzes cAMP to terminate the β-adrenergic response. One isoform of the cGMP-specific PDE is unique to the visual cells of the retina.

The PDE of the retina is a peripheral protein with its active site on the cytoplasmic side of the disk membrane. In the dark, a tightly bound inhibitory subunit very effectively suppresses the PDE activity. When T_α-GTP encounters the PDE, the inhibitory subunit leaves the enzyme and instead binds T_α, and the enzyme's activity immediately increases by several orders of magnitude. Each molecule of the active PDE degrades many molecules of cGMP to the biologically inactive 5′-GMP, lowering [cGMP] in the outer segment within a fraction of a second. At the new, lower [cGMP], the cGMP-gated ion channels close, blocking reentry of Na^+ and Ca^{2+} into the outer segment and hyperpolarizing the membrane of the rod or cone cell (step ❺). Through this process, the initial stimulus—a photon—changes the V_m of the cell. The brighter the illumination of the rod cell, the greater the hyperpolarization. This hyperpolarization is perceived by the integrating neurons of the retina, which pass the integrated signal on to the ganglion cells, which send axons via the optic nerve to the brain.

Several steps in the visual-transduction process result in a huge amplification of the signal. Each excited rhodopsin molecule activates at least 500 molecules of transducin, each of which can activate a molecule of

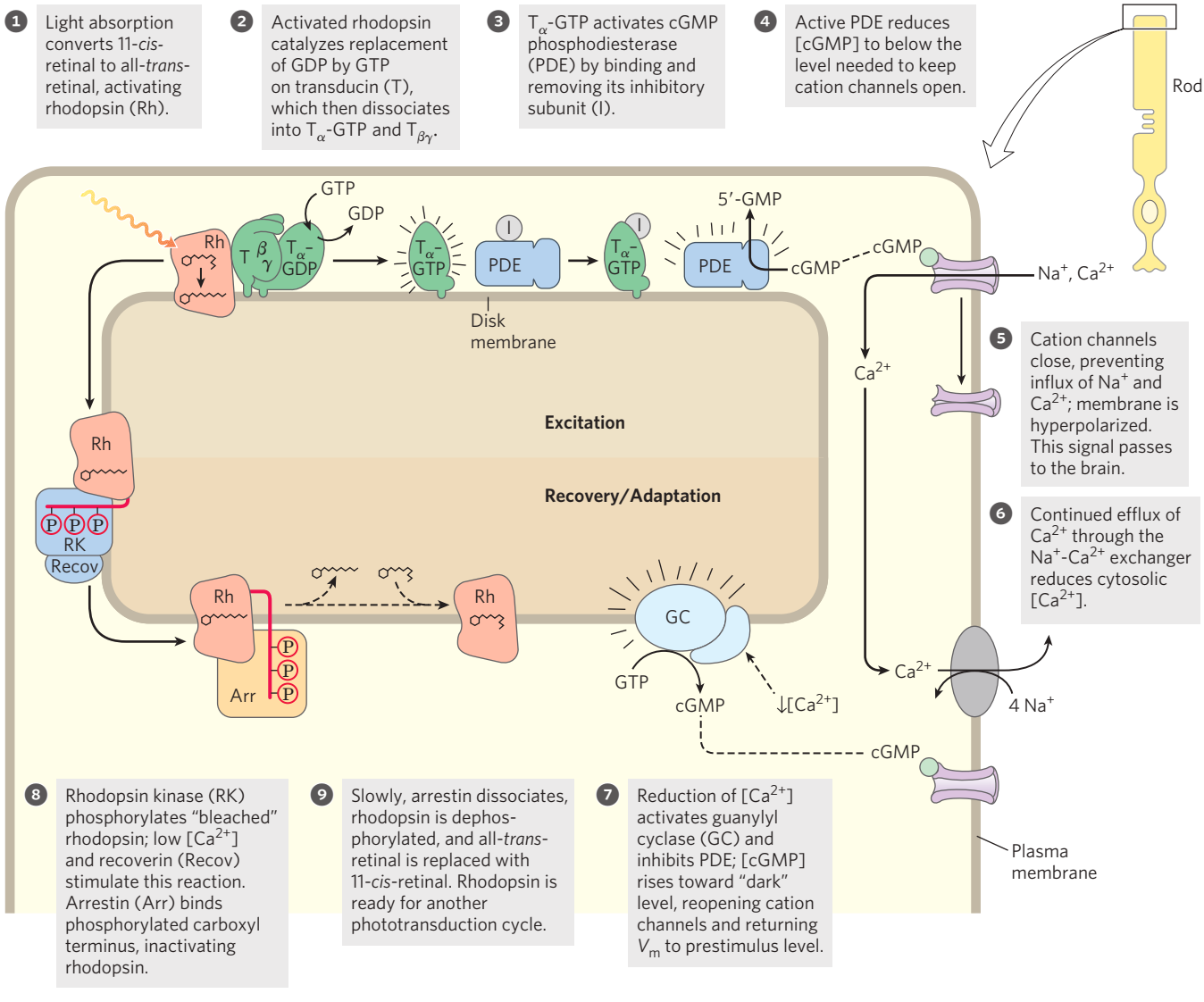


FIGURE 12–39 Molecular consequences of photon absorption by rhodopsin in the rod outer segment. The top half of the figure (steps ❶ to ❺) describes excitation; the bottom shows post-illumination steps: recovery (steps ❻ and ❼) and adaptation (steps ❽ and ❾).

the PDE. This phosphodiesterase has a remarkably high turnover number, each activated molecule hydrolyzing 4,200 molecules of cGMP per second. The binding of cGMP to cGMP-gated ion channels is cooperative, and a relatively small change in [cGMP] therefore registers as a large change in ion conductance. The result of these amplifications is exquisite sensitivity to light. Absorption of a single photon closes 1,000 or more ion channels and changes the cell's membrane potential by about 1 mV.

The Visual Signal Is Quickly Terminated

As your eyes move across this line, the retinal images of the first words disappear rapidly—before you see the next series of words. In that short interval, a great deal of biochemistry has taken place. Very shortly after illumination of the rod or cone cells stops, the photosensory system shuts off. The α subunit of transducin (with bound GTP) has intrinsic GTPase activity. Within milliseconds after the decrease in light intensity, GTP is hydrolyzed and T_α reassociates with $T_{\beta\gamma}$. The inhibitory subunit of the PDE, which had been bound to T_α-GTP, is released and reassociates with the enzyme, strongly inhibiting its activity.

To return [cGMP] to its "dark" level, the enzyme guanylyl cyclase converts GTP to cGMP (step ❼ in Fig. 12–39) in a reaction that is inhibited by high [Ca^{2+}] (>100 nM). Calcium levels drop during illumination, because the steady-state [Ca^{2+}] in the outer segment is the result of outward pumping of Ca^{2+} through the Na^+-Ca^{2+} exchanger of the plasma membrane (see Fig. 12–37) and influx of Ca^{2+} through open cGMP-gated channels. In the dark, this produces a [Ca^{2+}] of about 500 nM—enough to inhibit cGMP synthesis. After brief illumination, Ca^{2+} entry slows and [Ca^{2+}] declines (step ❻). The inhibition of guanylyl cyclase by Ca^{2+} is relieved, and the cyclase converts GTP to cGMP to return the system to its prestimulus state (step ❼).

Rhodopsin itself also undergoes changes in response to prolonged illumination. The conformational change induced by light absorption exposes several Thr and Ser residues in the carboxyl-terminal domain. These residues are quickly phosphorylated by **rhodopsin kinase** (step ❽ in Fig. 12–39), which is functionally and structurally homologous to the β-adrenergic kinase (βARK) that desensitizes the β-adrenergic receptor (Fig. 12–8). The Ca^{2+}-binding protein **recoverin** inhibits rhodopsin kinase at high [Ca^{2+}], but the inhibition is relieved when [Ca^{2+}] drops after illumination, as described above. The phosphorylated carboxyl-terminal domain of rhodopsin is bound by the protein **arrestin 1**, preventing further interaction between activated rhodopsin and transducin. Arrestin 1 is a close homolog of arrestin 2 (βarr; Fig. 12–8). On a relatively long time scale (seconds to minutes), the all-*trans*-retinal of an excited rhodopsin molecule is removed and replaced by 11-*cis*-retinal, to produce rhodopsin that is ready for another round of excitation (step ❾ in Fig. 12–39).

Cone Cells Specialize in Color Vision

Color vision involves a path of sensory transduction in cone cells essentially identical to that described above, but triggered by slightly different light receptors. Three types of cone cells are specialized to detect light from different regions of the spectrum, using three related photoreceptor proteins (opsins). Each cone cell expresses only one kind of opsin, but each type is closely related to rhodopsin in size, amino acid sequence, and presumably three-dimensional structure. The differences among the opsins, however, are great enough to place the chromophore, 11-*cis*-retinal, in three slightly different environments, with the result that the three photoreceptors have different absorption spectra **(Fig. 12–40)**. We discriminate colors and hues by integrating the output from the three types of cone cells, each containing one of the three photoreceptors.

Color blindness, such as the inability to distinguish red from green, is a fairly common, genetically inherited trait in humans. The various types of color blindness result from different opsin mutations. One form is due to loss of the red photoreceptor; affected individuals are **red$^-$ dichromats** (they see only two primary colors). Others lack the green pigment and are **green$^-$ dichromats**. In some cases, the red and green photoreceptors are present but have a changed amino acid sequence that causes a change in their absorption spectra, resulting in abnormal color vision. Depending on which pigment is altered, such individuals are **red-anomalous trichromats** or **green-anomalous trichromats**. Examination of the genes for the visual receptors has allowed the diagnosis of color blindness in a famous "patient" more than a century after his death (Box 12–4). ■

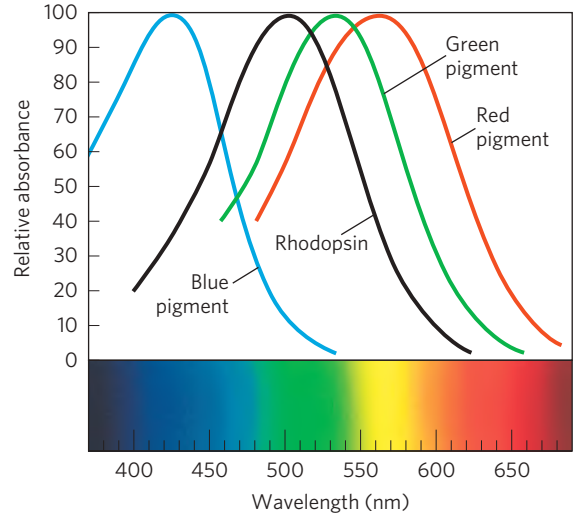


FIGURE 12–40 Absorption spectra of purified rhodopsin and the red, green, and blue receptors of cone cells. The receptor spectra, obtained from individual cone cells isolated from cadavers, peak at about 420, 530, and 560 nm, and the maximum absorption for rhodopsin is at about 500 nm. For reference, the visible spectrum for humans is about 380 to 750 nm.

BOX 12–4 MEDICINE Color Blindness: John Dalton's Experiment from the Grave

The chemist John Dalton (of atomic theory fame) was color-blind. He thought it probable that the vitreous humor of his eyes (the fluid that fills the eyeball behind the lens) was tinted blue, unlike the colorless fluid of normal eyes. He proposed that after his death, his eyes should be dissected and the color of the vitreous humor determined. His wish was honored. The day after Dalton's death in July 1844, Joseph Ransome dissected his eyes and found the vitreous humor to be perfectly colorless. Ransome, like many scientists, was reluctant to throw samples away. He placed Dalton's eyes in a jar of preservative, where they stayed for a century and a half (Fig. 1).

Then, in the mid-1990s, molecular biologists in England took small samples of Dalton's retinas and extracted DNA. Using the known gene sequences for the opsins of the red and green light receptors, they amplified the relevant sequences (using techniques described in Chapter 9) and determined that Dalton had the opsin gene for the red photopigment but lacked the opsin gene for the green photopigment. Dalton was a green-dichromat. So, 150 years after his death, the experiment Dalton started—by hypothesizing about the cause of his color blindness—was finally finished.

FIGURE 1 Dalton's eyes.

Vertebrate Olfaction and Gustation Use Mechanisms Similar to the Visual System

The sensory cells that detect odors and tastes have much in common with the rod and cone cells. Olfactory neurons have long thin cilia extending from one end of the cell into a mucous layer that overlays the cell. These cilia present a large surface area for interaction with olfactory signals. The receptors for olfactory stimuli are ciliary membrane proteins with the familiar GPCR structure of seven transmembrane α helices. The olfactory signal can be any one of the many volatile compounds for which there are specific receptor proteins. Our ability to discriminate odors stems from hundreds of different olfactory receptors in the tongue and nasal passages and from the brain's ability to integrate input from different types of olfactory receptors to recognize a "hybrid" pattern, extending our range of discrimination far beyond the number of receptors.

The olfactory stimulus arrives at the sensory cells by diffusion through the air. In the mucous layer covering the olfactory neurons, the odorant molecule binds directly to an olfactory receptor or to a specific binding protein that carries the odorant to a receptor **(Fig. 12–41)**. Interaction between odorant and receptor triggers a change in receptor conformation that results in the replacement of bound GDP by GTP on a G protein, G_{olf}, analogous to transducin and to G_s of the β-adrenergic system. The activated G_{olf} then activates adenylyl cyclase of the ciliary membrane, which synthesizes cAMP from ATP, raising the local [cAMP]. The cAMP-gated Na^+ and Ca^{2+} channels of the ciliary membrane open, and the influx of Na^+ and Ca^{2+} produces a small depolarization called the **receptor potential**. If a sufficient number of odorant molecules encounter receptors, the receptor potential is strong enough to cause the neuron to fire an action potential. This is relayed to the brain in several stages and registers as a specific smell. All these events occur within 100 to 200 ms.

When the olfactory stimulus is no longer present, the transducing machinery shuts itself off in several ways. A cAMP phosphodiesterase returns [cAMP] to the prestimulus level. G_{olf} hydrolyzes its bound GTP to GDP, thereby inactivating itself. Phosphorylation of the receptor by a specific kinase prevents its interaction with G_{olf}, by a mechanism analogous to that used to desensitize the β-adrenergic receptor and rhodopsin. And lastly, some odorants are enzymatically destroyed by oxidases.

The sense of taste in vertebrates reflects the activity of gustatory neurons clustered in taste buds on the surface of the tongue. In these sensory neurons, GPCRs are coupled to the heterotrimeric G protein **gustducin** (very similar to the transducin of rod and cone cells). Sweet-tasting molecules are those that bind receptors in "sweet" taste buds. When the molecule (tastant) binds, gustducin is activated by replacement of bound GDP with GTP and then stimulates cAMP production by adenylyl cyclase. The resulting elevation of [cAMP] activates PKA, which phosphorylates K^+ channels in the plasma membrane, causing them to close. Reduced efflux of K^+ depolarizes the cell **(Fig. 12–42)** sending an electrical signal to the brain. Other taste buds specialize in detecting bitter, sour, salty, or umami (savory) tastants, using various combinations of second messengers and ion channels in the transduction mechanisms.

