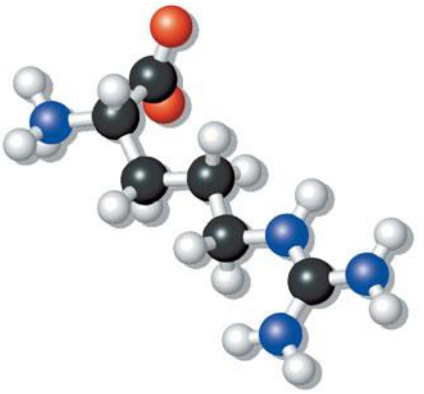


Amino Acids and the Primary Structures of Proteins

The relationship between structure and function is a fundamental part of biochemistry. In spite of its importance, we sometimes forget to mention structure-function relationships, thinking that the concept is obvious from the examples. In this book we will try and remind you from time to time how the study of structure leads to a better understanding of function. This is especially important when studying proteins.

In this chapter and the next one we will cover the basic rules of protein structure. In Chapters 5 and 6, we will learn how enzymes work and how their structure contributes to the mechanisms of enzyme action.

Before beginning, let's review the various kinds of proteins. The following list, although not exhaustive, covers most of the important biological functions of proteins:

1. Many proteins function as enzymes, the biochemical catalysts. Enzymes catalyze nearly all reactions that occur in living organisms.
2. Some proteins bind other molecules for storage and transport. For example, hemoglobin binds and transports O_2 and CO_2 in red blood cells and other proteins bind fatty acids and lipids.
3. Several types of proteins serve as pores and channels in membranes, allowing for the passage of small, charged molecules.
4. Some proteins, such as tubulin, actin, and collagen, provide support and shape to cells and hence to tissues and organisms.
5. Assemblies of proteins can do mechanical work, such as the movement of flagella, the separation of chromosomes at mitosis, and the contraction of muscles.
6. Many proteins play a role in information flow in the cell. Some are involved in translation whereas others play a role in regulating gene expression by binding to nucleic acids.
7. Some proteins are hormones, which regulate biochemical activities in target cells or tissues; other proteins serve as receptors for hormones.

"Amino acids are literally raining down from the sky, and if that's not a big deal then I don't know what is."

Max Bernstein,
SETI Institute

KEY CONCEPT

The functions of biochemical molecules can only be understood by knowing their structures.

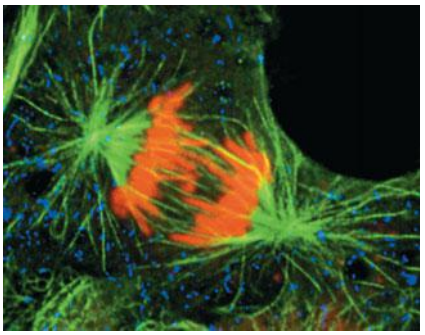
Top: L-Arginine, one of the 20 common amino acids.

KEY CONCEPT

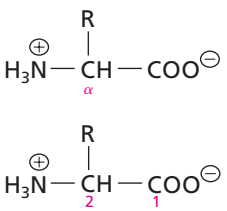
There are many different kinds of proteins with many different roles in metabolism and cell structure.

8. Proteins on the cell surface can act as receptors for various ligands and as modifiers of cell-cell interactions.
9. Some proteins have highly specialized functions. For example, antibodies defend vertebrates against bacterial and viral infections, and toxins, produced by bacteria, can kill larger organisms.

We begin our study of proteins by exploring the structures and chemical properties of their constituent amino acids. In this chapter we will also discuss the purification, analysis, and sequencing of polypeptides.



▲ **Spindle fibers.** Spindle fibers (green) help separate chromosomes at mitosis. The fibers are microtubules formed from the structural protein tubulin.



▲ **Numbering conventions for amino acids.** In traditional names, the carbon atoms adjacent to the carboxyl group are identified by the Greek letters α, β, γ, etc. In the official IUPAC/IUBMB chemical names or systematic names, the carbon atom in the carboxyl group is number 1 and the adjacent carbons are numbered sequentially. Thus, the α-carbon atom in traditional names is the carbon 2 atom in systematic names.

The IUPAC-IUBMB website for *Nomenclature and Symbolism for Amino Acids and Peptides* is: www.chem.qmul.ac.uk/iupac/AminoAcid/.

3.1 General Structure of Amino Acids

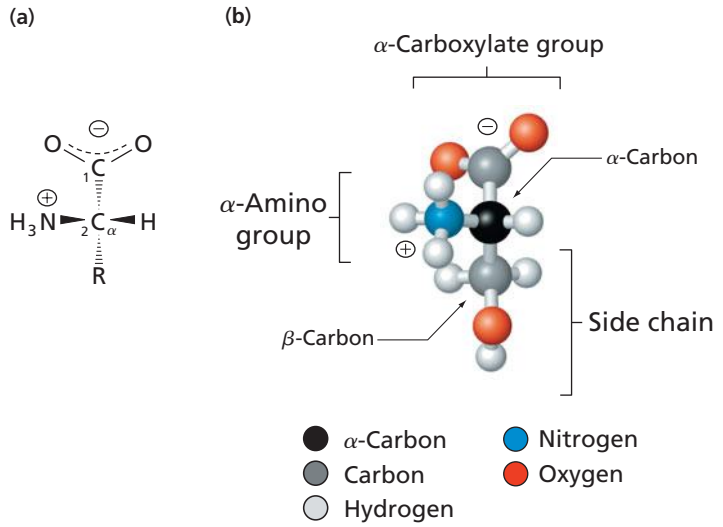
All organisms use the same 20 amino acids as building blocks for the assembly of protein molecules. These 20 amino acids are called the *common*, or *standard*, amino acids. Despite the limited number of amino acids, an enormous variety of different polypeptides can be produced by connecting the 20 common amino acids in various combinations.

Amino acids are called amino acids because they are amino derivatives of carboxylic acids. In the 20 common amino acids the amino group and the carboxyl group are bonded to the same carbon atom: the α-carbon atom. Thus, all of the standard amino acids found in proteins are α-amino acids. Two other substituents are bound to the α-carbon—a hydrogen atom and a side chain (R) that is distinctive for each amino acid. In the chemical names of amino acids, carbon atoms are identified by numbers, beginning with the carbon atom of the carboxyl group. [The correct chemical name, or systematic name, follows rules established by the International Union of Pure and Applied Chemistry (IUPAC) and the International Union of Biochemistry and Molecular Biology (IUBMB).] If the R group is —CH_3 then the *systematic name* for that amino acid would be 2-aminopropanoic acid. (Propanoic acid is CH_3—CH_2—COOH.) The *trivial name* for CH_3—$CH(NH_2)$—COOH is *alanine*. The old nomenclature uses Greek letters to identify the α-carbon atom and the carbon atoms of the side chain. This nomenclature identifies the carbon atom relative to the carboxyl group so the carbon atom of the carboxyl group is not specified, unlike in the systematic nomenclature, where this carbon atom is number 1 in the numbering system. Biochemists have traditionally used the old, alternate nomenclature.

Inside a cell, under normal physiological conditions, the amino group is protonated (—$NH_3^{\oplus}$) because the pK_a of this group is close to 9. The carboxyl group is ionized (—$COO^{\ominus}$) because the pK_a of that group is below 3, as we saw in Section 2.9. Thus, in the physiological pH range of 6.8 to 7.4, amino acids are **zwitterions**, or dipolar ions, even though their net charge may be zero. We will see in Section 3.4 that some side chains can also ionize. Biochemists always represent the structures of amino acids in the form that is biologically relevant which is why you will see the zwitterions in the following figures.

Figure 3.1a shows the general three-dimensional structure of an amino acid. Figure 3.1b shows a ball-and-stick model of a representative amino acid, serine, whose side chain is —CH_2OH. The first carbon atom that's directly bound to the carboxylate carbon is the α-carbon so the other carbon atoms of a side chain are sequentially labeled β, γ, δ, and ε, referring to carbons 3, 4, 5, and 6, respectively, in the newer convention. The systematic name for serine is 2-amino-3-hydroxypropanoic acid.

In 19 of the 20 common amino acids the α-carbon atom is **chiral**, or asymmetric, since it has four different groups bonded to it. The exception is glycine, whose R group is simply a hydrogen atom. The molecule is not chiral because the α-carbon atom is bonded to two identical hydrogen atoms. The 19 chiral amino acids can therefore exist as stereoisomers. **Stereoisomers** are compounds that have the same molecular formula but differ in the arrangement, or **configuration**, of their atoms in space. The two stereoisomers are distinct molecules that can't be easily converted from one form to the other since a change in configuration requires the breaking of one or more bonds. Amino acid stereoisomers are nonsuperimposable mirror images called **enantiomers**. Two of the 19 chiral amino acids, isoleucine and threonine, have two chiral carbon atoms each. Isoleucine and threonine can each form four different stereoisomers.



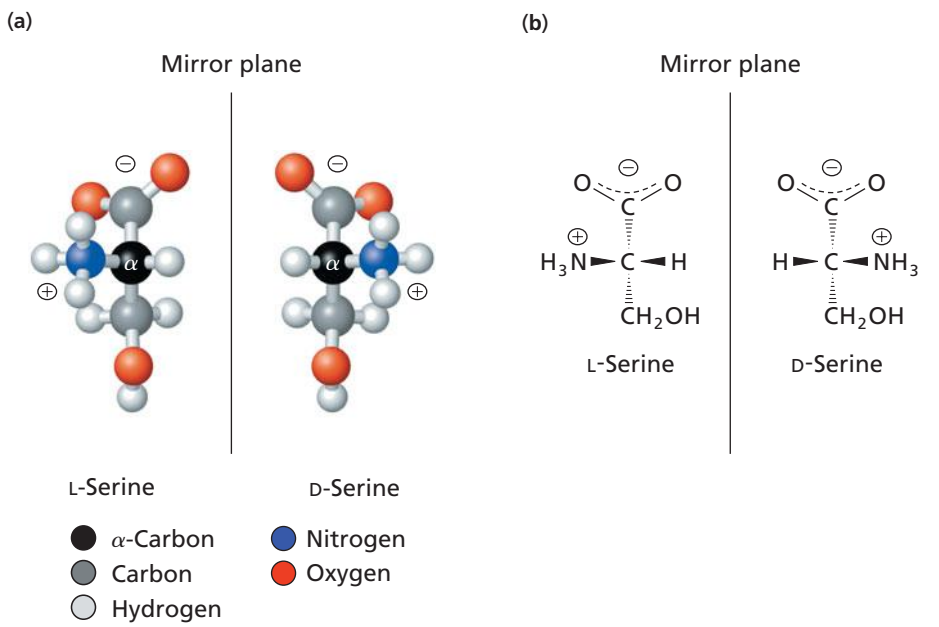
◀ **Figure 3.1**
Two representations of an L-amino acid at neutral pH. (a) General structure. An amino acid has a carboxylate group (whose carbon atom is designated C-1), an amino group, a hydrogen atom, and a side chain (or R group), all attached to C-2 (the α-carbon). Solid wedges indicate bonds above the plane of the paper; dashed wedges indicate bonds below the plane of the paper. The blunt ends of wedges are nearer the viewer than the pointed ends. **(b)** Ball-and-stick model of serine (whose R group is (—CH_2OH).

By convention, the mirror-image pairs of amino acids are designated D (for dextro, from the Latin *dexter*, "right") and L (for levo, from the Latin *laevus*, "left"). The configuration of the amino acid in Figure 3.1a is L and that of its mirror image is D. To assign the stereochemical designation, one draws the amino acid vertically with its α-carboxylate group at the top and its side chain at the bottom, both pointing away from the viewer. In this orientation, the α-amino group of the L isomer is on the left of the α-carbon, and that of the D isomer is on the right, as shown in Figure 3.2. (The four atoms attached to the α-carbon occupy the four corners of a tetrahedron much like the bonding of hydrogen atoms to oxygen in water, as shown in Figure 2.4.)

The 19 chiral amino acids used in the assembly of proteins are all of the L configuration, although a few D-amino acids occur in nature. By convention, amino acids are assumed to be in the L configuration unless specifically designated D. Often it is convenient to draw the structures of L-amino acids in a form that is stereochemically uncommitted, especially when a correct stereochemical representation is not critical to a given discussion.

The fact that all living organisms use the same standard amino acids in protein synthesis is evidence that all species on Earth are descended from a common ancestor. Like modern organisms, the last common ancestor (LCA) must have used L-amino

▲ **Meteorites and amino acids.** The Murchison meteorite fell in 1969 near Murchison, Australia. There are many similar carbonaceous meteorites and many of them contain spontaneously formed amino acids, including some of the common amino acids found in proteins. These amino acids are found in the meteorites as almost equal mixtures of the L and D configurations.



See Section 8.1 for a more complete description of the convention for displaying stereoisomers (Fischer projection).

◀ **Figure 3.2**
Mirror-image pairs of amino acids. (a) Ball-and-stick models of L-serine and D-serine. Note that the two molecules are not identical; they cannot be superimposed. **(b)** L-Serine and D-serine. The common amino acids all have the L configuration.

acids and not D-amino acids. Mixtures of L- and D-amino acids are formed under conditions that mimic those present when life first arose on Earth 4 billion years ago and both enantiomers are found in meteorites and in the vicinity of stars. It is not known how or why primitive life forms selected L-amino acids from the presumed mixture of the enantiomers present when life first arose. It's likely that the first proteins were composed of a small number of simple amino acids and selection of L-amino acids over D-amino acids was a chance event. Modern living organisms do not select L-amino acids from a mixture because only the L-amino acids are synthesized in sufficient quantities. Thus, the predominance of L-amino acids in modern species is due to the evolution of metabolic pathways that produce L-amino acids and not D-amino acids (Chapter 17).

3.2 Structures of the 20 Common Amino Acids

The structures of the 20 amino acids commonly found in proteins are shown in the following figures as Fischer projections. In Fischer projections, horizontal bonds at a chiral center extend toward the viewer and vertical bonds extend away (as in Figures 3.1 and 3.2). Examination of the structures reveals considerable variation in the side chains of the 20 amino acids. Some side chains are nonpolar and thus hydrophobic whereas others are polar or ionized at neutral pH and are therefore hydrophilic. The properties of the side chains greatly influence the overall three-dimensional shape, or conformation, of a protein. For example, most of the hydrophobic side chains of a water-soluble protein fold into the interior giving the protein a compact, globular shape.

Some nonstandard amino acids are described in Section 3.3.

Both the three-letter and one-letter abbreviations for each amino acid are shown in the figures. The three-letter abbreviations are self-evident but the one-letter abbreviations are less obvious. Several amino acids begin with the same letter so other letters of the alphabet have to be used in order to provide a unique label; for example, threonine = T, tyrosine = Y, and tryptophan = W. These labels have to be memorized.

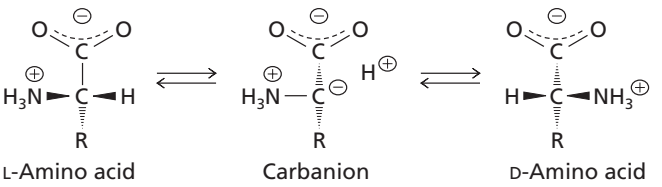
BOX 3.1 FOSSIL DATING BY AMINO ACID RACEMIZATION

Amino acids can spontaneously convert from the D configuration to the L configuration and *vice versa*. This is a chemical reaction that usually proceeds through a carbanion intermediate.

The racemization reaction is normally very slow but it can be sped up at high temperatures. For example, the half-life for conversion of L-aspartate to D-aspartate is about 30 days at 100°C. The half-life of this reaction at 37°C is about 350 years and at 18°C it's about 50,000 years.

The amino acid composition of mammalian tooth enamel can be used to determine the age of a fossil if the average temperature of the environment is known or can be estimated. When the amino acids are first synthesized they are exclusively of the L configuration. Over time, the amount of the D enantiomer increases and the D/L ratio can be measured very precisely.

Fossil dating by measuring amino acid racemization has been superceded by more reliable methods but it's an interesting example of a slow chemical reaction. Some organisms contain specific racemases that catalyze the interconversion of an L-amino acid and a D-amino acid; for example, bacteria have alanine racemase for converting L-alanine to D-alanine (see Section 8.7B). These enzymes catalyze thousands of reactions per second.



▲ The Badegoule Jaw from a stone age juvenile. *Homo sapiens* (Natural History Museum, Lyon, France)

It is important to learn the structures of the standard amino acids because we refer to them frequently in the chapters on protein structure, enzymes, and protein synthesis. In the following sections we have grouped the standard amino acids by their general properties and the chemical structures of their side chains. The side chains fall into the following chemical classes: aliphatic, aromatic, sulfur-containing, alcohols, positively charged, negatively charged, and amides. Of the 20 amino acids five are further classified as highly hydrophobic (blue) and seven are classified as highly hydrophilic (red). Understanding the classification of the R groups will simplify memorizing the structures and names.

A. Aliphatic R Groups

Glycine (Gly, G) is the smallest amino acid. Since its R group is simply a hydrogen atom, the α-carbon of glycine is not chiral. The two hydrogen atoms of the α-carbon of glycine impart little hydrophobic character to the molecule. We will see that glycine plays a unique role in the structure of many proteins because its side chain is small enough to fit into niches that cannot accommodate any other amino acid.

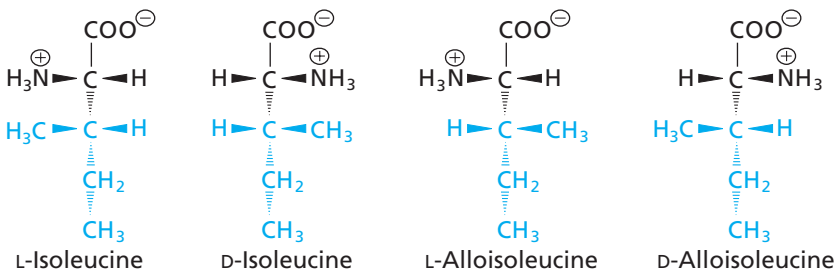
Four amino acids—alanine (Ala, A), valine (Val, V), leucine (Leu, L), and the structural isomer of leucine, isoleucine (Ile, I)—have saturated aliphatic side chains. The side chain of alanine is a methyl group whereas valine has a three-carbon branched side chain and leucine and isoleucine each contain a four-carbon branched side chain. Both the α- and β-carbon atoms of isoleucine are asymmetric. Because isoleucine has two chiral centers, it has four possible stereoisomers. The stereoisomer used in proteins is called L-isoleucine and the amino acid that differs at the β-carbon is called L-alloisoleucine (Figure 3.3). The other two stereoisomers are D-isoleucine and D-alloisoleucine.