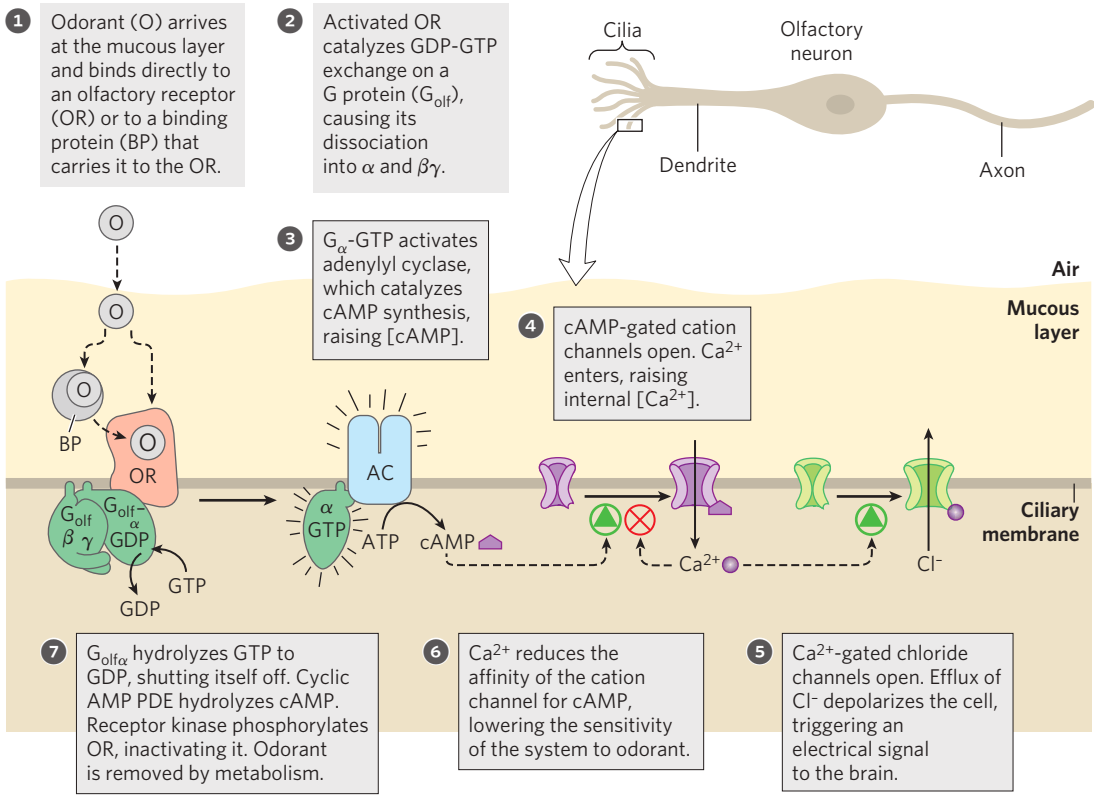


FIGURE 12–41 Molecular events of olfaction. These interactions occur in the cilia of olfactory receptor cells.

GPCRs of the Sensory Systems Share Several Features with GPCRs of Hormone Signaling Systems

We have now looked at several types of signaling systems (hormone signaling, vision, olfaction, and gustation) in which membrane receptors are coupled to second messenger–generating enzymes through G proteins. As we have intimated, signaling mechanisms must have arisen early in evolution; genomic studies have revealed hundreds of genes encoding GPCRs in vertebrates, arthropods (*Drosophila* and mosquito), and the roundworm *Caenorhabditis elegans*. Even the common baker's yeast *Saccharomyces* uses GPCRs and G proteins to detect the opposite mating type. Overall patterns have been conserved, and the introduction of variety has given modern organisms the ability to

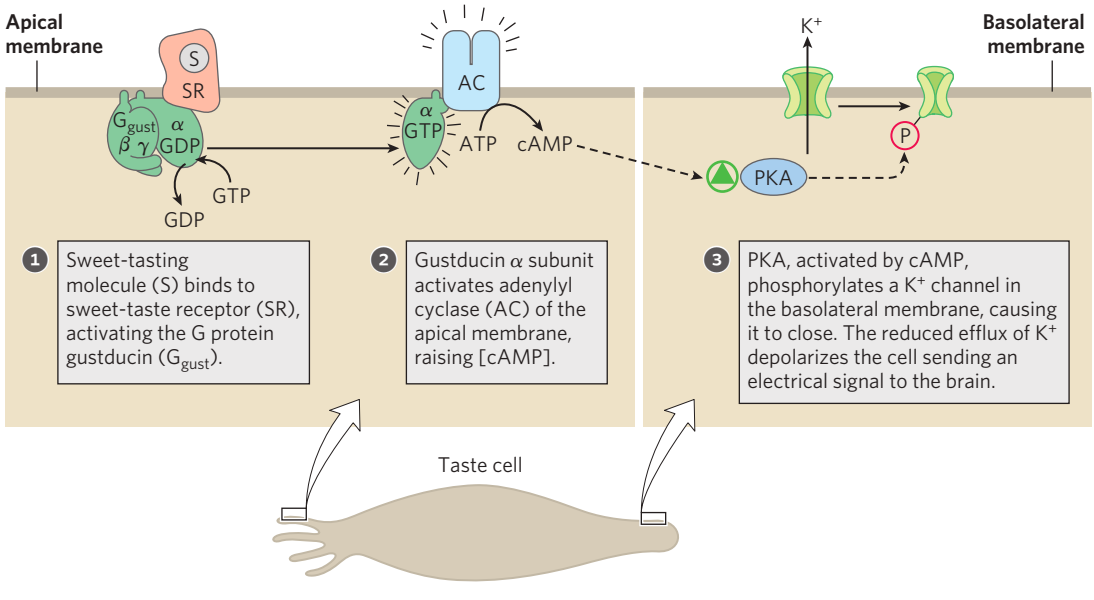


FIGURE 12–42 Transduction mechanism for sweet tastants.

respond to a wide range of stimuli (Table 12–8). Of the approximately 29,000 genes in the human genome, as many as 1,000 encode GPCRs, including hundreds for olfactory stimuli and many "orphan receptors" for which the natural ligand is not yet known.

All well-studied signal-transducing systems that act through heterotrimeric G proteins share some common features, which reflect their evolutionary relatedness **(Fig. 12–43)**. The receptors have seven transmembrane segments, a domain (generally the loop between transmembrane helices 6 and 7) that interacts with a G protein, and a carboxyl-terminal cytoplasmic domain that undergoes reversible phosphorylation on several Ser or Thr residues. The ligand-binding site (or, in the case of light reception, the light receptor) is buried deep in the membrane and includes residues from several of the transmembrane segments. Ligand binding (or light) induces a conformational change in the receptor, exposing a domain that can interact with a G protein. Heterotrimeric G proteins activate or inhibit effector enzymes (adenylyl cyclase, PDE, or PLC), which change the concentration of a second messenger (cAMP, cGMP, IP_3, or Ca^{2+}). In the hormone-detecting systems, the final output is an activated protein kinase that regulates some cellular process by phosphorylating a protein critical to that process. In sensory neurons, the output is a change in membrane potential and a consequent electrical signal that passes to another neuron in the pathway connecting the sensory cell to the brain.

All these systems self-inactivate. Bound GTP is converted to GDP by the intrinsic GTPase activity of G proteins, often augmented by GTPase-activating proteins (GAPs) or RGS proteins (regulators of G-protein

TABLE 12–8 Some Signals That Act through GPCRs

Amines	Somatostatin
Acetylcholine (muscarinic)	Tachykinin
Dopamine	Thyrotropin-releasing hormone
Epinephrine	Urotensin II
Histamine	**Protein hormones**
Serotonin	Follicle-stimulating hormone
Peptides	Gonadotropin
Angiotensin	Lutropin-choriogonadotropic hormone
Bombesin	Thyrotropin
Bradykinin	**Prostanoids**
Chemokine	Prostacyclin
Colecystokinin (CCK)	Prostaglandin
Endothelin	Thromboxane
Gonadotropin-releasing hormone	**Others**
Interleukin-8	Cannabinoids
Melanocortin	Lysosphingolipids
Neuropeptide Y	Melatonin
Neurotensin	Olfactory stimuli
Opioid	Rhodopsin
Orexin	

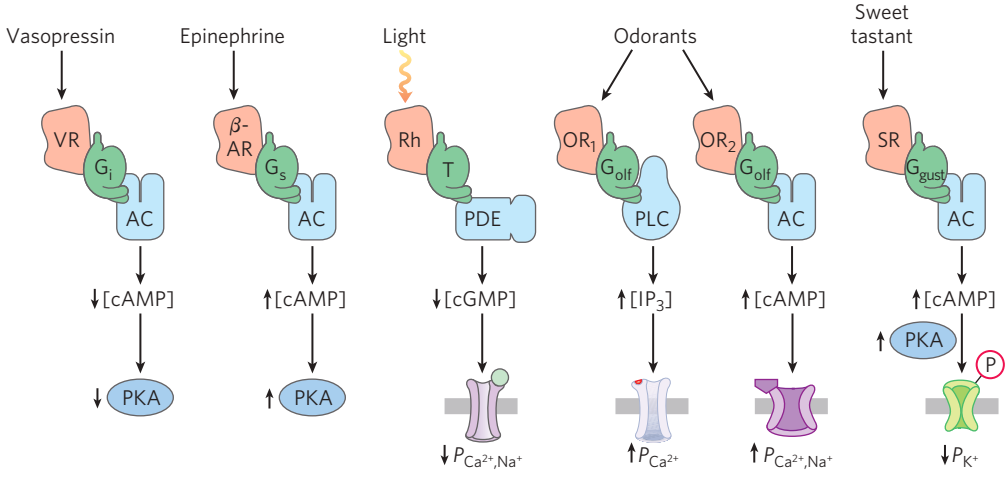


FIGURE 12–43 Common features of signaling systems that detect hormones, light, smells, and tastes. GPCRs provide signal specificity, and their interaction with G proteins provides signal amplification. Heterotrimeric G proteins activate effector enzymes: adenylyl cyclase (AC), phospholipase C (PLC), and phosphodiesterases (PDEs) that degrade cAMP or cGMP. Changes in concentration of the second messengers (cAMP, cGMP, IP_3) result in alterations of enzymatic activities by phosphorylation or alterations in the permeability (*P*) of surface membranes to Ca^{2+}, Na^+, and K^+. The resulting depolarization or hyperpolarization of the sensory cell (the signal) passes through relay neurons to sensory centers in the brain. In the best-studied cases, desensitization includes phosphorylation of the receptor and binding of a protein (arrestin) that interrupts receptor–G protein interactions. VR is the vasopressin receptor; β-AR is the β-adrenergic receptor. Other receptor and G-protein abbreviations are as used in earlier illustrations.

signaling; see Fig. 12–5 and Box 12–2, Fig. 4). In some cases, the effector enzymes that are the targets of modulation by G proteins also serve as GAPs. The desensitization mechanism involving phosphorylation of the carboxyl-terminal region followed by arrestin binding is widespread, and may be universal.

SUMMARY 12.10 Sensory Transduction in Vision, Olfaction, and Gustation

- Vision, olfaction, and gustation in vertebrates employ GPCRs, which act through heterotrimeric G proteins to change the V_m of a sensory neuron.
- In rod and cone cells of the retina, light activates rhodopsin, which activates the G protein transducin. The freed α subunit of transducin activates a cGMP phosphodiesterase, which lowers [cGMP] and thus closes cGMP-dependent ion channels in the outer segment of the neuron. The resulting hyperpolarization of the rod or cone cell carries the signal to the next neuron in the pathway, and eventually to the brain.
- In olfactory neurons, olfactory stimuli, acting through GPCRs and G proteins, trigger either an increase in [cAMP] (by activating adenylyl cyclase) or an increase in [Ca^{2+}] (by activating PLC). These second messengers affect ion channels and thus the V_m.
- Gustatory neurons have GPCRs that respond to tastants by altering levels of cAMP, which changes V_m by gating ion channels.
- There is a high degree of conservation of signaling proteins and transduction mechanisms across signaling systems and across species.

12.11 Regulation of the Cell Cycle by Protein Kinases

One of the most dramatic manifestations of signaling pathways is the regulation of the eukaryotic cell cycle. During embryonic growth and later development, cell division occurs in virtually every tissue. In the adult organism most tissues become quiescent. A cell's "decision" to divide or not is of crucial importance to the organism. When the regulatory mechanisms that limit cell division are defective and cells undergo unregulated division, the result is catastrophic—cancer. Proper cell division requires a precisely ordered sequence of biochemical events that assures every daughter cell a full complement of the molecules required for life. Investigations into the control of cell division in diverse eukaryotic cells have revealed universal regulatory mechanisms. Signaling mechanisms much like those discussed above are central in determining whether and when a cell undergoes cell division, and they also ensure orderly passage through the stages of the cell cycle.

The Cell Cycle Has Four Stages

Cell division accompanying mitosis in eukaryotes occurs in four well-defined stages **(Fig. 12–44)**. In the S (synthesis) phase, the DNA is replicated to produce copies for both daughter cells. In the G2 phase (G indicates the gap between divisions), new proteins are synthesized and the cell approximately doubles in size. In the M phase (mitosis), the maternal nuclear envelope breaks down, paired chromosomes are pulled to opposite poles of the cell, each set of daughter chromosomes is surrounded by a newly formed nuclear envelope, and cytokinesis pinches the cell in half, producing two daughter cells (see Fig. 24–24). In embryonic or rapidly proliferating tissue, each daughter cell divides again, but only after a waiting period (G1). In cultured animal cells the entire process takes about 24 hours.

After passing through mitosis and into G1, a cell either continues through another division or ceases to divide, entering a quiescent phase (G0) that may last hours, days, or the lifetime of the cell. When a cell in G0 begins to divide again, it reenters the division cycle through the G1 phase. Differentiated cells such as hepatocytes or adipocytes have acquired their specialized function and form; they remain in the G0 phase. Stem cells retain their potential to divide and to differentiate into any of a number of cell types.