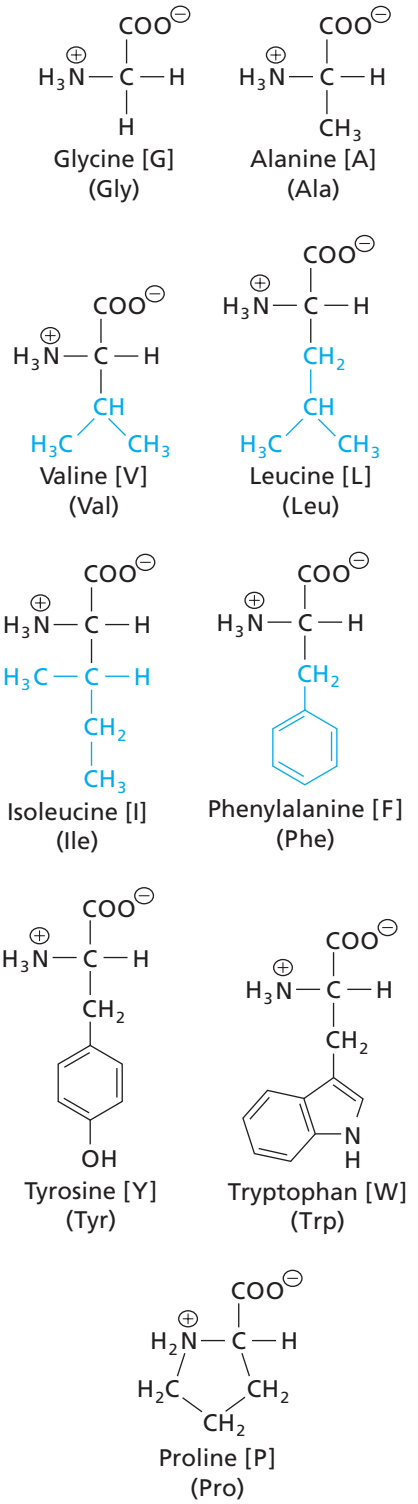
Alanine, valine, leucine, and isoleucine play an important role in establishing and maintaining the three-dimensional structures of proteins because of their tendency to cluster away from water. Valine, leucine, and isoleucine are known collectively as the branched chain amino acids because their side chains of carbon atoms contain branches. All three amino acids are highly hydrophobic and they share biosynthesis and degradation pathways (Chapter 17).

Proline (Pro, P) differs from the other 19 amino acids because its three-carbon side chain is bonded to the nitrogen of its α-amino group as well as to the α-carbon creating a cyclic molecule. As a result, proline contains a secondary rather than a primary amino group. The heterocyclic pyrrolidine ring of proline restricts the geometry of polypeptides sometimes introducing abrupt changes in the direction of the peptide chain. The cyclic structure of proline makes it much less hydrophobic than valine, leucine, and isoleucine.

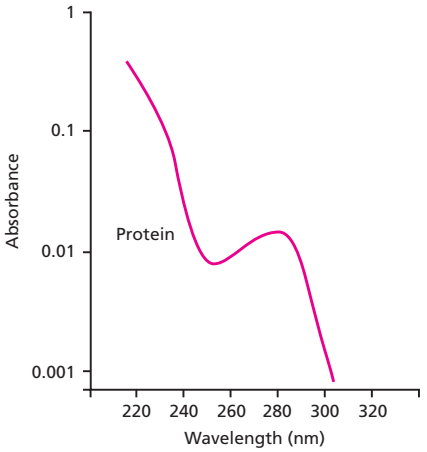
B. Aromatic R Groups

Phenylalanine (Phe, F), tyrosine (Tyr, Y), and tryptophan (Trp, W) have side chains with aromatic groups. Phenylalanine has a hydrophobic benzyl side chain. Tyrosine is structurally similar to phenylalanine except that the *para* hydrogen of phenylalanine is replaced in tyrosine by a hydroxyl group (—OH) making tyrosine a phenol. The hydroxyl group of tyrosine is ionizable but retains its hydrogen under normal physiological conditions. The side chain of tryptophan contains a bicyclic indole group. Tyrosine and

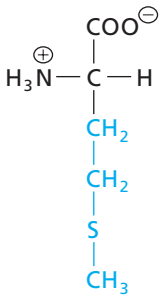




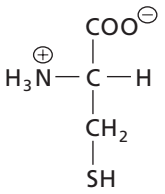
◀ **Figure 3.3**
Stereoisomers of isoleucine. Isoleucine and threonine are the only two common amino acids with more than one chiral center. The other DL pair of isoleucine isomers is called alloleucine. Note that in L-isoleucine the $—NH_3^{\oplus}$ and $—CH_3$ groups are both on the left in this projection, while in D-isoleucine they are both on the right, so that D-isoleucine and L-isoleucine are mirror images.



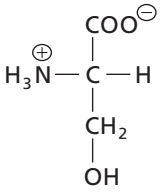
▲ **UV absorbance of proteins.** The peak of absorbance of most proteins peaks at 280 nm. Most of the absorbance is due to the presence of tryptophan and tyrosine residues in the protein.



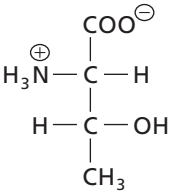
Methionine [M]
(Met)



Cysteine [C]
(Cys)



Serine [S]
(Ser)



Threonine [T]
(Thr)

▲ **A sulfur bridge.** Natural stone bridge, Puente del Inca, in Mendoza, Argentina. Over the years the bridge has been covered with sulfur deposits.

tryptophan are not as hydrophobic as phenylalanine because their side chains include polar groups (Table 3.1, page 62).

All three aromatic amino acids absorb ultraviolet (UV) light because, unlike the saturated aliphatic amino acids, the aromatic amino acids contain delocalized π-electrons. At neutral pH both tryptophan and tyrosine absorb light at a wavelength of 280 nm whereas phenylalanine is almost transparent at 280 nm and absorbs light weakly at 260 nm. Since most proteins contain tryptophan and tyrosine they will absorb light at 280 nm. Absorbance at 280 nm is routinely used to estimate the concentration of proteins in solutions.

C. R Groups Containing Sulfur

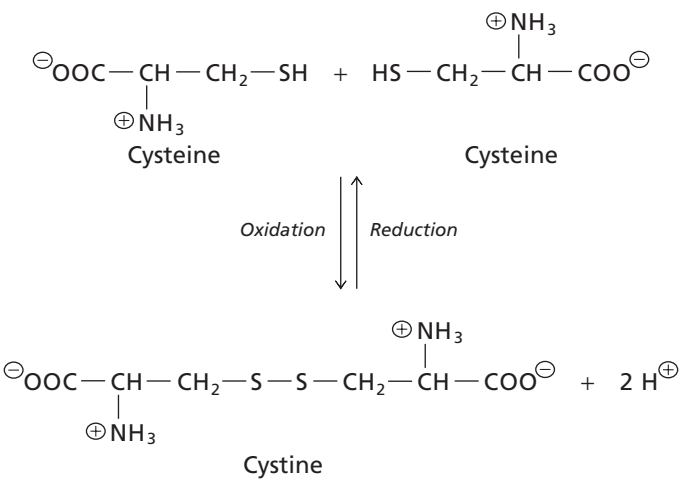
Methionine (Met, M) and cysteine (Cys, C) are the two amino acids whose side chains contain a sulfur atom. Methionine contains a nonpolar methyl thioether group in its side chain and this makes it one of the more hydrophobic amino acids. Methionine plays a special role in protein synthesis because it is almost always the first amino acid in a growing polypeptide chain. The structure of cysteine resembles that of alanine with a hydrogen atom replaced by a sulfhydryl group (—SH).

Although the side chain of cysteine is somewhat hydrophobic, it is also highly reactive. Because the sulfur atom is polarizable the sulfhydryl group of cysteine can form weak hydrogen bonds with oxygen and nitrogen. Moreover, the sulfhydryl group of cysteine residues in proteins can be a weak acid which allows it to lose its proton to become a negatively charged thiolate ion. (The pK_a of the sulfhydryl group of the free amino acid is 8.3 but this can range from 5-10 in proteins.)

A compound called cystine can be isolated when some proteins are hydrolyzed. Cystine is formed from two oxidized cysteine molecules linked by a disulfide bond (Figure 3.4). Oxidation of the sulfhydryl groups of cysteine molecules proceeds most readily at slightly alkaline pH values because the sulfhydryl groups are ionized at high pH. The two cysteine side chains must be adjacent in three-dimensional space in order to form a disulfide bond but they don't have to be close together in the amino acid sequence of the polypeptide chain. They may even be found in different polypeptide chains. Disulfide bonds, or disulfide bridges, may stabilize the three-dimensional structures of some proteins by covalently cross-linking cysteine residues in peptide chains. Most proteins do not contain disulfide bridges because conditions inside the cell do not favor oxidation; however, many secreted, or extracellular, proteins contain disulfide bridges.

D. Side Chains with Alcohol Groups

Serine (Ser, S) and threonine (Thr, T) have uncharged polar side chains containing β-hydroxyl groups. These alcohol groups give a hydrophilic character to the aliphatic



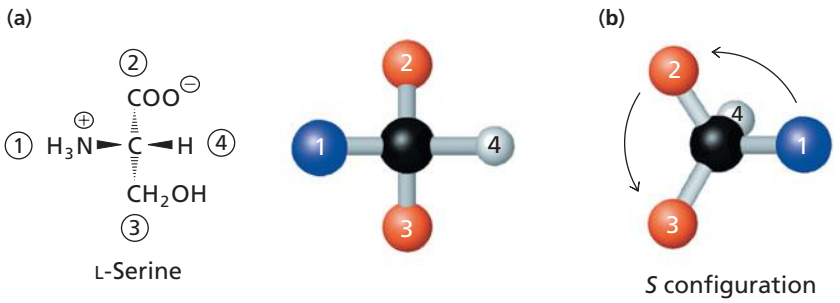
▲ **Figure 3.4**
Formation of cystine. When oxidation links the sulfhydryl groups of two cysteine molecules, the resulting compound is a disulfide called cystine.

BOX 3.2 AN ALTERNATIVE NOMENCLATURE

The *RS* system of configurational nomenclature is also sometimes used to describe the chiral centers of amino acids. The *RS* system is based on the assignment of a priority sequence to the four groups bound to a chiral carbon atom. Once assigned, the group priorities are used to establish the configuration of the molecule. Priorities are numbered 1 through 4 and are assigned to groups according to the following rules:

1. For atoms directly attached to the chiral carbon, the one with the lowest atomic mass is assigned the lowest priority (number 4).
2. If there are two identical atoms bound to the chiral carbon, the priority is decided by the atomic mass of the next atoms bound. For example, a $-CH_3$ group has a lower priority than a $-CH_2Br$ group because hydrogen has a lower atomic mass than bromine.
3. If an atom is bound by a double or triple bond, the atom is counted once for each formal bond. Thus, —CHO, with a double-bonded oxygen, has a higher priority than $-CH_2OH$. The order of priority for the most common groups, from lowest to highest, is —H, $-CH_3$, $-C_6H_5$, $-CH_2OH$, —CHO, —COOH, —COOR, $-NH_2$, —NHR, —OH, —OR, and —SH.

With these rules in mind, imagine the molecule as the steering wheel of a car, with the group of lowest priority (numbered 4) pointing away from you (like the steering column) and the other three groups arrayed around the rim of the steering wheel. Trace the rim of the wheel, moving from the group of highest priority to the group of lowest priority (1, 2, 3). If the movement is clockwise, the configuration is *R* (from the Latin *rectus*, "right-handed"). If the movement is counterclockwise, the configuration is *S* (from the Latin, *sinister*, "left-handed"). The figure demonstrates the assignment of *S* configuration to L-serine by the *RS* system. L-Cysteine has the opposite configuration, *R*. The DL system is used more often in biochemistry because not all amino acids found in proteins have the same *RS* designation.



◀ Assignment of configuration by the *RS* system. **(a)** Each group attached to a chiral carbon is assigned a priority based on atomic mass, 4 being the lowest priority. **(b)** By orienting the molecule with the priority 4 group pointing away (behind the chiral carbon) and tracing the path from the highest priority group to the lowest, the absolute configuration can be established. If the sequence 1, 2, 3 is clockwise, the configuration is *R*. If the sequence 1, 2, 3 is counterclockwise, the configuration is *S*. L-Serine has the *S* configuration.

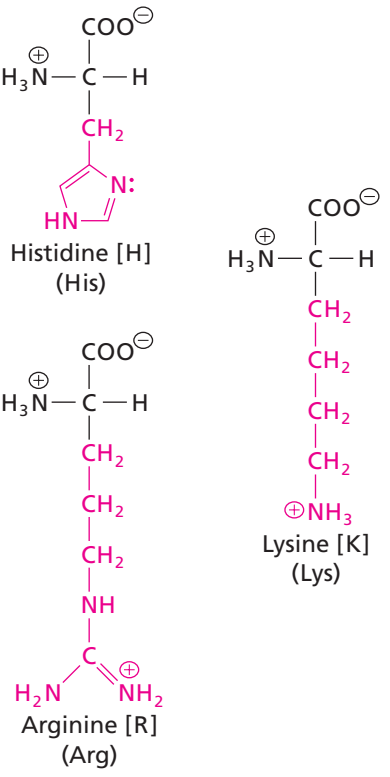
side chains. Unlike the more acidic phenolic side chain of tyrosine the hydroxyl groups of serine and threonine have the weak ionization properties of primary and secondary alcohols. The hydroxymethyl group of serine ($-CH_2OH$) does not appreciably ionize in aqueous solutions; nevertheless, this alcohol can react within the active sites of a number of enzymes as though it were ionized. Threonine, like isoleucine, has two chiral centers—the α- and β-carbon atoms. L-Threonine is the only one of the four stereoisomers that commonly occurs in proteins. (The other stereoisomers are called D-threonine, L-allothreonine, and D-allothreonine.)

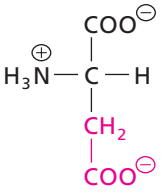
E. Positively Charged R Groups

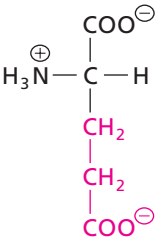
Histidine (His, H), lysine (Lys, K), and arginine (Arg, R) have hydrophilic side chains that are nitrogenous bases. The side chains can be positively charged at physiological pH.

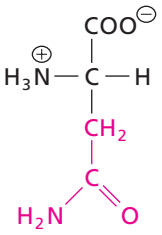
The side chain of histidine contains an imidazole ring substituent. The protonated form of this ring is called an imidazolium ion (Section 3.4). At pH 7 most histidines are neutral (base form) as shown in the accompanying figure but the form with a positively charged side chain is present and it becomes more common at slightly lower pH.

Lysine is a diamino acid with both α- and ε-amino groups. The ε-amino group exists as an alkylammonium ion ($-CH_2-NH_3^{\oplus}$) at neutral pH and confers a positive charge on proteins. Arginine is the most basic of the 20 amino acids because its




Aspartate [D]
(Asp)


Glutamate [E]
(Glu)


Asparagine [N]
(Asn)

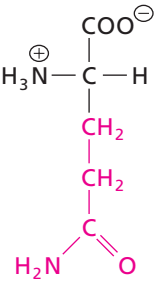

Glutamine [Q]
(Gln)

Table 3.1 Hydropathy scale

Amino acid	Free energy change of transfer[a] (kj mol^{-1})
Highly hydrophobic	
Isoleucine	3.1
Phenylalanine	2.5
Valine	2.3
Leucine	2.2
Methionine	1.1
Less hydrophobic	
Tryptophan	1.5[b]
Alanine	1.0
Glycine	0.67
Cysteine	0.17
Tyrosine	0.08
Proline	−0.29
Threonine	−0.75
Serine	−1.1
Highly hydrophilic	
Histidine	−1.7
Glutamate	−2.6
Asparagine	−2.7
Glutamine	−2.9
Aspartate	−3.0
Lysine	−4.6
Arginine	−7.5

[a]The free-energy change is for transfer of an amino acid residue from the interior of a lipid bilayer to water.

[b]On other scales, tryptophan has a lower hydropathy value.

[Adapted from Eisenberg, D., Weiss, R. M., Terwilliger, T. C., Wilcox, W. (1982). Hydrophobic moments in protein structure. *Faraday Symp. Chem. Soc.* 17:109–120.]

side-chain guanidinium ion is protonated under all conditions normally found within a cell. Arginine side chains also contribute positive charges in proteins.

F. Negatively Charged R Groups and Their Amide Derivatives

Aspartate (Asp, D) and glutamate (Glu, E) are dicarboxylic amino acids and have negatively charged hydrophilic side chains at pH 7. In addition to α-carboxyl groups, aspartate possesses a β-carboxyl group and glutamate possesses a γ-carboxyl group. Aspartate and glutamate confer negative charges on proteins because their side chains are ionized at pH 7. Aspartate and glutamate are sometimes called aspartic acid and glutamic acid but under most physiological conditions they are found as the conjugate bases and, like other carboxylates, have the suffix *-ate*. Glutamate is probably familiar as its monosodium salt, monosodium glutamate (MSG), which is used in food as a flavor enhancer.

Asparagine (Asn, N) and glutamine (Gln, Q) are the amides of aspartic acid and glutamic acid, respectively. Although the side chains of asparagine and glutamine are uncharged these amino acids are highly polar and are often found on the surfaces of proteins where they can interact with water molecules. The polar amide groups of asparagine and glutamine can also form hydrogen bonds with atoms in the side chains of other polar amino acids.

G. The Hydrophobicity of Amino Acid Side Chains

The various side chains of amino acids range from highly hydrophobic, through weakly polar, to highly hydrophilic. The relative hydrophobicity or hydrophilicity of each amino acid is called its **hydropathy**.

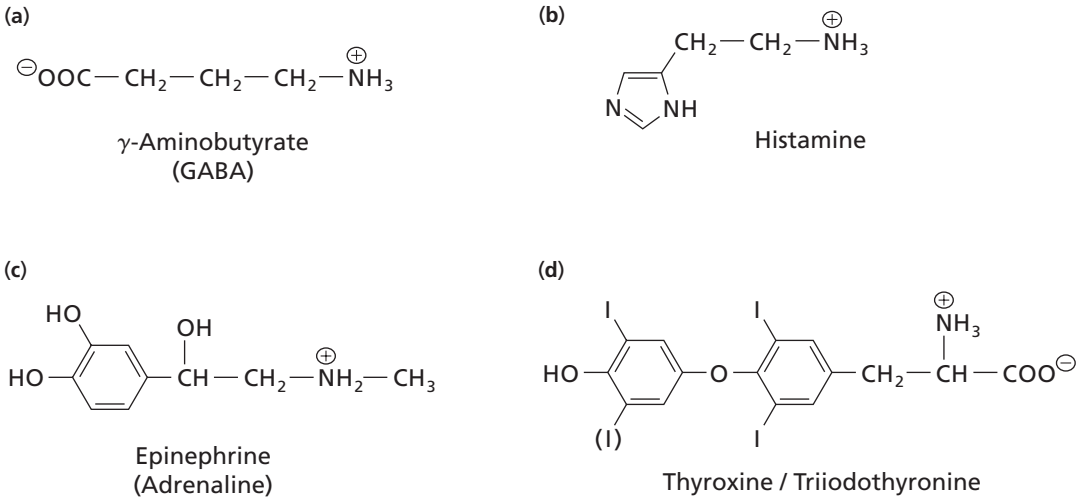
There are several ways of measuring hydropathy, but most of them rely on calculating the tendency of an amino acid to prefer a hydrophobic environment over a hydrophilic environment. A commonly used hydropathy scale is shown in Table 3.1. Amino acids with highly positive hydropathy values are considered hydrophobic whereas those with the largest negative values are hydrophilic. It is difficult to determine the hydropathy values of some amino acid residues that lie near the center of the scale. For example, there is disagreement over the hydropathy of the indole group of tryptophan and in some tables tryptophan has a much lower hydropathy value. Conversely, cysteine can have a higher hydropathy value in some tables.