Levels of Cyclin-Dependent Protein Kinases Oscillate

The timing of the cell cycle is controlled by a family of protein kinases with activities that change in response

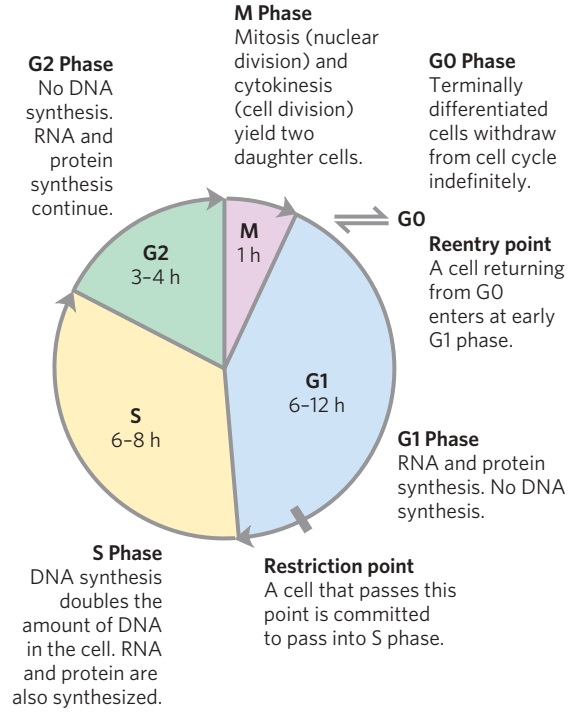
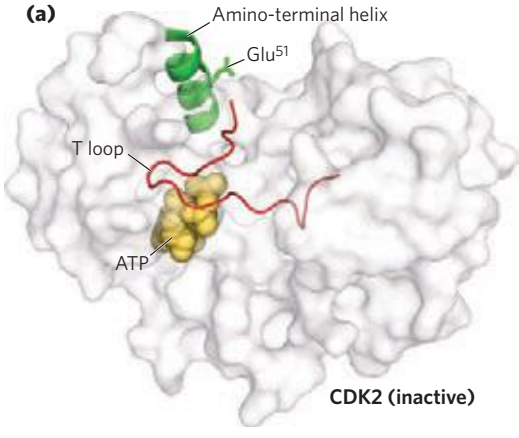


FIGURE 12–44 Eukaryotic cell cycle. The durations (in hours) of the four stages vary, but those shown are typical.

to cellular signals. By phosphorylating specific proteins at precisely timed intervals, these protein kinases orchestrate the metabolic activities of the cell to produce orderly cell division. The kinases are heterodimers with a regulatory subunit, **cyclin**, and a catalytic subunit, **cyclin-dependent protein kinase (CDK)**. In the absence of cyclin, the catalytic subunit is virtually inactive. When cyclin binds, the catalytic site opens up, a residue essential to catalysis becomes accessible **(Fig. 12–45)**, and the protein kinase activity of the catalytic subunit increases 10,000-fold. Animal cells have at least 10 different cyclins (designated A, B, and so forth) and at least 8 CDKs (CDK1 through CDK8), which act in various combinations at specific points in the cell cycle. Plants also use a family of CDKs to regulate their cell division in root and shoot meristems, the principal tissues in which division occurs.

In a population of animal cells undergoing synchronous division, some CDK activities show striking oscillations **(Fig. 12–46)**. These oscillations are the result of four mechanisms for regulating CDK activity: phosphorylation or dephosphorylation of the CDK, controlled degradation of the cyclin subunit, periodic synthesis of CDKs and cyclins, and the action of specific CDK-inhibiting proteins. The precisely timed activation and inactivation of a series of CDKs produce signals serving as a master clock that orchestrates the events in normal cell division and ensures that one stage is completed before the next begins.

Regulation of CDKs by Phosphorylation The activity of a CDK is strikingly affected by phosphorylation and dephosphorylation of two critical residues in the protein **(Fig. 12–47a)**. Phosphorylation of Tyr^{15} near the amino terminus by another protein kinase renders CDK2 inactive; the Ⓟ–Tyr residue is in the ATP-binding site of the kinase, and the negatively charged phosphate group blocks the entry of ATP. A specific phosphatase (a PTPase) dephosphorylates this Ⓟ–Tyr residue, permitting the binding of ATP. Phosphorylation of Thr^{160} in the "T loop" of CDK, catalyzed by yet another protein kinase, forces the T loop out of the substrate-binding cleft, permitting the binding of a specific downstream target protein and its phosphorylation by CDK (Fig. 12–45c).



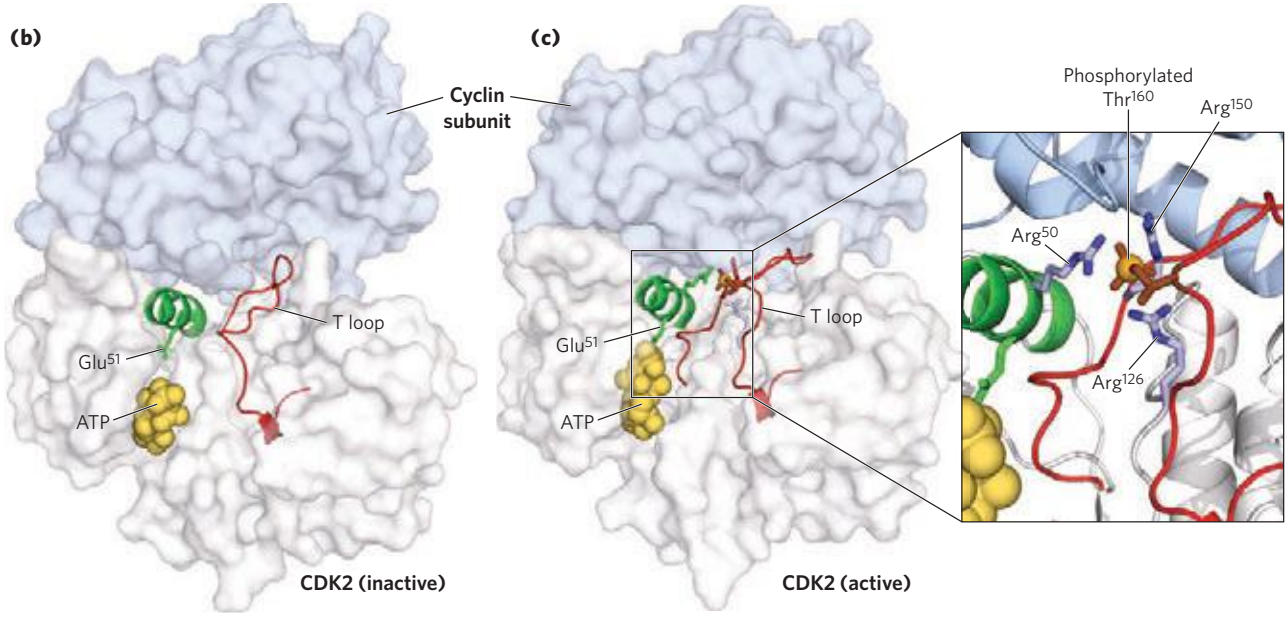


FIGURE 12–45 Activation of cyclin-dependent protein kinases (CDKs) by cyclin and phosphorylation. CDKs, a family of related enzymes, are active only when associated with cyclins, another protein family. The crystal structure of CDK2 with and without cyclin reveals the basis for this activation. **(a)** Without cyclin (PDB ID 1HCK), CDK2 folds so that one segment, the T loop, obstructs the binding site for protein substrates and thus inhibits protein kinase activity. The binding site for ATP is also near the T loop. **(b)** When cyclin binds (PDB ID 1FIN), it forces conformational changes that move the T loop away from the active site and reorient an amino-terminal helix, bringing a residue critical to catalysis (Glu^{51}) into the active site. **(c)** Phosphorylation of a Thr residue in the T loop produces a negatively charged residue that is stabilized by interaction with three Arg residues, holding CDK in its active conformation (PDB ID 1JST).

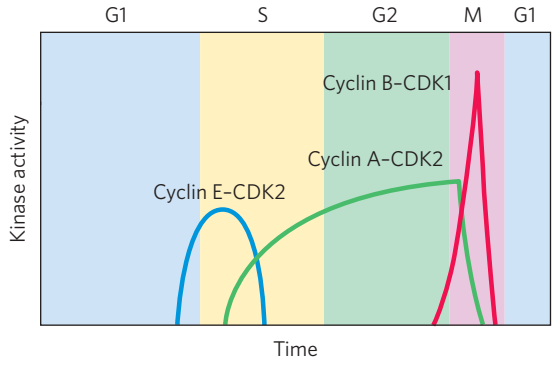


FIGURE 12–46 Variations in the activities of specific CDKs during the cell cycle in animals. Cyclin E-CDK2 activity peaks near the G1 phase–S phase boundary, when the active enzyme triggers synthesis of enzymes required for DNA synthesis (see Fig. 12-49). Cyclin A-CDK2 activity rises during the S and G2 phases, then drops sharply in the M phase, as cyclin B-CDK1 peaks.

The presence of single-strand breaks in DNA leads to arrest of the cell cycle in G2 by regulating a particular CDK. A specific protein kinase (called Rad3 in yeast), which is activated by single-strand breaks, triggers a cascade leading to the inactivation of the PTPase that dephosphorylates Tyr^{15} of CDK. The CDK remains inactive and the cell is arrested in G2, unable to divide until the DNA is repaired and the effects of the cascade are reversed.

Controlled Degradation of Cyclin Highly specific and precisely timed proteolytic breakdown of mitotic cyclins regulates CDK activity throughout the cell cycle. Progress through mitosis requires first the activation then the destruction of cyclins A and B, which activate the catalytic subunit of the M-phase CDK. These cyclins contain near their amino terminus the sequence –Arg–Thr–Ala–Leu–Gly–Asp–Ile–Gly–Asn–, the "destruction box," which targets them for degradation.

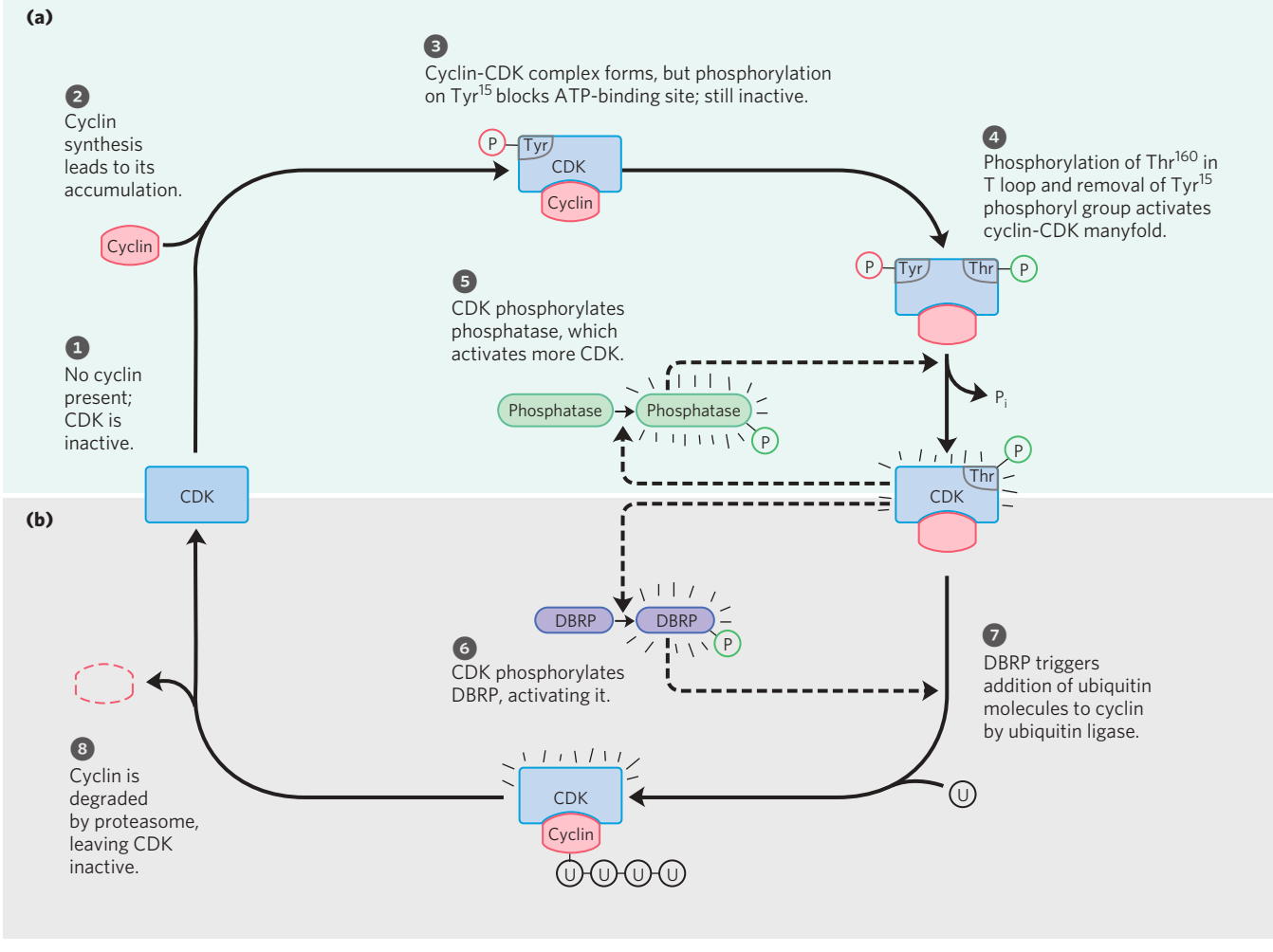


FIGURE 12–47 Regulation of CDK by phosphorylation and proteolysis. **(a)** The cyclin-dependent protein kinase activated at the time of mitosis (the M-phase CDK) has a "T loop" that can fold into the substrate-binding site. When Thr^{160} in the T loop is phosphorylated, the loop moves out of the substrate-binding site, activating the CDK manyfold (step ❹). **(b)** The active cyclin-CDK complex triggers its own inactivation by phosphorylation of DBRP (destruction box recognizing protein; step ❻). DBRP and ubiquitin ligase then attach several molecules of ubiquitin (U) to cyclin (step ❼), targeting it for destruction by proteasomes, proteolytic enzyme complexes (step ❽).

(This usage of "box" derives from the common practice, in diagramming the sequence of a nucleic acid or protein, of enclosing within a box a short sequence of nucleotide or amino acid residues with some specific function. It does not imply any three-dimensional structure.) The protein DBRP (*d*estruction *b*ox *r*ecognizing *p*rotein) recognizes this sequence and initiates the process of cyclin degradation by bringing together the cyclin and another protein, **ubiquitin**. Cyclin and activated ubiquitin are covalently joined by the enzyme ubiquitin ligase (Fig. 12–47b). Several more ubiquitin molecules are then appended, providing the signal for a proteolytic enzyme complex, or **proteasome**, to degrade cyclin.

What controls the timing of cyclin breakdown? A feedback loop occurs in the overall process shown in Figure 12–47. Increased CDK activity (step ❹) leads eventually to cyclin proteolysis (step ❽). Newly synthesized cyclin associates with and activates CDK, which phosphorylates and activates DBRP. Active DBRP then causes proteolysis of cyclin. The lowered cyclin level causes a decline in CDK activity, and the activity of DBRP also drops through slow, constant dephosphorylation and inactivation by a DBRP phosphatase. The cyclin level is ultimately restored by synthesis of new cyclin molecules.

The role of ubiquitin and proteasomes is not limited to the regulation of cyclin; as we shall see in Chapter 27, both also take part in the turnover of cellular proteins, a process fundamental to cellular housekeeping.

Regulated Synthesis of CDKs and Cyclins The third mechanism for changing CDK activity is regulation of the rate of synthesis of cyclin or CDK or both. For example, cyclin D, cyclin E, CDK2, and CDK4 are synthesized only when a specific transcription factor, E2F, is present in the nucleus to activate transcription of their genes. Synthesis of E2F is in turn regulated by extracellular signals such as **growth factors** and cytokines (developmental signals that induce cell division), compounds found to be essential for the division of mammalian cells in culture. They induce the synthesis of specific nuclear transcription factors essential to the production of the enzymes of DNA synthesis. Growth factors trigger phosphorylation of the nuclear proteins Jun and Fos, transcription factors that promote the synthesis of a variety of gene products, including cyclins, CDKs, and E2F. In turn, E2F controls production of several enzymes essential for the synthesis of deoxynucleotides and DNA, enabling cells to enter the S phase **(Fig. 12–48)**.

Inhibition of CDKs Finally, specific protein inhibitors bind to and inactivate specific CDKs. One such protein is p21, which we discuss below.

These four control mechanisms modulate the activity of specific CDKs that, in turn, control whether a cell will divide, differentiate, become permanently quiescent, or begin a new cycle of division after a period of quiescence. The details of cell cycle regulation, such as the number of different cyclins and kinases and the combinations in which they act, differ from species to species, but the basic mechanism has been conserved in the evolution of all eukaryotic cells.

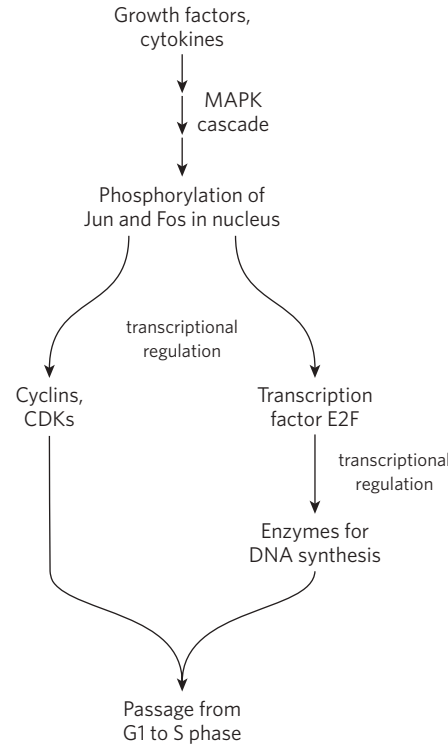


FIGURE 12–48 Regulation of cell division by growth factors. The path from growth factors to cell division leads through the enzyme cascade that activates MAPK; phosphorylation of the nuclear transcription factors Jun and Fos; and the activity of the transcription factor E2F, which promotes synthesis of several enzymes essential for DNA synthesis.

CDKs Regulate Cell Division by Phosphorylating Critical Proteins

We have examined how cells maintain close control of CDK activity, but how does the activity of CDK control the cell cycle? The list of target proteins that CDKs are known to act upon continues to grow, and much remains to be learned. But we can see a general pattern behind CDK regulation by inspecting the effect of CDKs on the structures of lamin and myosin and on the activity of retinoblastoma protein.

The structure of the nuclear envelope is maintained in part by highly organized meshworks of intermediate filaments composed of the protein lamin. Breakdown of the nuclear envelope before segregation of the sister chromatids in mitosis is partly due to the phosphorylation of lamin by a CDK, which causes lamin filaments to depolymerize.

A second kinase target is the ATP-driven contractile machinery (actin and myosin) that pinches a dividing cell into two equal parts during cytokinesis. After the division, CDK phosphorylates a small regulatory subunit of myosin, causing dissociation of myosin from actin

filaments and inactivating the contractile machinery. Subsequent dephosphorylation allows reassembly of the contractile apparatus for the next round of cytokinesis.

A third and very important CDK substrate is the **retinoblastoma protein, pRb**; when DNA damage is detected, this protein participates in a mechanism that arrests cell division in G1 **(Fig. 12–49)**. Named for the retinal tumor cell line in which it was discovered, pRb functions in most, perhaps all, cell types to regulate cell division in response to a variety of stimuli. Unphosphorylated pRb binds the transcription factor E2F; while bound to pRb, E2F cannot promote transcription of a group of genes necessary for DNA synthesis (the genes for DNA polymerase α, ribonucleotide reductase, and other proteins; see Chapter 25). In this state, the cell cycle cannot proceed from the G1 to the S phase, the step that commits a cell to mitosis and cell division. The pRb-E2F blocking mechanism is relieved when pRb is phosphorylated by cyclin E–CDK2, which occurs in response to a signal for cell division to proceed.

When the protein kinases ATM and ATR detect damage to DNA (signaled by the presence of the protein MRN at a double-strand break site), they phosphorylate p53, activating it to serve as a transcription factor that stimulates the synthesis of the protein p21 (Fig. 12–49). This protein inhibits the protein kinase activity of cyclin E–CDK2. In the presence of p21, pRb remains unphosphorylated and bound to E2F, blocking the activity of this transcription factor, and the cell cycle is arrested in G1. This gives the cell time to repair its DNA before entering the S phase, thereby avoiding the potentially disastrous transfer of a defective genome to one or both daughter cells. When the damage is too severe to allow effective repair, this same machinery triggers a process (apoptosis, described below) that leads to the death of the cell, preventing the possible development of a cancer.

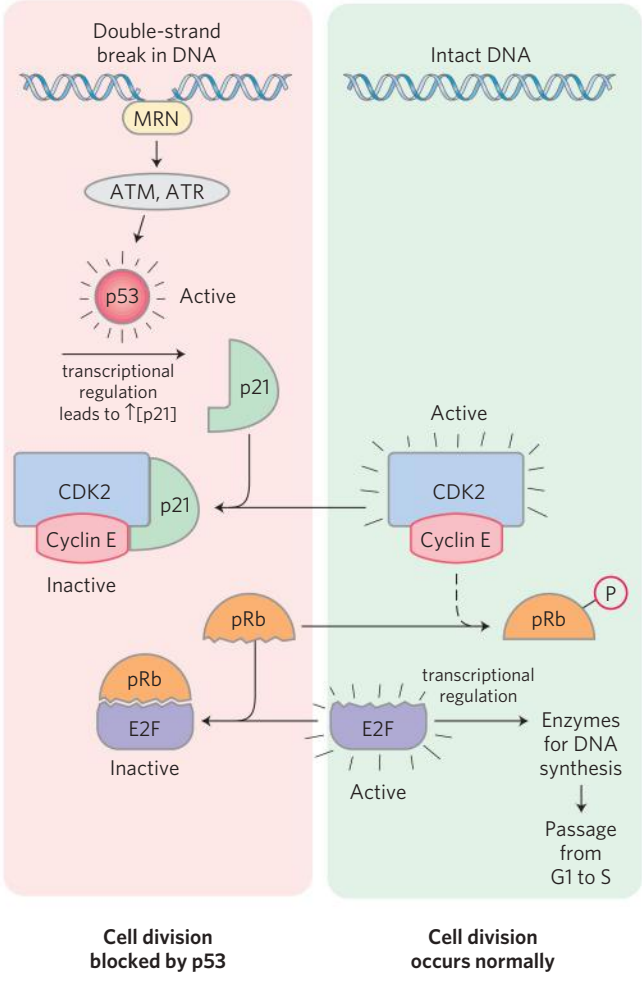


FIGURE 12–49 Regulation of passage from G1 to S by phosphorylation of pRb. Transcription factor E2F promotes transcription of genes for certain enzymes essential to DNA synthesis. The retinoblastoma protein, pRb, can bind E2F (lower left), inactivating it and preventing transcription of these genes. Phosphorylation of pRb by CDK2 prevents it from binding and inactivating E2F, and the genes are transcribed, allowing cell division. Damage to the cell's DNA (upper left) triggers a series of events that inactivate CDK2, blocking cell division. When the protein MRN detects damage to the DNA, it activates two protein kinases, ATM and ATR, and they phosphorylate and activate the transcription factor p53. Active p53 promotes the synthesis of another protein, p21, an inhibitor of CDK2. Inhibition of CDK2 stops the phosphorylation of pRb, which therefore continues to bind and inhibit E2F. With E2F inactivated, genes essential to cell division are not transcribed and cell division is blocked. When DNA has been repaired, this inhibition is released, and the cell divides.

SUMMARY 12.11 Regulation of the Cell Cycle by Protein Kinases

- Progression through the cell cycle is regulated by the cyclin-dependent protein kinases (CDKs), which act at specific points in the cycle, phosphorylating key proteins and modulating their activities. The catalytic subunit of CDKs is inactive unless associated with the regulatory cyclin subunit.
- The activity of a cyclin-CDK complex changes during the cell cycle through differential synthesis of CDKs, specific degradation of cyclin, phosphorylation and dephosphorylation of critical residues in CDKs, and binding of inhibitory proteins to specific cyclin-CDKs.
- Among the targets phosphorylated by cyclin-CDKs are proteins of the nuclear envelope and proteins required for cytokinesis and DNA repair.

12.12 Oncogenes, Tumor Suppressor Genes, and Programmed Cell Death

Tumors and cancer are the result of uncontrolled cell division. Normally, cell division is regulated by a family of extracellular growth factors, proteins that cause resting cells to divide and, in some cases, differentiate. The result is a precise balance between the formation of new cells (such as skin cells that die and are replaced every few weeks, or white blood cells that are replaced every few days) and cell destruction. When this balance is

disturbed by defects in regulatory proteins, the result is sometimes the formation of a clone of cells that divide repeatedly and without regulation (a tumor) until their presence interferes with the function of normal tissues—cancer. The direct cause is almost always a genetic defect in one or more of the proteins that regulate cell division. In some cases, a defective gene is inherited from one parent; in other cases, the mutation occurs when a toxic compound from the environment (a mutagen or carcinogen) or high-energy radiation interacts with the DNA of a single cell to damage it and introduce a mutation. In most cases there is both an inherited and an environmental contribution, and in most cases, more than one mutation is required to cause completely unregulated division and full-blown cancer.

Oncogenes Are Mutant Forms of the Genes for Proteins That Regulate the Cell Cycle

Oncogenes were originally discovered in tumor-causing viruses, then later found to be derived from genes in the animal host cells, **proto-oncogenes**, which encode growth-regulating proteins. During a viral infection, the host DNA sequence of a proto-oncogene is sometimes copied into the viral genome, where it proliferates with the virus. In subsequent viral infection cycles, the proto-oncogenes can become defective by truncation or mutation. Viruses, unlike animal cells, do not have effective mechanisms for correcting mistakes during DNA replication, so they accumulate mutations rapidly. When a virus carrying an oncogene infects a new host cell, the viral DNA (and oncogene) can be incorporated into the host cell's DNA, where it can now interfere with the regulation of cell division in the host cell. In an alternative, nonviral mechanism, a single cell in a tissue exposed to carcinogens may suffer DNA damage that renders one of its regulatory proteins defective, with the same effect as the oncogenic mechanism: failed regulation of cell division.

The mutations that produce oncogenes are genetically dominant; if either of a pair of chromosomes contains a defective gene, that gene product sends the signal "divide" and a tumor may result. The oncogenic defect can be in any of the proteins involved in communicating the "divide" signal. Oncogenes discovered thus far include those that encode secreted proteins, growth factors, transmembrane proteins (receptors), cytoplasmic proteins (G proteins and protein kinases), and the nuclear transcription factors that control the expression of genes essential for cell division (Jun, Fos).

Some oncogenes encode surface receptors with defective or missing signal-binding sites, such that their intrinsic Tyr kinase activity is unregulated. For example, the oncoprotein ErbB is essentially identical to the normal receptor for epidermal growth factor, except that ErbB lacks the amino-terminal domain that normally binds EGF **(Fig. 12–50)** and as a result sends the "divide" signal whether EGF is present or not. Mutations

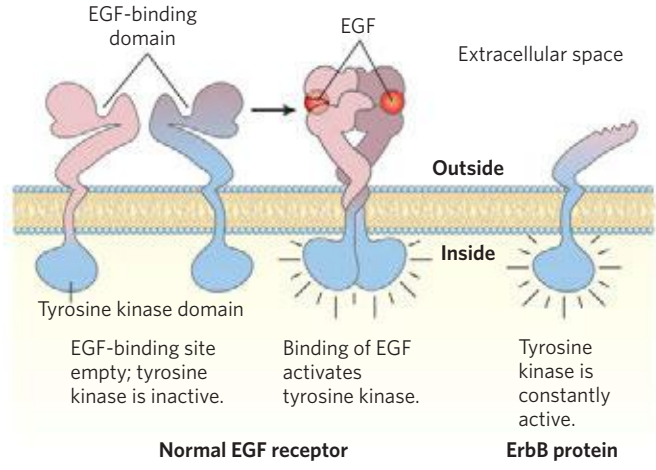


FIGURE 12–50 Oncogene-encoded defective EGF receptor. The product of the *erbB* oncogene (the ErbB protein) is a truncated version of the normal receptor for epidermal growth factor (EGF). Its intracellular domain has the structure normally induced by EGF binding, but the protein lacks the extracellular binding site for EGF. Unregulated by EGF, ErbB continuously signals cell division.

in *erbB2*, the gene for a receptor Tyr kinase related to ErbB, are commonly associated with cancers of the glandular epithelium in breast, stomach, and ovary. (For an explanation of the use of abbreviations in naming genes and their products, see Chapter 25.)

The prominent role played by protein kinases in signaling processes related to normal and abnormal cell division has made them a prime target in the development of drugs for the treatment of cancer (Box 12–5). Mutant forms of the G protein Ras are common in tumor cells. The *ras* oncogene encodes a protein with normal GTP binding but no GTPase activity. The mutant Ras protein is therefore always in its activated (GTP-bound) form, regardless of the signals arriving through normal receptors. The result can be unregulated growth. Mutations in *ras* are associated with 30% to 50% of lung and colon carcinomas and more than 90% of pancreatic carcinomas. ■

Defects in Certain Genes Remove Normal Restraints on Cell Division

Tumor suppressor genes encode proteins that normally restrain cell division. Mutation in one or more of these genes can lead to tumor formation. Unregulated growth due to defective tumor suppressor genes, unlike that due to oncogenes, is genetically recessive; tumors form only if *both* chromosomes of a pair contain a defective gene. This is because the function of these genes is to prevent cell division, and if either copy of the gene for such a protein is normal, the normal inhibition of division will take place. In a person who inherits one correct copy and one defective copy, every cell begins with one defective copy of the gene. If any one of those 10^{12} somatic cells undergoes mutation in the one good copy, a tumor may grow from that doubly

BOX 12–5 MEDICINE Development of Protein Kinase Inhibitors for Cancer Treatment

When a single cell divides without any regulatory limitation, it eventually gives rise to a clone of cells so large that it interferes with normal physiological functions (Fig. 1). This is cancer, a leading cause of death in the developed world, and increasingly so in the developing world. In all types of cancer, the normal regulation of cell division has become dysfunctional due to defects in one or more genes. For example, genes encoding proteins that normally send intermittent signals for cell division become oncogenes, producing constitutively active signaling proteins, or genes encoding proteins that normally restrain cell division (tumor suppressor genes) mutate to produce proteins that lack this braking function. In many tumors, both kinds of mutation have occurred.