Hydropathy is an important determinant of protein folding because hydrophobic side chains tend to be clustered in the interior of a protein and hydrophilic residues are usually found on the surface (Section 4.10). However, it is not yet possible to predict accurately whether a given residue will be found in the nonaqueous interior of a protein or on the solvent-exposed surface. On the other hand, hydropathy measurements of free amino acids can be successfully used to predict which segments of membrane-spanning proteins are likely to be embedded in a hydrophobic lipid bilayer (Chapter 9).

3.3 Other Amino Acids and Amino Acid Derivatives

More than 200 different amino acids are found in living organisms. In addition to the 20 common amino acids covered in the previous section there are three others that are incorporated into proteins during protein synthesis. The 21st amino acid is *N*-formylmethionine which serves as the initial amino acid during protein synthesis in bacteria (Section 22.5). The 22nd amino acid is selenocysteine which contains selenium in place of the sulfur of cysteine. It is incorporated into a few proteins in almost every species. Selenocysteine is formed from serine during protein synthesis. The 23rd amino acid is pyrrolysine, found in some species of archaebacteria. Pyrrolysine is a modified form of lysine that is synthesized before being added to a growing polypeptide chain by the translation machinery.

N-formylmethionine, selenocysteine, and pyrrolysine are incorporated at specific codons and that's why they are considered additions to the standard repertoire of protein precursors. Because of post-translational modifications many complete proteins have more than the standard 23 amino acids used in protein synthesis (see below).

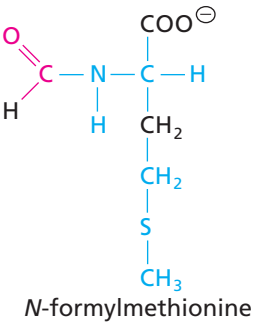


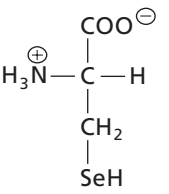
▲ **Figure 3.5**
Compounds derived from common amino acids. (a) γ-Aminobutyrate. a derivative of glutamate. **(b)** Histamine, a derivative of histidine. **(c)** Epinephrine, a derivative of tyrosine. **(d)** Thyroxine and triiodothyronine, derivatives of tyrosine. Thyroxine contains one more atom of iodine (in parentheses) than does triiodothyronine.

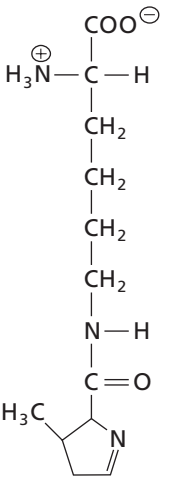
In addition to the common 23 amino acids that are incorporated into proteins, all species contain a variety of L-amino acids that are either precursors of the common amino acids or intermediates in other biochemical pathways. Examples are homocysteine, homoserine, ornithine, and citrulline (see Chapter 17). *S*-Adenosylmethionine (SAM) is a common methyl donor in many biochemical pathways (Section 7.2). Many species of bacteria and fungi synthesize D-amino acids that are used in cell walls and in complex peptide antibiotics such as actinomycin.

Several common amino acids are chemically modified to produce biologically important amines. These are synthesized by enzyme-catalyzed reactions that include decarboxylation and deamination. In the mammalian brain, for example, glutamate is converted to the neurotransmitter γ-aminobutyrate (GABA) (Figure 3.5a). Mammals can also synthesize histamine (Figure 3.5b) from histidine. Histamine controls the constriction of certain blood vessels and also the secretion of hydrochloric acid by the stomach. In the adrenal medulla, tyrosine is metabolized to epinephrine, also known as adrenaline (Figure 3.5c). Epinephrine and its precursor, norepinephrine (a compound whose amino group lacks a methyl substituent), are hormones that help regulate metabolism in mammals. Tyrosine is also the precursor of the thyroid hormones thyroxine and triiodothyronine (Figure 3.5d). Biosynthesis of the thyroid hormones requires iodide. Small amounts of sodium iodide are commonly added to table salt to prevent goiter, a condition of hypothyroidism caused by a lack of iodide in the diet.

Some amino acids are chemically modified after they have been incorporated into polypeptides. In fact, there are hundreds of known post-translational modifications. For example, some proline residues in the protein collagen are oxidized to form hydroxyproline residues (Section 4.11). Another common modification is the addition of complex carbohydrate chains—a process known as glycosylation (Chapters 8 and 22). Many proteins are phosphorylated, usually by the addition of phosphoryl groups to the side chains of serine, threonine, or tyrosine (histidine, lysine, cysteine, aspartate, and glutamate can also be phosphorylated). The oxidation of pairs of cysteine residues to form cystine also occurs after a polypeptide has been synthesized.


N-formylmethionine


Selenocysteine


Pyrrolysine

3.4 Ionization of Amino Acids

The physical properties of amino acids are influenced by the ionic states of the α-carboxyl and α-amino groups and of any ionizable groups in the side chains. Each ionizable group is associated with a specific pK_a value that corresponds to the pH at which the

BOX 3.3 COMMON NAMES OF AMINO ACIDS

Alanine: probably from aldehyde + "an" (for convenience) + amine (1849)

Arginine: crystallizes as a silver salt, from Latin *argentum* (silver) (1886)

Asparagine: first isolated from asparagus (1813)

Aspartate: similar to asparagine (1836)

Glutamate: first identified in the plant protein gluten (1866)

Glutamine: similar to glutamate (1866)

Glycine: from the Greek *glykys* (sweet), tastes sweet (1848)

Cysteine: from the Greek *kystis* (bladder), discovered in bladder stones (1882)

Histidine: first isolated from sturgeon sperm, named for the Greek *histidin* (tissue) (1896)

Isoleucine: isomer of leucine

Leucine: from the Greek *leukos* (white), forms white crystals (1820)

Lysine: product of protein hydrolysis, from the Greek *lysis* (loosening) (1891)

Methionine: side chain is a sulfur (Greek *theion*) atom with a methyl group (1928)

Phenylalanine: alanine with a phenyl group (1883)

Proline: a corrupted form of "pyrrolidine" because it forms a pyrrolidine ring (1904)

Serine: from the Latin *sericum* (silk), serine is common in silk (1865)

Threonine: similar to the four-carbon sugar threose (1936)

Tryptophan: isolated from a tryptic digest of protein 1 Greek *phanein* (to appear) (1890)

Tyrosine: found in cheese, from the Greek *tyros* (cheese) (1890)

Valine: derivative of valeric acid from the plant genus *Valeriana* (1906)

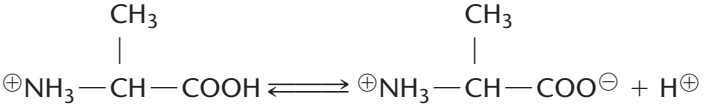
Sources: *Oxford English Dictionary* 2nd ed., and Leung, S.H. (2000) Amino acids, aromatic compounds, and carboxylic acids: how did they get their common names? *J. Chem. Educ.* 77: 48–49.

concentrations of the protonated and unprotonated forms are equal (Section 2.9). When the pH of the solution is below the pK_a the protonated form predominates and the amino acid is then a true acid that is capable of donating a proton. When the pH of the solution is above the pK_a of the ionizable group the unprotonated form of that group predominates and the amino acid exists as the conjugate base, which is a proton acceptor. Every amino acid has at least two pK_a values corresponding to the ionization of the α-carboxyl and α-amino groups. In addition, seven of the common amino acids have ionizable side chains with additional, measurable pK_a values. These values differ among the amino acids. Thus, at a given pH, amino acids frequently have different net charges. Many of the modified amino acids have additional ionizable groups contributing to the diversity of charged amino acid side chains in proteins. Phosphoserine and phosphotyrosine, for example, will be negatively charged.

Knowing the ionic states of amino acid side chains is important for two reasons. First, the charged state influences protein folding and the three-dimensional structure of proteins (Section 4.10). Second, an understanding of the ionic properties of amino acids in the active site of an enzyme helps one understand enzyme mechanisms (Chapter 6).

The pK_a values of amino acids are determined from titration curves such as those we saw in the previous chapter. The titration of alanine is shown in Figure 3.6. Alanine has two ionizable groups—the α-carboxyl and the protonated α-amino group. As more base is added to the solution of acid, the titration curve exhibits two pK_a values, at pH 2.4 and pH 9.9. Each pK_a value is associated with a buffering zone where the pH of the solution changes relatively little when more base is added.

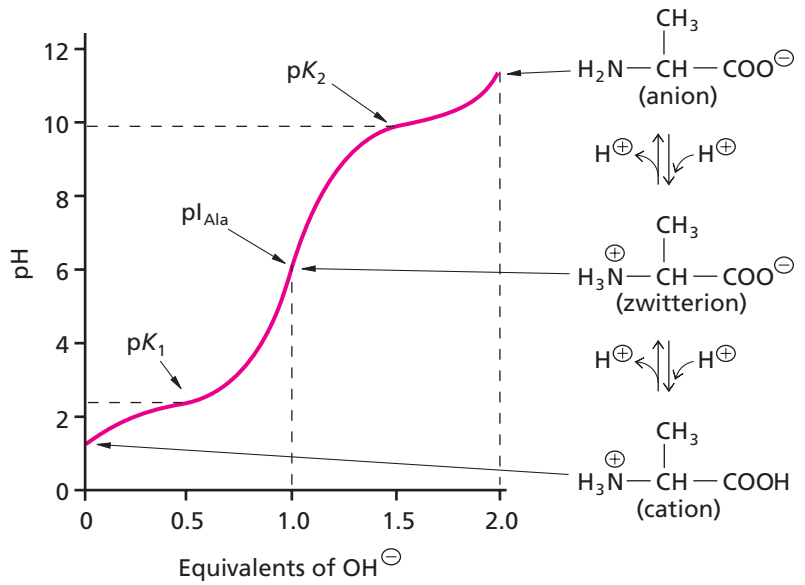
The pK_a of an ionizable group corresponds to a midpoint of its titration curve. It is the pH at which the concentration of the acid form (proton donor) exactly equals the concentration of its conjugate base (proton acceptor). In the example shown in Figure 3.6 the concentrations of the positively charged form of alanine and of the zwitterion are equal at pH 2.4.



(3.1)

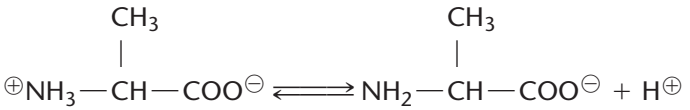
KEY CONCEPT

For every acid-base pair the p*Ka* is the pH at which the concentrations of the two forms are equal.



◀ **Figure 3.6**
Titration curve for alanine. The first pK_a value is 2.4; the second is 9.9. pI_{Ala} represents the isoelectric point of alanine.

At pH 9.9 the concentration of the zwitterion equals the concentration of the negatively charged form.



(3.2)

Note that in the acid–base pair shown in the first equilibrium (Reaction 3.1) the zwitterion is the conjugate base of the acid form of alanine. In the second acid–base pair (Reaction 3.2) the zwitterion is the proton donor, or conjugate acid, of the more basic form that predominates at higher pH.

One can deduce that the net charge on alanine molecules at pH 2.4 averages +0.5 because there are equal amounts of neutral zwitterion (+/–) and cation (+). The net charge at pH 9.9 averages –0.5. Midway between pH 2.4 and pH 9.9, at pH 6.15, the average net charge on alanine molecules in solution is zero. For this reason, pH 6.15 is referred to as the isoelectric point (pI), or isoelectric pH, of alanine. If alanine were placed in an electric field at a pH below its pI it would carry a net positive charge (in other words, its cationic form would predominate), and it would therefore migrate toward the cathode (the negative electrode). At a pH higher than its pI alanine would carry a net negative charge and would migrate toward the anode (the positive electrode). At its isoelectric point (pH = 6.15) alanine would not migrate in either direction.

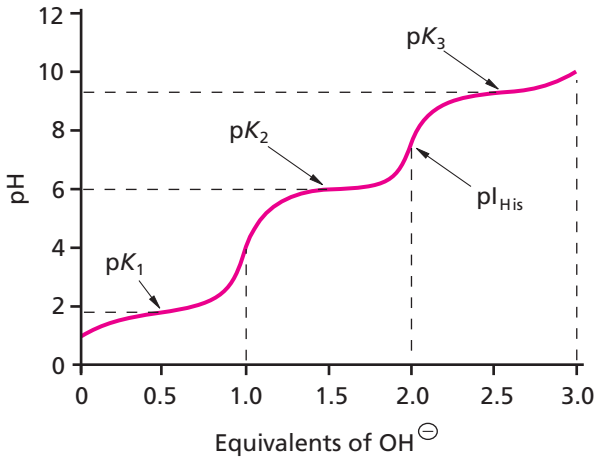
Histidine contains an ionizable side chain. The titration curve for histidine contains an additional inflection point that corresponds to the pK_a of its side chain (Figure 3.7a).

KEY CONCEPT

The ionic state of a particular amino acid side chain is determined by its pK_a value and the pH of the local environment.

▼ **Figure 3.7**
Ionization of histidine. (a) Titration curve for histidine. The three pK_a values are 1.8, 6.0, and 9.3. pI_{His} represents the isoelectric point of histidine. **(b)** Deprotonation of the imidazolium ring of the side chain of histidine.

(a)



(b)

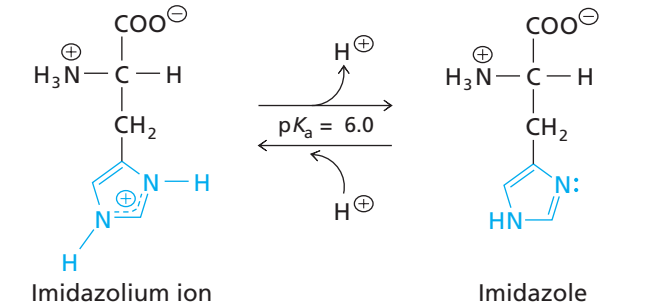


Table 3.2 pK_a values of acidic and basic constituents of free amino acids at 25°C

Amino acid	p*K*a value		
	Carboxyl group	Amino group	Side chain
Glycine	2.4	9.8	
Alanine	2.4	9.9	
Valine	2.3	9.7	
Leucine	2.3	9.7	
Isoleucine	2.3	9.8	
Methionine	2.1	9.3	
Proline	2.0	10.6	
Phenylalanine	2.2	9.3	
Tryptophan	2.5	9.4	
Serine	2.2	9.2	
Threonine	2.1	9.1	
Cysteine	1.9	10.7	8.4
Tyrosine	2.2	9.2	10.5
Asparagine	2.1	8.7	
Glutamine	2.2	9.1	
Aspartic acid	2.0	9.9	3.9
Glutamic acid	2.1	9.5	4.1
Lysine	2.2	9.1	10.5
Arginine	1.8	9.0	12.5
Histidine	1.8	9.3	6.0

As is the case with alanine, the first pK_a (1.8) represents the ionization of the α-COOH carboxyl group and the most basic pK_a value (9.3) represents the ionization of the α-amino group. The middle pK_a (6.0) corresponds to the deprotonation of the imidazolium ion of the side chain of histidine (Figure 3.7b). At pH 7.0 the ratio of imidazole (conjugate base) to imidazolium ion (conjugate acid) is 10:1. Thus, the protonated and neutral forms of the side chain of histidine are both present in significant concentrations near physiological pH. A given histidine side chain in a protein may be either protonated or unprotonated depending on its immediate environment within the protein. In other words, the actual pK_a value of the side-chain group may not be the same as its value for the free amino acid in solution. This property makes the side chain of histidine ideal for the transfer of protons within the catalytic sites of enzymes. (A famous example is described in Section 6.7c.)

The isoelectric point of an amino acid that contains only two ionizable groups (the α-amino and the α-carboxyl groups) is the arithmetic mean of its two pK_a values (i.e., pI = (pK_1 + pK_2)/2). However, for an amino acid that contains three ionizable groups, such as histidine, one must assess the net charge of each ionic species. The isoelectric point for histidine lies between the pK_a values on either side of the species with no net charge, that is, midway between 6.0 and 9.3, or 7.65.

As shown in Table 3.2 the pK_a values of the α-carboxyl groups of free amino acids range from 1.8 to 2.5. These values are lower than those of typical carboxylic acids such as acetic acid (pK_a = 4.8) because the neighboring —$NH_3^{\oplus}$ group withdraws electrons from the carboxylic acid group and this favors the loss of a proton from the α-carboxyl group. The side chains, or R groups, also influence the pK_a value of the α-carboxyl group which is why different amino acids have different pK_a values. (We have just seen that the values for histidine and alanine are not the same.)

The α-COOH group of an amino acid is a weak acid. We can use the Henderson–Hasselbalch equation (Section 2.9) to calculate the fraction of the group that is ionized at any given pH.

$$\text{pH} = \text{p}K_a + \log\frac{[\text{proton acceptor}]}{[\text{proton donor}]} \qquad \textbf{(3.3)}$$

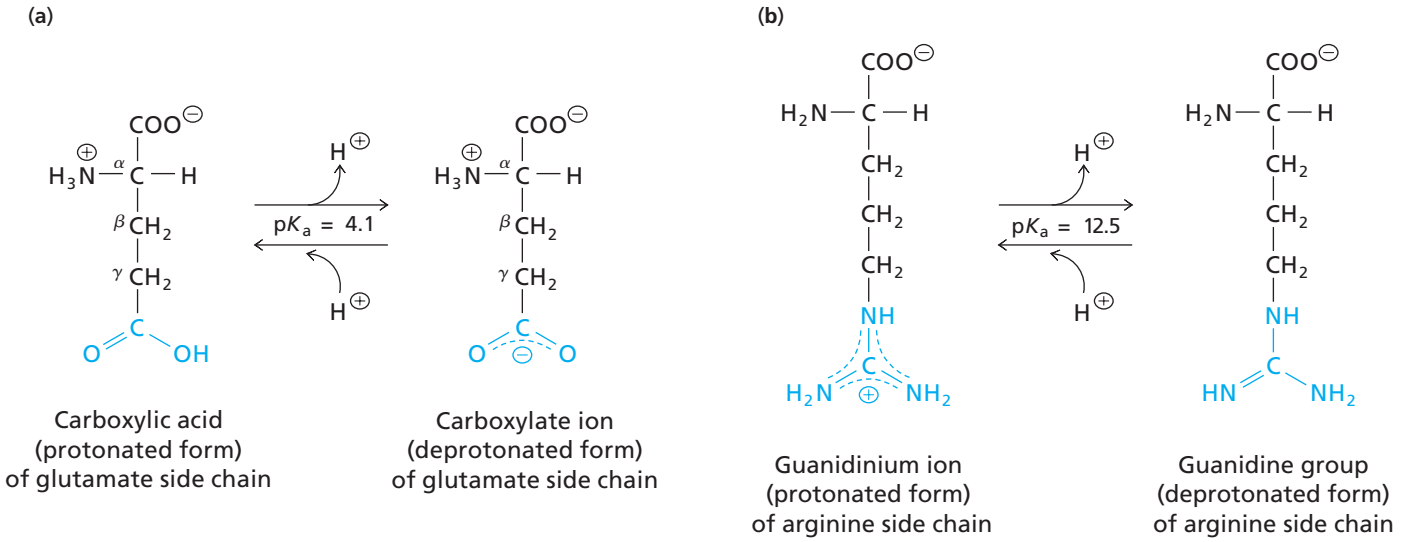
For a typical amino acid whose α-COOH group has a pK_a of 2.0, the ratio of proton acceptor (carboxylate anion) to proton donor (carboxylic acid) at pH 7.0 can be calculated using the Henderson–Hasselbalch equation.

$$7.0 = 2.0 + \log\frac{[\text{RCOO}^{\ominus}]}{[\text{RCOOH}]} \qquad \textbf{(3.4)}$$

In this case, the ratio of carboxylate anion to carboxylic acid is 100,000:1. This means that under the conditions normally found inside a cell the carboxylate anion is the predominant species.