Many oncogenes and tumor suppressor genes encode protein kinases or proteins that act in pathways upstream from protein kinases. It is therefore reasonable to hope that specific inhibitors of protein kinases could prove valuable in the treatment of cancer. For example, a mutant form of the EGF receptor is a constantly active receptor Tyr kinase (RTK), signaling cell division whether EGF is present or not (see Fig. 12–50). In about 30% of all women with invasive breast cancer, a mutation in the receptor gene HER2/neu yields an RTK with activity increased up to 100-fold. Another RTK, **vascular endothelial growth factor receptor (VEGFR)**, must be activated for the formation of new blood vessels (angiogenesis) to provide a solid tumor with its own blood supply, and inhibition of VEGFR might starve a tumor of essential nutrients. Nonreceptor Tyr kinases can also mutate, resulting in constant signaling and unregulated cell division. For example, the oncogene *Abl* (from the *Ab*elson *l*eukemia virus) is associated with acute myeloid leukemia, a relatively rare blood disease (~5,000 cases a year in the United States). Another group of oncogenes encode unregulated cyclin-dependent protein kinases. In each of these cases, specific protein kinase inhibitors might be valuable chemotherapeutic agents in the treatment of disease. Not surprisingly, huge efforts are under way to develop such inhibitors. How should one approach this challenge?

FIGURE 1 Unregulated division of a single cell in the colon led to a primary cancer that metastasized to the liver. Secondary cancers are seen as white patches in this liver obtained at autopsy.

Protein kinases of all types show striking conservation of structure at the active site. All share with the prototypical PKA structure the features shown in Figure 2: two lobes that enclose the active site, with a P loop that helps to align and bind the phosphoryl groups of ATP, an activation loop that moves to open the active site to the protein substrate, and a C helix that changes position as the enzyme is activated, bringing the residues in the substrate-binding cleft into their binding positions.

The simplest protein kinase inhibitors are ATP analogs that occupy the ATP-binding site but cannot serve as phosphoryl group donors. Many such compounds are known, but their clinical usefulness is limited by their lack of selectivity—they inhibit virtually all protein kinases and would produce unacceptable side effects. More selectivity

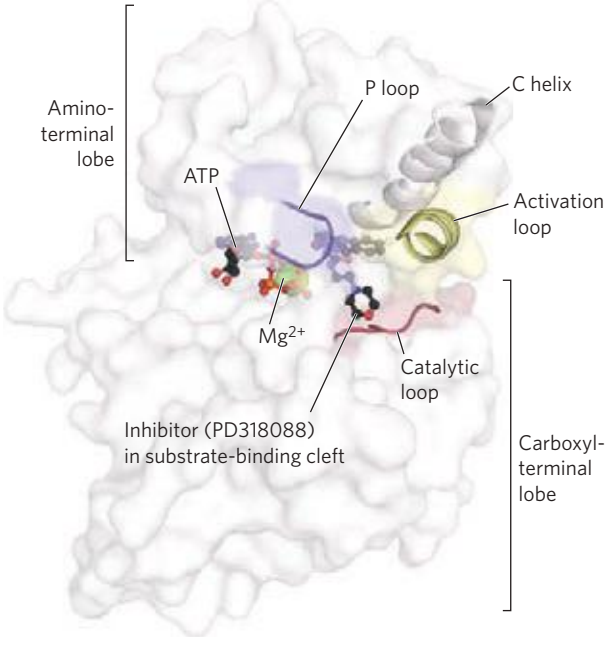


FIGURE 2 Conserved features of the active site of protein kinases (PDB ID 1S9I). The amino-terminal and carboxyl-terminal lobes surround the active site of the enzyme, near the catalytic loop and the site where ATP binds. The activation loop of this and many other kinases undergoes phosphorylation, then moves away from the active site to expose the substrate-binding cleft, which in this image is occupied by a specific inhibitor of this enzyme, PD318088. The P loop is essential in the binding of ATP, and the C helix must also be correctly aligned for ATP binding and kinase activity.

is seen with compounds that fill part of the ATP-binding site but also interact outside this site, with parts of the protein unique to the target protein kinase. A third possible strategy is based on the fact that although the active conformations of all protein kinases are similar, their inactive conformations are not. Drugs that target the inactive conformation of a specific protein kinase and prevent its conversion to the active form may have a higher specificity of action. A fourth approach employs the great specificity of antibodies. For example, monoclonal antibodies (p. 178) that bind the extracellular portions of specific RTKs could eliminate the receptors' kinase activity by preventing dimerization or by causing their removal from the cell surface. In some cases, an antibody selectively binding to the surface of cancer cells could cause the immune system to attack those cells.

The search for drugs active against specific protein kinases has yielded encouraging results. For example, imatinib mesylate (Gleevec; Fig. 3a), one of the small-molecule inhibitors, has proved nearly 100% effective in bringing about remission in patients with early-stage chronic myeloid leukemia. Erlotinib (Tarceva; Fig. 3b), which targets EGFR, is effective against advanced non-small-cell lung cancer (NSCLC). Because many cell-division signaling systems involve more than one protein kinase, inhibitors that act on several protein kinases may be useful in the treatment of cancer. Sunitinib (Sutent) and sorafenib (Nexavar) target several protein kinases, including VEGFR and PDGFR. These two drugs are in clinical use for patients with gastrointestinal stromal tumors and advanced renal cell carcinoma, respectively. Trastuzumab (Herceptin), cetuximab (Erbitux), and bevacizumab (Avastin) are monoclonal antibodies that target HER2/neu, EGFR, and VEGFR, respectively; all three drugs are in clinical use for certain types of cancer. Detailed knowledge of the structure around the ATP-binding site makes it possible to design drugs that inhibit a *specific* protein kinase by (1) blocking the critical ATP-binding site, while (2) interacting with residues around that site that are *unique* to that particular protein kinase.

At least a hundred more compounds are in preclinical trials. Among the drugs being evaluated are some obtained from natural sources and some produced by synthetic chemistry. Indirubin is a component of a Chinese herbal preparation traditionally used to treat certain leukemias; it inhibits CDK2 and CDK5. Roscovatine (Fig. 3d), a substituted adenine, has a benzyl ring that makes it highly specific as an inhibitor of CDK2. With several hundred potential anti-cancer drugs heading toward clinical testing, it is realistic to hope that some will prove more effective or more target-specific than those now in use.

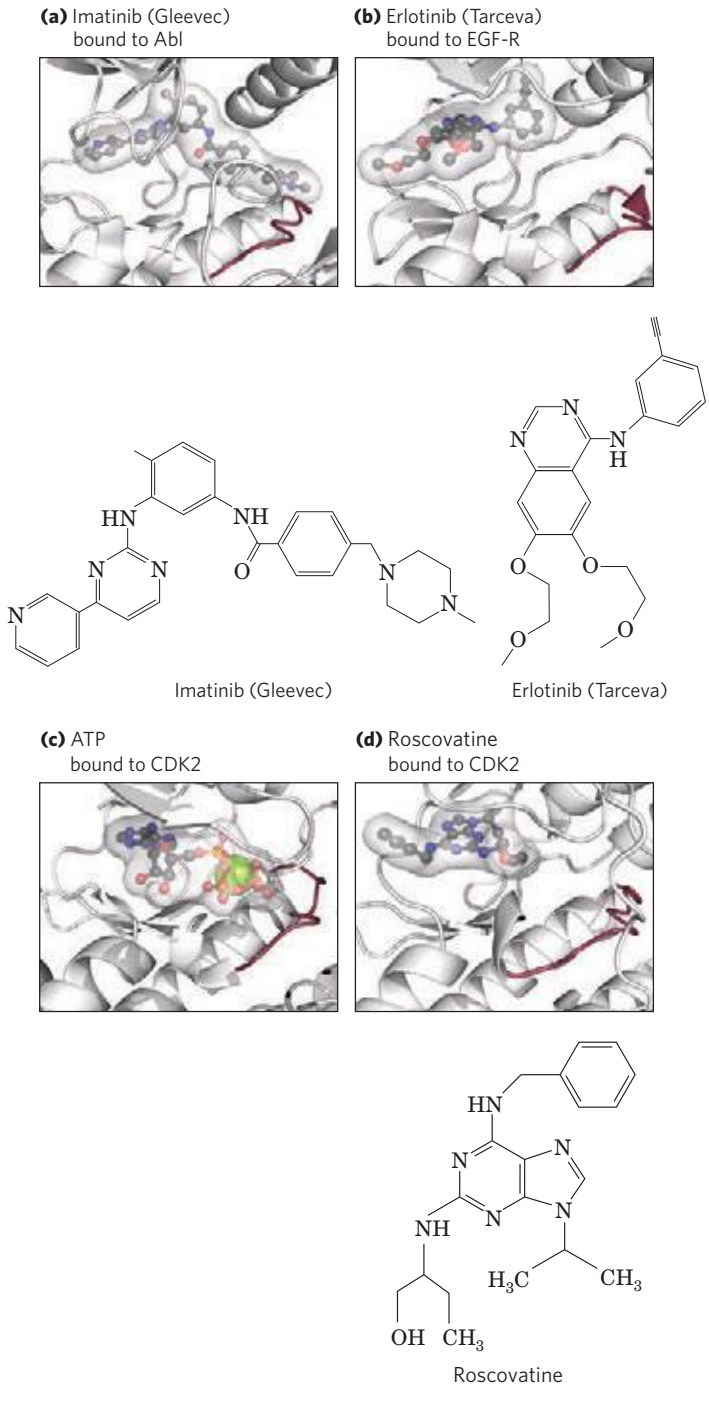


FIGURE 3 Some protein kinase inhibitors now in clinical trials or clinical use, showing their binding to the target protein. **(a)** Imatinib binds to the Abl oncogene kinase active site (PDB ID 1IEP); it occupies both the ATP-binding site and a region adjacent to that site. **(b)** Erlotinib binds to the active site of EGFR (PDB ID 1M17). **(c), (d)** Roscovatine is an inhibitor of the cyclin-dependent kinase CDK2; shown here are normal Mg-ATP binding (c) at the active site (PDB ID 1S9I) and roscovatine binding (d), which prevents the binding of ATP (PDB ID 2A4L).

mutant cell. Mutations in both copies of the genes for pRb, p53, or p21 yield cells in which the normal restraint on cell division is lost and a tumor forms.

Retinoblastoma occurs in children and causes blindness if not surgically treated. The cells of a retinoblastoma have two defective versions of the *Rb* gene (two defective alleles). Very young children who develop retinoblastoma commonly have multiple tumors in both eyes. These children have inherited one defective copy of the *Rb* gene, which is present in every cell; each tumor is derived from a single retinal cell that has undergone a mutation in its one good copy of the *Rb* gene. (A fetus with two mutant alleles in every cell is nonviable.) People with retinoblastoma who survive childhood also have a high incidence of cancers of the lung, prostate, and breast later in life.

A far less likely event is that a person born with two good copies of the *Rb* gene will have independent mutations in both copies in the *same* cell. Some individuals do develop retinoblastomas later in childhood, usually with only one tumor in one eye. These individuals were presumably born with two good copies (alleles) of *Rb* in every cell, but both *Rb* alleles in a single retinal cell have undergone mutation, leading to a tumor. After about age three, retinal cells stop dividing, and retinoblastomas at later ages are quite rare.

Stability genes (also called caretaker genes) encode proteins that function in the repair of major genetic defects that result from aberrant DNA replication, ionizing radiation, or environmental carcinogens. Mutations in these genes lead to a high frequency of unrepaired damage (mutations) in other genes, including proto-oncogenes and tumor suppressor genes, and thus to cancer. Among the stability genes are *ATM* (see Fig. 12–49); the *XP* gene family, in which mutations lead to xeroderma pigmentosum; and the *BRCA1* genes associated with some types of breast cancer (see Box 25–1). Mutations in the gene for p53 also cause tumors; in more than 90% of human cutaneous squamous cell carcinomas (skin cancers) and in about 50% of all other human cancers, *p53* is defective. Those very rare individuals who *inherit* one defective copy of *p53* commonly have the Li-Fraumeni cancer syndrome, with multiple cancers (of the breast, brain, bone, blood, lung, and skin) occurring at high frequency and at an early age. The explanation for multiple tumors in this case is the same as that for *Rb* mutations: an individual born with one defective copy of *p53* in every somatic cell is likely to suffer a second *p53* mutation in more than one cell during his or her lifetime.

In summary, then, three classes of defects can contribute to the development of cancer: oncogenes, in which the defect is the equivalent of a car's accelerator pedal being stuck down, with the engine racing; mutated tumor suppressor genes, in which the defect leads to the equivalent of brake failure; and mutated stability genes, with the defect leading to unrepaired damage to the cell's replication machinery, the equivalent of an unskilled car mechanic.

Mutations in oncogenes and tumor suppressor genes do not have an all-or-none effect. In some cancers, perhaps in all, the progression from a normal cell to a malignant tumor requires an accumulation of mutations (sometimes over several decades), none of which, alone, is responsible for the end effect. For example, the development of colorectal cancer has several recognizable stages, each associated with a mutation **(Fig. 12–51)**. If an epithelial cell in the colon undergoes mutation of both copies of the tumor suppressor gene *APC* (adenomatous polyposis coli), it begins to divide faster than normal and produces a clone of itself, a benign polyp (early adenoma). For reasons not yet known, the *APC* mutation results in chromosomal instability, and whole regions of a chromosome are lost or rearranged during cell division. This instability can lead to another mutation, commonly in *ras*, that converts the clone into an intermediate adenoma. A third mutation (often in the tumor suppressor gene *DCC*) leads to a late adenoma. Only when both copies of *p53* become defective does this cell mass become a carcinoma—a malignant, life-threatening tumor. The full sequence therefore requires at least seven genetic "hits": two on each of three tumor suppressor genes (*APC*, *DCC*, and *p53*) and one on the proto-oncogene *ras*. There are probably several other routes to colorectal cancer as well, but the principle that full malignancy results only from multiple mutations is likely to hold true for all of them. When a polyp is detected in the early adenoma stage and the cells containing the first mutations are removed surgically, late adenomas and carcinomas will not develop; hence the importance of early detection. Cells and organisms, too, have their early detection systems. For example, the ATM and ATR proteins described in Section 12.11 can detect DNA damage too extensive to be repaired effectively. They then trigger, through a pathway that includes p53, the process of apoptosis, in which a cell that has become dangerous to the organism kills itself. ■

Apoptosis Is Programmed Cell Suicide

Many cells can precisely control the time of their own death by the process of **programmed cell death**, or **apoptosis** (app′-a-toe′-sis; from the Greek for "dropping off," as in leaves dropping in the fall). One trigger for apoptosis is irreparable damage to DNA. Programmed cell death also occurs during the development of an embryo, when some cells must die to give a tissue or organ its final shape. Carving fingers from stubby limb buds requires the precisely timed death of cells between developing finger bones. During development of the nematode *C. elegans* from a fertilized egg, exactly 131 cells (of a total of 1,090 somatic cells in the embryo) must undergo programmed death in order to construct the adult body.