The α-amino group of a free amino acid can exist as a free amine, —NH_2 (proton acceptor) or as a protonated amine, —$NH_3^{\oplus}$ (proton donor). The pK_a values range from 8.7 to 10.7 as shown in Table 3.2. For an amino acid whose α-amino group has a pK_a value of 10.0 the ratio of proton acceptor to proton donor is 1:1000 at pH 7.0. In other words, under physiological conditions the α-amino group is mostly protonated and positively charged. These calculations verify our earlier statement that free amino acids exist predominantly as zwitterions at neutral pH. They also show that it is inappropriate to draw the structure of an amino acid with both —COOH and —NH groups since there is no pH at which a significant number of molecules contain a protonated carboxyl group and an unprotonated amino group (see Problem 19). Note that the secondary amino group of proline (pK_a = 10.6) is also protonated at neutral pH so proline—despite the bonding of the side chain to the α-amino group—is also zwitterionic at pH 7.

The seven standard amino acids with readily ionizable groups in their side chains are aspartate, glutamate, histidine, cysteine, tyrosine, lysine, and arginine. Ionization of these groups obeys the same principles as ionization of the α-carboxyl and α-amino groups and the Henderson–Hasselbalch equation can be applied to each ionization. The ionization of the γ-carboxyl group of glutamate (pK_a = 4.1) is shown in Figure 3.8a.



▲ **Figure 3.8**
Ionization of amino acid side chains. (a) Ionization of the protonated γ-carboxyl group of glutamate. The negative charge of the carboxylate anion is delocalized. **(b)** Deprotonation of the guanidinium group of the side chain of arginine. The positive charge is delocalized.

Note that the γ-carboxyl group is further removed from the influence of the α-ammonium ion and behaves as a weak acid with a pK_a of 4.1. This makes it similar in strength to acetic acid ($pK_a = 4.8$) whereas the α-carboxyl group is a stronger acid ($pK_a = 2.1$). Figure 3.8b shows the deprotonation of the guanidinium group of the side chain of arginine in a strongly basic solution. Charge delocalization stabilizes the guanidinium ion contributing to its high pK_a value of 12.5.

As mentioned earlier, the pK_a values of ionizable side chains in proteins can differ from those of the free amino acids. Two factors cause this perturbation of ionization constants. First, α-amino and α-carboxyl groups lose their charges once they are linked by peptide bonds in proteins—consequently, they exert weaker inductive effects on their neighboring side chains. Second, the position of an ionizable side chain within the three dimensional structure of a protein can affect its pK_a. For example, the enzyme ribonuclease A has four histidine residues but the side chain of each residue has a slightly different pK_a as a result of differences in their immediate surroundings, or microenvironments.

3.5 Peptide Bonds Link Amino Acids in Proteins

The linear sequence of amino acids in a polypeptide chain is called the **primary structure** of a protein. Higher levels of structure are referred to as secondary, tertiary, and quaternary. The structure of proteins is covered more thoroughly in the next chapter but it's important to understand peptide bonds and primary structure before discussing some of the remaining topics in this chapter.

The structure of peptide bonds is described in Section 4.3.

The linkage formed between amino acids is an amide bond called a **peptide bond** (Figure 3.9). This linkage can be thought of as the product of a simple condensation reaction between the α-carboxyl group of one amino acid and the α-amino group of another. A water molecule is lost from the condensing amino acids in the reaction. (Recall from Section 2.6 that such simple condensation reactions are extremely unfavorable in aqueous solutions due to the huge excess of water molecules. The actual pathway of protein synthesis involves reactive intermediates that overcome this limitation.) Unlike the carboxyl and amino groups of free amino acids in solution the groups involved in peptide bonds carry no ionic charges.

Protein synthesis (translation) is described in Chapter 22.

Linked amino acids in a polypeptide chain are called amino acid *residues*. The names of residues are formed by replacing the ending *-ine* or *-ate* with *-yl*. For example, a glycine residue in a polypeptide is called glycyl and a glutamate residue is called glutamyl.

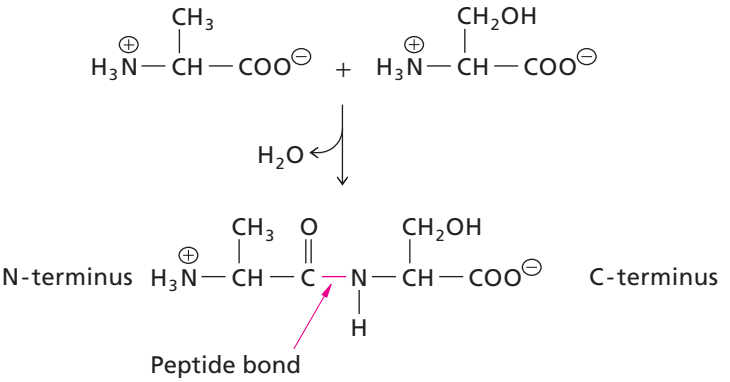


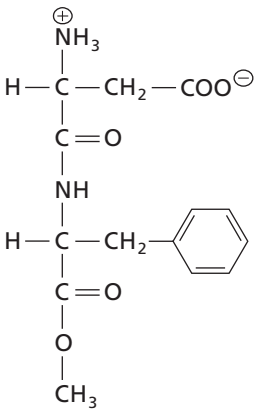
Figure 3.9 ►
Peptide bond between two amino acids. The structure of the peptide linkage can be viewed as the product of a condensation reaction in which the α-carboxyl group of one amino acid condenses with the α-amino group of another amino acid. The result is a dipeptide in which the amino acids are linked by a peptide bond. Here, alanine is condensed with serine to form alanylserine.

In the cases of asparagine, glutamine, and cysteine, *-yl* replaces the final *-e* to form asparaginyl, glutaminyl, and cysteinyl, respectively. The *-yl* ending indicates that the residue is an acyl unit (a structure that lacks the hydroxyl of the carboxyl group). The dipeptide in Figure 3.9 is called alanylserine because alanine is converted to an acyl unit but the amino acid serine retains its carboxyl group.

The free amino group and free carboxyl group at the opposite ends of a peptide chain are called the N-terminus (amino terminus) and the C-terminus (carboxyl terminus), respectively. At neutral pH each terminus carries an ionic charge. By convention, amino acid residues in a peptide chain are numbered from the N-terminus to the C-terminus and are usually written from left to right. This convention corresponds to the direction of protein synthesis (Section 22.6). Synthesis begins with the N-terminal amino acid—almost always methionine (Section 22.5)—and proceeds sequentially toward the C-terminus by adding one residue at a time.

Both the standard three-letter abbreviations for the amino acids (e.g., Gly–Arg–Phe–Ala–Lys) and the one-letter abbreviations (e.g., GRFAK) are used to describe the sequence of amino acid residues in peptides and polypeptides. It's important to know both abbreviation systems. The terms *dipeptide*, *tripeptide*, *oligopeptide*, and *polypeptide* refer to chains of two, three, several (up to about 20), and many (usually more than 20) amino acid residues, respectively. A dipeptide contains one peptide bond, a tripeptide contains two peptide bonds, and so on. As a general rule, each peptide chain, whatever its length, possesses one free α-amino group and one free α-carboxyl group. (Exceptions include covalently modified terminal residues and circular peptide chains.) Note that the formation of a peptide bond eliminates the ionizable α-carboxyl and α-amino groups found in free amino acids. As a result, most of the ionic charges associated with a protein molecule are contributed by the side chains of the amino acids. This means that the solubility and ionic properties of a protein are largely determined by its amino acid composition. Furthermore, the side chains of the residues interact with each other and these interactions contribute to the three dimensional shape and stability of a protein molecule (Chapter 4).

Some peptides are important biological compounds and the chemistry of peptides is an active area of research. Several hormones are peptides; for example, endorphins are the naturally occurring molecules that modulate pain in vertebrates. Some very simple peptides are useful as food additives; for example, the sweetening agent aspartame is the methyl ester of aspartylphenylalanine (Figure 3.10). Aspartame is about 200 times sweeter than table sugar and is widely used in diet drinks. There are also many peptide toxins such as those found in snake venom and poisonous mushrooms.



▲ **Figure 3.10**
Aspartame (aspartylphenylalanine methyl ester).

3.6 Protein Purification Techniques

In order to study a particular protein in the laboratory it must be separated from all other cell components including other, similar proteins. Few analytical techniques will work with crude mixtures of cellular proteins because they contain hundreds (or thousands) of different proteins. The purification steps are different for each protein. They are worked

out by trying a number of different techniques until a procedure is developed that reproducibly yields highly purified protein that is still biologically active. Purification steps usually exploit minor differences in the solubilities, net charges, sizes, and binding specificities of proteins. In this section, we consider some of the common methods of protein purification. Most purification techniques are performed at 0°C to 4°C to minimize temperature-dependent processes such as protein degradation and denaturation (unfolding).

The first step in protein purification is to prepare a solution of proteins. The source of a protein is often whole cells in which the target protein accounts for less than 0.1% of the total dry weight. Isolation of an intracellular protein requires that cells be suspended in a buffer solution and homogenized, or disrupted into cell fragments. Under these conditions most proteins dissolve. (Major exceptions include membrane proteins which require special purification procedures.) Let's assume that the desired protein is one of many proteins in this solution.

One of the first steps in protein purification is often a relatively crude separation that makes use of the different solubilities of proteins in salt solutions. Ammonium sulfate is frequently used in such fractionations. Enough ammonium sulfate is mixed with the solution of proteins to precipitate the less soluble impurities, which are removed by centrifugation. The target protein and other more soluble proteins remain in the fluid called the supernatant fraction. Next, more ammonium sulfate is added to the supernatant fraction until the desired protein is precipitated. The mixture is centrifuged, the fluid removed, and the precipitate dissolved in a minimal volume of buffer solution. Typically, fractionation using ammonium sulfate gives a two- to threefold purification (i.e., one-half to two-thirds of the unwanted proteins have been removed from the resulting enriched protein fraction). At this point the solvent containing residual ammonium sulfate is exchanged by dialysis for a buffer solution suitable for chromatography.

In dialysis, a protein solution is sealed in a cylinder of cellophane tubing and suspended in a large volume of buffer. The cellophane membrane is semipermeable—high molecular weight proteins are too large to pass through the pores of the membrane so proteins remain inside the tubing while low molecular weight solutes (including, in this case, ammonium and sulfate ions) diffuse out and are replaced by solutes in the buffer.

Column chromatography is often used to separate a mixture of proteins. A cylindrical column is filled with an insoluble material such as substituted cellulose fibers or synthetic beads. The protein mixture is applied to the column and washed through the matrix of insoluble material by the addition of solvent. As solvent flows through the column the eluate (the liquid emerging from the bottom of the column) is collected in many fractions, a few of which are represented in Figure 3.11a. The rate at which proteins travel through the matrix depends on interactions between matrix and protein. For a given column different proteins are eluted at different rates. The concentration of protein in each fraction can be determined by measuring the absorbance of the eluate at a wavelength of 280 nm (Figure 3.11b). (Recall from Section 3.2B that at neutral pH, tyrosine and tryptophan absorb UV light at 280 nm.) To locate the target protein the fractions containing protein must then be assayed, or tested, for biological activity or some other characteristic property. Column chromatography may be performed under high pressure using small, tightly packed columns with solvent flow controlled by a computer. This technique is called **HPLC**, for high-performance liquid chromatography.

Chromatographic techniques are classified according to the type of matrix. In **ion-exchange chromatography** the matrix carries positive charges (anion-exchange resins) or negative charges (cation-exchange resins). Anion-exchange matrices bind negatively charged proteins retaining them in the matrix for subsequent elution. Conversely, cation-exchange materials bind positively charged proteins. The bound proteins can be serially eluted by gradually increasing the salt concentration in the solvent. As the salt concentration is increased it eventually reaches a concentration where the salt ions outcompete proteins in binding to the matrix. At this concentration the protein is released and is collected in the eluate. Individual bound proteins are eluted at different salt concentrations and this fractionation makes ion-exchange chromatography a powerful tool in protein purification.

Gel-filtration chromatography separates proteins on the basis of molecular size. The gel is a matrix of porous beads. Proteins that are smaller than the average pore size

▲ There is only one correct way to write the sequence of a polypeptide- from N-teminus to C-terminus.

▲ **Green mamba (*Dendroapsis angusticeps*).** One of the toxins in the venom of this poisonous snake is a large peptide with the sequence MICYSHKTPQPSATITCEEKT-CYKKSVRKL PAVVAGRGCGCPSKEMLVAIH CCRSDKCNE [Viljoen and Botes (1974). *J.Biol.Chem.* 249:366]

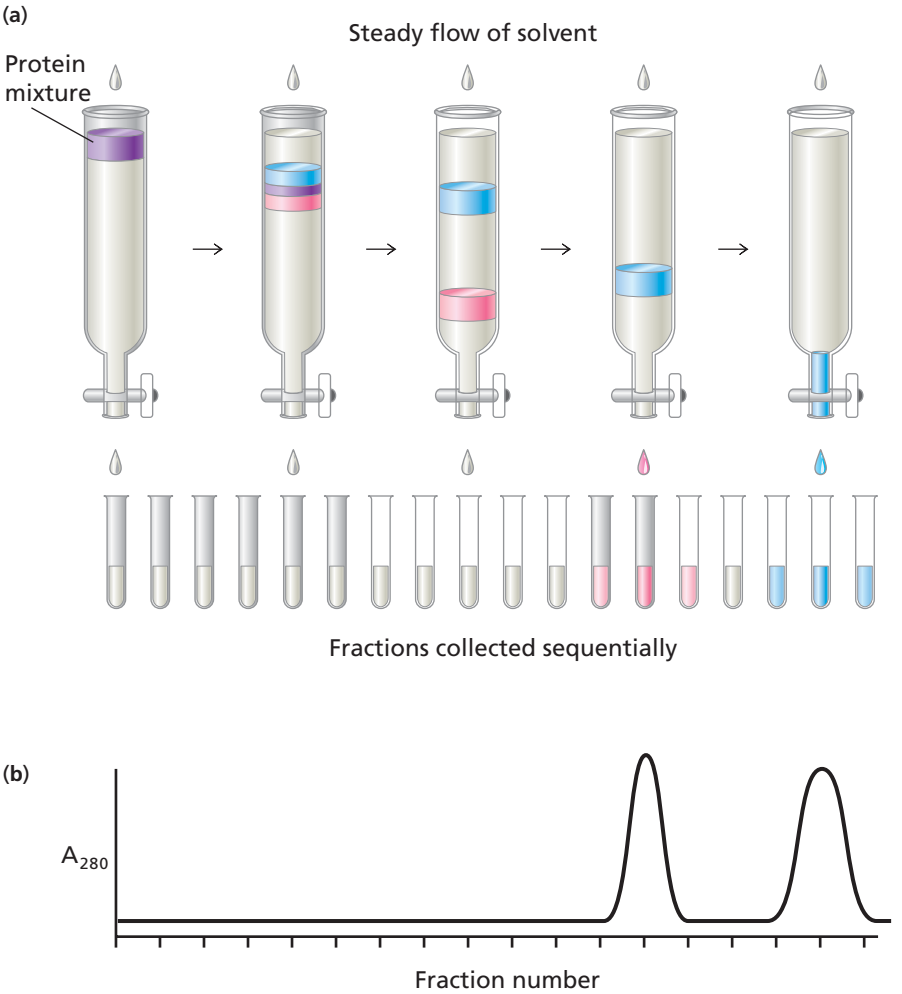
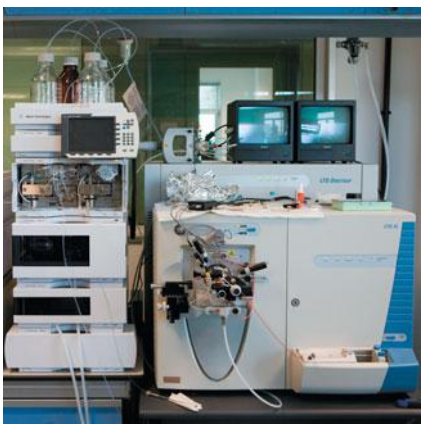


Figure 3.11 ►
Column chromatography. (a) A mixture of proteins is added to a column containing a solid matrix. Solvent then flows into the column from a reservoir. Washed by solvent, different proteins (represented by red and blue bands) travel through the column at different rates, depending on their interactions with the matrix. Eluate is collected in a series of fractions, a few of which are shown. **(b)** The protein concentration of each fraction is determined by measuring the absorbance at 280 nm. The peaks correspond to the elution of the protein bands shown in (a). The fractions are then tested for the presence of the target protein.



▲ A typical high-performance liquid chromatography (HPLC) system in a research lab (left). The large instrument on the right is a mass spectrometer (Istituto di Ricerche Farmacologiche, Milan, Italy)

penetrate much of the internal volume of the beads and are therefore retarded by the matrix as the buffer solution flows through the column. The smaller the protein, the later it elutes from the column. Fewer of the pores are accessible to larger protein molecules. Consequently, the largest proteins flow past the beads and elute first.

Affinity chromatography is the most selective type of column chromatography. It relies on specific binding interactions between the target protein and some other molecule that is covalently bound to the matrix of the column. The molecule bound to the matrix may be a substance or a ligand that binds to a protein *in vivo*, an antibody that recognizes the target protein, or another protein that is known to interact with the target protein inside the cell. As a mixture of proteins passes through the column only the target protein specifically binds to the matrix. The column is then washed with buffer several times to rid it of nonspecifically bound proteins. Finally, the target protein can be eluted by washing the column with a solvent containing a high concentration of salt that disrupts the interaction between the protein and column matrix. In some cases, bound protein can be selectively released from the affinity column by adding excess ligand to the elution buffer. The target protein preferentially binds to the ligand in solution instead of the lower concentration of ligand that is attached to the insoluble matrix of the column. This method is most effective when the ligand is a small molecule. Affinity chromatography alone can sometimes purify a protein 1000- to 10,000-fold.