Apoptosis also has roles in processes other than development. If a developing antibody-producing cell

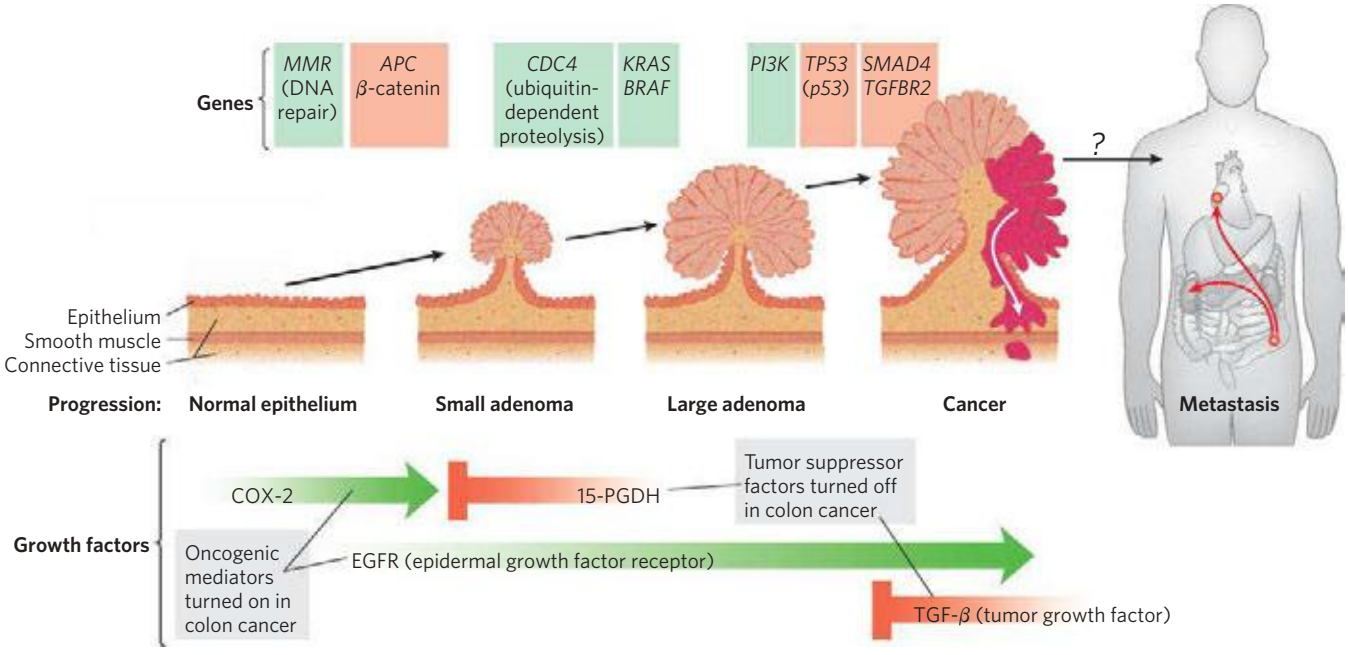


FIGURE 12–51 Multistep transition from normal epithelial cell to colorectal cancer. Serial mutations in oncogenes (green) or tumor suppressor genes (red) lead to progressively less control of cell division, until finally an active tumor forms, which can sometimes metastasize (spread from the initial site to other regions of the body). Mutation of the *MMR* gene leads to defective DNA repair and consequently to a higher rate of mutation. Mutations in both copies of the tumor suppressor gene *APC* lead to benign clusters of epithelial cells that multiply too rapidly (early adenoma). The *CDC4* oncogene results in defective ubiquitination, which is essential to the regulation of cyclin-dependent kinases (see Fig. 12–47). The oncogenes *KRAS* and *BRAF* encode ras and raf proteins (see Fig. 12–15), and this further disruption of signaling leads to the formation of a large adenoma, which may be detected by colonoscopy as a benign polyp. Oncogenic mutations in the *PI3K* gene that encodes the enzyme phosphoinositide-3 kinase, or in *PTEN*, which regulates the synthesis of this enzyme, lead to a further strengthening of the signal: divide now. When a cell in one of the polyps undergoes further mutations, in the tumor suppressor genes *DCC* and *p53* (see Fig. 12–49) for example, increasingly aggressive tumors form. Finally, mutations in other tumor suppressor genes such as *SMAD4* lead to a malignant tumor and sometimes to a metastatic tumor that can spread to other tissues. A second type of mutation that can add to the deleterious effects is one that affects the production or action of growth factors or their receptors (bottom). Mutations in EGFR (epidermal growth factor receptor) or TGF-β (transforming growth factor-β) favor uncontrolled growth, as do mutations in the enzymes that produce certain prostaglandins (COX-2; cyclooxygenase; see pp. 845–846) or 15-PGDH (15-hydroxyprostaglandin dehydrogenase). Most malignant tumors of other tissues probably result from a series of mutations such as this, although not necessarily these particular genes, or in this order.

generates antibodies against a protein or glycoprotein normally present in the body, that cell undergoes programmed death in the thymus gland—an essential mechanism for eliminating anti-self antibodies (the cause of many autoimmune diseases). The monthly sloughing of cells of the uterine wall (menstruation) is another case of apoptosis mediating normal cell death. The dropping of leaves in the fall is the result of apoptosis in specific cells of the stem. Sometimes cell suicide is not programmed but occurs in response to biological circumstances that threaten the rest of the organism. For example, a virus-infected cell that dies before completion of the infection cycle prevents spread of the virus to nearby cells. Severe stresses such as heat, hyperosmolarity, UV light, and gamma irradiation also trigger cell suicide; presumably the organism is better off with any aberrant, potentially mutated cells dead.

The regulatory mechanisms that trigger apoptosis involve some of the same proteins that regulate the cell cycle. The signal for suicide often comes from outside, through a surface receptor. Tumor necrosis factor (TNF), produced by cells of the immune system, interacts with cells through specific TNF receptors. These receptors have TNF-binding sites on the outer face of the plasma membrane and a "death domain" (~80 amino acid residues) that carries the self-destruct signal through the membrane to cytosolic proteins such as TRADD (*T*NF *r*eceptor–*a*ssociated *d*eath *d*omain) **(Fig. 12–52)**. Another receptor, Fas, has a similar death domain that allows it to interact with the cytosolic protein FADD (*F*as-*a*ssociated *d*eath *d*omain), which activates the cytosolic protease caspase 8. This enzyme belongs to a family of proteases that participate in apoptosis; all are synthesized as inactive proenzymes, all have a critical Cys residue at the active site, and all hydrolyze their target proteins on the carboxyl-terminal side of specific Asp residues (hence the name caspase, from *C*ys and *Asp*).

When caspase 8, an "initiator" caspase, is activated by an apoptotic signal carried through FADD, it further self-activates by cleaving its own proenzyme form. Mitochondria are one target of active caspase 8. The protease

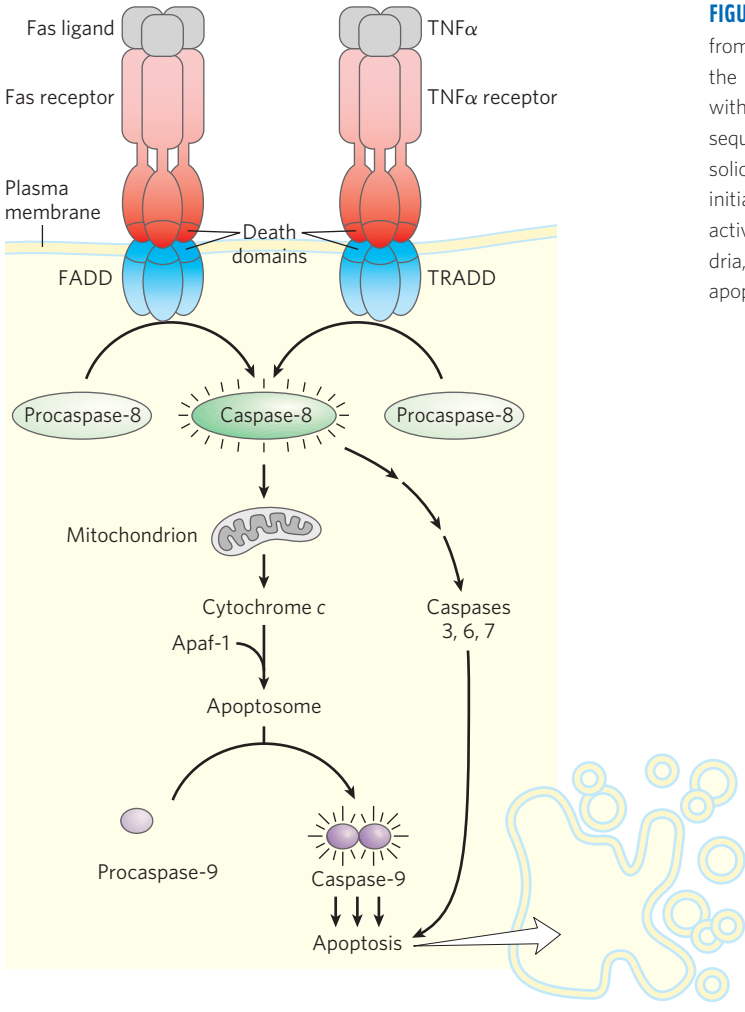


FIGURE 12–52 Initial events of apoptosis. Apoptosis-triggering signals from outside the cell (Fas and TNFα) bind to their specific receptors in the plasma membrane (FasR, TNFαR). The occupied receptors interact with the cytosolic proteins FADD and TRADD through an 80 amino acid sequence called the "death domain" on both the receptors and these cytosolic targets, which become activated. Activation of FADD and TRADD initiates a proteolytic cascade that leads to apoptosis. FADD and TRADD activate caspase-8, which acts to release cytochrome *c* from mitochondria, which, in concert with protein Apaf-1, activates caspase-9, triggering apoptosis.

causes the release of certain proteins contained between the inner and outer mitochondrial membranes: cytochrome *c* (Chapter 19) and several "effector" caspases. Cytochrome *c* binds to the proenzyme form of the effector enzyme caspase 9 and stimulates its proteolytic activation. The activated caspase 9 in turn catalyzes wholesale destruction of cellular proteins—a major cause of apoptotic cell death. One specific target of caspase action is a caspase-activated deoxyribonuclease.

In apoptosis, the monomeric products of protein and DNA degradation (amino acids and nucleotides) are released in a controlled process that allows them to be taken up and reused by neighboring cells. Apoptosis thus allows the organism to eliminate a cell that is unneeded or potentially dangerous without wasting its components.

SUMMARY 12.12 Oncogenes, Tumor Suppressor Genes, and Programmed Cell Death

- Oncogenes encode defective signaling proteins. By continually giving the signal for cell division, they lead to tumor formation. Oncogenes are genetically dominant and may encode defective growth factors, receptors, G proteins, protein kinases, or nuclear regulators of transcription.
- Tumor suppressor genes encode regulatory proteins that normally inhibit cell division; mutations in these genes are genetically recessive but can lead to tumor formation.
- Cancer is generally the result of an accumulation of mutations in oncogenes and tumor suppressor genes.
- When stability genes, which encode proteins necessary for the repair of genetic damage, are mutated, other mutations go unrepaired, including mutations in proto-oncogenes and tumor suppressor genes that can lead to cancer.
- Apoptosis is programmed and controlled cell death that functions during normal development and adulthood to get rid of unnecessary, damaged, or infected cells. Apoptosis can be triggered by extracellular signals such as TNF, acting through plasma membrane receptors.

Key Terms

Terms in bold are defined in the glossary.

signal transduction 433
specificity 433
cooperativity 434
amplification 434
enzyme cascade 434
modularity 434
scaffold proteins 434
desensitization 434
integration 434
Scatchard analysis 435
G protein–coupled receptors (GPCRs) 437
guanosine nucleotide–binding proteins 437
G proteins 437
second messenger 437
agonist 438
antagonist 438
β-adrenergic receptors 438
heptahelical receptors 438
stimulatory G protein (G_s) 438
adenylyl cyclase 438
cAMP-dependent protein kinase (protein kinase A; PKA) 438
P loop 441
GTPase activator protein (GAP) 442
regulator of G protein signaling (RGS) 442
guanosine nucleotide–exchange factor (GEF) 442
consensus sequence 443
β-arrestin (βarr; arrestin 2) 446
G protein–coupled receptor kinases (GRKs) 446
cAMP response element binding protein (CREB) 446
inhibitory G protein (G_i) 446
adaptor proteins 446
AKAPs (A kinase anchoring proteins) 447
phospholipase C (PLC) 447
inositol 1,4,5-trisphosphate (IP_3) 447
green fluorescent protein (GFP) 448
fluorescence resonance energy transfer (FRET) 448
protein kinase C (PKC) 450
calmodulin (CaM) 451
Ca^{2+}/calmodulin-dependent protein kinases (CaM kinases) 451
receptor Tyr kinase (RTK) 453
autophosphorylation 453
SH2 domain 454
small G proteins 455
MAPKs 455
cytokine 457
guanosine 3′,5′-cyclic monophosphate (cyclic GMP; cGMP) 459
cGMP-dependent protein kinase (protein kinase G; PKG) 459
atrial natriuretic factor (ANF) 459
NO synthase 460
angina pectoris 460
PTB domains 461
voltage-gated ion channels 465
nicotinic acetylcholine receptor 467
ionotropic 468
integrin 470
hormone response element (HRE) 471
two-component signaling systems 473
receptor histidine kinase 473
response regulator 473
receptorlike kinase (RLK) 476
retinal 477
rhodopsin 477
opsin 477
transducin 478
rhodopsin kinase 480
receptor potential 481
gustducin 481
cyclin 485
cyclin-dependent protein kinase (CDK) 485
ubiquitin 487
proteasome 487
growth factors 487
retinoblastoma protein (pRb) 488
oncogene 489
proto-oncogene 489
tumor suppressor gene 489
programmed cell death 492
apoptosis 492

Further Reading

General

Cohen, P. (2000) The regulation of protein function by multisite phosphorylation—a 25 year update. *Trends Biochem. Sci.* **25**, 596–601.
Historical account of protein phosphorylation.

Giepmans, B.N.G., Adams, S.R., Ellisman, M.H., & Tsien, R.Y. (2006) The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217–224.
A short, intermediate-level review of FRET.

Gomperts, B., Kramer, I.M., & Tatham, P.E.R. (2009) *Signal Transduction,* 2nd edn, Academic Press, New York.
A beautifully illustrated and clear description of all of the signaling systems discussed in this chapter.

Marks, F., Klingmüller, U., & Müller-Decker, K. (2009) *Cellular Signal Processing: An Introduction to the Molecular Mechanisms of Signal Transduction*, Garland Science, New York.

Pawson, T. & Scott, J.D. (2005) Protein phosphorylation in signaling—50 years and counting. *Trends Biochem. Sci.* **30**, 286–290.

G Protein–Coupled Receptors (GPCRs)

Aktories, K. (2011) Bacterial protein toxins that modify host regulatory GTPases. *Nat. Rev. Microbiol.* **9**, 487–498.