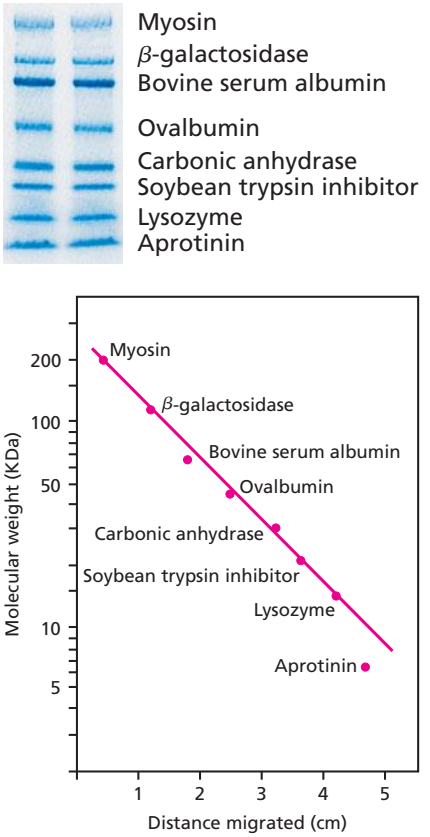
3.7 Analytical Techniques

Electrophoresis separates proteins based on their migration in an electric field. In **polyacrylamide gel electrophoresis (PAGE)** protein samples are placed on a highly cross-linked gel matrix of polyacrylamide and an electric field is applied. The matrix is

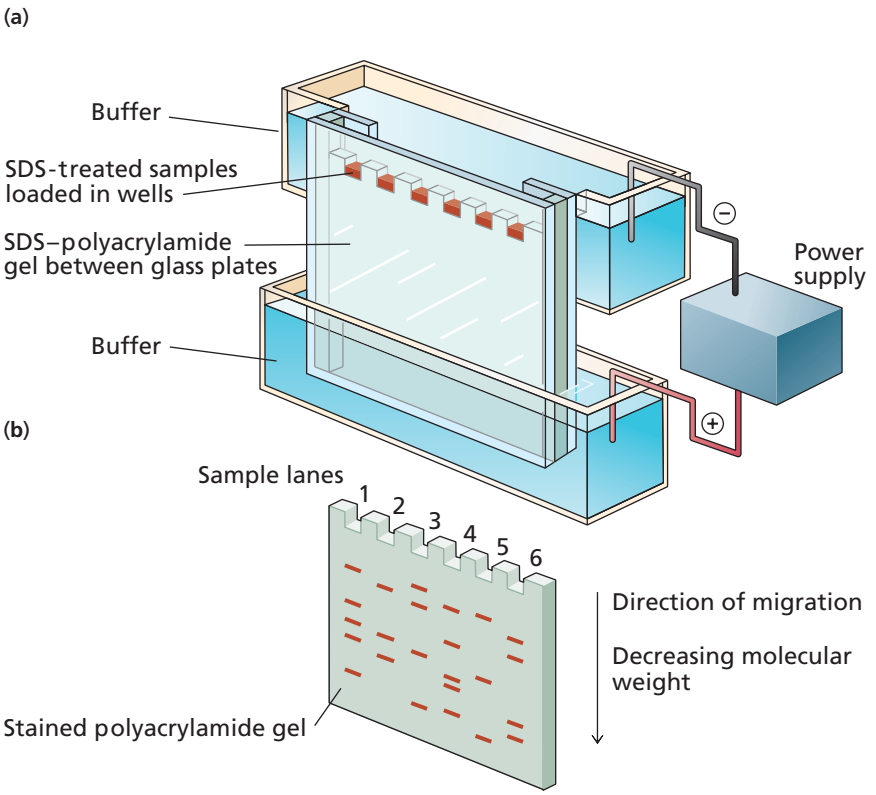
buffered to a mildly alkaline pH so that most proteins are anionic and migrate toward the anode. Typically, several samples are run at once together with a reference sample. The gel matrix retards the migration of large molecules as they move in the electric field. Hence, proteins are fractionated on the basis of both charge and mass.

A modification of the standard electrophoresis technique uses the negatively charged detergent sodium dodecyl sulfate (SDS) to overwhelm the native charge on proteins so that they are separated on the basis of mass only. SDS–polyacrylamide gel electrophoresis (**SDS–PAGE**) is used to assess the purity and to estimate the molecular weight of a protein. In SDS–PAGE the detergent is added to the polyacrylamide gel as well as to the protein samples. A reducing agent is also added to the samples to reduce any disulfide bonds. The dodecyl sulfate anion, which has a long hydrophobic tail ($CH_3(CH_2)_{11}OSO_3^{\ominus}$, Figure 2.8) binds to hydrophobic side chains of amino acid residues in the polypeptide chain. SDS binds at a ratio of approximately one molecule for every two residues of a typical protein. Since larger proteins bind proportionately more SDS the charge-to-mass ratios of all treated proteins are approximately the same. All the SDS–protein complexes are highly negatively charged and move toward the anode as diagrammed in Figure 3.12a. However, their rate of migration through the gel is inversely proportional to the logarithm of their mass—larger proteins encounter more resistance and therefore migrate more slowly than smaller proteins. This sieving effect differs from gel-filtration chromatography because in gel filtration larger molecules are excluded from the pores of the gel and hence travel faster. In SDS–PAGE all molecules penetrate the pores of the gel so the largest proteins travel most slowly. The protein bands that result from this differential migration (Figure 3.13) can be visualized by staining. Molecular weights of unknown proteins can be estimated by comparing their migration to the migration of reference proteins on the same gel.

Although SDS–PAGE is primarily an analytical tool, it can be adapted for purifying proteins. Denatured proteins can be recovered from SDS–PAGE by cutting out the bands of a gel. The protein is then electroeluted by applying an electric current to allow the protein to migrate into a buffer solution. After concentration and the removal of salts such protein preparations can be used for structural analysis, preparation of antibodies, or other purposes.



▲ **Figure 3.13**
Proteins separated on an SDS–polyacrylamide gel. (a) Stained proteins after separation. The high molecular weight proteins are at the top of the gel. **(b)** Graph showing the relationship between the molecular weight of a protein and the distance it migrates in the gel.



◀ **Figure 3.12**
SDS–PAGE. (a) An electrophoresis apparatus includes an SDS–polyacrylamide gel between two glass plates and buffer in the upper and lower reservoirs. Samples are loaded into the wells of the gel, and voltage is applied. Because proteins complexed with SDS are negatively charged, they migrate toward the anode. **(b)** The banding pattern of the proteins after electrophoresis can be visualized by staining. The smallest proteins migrate fastest, so the proteins of lowest molecular weight are at the bottom of the gel.

Mass spectrometry, as the name implies, is a technique that determines the mass of a molecule. The most basic type of mass spectrometer measures the time that it takes for a charged gas phase molecule to travel from the point of injection to a sensitive detector. This time depends on the charge of a molecule and its mass and the result is reported as the mass/charge ratio. The technique has been used in chemistry for almost 100 years but its application to proteins was limited because, until recently, it was not possible to disperse charged protein molecules into a gaseous stream of particles.

This problem was solved in the late 1980s with the development of two new types of mass spectrometry. In **electrospray mass spectrometry** the protein solution is pumped through a metal needle at high voltage to create tiny droplets. The liquid rapidly evaporates in a vacuum and the charged proteins are focused on a detector by a magnetic field. The second new technique is called **matrix-assisted laser desorption ionization (MALDI)**. In this method the protein is mixed with a chemical matrix and the mixture is precipitated on a metal substrate. The matrix is a small organic molecule that absorbs light at a particular wavelength. A laser pulse at the absorption wavelength imparts energy to the protein molecules via the matrix. The proteins are instantly released from the substrate (desorbed) and directed to the detector (Figure 3.14). When time-of-flight (TOF) is measured, the technique is called MALDI–TOF.

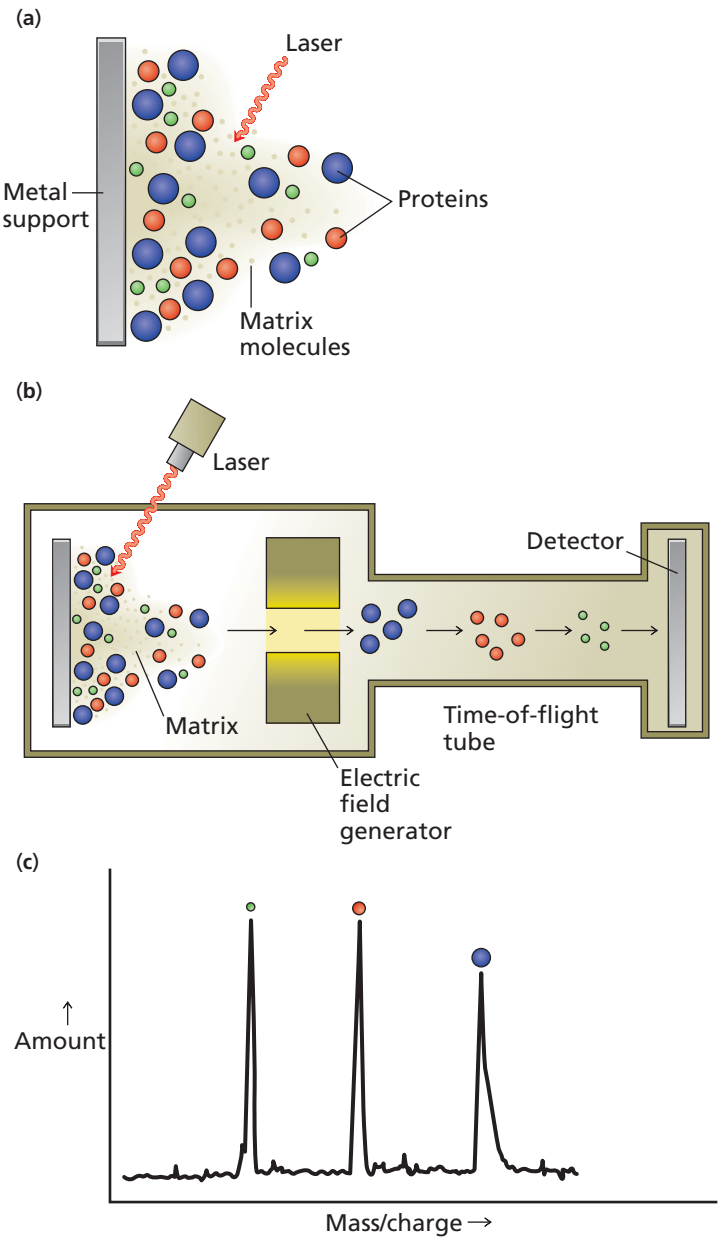


Figure 3.14 ►
MALDI–TOF mass spectrometry. (a) A burst of light releases proteins from the matrix. **(b)** Charged proteins are directed toward the detector by an electric field. **(c)** The time of arrival at the detector depends on the mass and the charge of the protein.

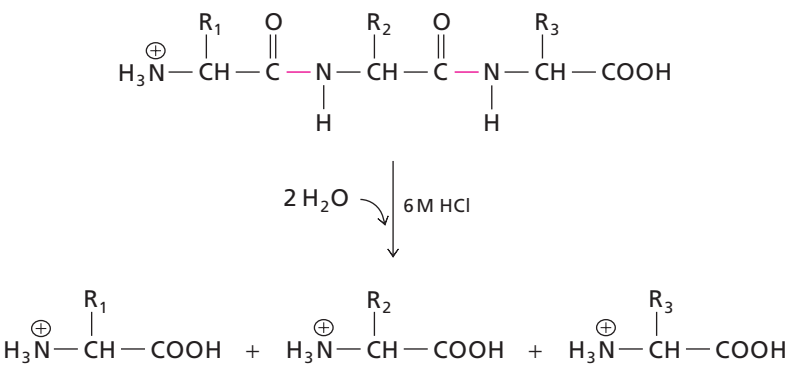
The raw data from a mass spectrometry experiment can be quite simple as shown in Figure 3.14. There, a single species with one positive charge is detected so the mass/charge ratio gives the mass directly. In other cases the spectra can be more complicated, especially in electrospray mass spectrometry. Often there are several different charged species and the correct mass has to be calculated by analyzing a collection of molecules with charges of +1, +2, +3, etc. The spectrum can be daunting when the source is a mixture of different proteins. Fortunately, there are sophisticated computer programs that can analyze the data and calculate the correct masses. The current popularity of mass spectrometry owes as much to the development of this software as it does to the new hardware and new methods of sample preparation.

Mass spectrometry is very sensitive and highly accurate. Often the mass of a protein can be obtained from picomole (10^{-12} mol) quantities that are isolated from an SDS–PAGE gel. The correct mass can be determined with an accuracy of less than the mass of a single proton.

3.8 Amino Acid Composition of Proteins

Once a protein has been isolated its amino acid composition can be determined. First, the peptide bonds of the protein are cleaved by acid hydrolysis, typically using 6 M HCl (Figure 3.15). Next, the hydrolyzed mixture, or hydrolysate, is subjected to a chromatographic procedure in which each of the amino acids is separated and quantitated, a process called **amino acid analysis**. One method of amino acid analysis involves treatment of the protein hydrolysate with phenylisothiocyanate (PITC) at pH 9.0 to generate phenylthiocarbamoyl (PTC)–amino acid derivatives (Figure 3.16). The PTC–amino acid mixture is then subjected to HPLC in a column of fine silica beads to which short hydrocarbon chains have been attached. The amino acids are separated by the hydrophobic properties of their side chains. As each PTC–amino acid derivative is eluted it is detected and its concentration is determined by measuring the absorbance of the eluate at 254 nm (the peak absorbance of the PTC moiety). Since different PTC–amino acid derivatives are eluted at different rates the time at which an amino acid derivative elutes from the column identifies the amino acid relative to known standards. The amount of each amino acid in the hydrolysate is proportional to the area under its peak. With this method, amino acid analysis can be performed on samples as small as 1 picomole of a protein that contains approximately 200 residues.

Despite its usefulness, acid hydrolysis cannot yield a complete amino acid analysis. Since the side chains of asparagine and glutamine contain amide bonds the acid used to cleave the peptide bonds of the protein also converts asparagine to aspartic acid and glutamine to glutamic acid. Other limitations of the acid hydrolysis method include small losses of serine, threonine, and tyrosine. In addition, the side chain of tryptophan is almost totally destroyed by acid hydrolysis. There are several ways of overcoming these limitations. For example, proteins can be hydrolyzed to amino acids by enzymes

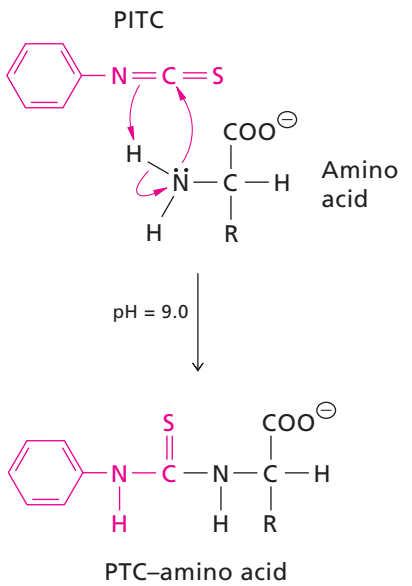


▲ **Figure 3.15**
Acid-catalyzed hydrolysis of a peptide. Incubation with 6 M HCl at 110°C for 16 to 72 hours releases the constituent amino acids of a peptide.

John B. Fenn (1917–)

Koichi Tanaka (1959–)

▲ John B. Fenn and Koichi Tanaka were awarded the Nobel Prize in Chemistry in 2002 "for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules."



▲ **Figure 3.16**
Amino acid treated with phenylisothiocyanate (PITC). The α-amino group of an amino acid reacts with phenylisothiocyanate to give a phenylthiocarbamoyl–amino acid (PTC–amino acid).

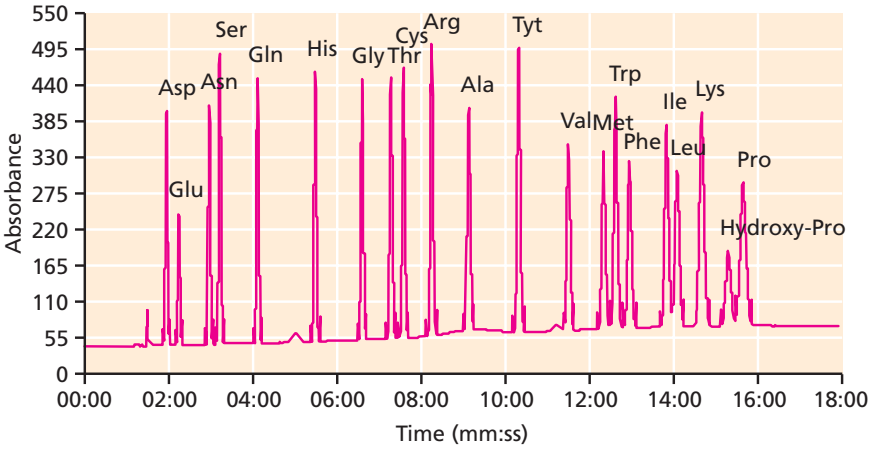


Figure 3.17 ►
HPLC separation of amino acids. Amino acids obtained from the enzymatic hydrolysis of a protein are treated with o-phthalaldehyde and separated by HPLC.

The frequency of amino acids in proteins is correlated with the number of codons for each amino acid (Section 22.1)

instead of using acid hydrolysis. The free amino acids are then attached to a chemical that absorbs light in the ultraviolet and the derivatized amino acids are analyzed by HPLC (Figure 3.17).

Using various analytical techniques the complete amino acid compositions of many proteins have been determined. Dramatic differences in composition have been found, illustrating the tremendous potential for diversity based on different combinations of the 20 amino acids.

The amino acid composition (and sequence) of proteins can also be determined from the sequence of its gene. In fact, these days it is often much easier to clone and sequence DNA than it is to purify and sequence a protein. Table 3.3 shows the average frequency of amino acid residues in more than 1000 different proteins whose sequences are deposited in protein databases. The most common amino acids are leucine, alanine, and glycine, followed by serine, valine, and glutamate. Tryptophan, cysteine, and histidine are the least abundant amino acids in typical proteins.

If you know the amino acid composition of a protein you can calculate the molecular weight using the molecular weights of the amino acids in Table 3.4. Be sure to subtract the molecular weight of one water molecule for each peptide bond (Section 3.5). You can get a rough estimate of the molecular weight of a protein by using the average molecular weight of a residue (= 110). Thus, a protein of 650 amino acid residues has an approximate relative molecular mass of 71,500 (M_r = 71,500).