Awais, M. & Ozawa, T. (2011) Illuminating intracellular signaling and molecules for single cell analyisis. *Mol. Biosyst.* **7**, 1376–1387.
Intermediate-level review of methods for using fluorescent probes in cell biology and biochemistry.

Beene, D.L. & Scott, J.D. (2007) A-kinase anchoring proteins take shape. *Curr. Opin. Cell Biol.* **19**, 192–198.

Birnbaumer, L. (2007) The discovery of signal transduction by G proteins: a personal account and an overview of the initial findings and contributions that led to our present understanding. *Biochim. Biophys. Acta Biomembr.* **1768**, 756–771.

Escribá, P.V. (2007) G protein–coupled receptors, signaling mechanisms and pathophysiological relevance. *Biochim. Biophys. Acta Biomembr.* **1768**, 747.
The editorial introduction to a series of 20 papers on GPCRs.

Francis, S.H., Blount, M.A., & Corbin, J.D. (2011) Mammalian cyclic nucleotide phosphodiesterases: molecular mechanisms and physiological function. *Physiol. Rev.* **91**, 651–690.

Kobilka, B.K. (2011) Structural insights into adrenergic receptor function and pharmacology. *Trends Pharmacol. Sci.* **32**, 213–218.
Intermediate-level review.

Kremers, G.-J., Gilbert, S.G., Cranfill, P.J., Davidson, M.W., & Piston, D.W. (2011) Fluorescent proteins at a glance. *J. Cell Sci.* **124**, 157–160.
Brief review of methodology for using fluorescently labeled proteins.

Lefkowitz, R.J. (2007) Introduction to special section on β-arrestins. *Annu. Rev. Physiol.* **69**.
This introduces five excellent advanced reviews on the roles of arrestin.

Malumbres, M. (2011) Physiological relevance of cell cycle kinases. *Physiol. Rev.* **91**, 973–1007.

Advanced and extensive review of the roles of protein kinases in regulating the cell cycle and their roles in cancer.

Pearce, L.R., Komander, D., & Alessi, D.R. (2010) The nuts and bolts of AGC protein kinases. *Nat. Rev. Mol. Cell Biol.* **11**, 9–22.

Advanced review of the PKA, PKG, PKC family of kinases.

Pinna, L.A. & Ruzzene, M. (1996) How do protein kinases recognize their substrates? *Biochim. Biophys. Acta* **1314**, 191–225.

Advanced review of the factors, including consensus sequences, that give protein kinases their specificity.

Rasmussen, S.G.F., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, K.Y., Kobilka, T.S., Thian, F.S., Chae, P.S., Pardon, E., Calinski, D., et al. (2011) Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555.

The definitive determination of the structure and interaction of receptor and G protein.

Receptor Enzymes

Garbers, D.L., Chrisman, T.D., Wiegn, P., Katafuchi, T., Albanesi, J.P., Bielinski, V., Barylko, B., Redfield, M.M., & Burnett, J.C., Jr. (2006) Membrane guanylyl cyclase receptors: an update. *Trends Endocrinol. Metab.* **17**, 251–258.

Karnoub, A.E. & Weinberg, R.A. (2008) Ras oncogenes: split personalities. *Nat. Rev. Mol. Cell Biol.* **9**, 517–531.

Lemmon, M.A. & Schlessinger, J. (2010) Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134.

Misono, K.S., Philo, J.S., Arakawa, T., Ogata, C.M., Qiu, Y., Ogawa, H., & Young, H.S. (2011) Structure, signaling mechanism and regulation of the natriuretic peptide receptor guanylate cyclase. *FEBS J.* **278**, 1818–1829.

Adaptor Proteins and Membrane Rafts

Dehmelt, L. & Bastiaens, P.I.H. (2010) Spatial organization of intracellular communication: insights from imaging. *Nat. Rev. Mol. Cell Biol.* **11**, 440–452.

Good, M.C., Zalatan, J.G., & Lim, W.A. (2011) Scaffold proteins: hubs for controlling the flow of cellular information. *Science* **332**, 680–686.

Intermediate-level review of scaffold structure and function.

Schwarz-Romond, T. & Gorski, S.A. (2010) Focus on the spatial organization of signaling. *EMBO J.* **29**, 2675–2676.

Editorial introduction to a collection of eight excellent reviews of spatial aspects of signaling.

Smith, F.D. & Scott, J.D. (2006) Anchored cAMP signaling: onward and upward. A short history of compartmentalized cAMP signal transduction. *Eur. J. Cell Biol.* **85**, 582–592.

Short, intermediate-level review introducing an entire journal issue on the subject of AKAPs and cAMP signaling.

Receptor Ion Channels

See also Chapter 11, Further Reading, Ion Channels.

Ashcroft, F.M. (2006) From molecule to malady. *Nature* **440**, 440–447.

Short, intermediate-level review of human diseases associated with defects in ion channels.

Changeux, J.-P. (2010) Allosteric receptors: from electric organ to cognition. *Annu. Rev. Pharmacol. Toxicol.* **50**, 1–38.

Grigoryan, G., Moore, D.T., & DeGrado, W.F. (2011) Transmembrane communication: general principles and lessons from the structure and function of the M2 proton channel, K^+ channels, and integrin receptors. *Annu. Rev. Biochem.* **80**, 211–237.

Advanced, rewarding review.

Tombola, F., Pathak, M.M., & Isacoff, E.Y. (2006) How does voltage open an ion channel? *Annu. Rev. Cell Dev. Biol.* **22**, 23–52.

Calcium Ions in Signaling

Berridge, M.J. (2009) Inositol trisphosphate and calcium signalling mechanisms. *Biochim. Biophys. Acta Mol. Cell Res.* **1793**, 933–940.

Bunney, T.D. & Katan, M. (2011) PLC regulation: emerging pictures for molecular mechanisms. *Trends Biochem. Sci.* **36**, 88–96.

Intermediate review of phospholipase C in Ca^{2+}/IP_3 signaling.

Chazin, W.J. (2011) Relating form and function of EF-hand calcium binding proteins. *Acc. Chem. Res.* **44**, 171–179.

Chin, D. & Means, A.R. (2000) Calmodulin: a prototypical calcium receptor. *Trends Cell Biol.* **10**, 322–328.

Haiech, J., Heizmann, C., & Krebs, J. (eds). (2009) 10th European Symposium on Calcium. *Biochim. Biophys. Acta Mol. Cell Res.* **1793**, 931–1114.

All the papers in this volume deal with calcium signaling.

Parekh, A.B. (2011) Decoding cytosolic Ca^{2+} oscillations. *Trends Biochem. Sci.* **36**, 78–87.

Integrins

Harburger, D.S. & Calderwood, D.A. (2009) Integrin signaling at a glance. *J. Cell Sci.* **122**, 159–163.

Poster format review of integrin signaling.

Valdembri, D., Sandri, C., Santambrogio, M., & Serini, G. (2011) Regulation of integrins by conformation and traffic: it takes two to tango. *Mol. Biosyst.* **7**, 2539–2546.

Steroid Hormone Receptors and Action

Biggins, J.B. & Koh, J.T. (2007) Chemical biology of steroid and nuclear hormone receptors. *Curr. Opin. Chem. Biol.* **11**, 99–110.

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.

Signaling in Plants and Bacteria

Chen, Y.F., Etheridge, N., & Schaller, G.E. (2005) Ethylene signal transduction. *Ann. Botany* **95**, 901–915.

Intermediate-level review.

Clouse, S.D. (2011) Brassinosteroid signal transduction: from receptor kinase activation to transcriptional networks regulating plant development. *Plant Cell* **23**, 1219–1230.

Cutler, S.R., Rodriguez, P.L., Finkelstein, R.R., & Abrams, S. (2010) Abscisic acid: emergence of a core signaling network. *Annu. Rev. Plant Biol.* **61**, 651–679.

Dodd, A.N., Kudla, J., & Sanders, D. (2010) The language of calcium signaling. *Annu. Rev. Plant Biol.* **61**, 593–620.

Ferreira, F.J. & Kieber, J.J. (2005) Cytokinin signaling. *Curr. Opin. Plant Biol.* **8**, 518–525.

Paciorek, T. & Friml, J. (2006) Auxin signaling. *J. Cell Sci.* **119**, 1199–1202.

Pauwels, L. & Goossens, A. (2011) The JAZ proteins: a crucial interface in the jasmonate signaling cascade. *Plant Cell* **23**, 3089–3100.

Perry, J., Koteva, K., & Wright, G. (2011) Receptor domains of two-component signal transduction systems. *Mol. Biosyst.* **7**, 1388–1398.

Rodriguez, M.C.S., Petersen, M., & Mundy, J. (2010) Mitogen-activated protein kinase signaling in plants. *Annu. Rev. Plant Biol.* **61**, 621–649.

Stepanova, A.N. & Alonso, J.M. (2009) Ethylene signaling and response: where different regulatory modules meet. *Curr. Opin. Plant Biol.* **12**, 548–555.

Wang, Z. (2010) From receptors to responses (in plants). *Curr. Opin. Plant Biol.* **13**, 485–488.

Editorial introduction to a series of papers on plant signaling, all published in this issue.

Vision, Olfaction, and Gustation

Kaupp, U.B. (2010) Olfactory signalling in vertebrates and insects: differences and commonalities. *Nat. Rev. Neurosci.* **11**, 188–200.

Smith, S.O. (2010) Structure and activation of the visual pigment rhodopsin. *Annu. Rev. Biophys.* **39**, 309–328.
Advanced review.

Yarmolinsky, D.A., Zuker, C.S., & Ryba, N.J.P. (2009) Common sense about taste: from mammals to insects. *Cell* **139**, 234–244.

Cell Cycle and Cancer

Bublil, E.M. & Yarden, Y. (2007) The EGF receptor family: spearheading a merger of signaling and therapeutics. *Curr. Opin. Cell Biol.* **19**, 124–134.

Clarke, P.R. & Allan, L.A. (2009) Cell-cycle control in the face of damage—a matter of life or death. *Trends Cell Biol.* **19**, 89–98.
Intermediate review of roles of ATP, p53, and other regulatory proteins in the cell cycle.

Dorsam, R.T. & Gutkind, J.S. (2007) G-protein-coupled receptors and cancer. *Nat. Rev. Cancer* **7**, 79–94.

Levine, A.J. & Oren, M. (2009) The first 30 years of p53: growing ever more complex. *Nat. Rev. Cancer* **9**, 749–758.
Historical perspective on the studies of p53 and cancer.

Ma, H.T. & Poon, R.Y.C. (2011) How protein kinases co-ordinate mitosis in animal cells. *Biochem. J.* **435**, 17–31.

Malumbres, M. (2011) Physiological relevance of cell cycle kinases *Physiol. Rev.* **91**, 973–1007.
Advanced review of roles of CDKs and other protein kinases in controlling the cell cycle.

Markowitz, S.D. & Bertagnolli, M.M. (2009) Molecular basis of colorectal cancer. *N. Engl. J. Med.* **361**, 2449–2460.

Novak, B., Kapuy, O., Domingo-Sananes, M.R., & Tyson, J.J. (2010) Regulated protein kinases and phosphatases in cell cycle decisions. *Curr. Opin. Cell Biol.* **22**, 801–808.

Pylayeva-Gupta, Y., Grabocka, E., & Bar-Sagi, D. (2011) RAS oncogenes: weaving a tumorigenic web. *Nat. Rev. Cancer* **11**, 761–774.

Apoptosis

Wylie, A.H. (2010) "Where, O death, is thy sting?" A brief review of apoptosis biology. *Mol. Neurobiol.* **42**, 4–9.

Problems

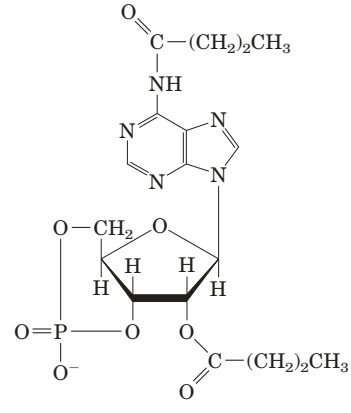
1. Hormone Experiments in Cell-Free Systems In the 1950s, Earl W. Sutherland, Jr., and his colleagues carried out pioneering experiments to elucidate the mechanism of action of epinephrine and glucagon. Given what you have learned in this chapter about hormone action, interpret each of the experiments described below. Identify substance X and indicate the significance of the results.

(a) Addition of epinephrine to a homogenate of normal liver resulted in an increase in the activity of glycogen phosphorylase. However, when the homogenate was first centrifuged at a high speed and epinephrine or glucagon was added to the clear supernatant fraction that contains phosphorylase, no increase in the phosphorylase activity occurred.

(b) When the particulate fraction from the centrifugation in (a) was treated with epinephrine, substance X was produced. The substance was isolated and purified. Unlike epinephrine, substance X activated glycogen phosphorylase when added to the clear supernatant fraction of the centrifuged homogenate.

(c) Substance X was heat stable; that is, heat treatment did not affect its capacity to activate phosphorylase. (Hint: Would this be the case if substance X were a protein?) Substance X was nearly identical to a compound obtained when pure ATP was treated with barium hydroxide. (Fig. 8–6 will be helpful.)

2. Effect of Dibutyryl cAMP versus cAMP on Intact Cells The physiological effects of epinephrine should in principle be mimicked by addition of cAMP to the target cells. In practice, addition of cAMP to intact target cells elicits only a minimal physiological response. Why? When the structurally related derivative dibutyryl cAMP (shown below) is added to intact cells, the expected physiological response is readily apparent. Explain the basis for the difference in cellular response to these two substances. Dibutyryl cAMP is widely used in studies of cAMP function.



Dibutyryl cAMP
(N^6,$O^{2'}$-Dibutyryl adenosine 3′,5′-cyclic monophosphate)

3. Effect of Cholera Toxin on Adenylyl Cyclase The gram-negative bacterium *Vibrio cholerae* produces a protein, cholera toxin (M_r 90,000), that is responsible for the characteristic symptoms of cholera: extensive loss of body water and Na^+ through continuous, debilitating diarrhea. If body fluids and Na^+ are not replaced, severe dehydration results; untreated, the disease is often fatal. When the cholera toxin gains access to the human intestinal tract, it binds tightly to specific sites in the plasma membrane of the epithelial cells lining the small intestine, causing adenylyl cyclase to undergo prolonged activation (hours or days).

(a) What is the effect of cholera toxin on [cAMP] in the intestinal cells?

(b) Based on the information above, suggest how cAMP normally functions in intestinal epithelial cells.

(c) Suggest a possible treatment for cholera.

4. Mutations in PKA Explain how mutations in the R or C subunit of cAMP-dependent protein kinase (PKA) might lead to (a) a constantly active PKA or (b) a constantly inactive PKA.