Table 3.3 Amino acid compositions of proteins

Amino acid	Frequency in proteins (%)
Highly hydrophobic	
Ile (I)	5.2
Val (V)	6.6
Leu (L)	9.0
Phe (F)	3.9
Met (M)	2.4
Less hydrophobic	
Ala (A)	8.3
Gly (G)	7.2
Cys (C)	1.7
Trp (W)	1.3
Tyr (Y)	3.2
Pro (P)	5.1
Thr (T)	5.8
Ser (S)	6.9
Highly hydrophilic	
Asn (N)	4.4
Gln (Q)	4.0
Acidic	
Asp (D)	5.3
Glu (E)	6.2
Basic	
His (H)	2.2
Lys (K)	5.7
Arg (R)	5.7

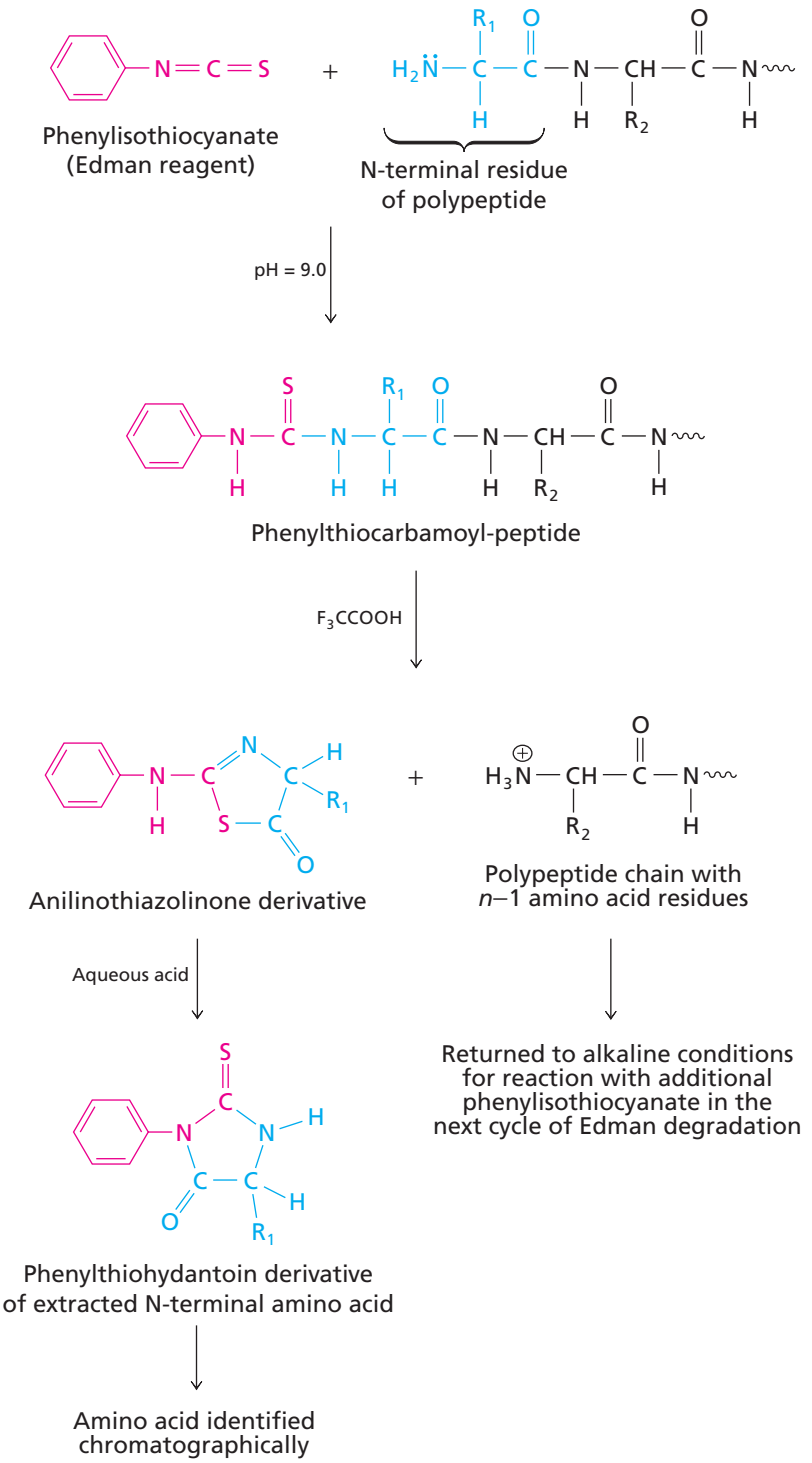
3.9 Determining the Sequence of Amino Acid Residues

Amino acid analysis provides information on the composition of a protein but not its primary structure (sequence of residues). In 1950, Pehr Edman developed a technique that permits removal and identification of one residue at a time from the N-terminus of a protein. The **Edman degradation procedure** involves treating a protein at pH 9.0 with PITC, also known as the Edman reagent. (Recall that PITC can also be used in the measurement of free amino acids as shown in Figure 3.16.) PITC reacts with the free N-terminus of the chain to form a phenylthiocarbamoyl derivative, or PTC-peptide (Figure 3.18, on the next page). When the PTC-peptide is treated with an anhydrous acid, such as trifluoroacetic acid the peptide bond of the N-terminal residue is selectively cleaved releasing an anilinothiazolinone derivative of the residue. This derivative can be extracted with an organic solvent, such as butyl chloride, leaving the remaining peptide in the aqueous phase. The unstable anilinothiazolinone derivative is then treated with aqueous acid which converts it to a stable phenylthiohydantoin derivative of the amino acid that had been the N-terminal residue (PTH–amino acid). The polypeptide chain in the aqueous phase, now one residue shorter (residue 2 of the original protein is now the N-terminus), can be adjusted back to pH 9.0 and treated again with PITC. The entire procedure can be repeated serially using an automated instrument known as a sequenator. Each cycle yields a PTH–amino acid that can be identified chromatographically, usually by HPLC.

The yield of the Edman degradation procedure under carefully controlled conditions approaches 100% and a few picomoles of sample protein can yield sequences of 30 residues or more before further measurement is obscured by the increasing concentration of unrecovered sample from previous cycles of the procedure. For example, if the Edman degradation procedure had an efficiency of 98% the cumulative yield at the 30th cycle would be 0.98^{30}, or 0.55. In other words, only about half of the PTH–amino acids generated in the 30th cycle would be derived from the 30th residue from the N-terminus.

Table 3.4 Molecular weights of amino acids

Amino acid	M_r
Ala(A)	89
Arg(R)	174
Asn(N)	132
Asp(D)	133
Cys(C)	121
Gln(O)	146
Glu(E)	147
Gly(G)	75
His(H)	155
He(I)	131
Leu(L)	131
Lys(K)	146
Met(M)	149
Phe(F)	165
Pro(P)	115
Ser(S)	105
Thr(T)	119
Trp(W)	204
Tyr(Y)	181
Val(V)	117



◀ **Figure 3.18**
Edman degradation procedure. The N-terminal residue of a polypeptide chain reacts with phenylisothiocyanate to give a phenylthiocarbamoyl–peptide. Treating this derivative with trifluoroacetic acid (F_3CCOOH) releases an anilinothiazolinone derivative of the N-terminal amino acid residue. The anilinothiazolinone is extracted and treated with aqueous acid, which rearranges the derivative to a stable phenylthiohydantoin derivative that can then be identified chromatographically. The remainder of the polypeptide chain, whose new N-terminal residue was formerly in the second position, is subjected to the next cycle of Edman degradation.

3.10 Protein Sequencing Strategies

Most proteins contain too many residues to be completely sequenced by Edman degradation proceeding only from the N-terminus. Therefore, proteases (enzymes that catalyze the hydrolysis of peptide bonds in proteins) or certain chemical reagents are used to selectively cleave some of the peptide bonds of a protein. The smaller peptides formed are then isolated and subjected to sequencing by the Edman degradation procedure.

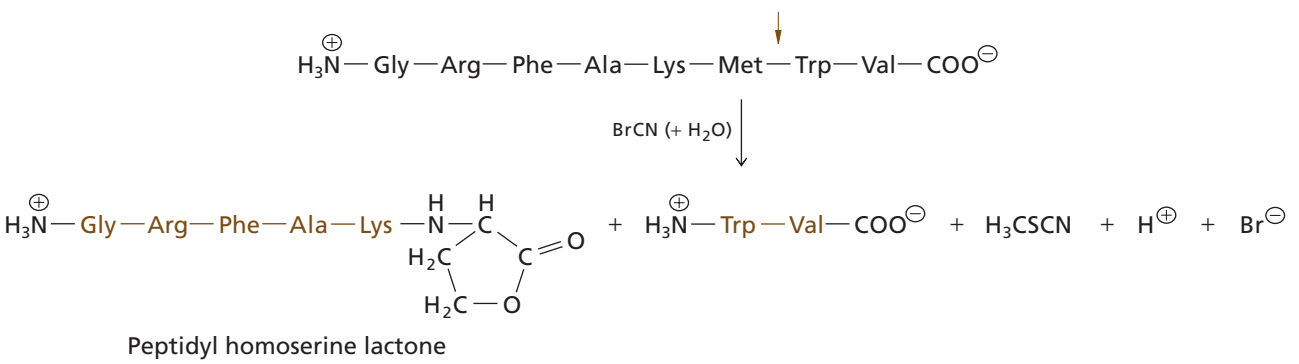
The chemical reagent cyanogen bromide (CNBr) reacts specifically with methionine residues to produce peptides with C-terminal homoserine lactone residues and new N-terminal residues (Figure 3.19). Since most proteins contain relatively few methionine residues treatment with CNBr usually produces only a few peptide fragments. For example, reaction of CNBr with a polypeptide chain containing three internal methionine residues should generate four peptide fragments. Each fragment can then be sequenced from its N-terminus.

Many different proteases can be used to generate fragments for protein sequencing. For example, trypsin specifically catalyzes the hydrolysis of peptide bonds on the carbonyl side of lysine and arginine residues both of which bear positively charged side chains (Figure 3.20a). *Staphylococcus aureus* V8 protease catalyzes the cleavage of peptide bonds on the carbonyl side of negatively charged residues (glutamate and aspartate); under appropriate conditions (50 mM ammonium bicarbonate), it cleaves only glutamyl bonds. Chymotrypsin, a less specific protease, preferentially catalyzes the hydrolysis of peptide bonds on the carbonyl side of uncharged residues with aromatic or bulky hydrophobic side chains, such as phenylalanine, tyrosine, and tryptophan (Figure 3.20b).

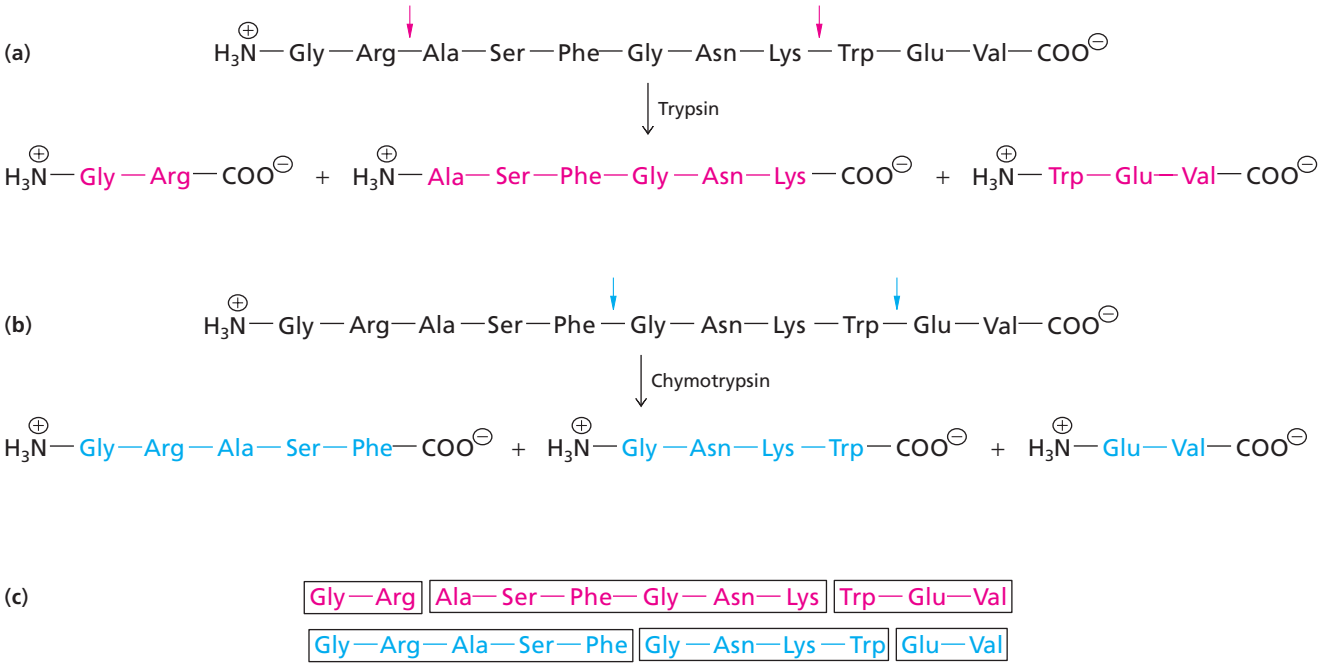
By judicious application of cyanogen bromide, trypsin, *S. aureus* V8 protease, and chymotrypsin to individual samples of a large protein one can generate many peptide fragments of various sizes. These fragments can then be separated and sequenced by Edman degradation. In the final stage of sequence determination the amino acid sequence of a large polypeptide chain can be deduced by lining up matching sequences of overlapping peptide fragments as illustrated in Figure 3.20c. When referring to an amino acid residue whose position in the sequence is known it is customary to follow the residue abbreviation with its sequence number. For example, the third residue of the peptide shown in Figure 3.20 is called Ala-3.

The process of generating and sequencing peptide fragments is especially important in obtaining information about the sequences of proteins whose N-termini are blocked. For example, the N-terminal α-amino groups of many bacterial proteins are formylated and do not react at all when subjected to the Edman degradation procedure. Peptide fragments with unblocked N-termini can be produced by selective cleavage and then separated and sequenced so that at least some of the internal sequence of the protein can be obtained.

For proteins that contain disulfide bonds, the complete covalent structure is not fully resolved until the positions of the disulfide bonds have been established. The positions of the disulfide cross-links can be determined by fragmenting the intact protein, isolating the peptide fragments, and determining which fragments contain cystine residues. The task of determining the positions of the cross-links becomes quite complicated when the protein contains several disulfide bonds.



▼ **Figure 3.19**
Protein cleavage by cyanogen bromide (CNBr). Cyanogen bromide cleaves polypeptide chains at the C-terminal side of methionine residues. The reaction produces a peptidyl homoserine lactone and generates a new N-terminus.



▲ **Figure 3.20**
Cleavage and sequencing of an oligopeptide. **(a)** Trypsin catalyzes cleavage of peptides on the carbonyl side of the basic residues arginine and lysine. **(b)** Chymotrypsin catalyzes cleavage of peptides on the carbonyl side of uncharged residues with aromatic or bulky hydrophobic side chains, including phenylalanine, tyrosine, and tryptophan. **(c)** By using the Edman degradation procedure to determine the sequence of each fragment (highlighted in boxes) and then lining up the matching sequences of overlapping fragments, one can determine the order of the fragments and thus deduce the sequence of the entire oligopeptide.

Deducing the amino acid sequence of a particular protein from the sequence of its gene (Figure 3.21) overcomes some of the technical limitations of direct analytical techniques. For example, the amount of tryptophan can be determined and aspartate and asparagine residues can be distinguished because they are encoded by different codons. However, direct sequencing of proteins is still important since it is the only way of determining whether modified amino acids are present or whether amino acid residues have been removed after protein synthesis is complete.

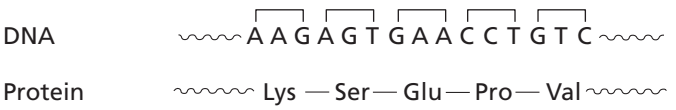
Researchers frequently want to identify a particular unknown protein. Let's say you have displayed human serum proteins on an SDS gel and you note the presence of a protein band at 67 KDa. What is that protein? Two recent developments have made the job of identifying unknown proteins much easier—sensitive mass spectrometry and genome sequences. Let's see how they work.

First, you isolate the protein by cutting out the unknown protein band and eluting the 67 KD protein. The next step is to digest the protein with a protease that cuts at specific sites. Let's say you choose trypsin, an enzyme that cleaves the peptide bond following arginine (R) or lysine (K) residues. After digestion with trypsin you end up with several dozen peptide fragments all of which end with arginine or lysine.

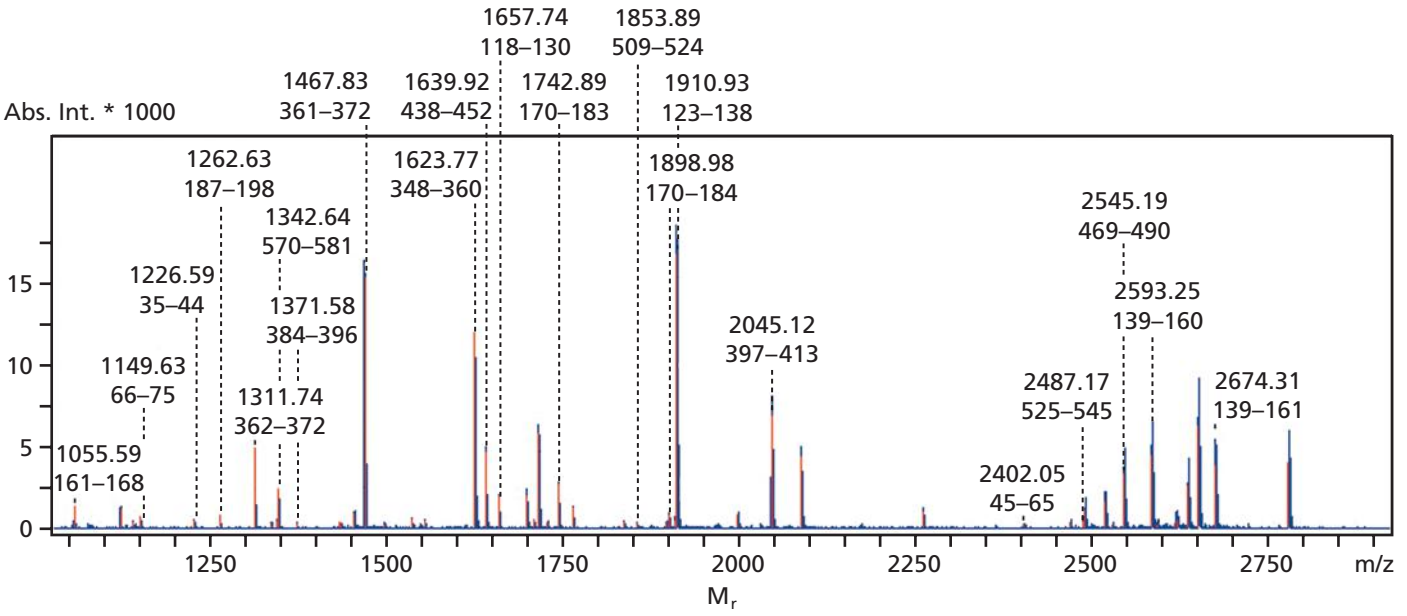
Next, you subject the peptide mixture to mass spectrometry choosing a method such as MALDI–TOF where the precise molecular weights of the peptides can be determined. The resulting spectrum is shown in Figure 3.22. You now have a "fingerprint" of the unknown protein corresponding to the molecular weights of all the trypsin digestion products.