5. Therapeutic Effects of Albuterol The respiratory symptoms of asthma result from constriction of the bronchi and bronchioles of the lungs, caused by contraction of the smooth muscle of their walls. This constriction can be reversed by raising [cAMP] in the smooth muscle. Explain the therapeutic effects of albuterol, a β-adrenergic agonist taken (by inhalation) for asthma. Would you expect this drug to have

any side effects? How might one design a better drug that does not have these effects?

6. Termination of Hormonal Signals Signals carried by hormones must eventually be terminated. Describe several different mechanisms for signal termination.

7. Using FRET to Explore Protein-Protein Interactions in Vivo Figure 12–8 shows the interaction between β-arrestin and the β-adrenergic receptor. How would you use FRET (see Box 12–3) to demonstrate this interaction in living cells? Which proteins would you fuse? Which wavelengths would you use to illuminate the cells, and which would you monitor? What would you expect to observe if the interaction occurred? If it did not occur? How might you explain the failure of this approach to demonstrate this interaction?

8. EGTA Injection EGTA (ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid) is a chelating agent with high affinity and specificity for Ca^{2+}. By microinjecting a cell with an appropriate Ca^{2+}-EGTA solution, an experimenter can prevent cytosolic [Ca^{2+}] from rising above 10^{-7} M. How would EGTA microinjection affect a cell's response to vasopressin (see Table 12–4)? To glucagon?

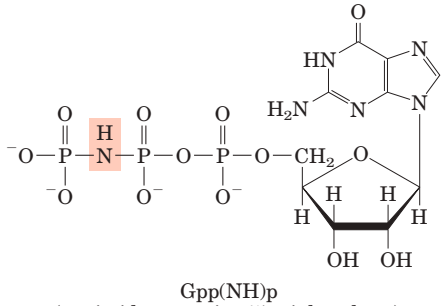
9. Amplification of Hormonal Signals Describe all the sources of amplification in the insulin receptor system.

10. Mutations in *ras* How would a mutation in *ras* that leads to formation of a Ras protein with no GTPase activity affect a cell's response to insulin?

11. Differences among G Proteins Compare the G proteins G_s, which acts in transducing the signal from β-adrenergic receptors, and Ras. What properties do they share? How do they differ? What is the functional difference between G_s and G_i?

12. Mechanisms for Regulating Protein Kinases Identify eight general types of protein kinases found in eukaryotic cells, and explain what factor is *directly* responsible for activating each type.

13. Nonhydrolyzable GTP Analogs Many enzymes can hydrolyze GTP between the β and γ phosphates. The GTP analog β,γ-imidoguanosine 5′-triphosphate (Gpp(NH)p), shown below, cannot be hydrolyzed between the β and γ phosphates.



Gpp(NH)p
(β,γ-imidoguanosine 5′-triphosphate)

Predict the effect of microinjection of Gpp(NH)p into a myocyte on the cell's response to β-adrenergic stimulation.

14. Use of Toxin Binding to Purify a Channel Protein α-Bungarotoxin is a powerful neurotoxin found in the venom of a poisonous snake (*Bungarus multicinctus*). It binds with high specificity to the nicotinic acetylcholine receptor (AChR) protein and prevents the ion channel from opening. This interaction was used to purify AChR from the electric organ of torpedo fish.

(a) Outline a strategy for using α-bungarotoxin covalently bound to chromatography beads to purify the AChR protein. (Hint: See Fig. 3–17c.)

(b) Outline a strategy for the use of [^{125}I]α-bungarotoxin to purify the AChR protein.

15. Resting Membrane Potential A variety of unusual invertebrates, including giant clams, mussels, and polychaete worms, live on the fringes of deep-sea hydrothermal vents, where the temperature is 60 °C.

(a) The adductor muscle of a giant clam has a resting membrane potential of −95 mV. Given the intracellular and extracellular ionic compositions shown below, would you have predicted this membrane potential? Why or why not?

	Concentration (mM)	
Ion	**Intracellular**	**Extracellular**
Na^+	50	440
K^+	400	20
Cl^-	21	560
Ca^{2+}	0.4	10

(b) Assume that the adductor muscle membrane is permeable to only one of the ions listed above. Which ion could determine the V_m?

16. Membrane Potentials in Frog Eggs Fertilization of a frog oocyte by a sperm cell triggers ionic changes similar to those observed in neurons (during movement of the action potential) and initiates the events that result in cell division and development of the embryo. Oocytes can be stimulated to divide without fertilization, by suspending them in 80 mM KCl (normal pond water contains 9 mM KCl).

(a) Calculate how much the change in extracellular [KCl] changes the resting membrane potential of the oocyte. (Hint: Assume the oocyte contains 120 mM K^+ and is permeable *only* to K^+.) Assume a temperature of 20 °C.

(b) When the experiment is repeated in Ca^{2+}-free water, elevated [KCl] has no effect. What does this suggest about the mechanism of the KCl effect?

17. Excitation Triggered by Hyperpolarization In most neurons, membrane *depolarization* leads to the opening of voltage-dependent ion channels, generation of an action potential, and ultimately an influx of Ca^{2+}, which causes release of neurotransmitter at the axon terminus. Devise a cellular strategy by which *hyperpolarization* in rod cells could produce excitation of the visual pathway and passage of visual signals to the brain. (Hint: The neuronal signaling pathway in higher organisms consists of a *series* of neurons that relay information to the brain (see Fig. 12–36). The signal released by one neuron can be either excitatory or inhibitory to the following, postsynaptic neuron.)

18. Genetic "Channelopathies" There are many genetic diseases that result from defects in ion channels. For each of the following, explain how the molecular defect might lead to the symptoms described.

(a) A loss-of-function mutation in the gene encoding the α subunit of the cGMP-gated cation channel of retinal cone cells leads to a complete inability to distinguish colors.

(b) Loss-of-function alleles of the gene encoding the α subunit of the ATP-gated K^+ channel shown in Figure 23–28 lead to a condition known as congenital hyperinsulinism—persistently high levels of insulin in the blood.

(c) Mutations affecting the β subunit of the ATP-gated K^+ channel that prevent ATP binding lead to neonatal diabetes—persistently low levels of insulin in the blood in newborn babies.

19. Visual Desensitization Oguchi disease is an inherited form of night blindness. Affected individuals are slow to recover vision after a flash of bright light against a dark background, such as the headlights of a car on the freeway. Suggest what the molecular defect(s) might be in Oguchi disease. Explain in molecular terms how this defect would account for night blindness.

20. Effect of a Permeant cGMP Analog on Rod Cells An analog of cGMP, 8-Br-cGMP, will permeate cellular membranes, is only slowly degraded by a rod cell's PDE activity, and is as effective as cGMP in opening the gated channel in the cell's outer segment. If you suspended rod cells in a buffer containing a relatively high [8-Br-cGMP], then illuminated the cells while measuring their membrane potential, what would you observe?

21. Hot and Cool Taste Sensations The sensations of heat and cold are transduced by a group of temperature-gated cation channels. For example, TRPV1, TRPV3, and TRPM8 are usually closed, but open under the following conditions: TRPV1 at $\geq$43°C; TRPV3 at $\geq$33°C; and TRPM8 at $<$25°C. These channels are expressed in sensory neurons known to be responsible for temperature sensation.

(a) Propose a reasonable model to explain how exposing a sensory neuron containing TRPV1 to high temperature leads to a sensation of heat.

(b) Capsaicin, one of the active ingredients in "hot" peppers, is an agonist of TRPV1. Capsaicin shows 50% activation of the TRPV1 response at a concentration (i.e., it has an EC_{50}) of 32 nM. Explain why even a very few drops of hot pepper sauce can taste very "hot" without actually burning you.

(c) Menthol, one of the active ingredients in mint, is an agonist of TRPM8 (EC_{50} = 30 μM) and TRPV3 (EC_{50} = 20 mM). What sensation would you expect from contact with low levels of menthol? With high levels?

22. Oncogenes, Tumor-Suppressor Genes, and Tumors For each of the following situations, provide a plausible explanation for how it could lead to unrestricted cell division.

(a) Colon cancer cells often contain mutations in the gene encoding the prostaglandin E_2 receptor. PGE_2 is a growth factor required for the division of cells in the gastrointestinal tract.

(b) Kaposi sarcoma, a common tumor in people with untreated AIDS, is caused by a virus carrying a gene for a protein similar to the chemokine receptors CXCR1 and CXCR2. Chemokines are cell-specific growth factors.

(c) Adenovirus, a tumor virus, carries a gene for the protein E1A, which binds to the retinoblastoma protein, pRb. (Hint: See Fig. 12–49.)

(d) An important feature of many oncogenes and tumor suppressor genes is their cell-type specificity. For example, mutations in the PGE_2 receptor are not typically found in lung tumors. Explain this observation. (Note that PGE_2 acts through a GPCR in the plasma membrane.)

23. Mutations in Tumor-Suppressor Genes and Oncogenes Explain why mutations in tumor-suppressor genes are recessive (both copies of the gene must be defective for the regulation of cell division to be defective) whereas mutations in oncogenes are dominant.

24. Retinoblastoma in Children Explain why some children with retinoblastoma develop multiple tumors of the retina in both eyes, whereas others have a single tumor in only one eye.

25. Specificity of a Signal for a Single Cell Type Discuss the validity of the following proposition. A signaling molecule (hormone, growth factor, or neurotransmitter) elicits identical responses in different types of target cells if they contain identical receptors.

Data Analysis Problem

26. Exploring Taste Sensation in Mice Figure 12–42 shows the signal-transduction pathway for sweet taste in mammals. Pleasing tastes are an evolutionary adaptation to encourage animals to consume nutritious foods. Zhao and coauthors (2003) examined the two major pleasurable taste sensations: sweet and umami. Umami is a "distinct savory taste" triggered by amino acids, especially aspartate and glutamate, and probably encourages animals to consume protein-rich foods. Monosodium glutamate (MSG) is a flavor enhancer that exploits this sensitivity.

At the time the article was published, specific taste receptor proteins (labeled SR in Fig. 12–42) for sweet and umami had been tentatively characterized. Three such proteins were known—T1R1, T1R2, and T1R3—which function as heterodimeric receptor complexes: T1R1-T1R3 was tentatively identified as the umami receptor, and T1R2-T1R3 as the sweet receptor. It was not clear how taste sensation was encoded and sent to the brain, and two possible models had been suggested. In the cell-based model, individual taste-sensing cells express only one kind of receptor; that is, there are "sweet cells," "bitter cells," "umami cells," and so on, and each type of cell sends its information to the brain via a different nerve. The brain "knows" which taste is detected by the identity of the nerve fiber that transmits the message. In the receptor-based model, individual taste-sensing cells have

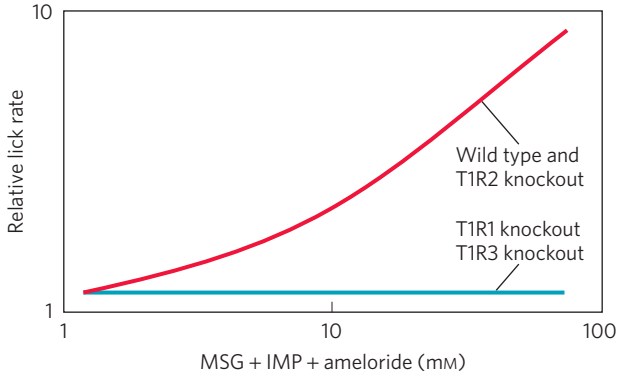
several kinds of receptors and send different messages along the same nerve fiber to the brain, the message depending on which receptor is activated. Also unclear at the time was whether there was any interaction between the different taste sensations, or whether parts of one taste-sensing system were required for other taste sensations.

(a) Previous work had shown that different taste receptor proteins are expressed in nonoverlapping sets of taste receptor cells. Which model does this support? Explain your reasoning.

Zhao and colleagues constructed a set of "knockout mice"—mice homozygous for loss-of-function alleles for one of the three receptor proteins, T1R1, T1R2, or T1R3—and double-knockout mice with nonfunctioning T1R2 and T1R3. The researchers measured the taste perception of these mice by measuring their "lick rate" of solutions containing different taste molecules. Mice will lick the spout of a feeding bottle with a pleasant-tasting solution more often than one with an unpleasant-tasting solution. The researchers measured relative lick rates: how often the mice licked a sample solution compared with water. A relative lick rate of 1 indicated no preference; <1, an aversion; and >1, a preference.

(b) All four types of knockout strains had the same responses to salt and bitter tastes as did wild-type mice. Which of the above issues did this experiment address? What do you conclude from these results?

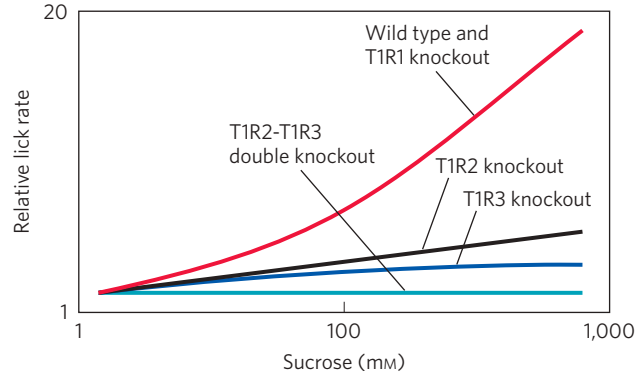
The researchers then studied umami taste reception by measuring the relative lick rates of the different mouse strains with different quantities of MSG in the feeding solution. Note that the solutions also contained inosine monophosphate (IMP), a strong potentiator of umami taste reception (and a common ingredient in ramen soups, along with MSG), and ameloride, which suppresses the pleasant salty taste imparted by the sodium of MSG. The results are shown in the graph.



(c) Are these data consistent with the umami taste receptor consisting of a heterodimer of T1R1 and T1R3? Why or why not?

(d) Which model(s) of taste encoding does this result support? Explain your reasoning.

Zhao and coworkers then performed a series of similar experiments using sucrose as a sweet taste. These results are shown below.



(e) Are these data consistent with the sweet taste receptor consisting of a heterodimer of T1R2 and T1R3? Why or why not?

(f) There were some unexpected responses at very high sucrose concentrations. How do these complicate the idea of a heterodimeric system as presented above?

In addition to sugars, humans also taste other compounds (e.g., the peptides monellin and aspartame) as sweet; mice do not taste these as sweet. Zhao and coworkers inserted into TIR2-knockout mice a copy of the human T1R2 gene under the control of the mouse T1R2 promoter. These modified mice now tasted monellin and saccharin as sweet. The researchers then went further, adding to T1R1-knockout mice the RASSL protein—a G protein–linked receptor for the synthetic opiate spiradoline; the RASSL gene was under the control of a promoter that could be induced by feeding the mice tetracycline. These mice did not prefer spiradoline in the absence of tetracycline; in the presence of tetracycline, they showed a strong preference for nanomolar concentrations of spiradoline.

(g) How do these results strengthen Zhao and coauthors' conclusions about the mechanism of taste sensation?

Reference

Zhao, G.Q., Zhang, Y., Hoon, M.A., Chandrashekar, J., Erlenbach, I., Ryba, N.J.P., & Zuker, C. (2003) The receptors for mammalian sweet and umami taste. *Cell* **115**, 255–266.