In many labs the technique of chemical sequencing using Edman degradation has been replaced by methods using the mass spectrometer. If you wanted to determine the sequences of each peptide shown in Figure 3.22 your next step would be to fragment each peptide into various sized pieces and measure the precise molecular weight of each fragment in the mass spectrometer.

The data can be used to determine the sequence of the peptide. For example, take the tryptic peptide of M_r = 1226.59 shown in Figure 3.22. One of the large pieces produced by fragmenting this peptide has a molecular weight of 1079.5. The difference



◀ **Figure 3.21**
Sequences of DNA and protein. The amino acid sequence of a protein can be deduced from the sequence of nucleotides in the corresponding gene. A sequence of three nucleotides specifies one amino acid. A, C, G, and T represent the nucleotide residues of DNA.



▲ **Figure 3.22**
Tryptic fingerprint of a 67 kDa serum protein. The numbers over each peak are the mass of the fragment. The number below each mass refer to the residues in Figure 3.23 (Adapted from Detlevuvkaw, Wikipedia entry on peptide mass fingerprinting)

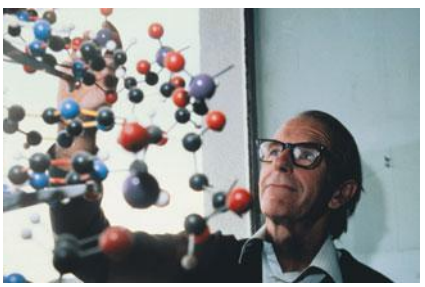
corresponds to a Phe (F) residue (1226.6 − 1079.5 = 147.1), meaning that Phe (F) is the residue at one end of the tryptic peptide. Another large fragment might have a molecular weight of 1098.5 and the difference (1226.6 − 1098.1) is the exact molecular weight of a Lys (K) residue. Thus, Lys (K) is the residue at the other end of the peptide. This has to be the C-terminal end since you know that trypsin cleaves after lysine or arginine residues. You can get the exact sequence of the peptide by analyzing the masses of all fragments in this manner. One of them will have a molecular weight of 258.0 and that is almost certainly the dipeptide Glu-Glu (EE). (The actual analysis is a bit more complicated than this but the principle is the same.)

But it's often not necessary to do the second mass spectrometry analysis in order to identify an unknown protein. Since your unkown protein is from a species whose genome has been sequenced you can simply compare the tryptic fingerprint to the predicted fingerprints of all the proteins encoded by all the genes in the genome. The database consists of a collection of hypothetical peptides produced by analyzing the amino acid sequence of each protein including proteins of unknown function that are known only from their sequence. In most cases your collection of peptide masses from the unknown protein will match only one protein from one of the genes in the database.

In this case, the match is to human serum albumin, a well known serum protein (Figure 3.23). The masses of several of the peptides correspond to the predicted masses of the peptides identified in red in the sequence. Take, for example, the peptide of M_r = 1226.59 in the output from the tryptic fingerprint. This is exactly the predicted mass of the peptide from residues 35–44 (FKDLGEENFK). (Note that the first trypsin cleavage site follows the arginine residue at position 34 and the second cleavage site is after the lysine residue at position 44.)

A single match is not sufficient to identify an unknown protein. In the example shown here there are 21 peptide fragments that match the amino acid sequence of human serum albumin and this is more than sufficient to uniquely identify the protein.

In 1953, Frederick Sanger was the first scientist to determine the complete sequence of a protein (insulin). In 1958, he was awarded a Nobel Prize for this work. Twenty-two years later, Sanger won a second Nobel Prize for pioneering the sequencing of nucleic acids. Today we know the amino acid sequences of thousands of different proteins. These sequences not only reveal details of the structure of individual proteins but also allow researchers to identify families of related proteins and to predict the three-dimensional structure, and sometimes the function, of newly discovered proteins.



▲ **Frederick Sanger (1918–)** Sanger won the Nobel Prize in Chemistry in 1958 for his work on sequencing proteins. He was awarded a second Nobel Prize in Chemistry in 1980 for developing methods of sequencing DNA.

▲ **Figure 3.23**
The sequence of human serum albumin. Red residues highlight predicted tryptic peptides and the ones identified in the tryptic fingerprint (Figure 3.22) are underlined.

3.11 Comparisons of the Primary Structures of Proteins Reveal Evolutionary Relationships

In many cases workers have obtained sequences of the same protein from a number of different species. The results show that closely related species contain proteins with very similar amino acid sequences and that proteins from distantly related species are much less similar in sequence. The differences reflect evolutionary change from a common ancestral protein sequence. As more and more sequences were determined it soon became clear that one could construct a tree of similarities and this tree closely resembled the phylogenetic trees constructed from morphological comparisons and the fossil record. The evidence from molecular data was producing independent confirmation of the history of life.

The first sequence-based trees were published almost 50 years ago. One of the earliest examples was the tree for cytochrome *c*—a single polypeptide chain of approximately 104 residues. It provides us with an excellent example of evolution at the molecular level. Cytochrome *c* is found in all aerobic organisms and the protein sequences from distantly related species, such as mammals and bacteria, are similar enough to confidently conclude that the proteins are **homologous**. (Different proteins and genes are defined as homologues if they have descended from a common ancestor. The evidence for homology is based on sequence similarity.)

The first step in revealing evolutionary relationships is to align the amino acid sequences of proteins from a number of species. Figure 3.24 shows an example of such an alignment for cytochrome *c*. The alignment reveals a remarkable conservation of residues at certain positions. For example, every sequence contains a proline at position 30 and a methionine at position 80. In general, conserved residues contribute to the structural stability of the protein or are essential for its function.

There is selection against any amino acid substitutions at these invariant positions. A limited number of substitutions are observed at other sites. In most cases, the allowed substitutions are amino acid residues with similar properties. For example, position 20 can be occupied by leucine, isoleucine, or valine—these are all hydrophobic residues. Similarly, many sites can be occupied by a number of different polar residues. Some positions are highly variable—residues at these sites contribute very little to the structure and function of the protein. The majority of observed amino acid substitutions in homologous proteins are neutral with respect to natural selection. The fixation of substitutions at such positions during evolution is due to random genetic drift and the phylogenetic tree represents proteins that have the same fuction even though they have different amino acid sequences.

The function of cytochrome *c* is described in Section 14.7.

KEY CONCEPT

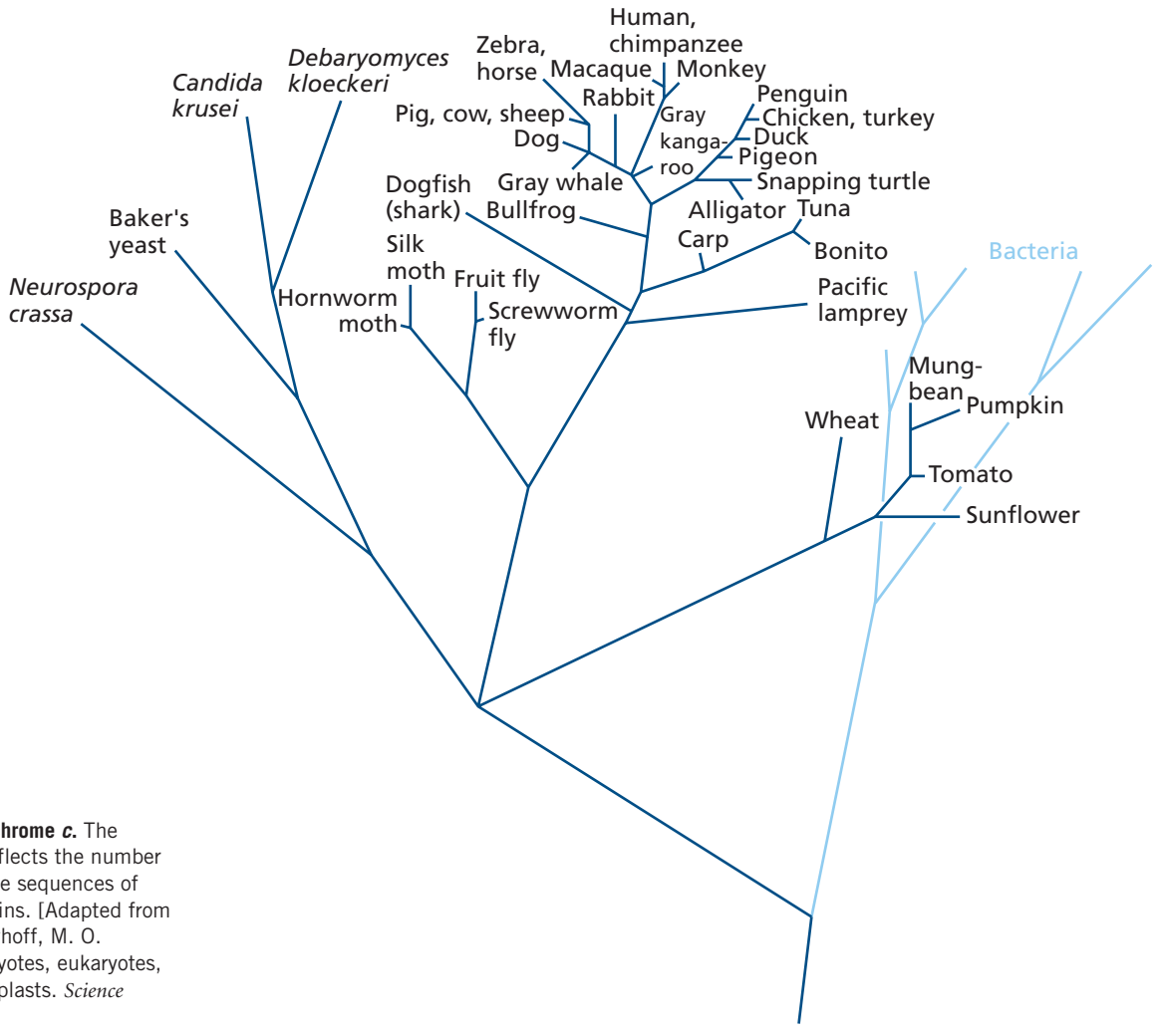
***Homology* is a conclusion that is based on evidence such as sequence similarity. Homologous proteins descend from a common ancestor. There are degrees of sequence similarity (e.g., 75% identity), but homology is an all-or-nothing conclusion. Something is either homologous or it isn't.**

Figure 3.24 ►
Cytochrome *c* sequences. The sequences of cytochrome *c* proteins from various species are aligned to show their similarities. In some cases, gaps (signified by hyphens) have been introduced to improve the alignment. The gaps represent deletions and insertions in the genes that encode these proteins. For some species, additional residues at the ends of the sequence have been omitted. Hydrophobic residues are blue and polar residues are red.

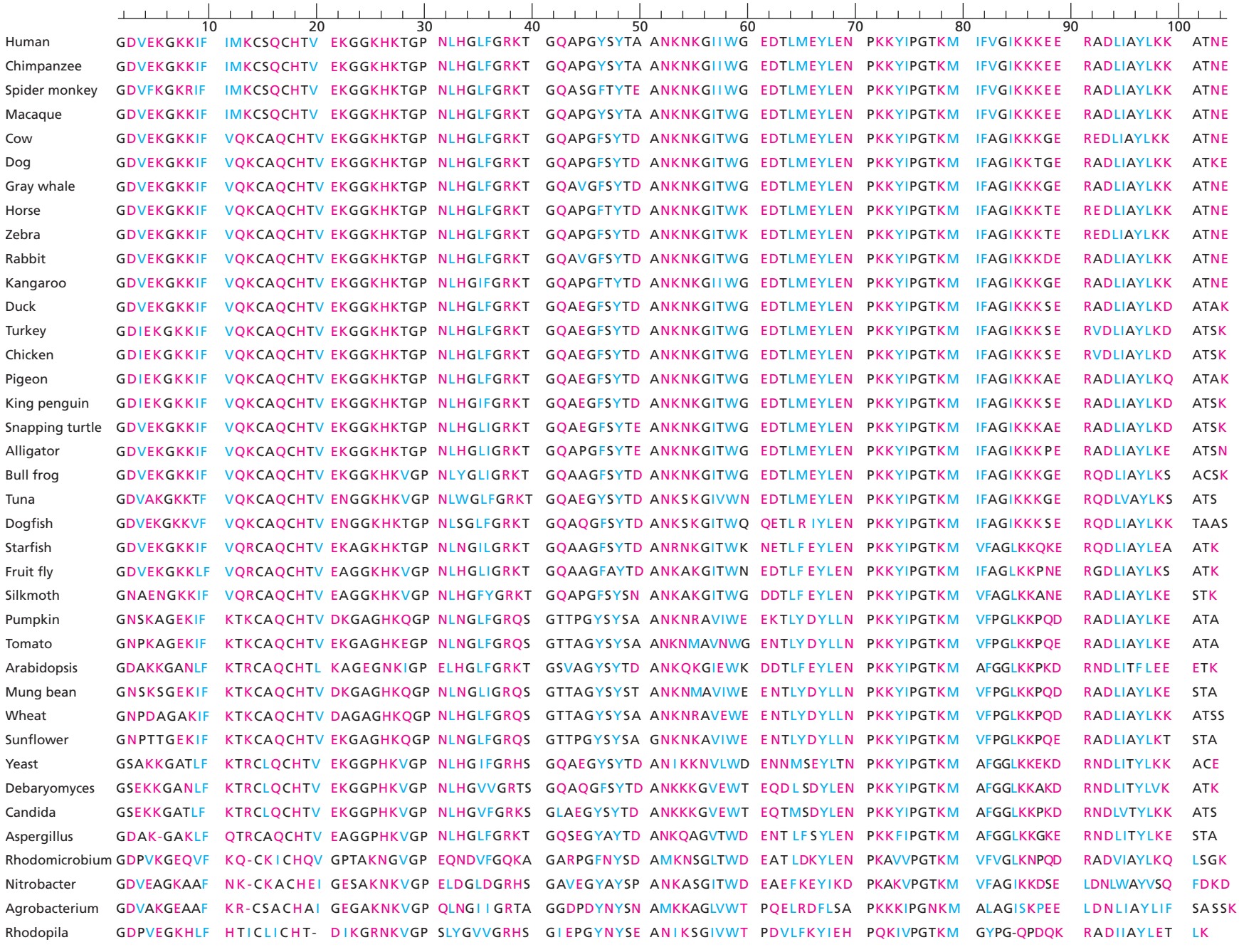
The cytochrome *c* sequences of humans and chimpanzees are identical. This is a reflection of their close evolutionary relationship. The monkey and macaque sequences are very similar to the human and chimpanzee sequences as expected since all four species are primates. Similarly, the sequences of the plant cytochrome *c* molecules resemble each other much more than they resemble any of the other sequences.

Figure 3.25 illustrates the similarities between cytochrome *c* sequences in different species by depicting them as a tree whose branches are proportional in length to the number of differences in the amino acid sequences of the protein. Species that are closely related cluster together on the same branches of the tree because their proteins are very similar. At great evolutionary distances the number of differences may be very large. For example, the bacterial sequences differ substantially from the eukaryotic sequences reflecting divergence from a common ancestor that lived several billion years ago. The tree clearly reveals the three main kingdoms of eukaryotes—fungi, animals, and plants. (Protist sequences are not included in this tree in order to make it less complicated.)

Note that every species has changed since divurging from their common ancastor.



► Figure 3.25
Phylogenetic tree for cytochrome *c*. The length of the branches reflects the number of differences between the sequences of many cytochrome *c* proteins. [Adapted from Schwartz, R. M., and Dayhoff, M. O. (1978). Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* 199:395–403.]

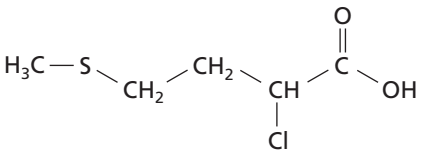

10 20 30 40 50 60 70 80 90 100
Human GDVEKGKKIF IMKCSQCHTV EKGGKHKTGP NLHGLFGRKT GQAPGYSYTA ANKNKGIIWG EDTLMEYLEN PKKYIPGTKM IFVGIKKKEE RADLIAYLKK ATNE
Chimpanzee GDVEKGKKIF IMKCSQCHTV EKGGKHKTGP NLHGLFGRKT GQAPGYSYTA ANKNKGIIWG EDTLMEYLEN PKKYIPGTKM IFVGIKKKEE RADLIAYLKK ATNE
Spider monkey GDVFKGKRIF IMKCSQCHTV EKGGKHKTGP NLHGLFGRKT GQASGFTYTE ANKNKGIIWG EDTLMEYLEN PKKYIPGTKM IFVGIKKKEE RADLIAYLKK ATNE
Macaque GDVEKGKKIF IMKCSQCHTV EKGGKHKTGP NLHGLFGRKT GQAPGYSYTA ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFVGIKKKEE RADLIAYLKK ATNE
Cow GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAPGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKGE REDLIAYLKK ATNE
Dog GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAPGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKTGE RADLIAYLKK ATKE
Gray whale GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAVGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKGE RADLIAYLKK ATNE
Horse GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAPGFTYTD ANKNKGITWK EDTLMEYLEN PKKYIPGTKM IFAGIKKKTE REDLIAYLKK ATNE
Zebra GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAPGFSYTD ANKNKGITWK EDTLMEYLEN PKKYIPGTKM IFAGIKKKTE REDLIAYLKK ATNE
Rabbit GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAVGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKDE RADLIAYLKK ATNE
Kangaroo GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGIFGRKT GQAPGFTYTD ANKNKGIIWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKGE RADLIAYLKK ATNE
Duck GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAEGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKSE RADLIAYLKD ATAK
Turkey GDIEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAEGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKSE RVDLIAYLKD ATSK
Chicken GDIEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAEGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKSE RVDLIAYLKD ATSK
Pigeon GDIEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAEGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKAE RADLIAYLKQ ATAK
King penguin GDIEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGIFGRKT GQAEGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKSE RADLIAYLKD ATSK
Snapping turtle GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLIGRKT GQAEGFSYTE ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKAE RADLIAYLKD ATSK
Alligator GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLIGRKT GQAPGFSYTE ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKPE RADLIAYLKE ATSN
Bull frog GDVEKGKKIF VQKCAQCHTV EKGGKHKVGP NLYGLIGRKT GQAAGFSYTD ANKNKGITWG EDTLMEYLEN PKKYIPGTKM IFAGIKKKGE RQDLIAYLKS ACSK
Tuna GDVAKGKKTF VQKCAQCHTV ENGGKHKVGP NLWGLFGRKT GQAEGYSYTD ANKSKGIVWN EDTLMEYLEN PKKYIPGTKM IFAGIKKKGE RQDLVAYLKS ATS
Dogfish GDVEKGKKVF VQKCAQCHTV ENGGKHKTGP NLSGLFGRKT GQAQGFSYTD ANKSKGITWQ QETLRIYLEN PKKYIPGTKM IFAGIKKKSE RQDLIAYLKK TAAS
Starfish GDVEKGKKIF VQRCAQCHTV EKAGKHKTGP NLNGILGRKT GQAAGFSYTD ANRNKGITWK NETLFEYLEN PKKYIPGTKM VFAGLKKQKE RQDLIAYLEA ATK
Fruit fly GDVEKGKKLF VQRCAQCHTV EAGGKHKVGP NLHGLIGRKT GQAAGFAYTD ANKAKGITWN EDTLFEYLEN PKKYIPGTKM IFAGLKKPNE RGDLIAYLKS ATK
Silkmoth GNAENGKKIF VQRCAQCHTV EAGGKHKVGP NLHGFYGRKT GQAPGFSYSN ANKAKGITWG DDTLFEYLEN PKKYIPGTKM VFAGLKKANE RADLIAYLKE STK
Pumpkin GNSKAGEKIF KTKCAQCHTV DKGAGHKQGP NLNGLFGRQS GTTPGYSYSA ANKNRAVIWE EKTLYDYLLN PKKYIPGTKM VFPGLKKPQD RADLIAYLKE ATA
Tomato GNPKAGEKIF KTKCAQCHTV EKGAGHKEGP NLNGLFGRQS GTTAGYSYSA ANKNMAVNWG ENTLYDYLLN PKKYIPGTKM VFPGLKKPQE RADLIAYLKE ATA
Arabidopsis GDAKKGANLF KTRCAQCHTL KAGEGNKIGP ELHGLFGRKT GSVAGYSYTD ANKQKGIEWK DDTLFEYLEN PKKYIPGTKM AFGGLKKPKD RNDLITFLEE ETK
Mung bean GNSKSGEKIF KTKCAQCHTV DKGAGHKQGP NLNGLIGRQS GTTAGYSYST ANKNMAVIWE ENTLYDYLLN PKKYIPGTKM VFPGLKKPQD RADLIAYLKE STA
Wheat GNPDAGAKIF KTKCAQCHTV DAGAGHKQGP NLHGLFGRQS GTTAGYSYSA ANKNRAVEWE ENTLYDYLLN PKKYIPGTKM VFPGLKKPQD RADLIAYLKK ATSS
Sunflower GNPTTGEKIF KTKCAQCHTV EKGAGHKQGP NLNGLFGRQS GTTPGYSYSA GNKNKAVIWE ENTLYDYLLN PKKYIPGTKM VFPGLKKPQE RADLIAYLKT STA
Yeast GSAKKGATLF KTRCLQCHTV EKGGPHKVGP NLHGIFGRHS GQAEGYSYTD ANIKKNVLWD ENNMSEYLTN PKKYIPGTKM AFGGLKKEKD RNDLITYLKK ACE
Debaryomyces GSEKKGANLF KTRCLQCHTV EKGGPHKVGP NLHGVVGRTS GQAQGFSYTD ANKKKGVEWT EQDLSDYLEN PKKYIPGTKM AFGGLKKAKD RNDLITYLVK ATK
Candida GSEKKGATLF KTRCLQCHTV EKGGPHKVGP NLHGVFGRKS GLAEGYSYTD ANKKKGVEWT EQTMSDYLEN PKKYIPGTKM AFGGLKKPKD RNDLVTYLKK ATS
Aspergillus GDAK-GAKLF QTRCAQCHTV EAGGPHKVGP NLHGLFGRKT GQSEGYAYTD ANKQAGVTWD ENTLFSYLEN PKKFIPGTKM AFGGLKKGKE RNDLITYLKE STA
Rhodomicrobium GDPVKGEQVF KQ-CKICHQV GPTAKNGVGP EQNDVFGQKA GARPGFNYSD AMKNSGLTWD EATLDKYLEN PKAVVPGTKM VFVGLKNPQD RADVIAYLKQ LSGK
Nitrobacter GDVEAGKAAF NK-CKACHEI GESAKNKVGP ELDGLDGRHS GAVEGYAYSP ANKASGITWD EAEFKEYIKD PKAKVPGTKM VFAGIKKDSE LDNLWAYVSQ FDKD
Agrobacterium GDVAKGEAAF KR-CSACHAI GEGAKNKVGP QLNGIIGRTA GGDPDYNYSN AMKKAGLVWT PQELRDFLSA PKKKIPGNKM ALAGISKPEE LDNLIAYLIF SASSK
Rhodopila GDPVEGKHLF HTICLICHT- DIKGRNKVGP SLYGVVGRHS GIEPGYNYSE ANIKSGIVWT PDVLFKYIEH PQKIVPGTKM GYPG-QPDQK RADIIAYLET LK

Summary

1. Proteins are made from 20 standard amino acids each of which contains an amino group, a carboxyl group, and a side chain, or R group. Except for glycine, which has no chiral carbon, all amino acids in proteins are of the L configuration.
2. The side chains of amino acids can be classified according to their chemical structures—aliphatic, aromatic, sulfur containing, alcohols, bases, acids, and amides. Some amino acids are further classified as having highly hydrophobic or highly hydrophilic side chains. The properties of the side chains of amino acids are important determinants of protein structure and function.
3. Cells contain additional amino acids that are not used in protein synthesis. Some amino acids can be chemically modified to produce compounds that act as hormones or neurotransmitters. Some amino acids are modified after incorporation into polypeptides.
4. At pH 7, the α-carboxyl group of an amino acid is negatively charged (—$COO^{\ominus}$) and the α-amino group is positively charged (—$NH_3^{\oplus}$). The charges of ionizable side chains depend on both the pH and their pK_a values.
5. Amino acid residues in proteins are linked by peptide bonds. The sequence of residues is called the primary structure of the protein.
6. Proteins are purified by methods that take advantage of the differences in solubility, net charge, size, and binding properties of individual proteins.
7. Analytical techniques such as SDS–PAGE and mass spectrometry reveal properties of proteins such as molecular weight.
8. The amino acid composition of a protein can be determined quantitatively by hydrolyzing the peptide bonds and analyzing the hydrolysate chromatographically.
9. The sequence of a polypeptide chain can be determined by the Edman degradation procedure in which the N-terminal residues are successively cleaved and identified.
10. Proteins with very similar amino acid sequences are homologous—they descend from a common ancestor.
11. A comparison of sequences from different species reveals evolutionary relationships.

Problems

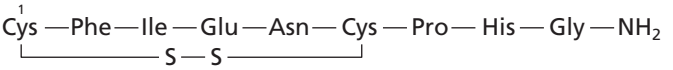
1. Draw and label the stereochemical structure of L-cysteine. Indicate whether it is *R* or *S* by referring to Box 3.2 on page 61.
2. Show that the Fischer projection of the common form of threonine (page 60) corresponds to 2*S*, 3*R*-threonine. Draw and name the three other isomers of threonine.
3. Histamine dihydrochloride is administered to melanoma (skin cancer) patients in combination with anticancer drugs because it makes the cancer cells more receptive to the drugs. Draw the chemical structure of histamine dihydrochloride.
4. Dried fish treated with salt and nitrite has been found to contain the mutagen 2-chloro-4-methylthiobutanoic acid (CMBA). From what amino acid is CMBA derived?



5. For each of the following modified amino acid side chains, identify the amino acid from which it was derived and the type of chemical modification that has occurred.
   (a) —$CH_2OPO_3^{2-}$
   (b) —$CH_2CH(COO^{\ominus})_2$
   (c) —$(CH_2)_4$—NH—C(O)CH_3
6. The tripeptide glutathione (GSH) (γ-Glu-Cys-Gly) serves a protective function in animals by destroying toxic peroxides that are generated during aerobic metabolic processes. Draw the chemical structure of glutathione. Note: The γ symbol indicates that the peptide bond between Glu and Cys is formed between the γ-carboxyl of Glu and the amino group of Cys.
7. Melittin is a 26-residue polypeptide found in bee venom. In its monomeric form, melittin is thought to insert into lipid-rich membrane structures. Explain how the amino acid sequence of melittin accounts for this property.

$$H_3\overset{\oplus}{N}\text{-}\overset{1}{\text{Gly}}\text{-Ile-Gly-Ala-Val-Leu-Lys-Val-Leu-Thr-Gly-Leu}$$
$$\text{Pro-Ala-Leu-Ile-Ser-Trp-Ile-Lys-Arg-Lys-Arg-Gln-}\underset{26}{\text{Gln}}\text{-}NH_2$$

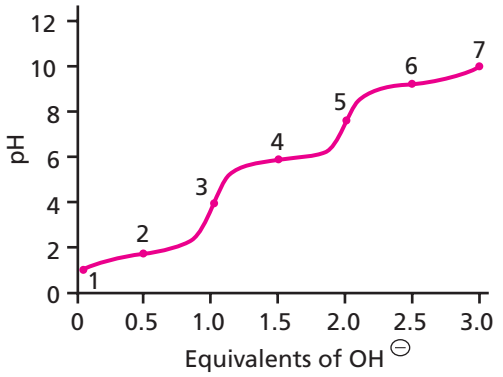
8. Calculate the isoelectric points of (a) arginine and (b) glutamate.
9. Oxytocin is a nonapeptide (a nine-residue peptide) hormone involved in the milk-releasing response in lactating mammals. The sequence of a synthetic version of oxytocin is shown below. What is the net charge of this peptide at (a) pH 2.0, (b) pH 8.5, and (c) pH 10.7? Assume that the ionizable groups have the pK_a values listed in Table 3.2. The disulfide bond is stable at pH 2.0, pH 8.5, and pH 10.7. Note that the C-terminus is amidated.



10. Draw the following structures for compounds that would occur during the Edman degradation procedure: (a) PTC-Leu-Ala, (b) PTH-Ser, (c) PTH-Pro.
11. Predict the fragments that will be generated from the treatment of the following peptide with (a) trypsin, (b) chymotrypsin, and (c) *S. aureus* V8 protease.

Gly-Ala-Trp-Arg-Asp-Ala-Lys-Glu-Phe-Gly-Gln

12. The titration curve for histidine is shown below. The pK_a values are 1.8 (—COOH), 6.0 (side chain), and 9.3 (—$NH_3^{\oplus}$).



(a) Draw the structure of histidine at each stage of ionization.
(b) Identify the points on the titration curve that correspond to the four ionic species.
(c) Identify the points at which the average net charge is +2, +0.5 and −1.
(d) Identify the point at which the pH equals the pK_a of the side chain.
(e) Identify the point that indicates complete titration of the side chain.
(f) In what pH ranges would histidine be a good buffer?

13. You have isolated a decapeptide (a 10-residue peptide) called FP, which has anticancer activity. Determine the sequence of the peptide from the following information. (Note that amino acids are separated by commas when their sequence is not known.)

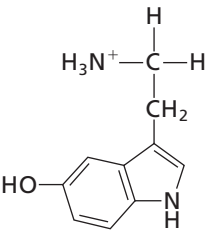
(a) One cycle of Edman degradation of intact FP yields 2 mol of PTH-aspartate per mole of FP.
(b) Treatment of a solution of FP with 2-mercaptoethanol followed by the addition of trypsin yields three peptides with the composition (Ala, Cys, Phe), (Arg, Asp), and (Asp, Cys, Gly, Met, Phe). The intact (Ala, Cys, Phe) peptide yields PTH-cysteine in the first cycle of Edman degradation.
(c) Treatment of 1 mol of FP with carboxypeptidase (which cleaves the C-terminal residue from peptides) yields 2 mol of phenylalanine.
(d) Treatment of the intact pentapeptide (Asp, Cys, Gly, Met, Phe) with CNBr yields two peptides with the composition (homoserine lactone, Asp) and (Cys, Gly, Phe). The (Cys, Gly, Phe) peptide yields PTH-glycine in the first cycle of Edman degradation.

14. A portion of the amino acid sequences for cytochrome *c* from the alligator and bullfrog are given (from Figure 3.24).

	Amino acids 31-50	
Alligator:	NLHGLIGRKT	GQAPGFSYTE
Bullfrog:	NLYGLIGRKT	GQAAGFSYTD

(a) Give an example of a substitution involving similar amino acids.
(b) Give an example of a more radical substitution.

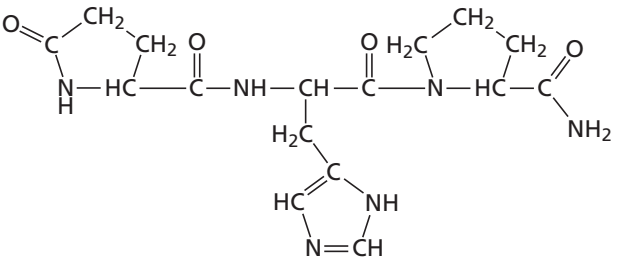
15. Several common amino acids are modified to produce biologically important amines. Serotonin is a biologically important neurotransmitter synthesized in the brain. Low levels of serotonin in the brain have been linked to conditions such as depression, aggression, and hyperactivity. From what amino acid is serotonin derived? Identify the differences in structure between the amino acid and serotonin.



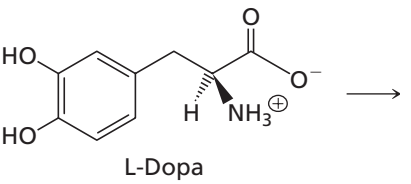
Serotonin

16. The structure of thyrotropin-releasing hormone (TRH) is shown below. TRH is a peptide hormone originally isolated from the extracts of hypothalamus.

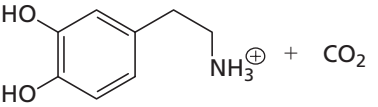
(a) How many peptide bonds are present in TRH?
(b) From what tripeptide is TRH derived?
(c) What result do the modifications have on the charges of the amino and carboxyl-terminal groups?



17. Chirality plays a major role in the development of new pharmaceuticals. People with Parkinson's disease have depleted amounts of dopamine in their brains. In an effort to increase the amount of dopamine in patients, they are given the drug L-dopa which is converted to dopamine in the brain. L-Dopa is marketed in an enantiomerically pure form. (a) Give the *RS* designation for L-dopa. (b) From which amino acid are both L-dopa and dopamine derived?



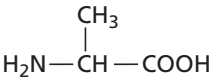
L-Dopa



Dopamine

18. Generations of biochemistry students have encountered a question like the one below on their final exam.

Calculate the approximate concentration of the uncharged form of alanine (see below) in a 0.01M solution of alanine at (a) pH 2.4 (b) pH 6.15 and (c) pH 9.9.



Can you answer the question without peeking at the solution?

19. A solution of 0.01M alanine is adjusted to pH 2.4 by adding NaOH. What is the concentration of the zwitterion in this solution? What would it be if the pH was 4.0?

Selected Readings

General

Creighton, T. E. (1993). *Proteins: Structures and Molecular Principles*, 2nd ed. (New York: W. H. Freeman), pp. 1–48.

Greenstein, J. P., and Winitz, M. (1961). *Chemistry of the Amino Acids* (New York: John Wiley & Sons).

Kreil, G. (1997). D-Amino Acids in Animal Peptides. *Annu. Rev. Biochem.* 66:337–345.

Meister, A. (1965). *Biochemistry of the Amino Acids*, 2nd ed. (New York: Academic Press).

Protein Purification and Analysis

Hearn, M. T. W. (1987). General strategies in the separation of proteins by high-performance liquid chromatographic methods. *J. Chromatogr.* 418:3–26.

Mann, M., Hendrickson, R.C., and Pandry, A. (2001) Analysis of Proteins and Proteomes by Mass Spectrometry. *Annu. Rev. Biochem.* 70:437–473.

Sherman, L. S., and Goodrich, J. A. (1985). The historical development of sodium dodecyl sulphate–polyacrylamide gel electrophoresis. *Chem. Soc. Rev.* 14:225–236.

Stellwagen, E. (1990). Gel filtration. *Methods Enzymol.* 182:317–328.

Amino Acid Analysis and Sequencing

Doolittle, R. F. (1989). Similar amino acid sequences revisited. *Trends Biochem. Sci.* 14:244–245.

Han, K. -K., Belaiche, D., Moreau, O., and Briand, G. (1985). Current developments in stepwise Edman degradation of peptides and proteins. *Int. J. Biochem.* 17:429–445.

Hunkapiller, M. W., Strickler, J. E., and Wilson, K. J. (1984). Contemporary methodology for protein structure determination. *Science* 226:304–311.

Ozols, J. (1990). Amino acid analysis. *Methods Enzymol.* 182:587–601.

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

Proteins: Three-Dimensional Structure and Function

We saw in the previous chapter that a protein can be described as a chain of amino acids joined by peptide bonds in a specific sequence. However, polypeptide chains are not simply linear but are also folded into compact shapes that contain coils, zigzags, turns, and loops. Over the last 50 years the three-dimensional shapes, or conformations, of thousands of proteins have been determined. A **conformation** is a spatial arrangement of atoms that depends on the rotation of a bond or bonds. The *conformation* of a molecule, such as a protein, can change without breaking covalent bonds whereas the various **configurations** of a molecule can be changed only by breaking and re-forming covalent bonds. (Recall that the L and D forms of amino acids represent different configurations.) Each protein has an astronomical number of potential conformations. Since every amino acid residue has a number of possible conformations and since there are many residues in a protein. Nevertheless, under physiological conditions most proteins fold into a single stable shape known as its native conformation. A number of factors constrain rotation around the covalent bonds in a polypeptide chain in its native conformation. These include the presence of hydrogen bonds and other weak interactions between amino acid residues. The biological function of a protein depends on its native three-dimensional conformation.

A protein may be a single polypeptide chain or it may be composed of several polypeptide chains bound to each other by weak interactions. As a general rule, each polypeptide chain is encoded by a single gene although there are some interesting exceptions to this rule. The size of genes and the polypeptides they encode can vary by more than an order of magnitude. Some polypeptides contain only 100 amino acid residues with a relative molecular mass of about 11,000 (M_r = 11,000) (Recall that the average relative molecular mass of an amino acid residue of a protein is 110.) On the other hand, some very large polypeptide chains contain more than 2000 amino acid residues (M_r = 220,000).

From the intensity of the spots near the centre, we can infer that the protein molecules are relatively dense globular bodies, perhaps joined together by valency bridges, but in any event separated by relatively large spaces which contain water. From the intensity of the more distant spots, it can be inferred that the arrangement of atoms inside the protein molecule is also of a perfectly definite kind, although without the periodicities characterising the fibrous proteins. The observations are compatible with oblate spheroidal molecules of diameters about 25 A. and 35 A., arranged in hexagonal screw-axis. . . . At this stage, such ideas are merely speculative, but now that a crystalline protein has been made to give X-ray photographs, it is clear that we have the means of checking them and, by examining the structure of all crystalline proteins, arriving at a far more detailed conclusion about protein structure than previous physical or chemical methods have been able to give.

—Dorothy Crowfoot Hodgkin (1934)

Top: Bighorn sheep. The skin, wool, and horns are composed largely of fibrous proteins.

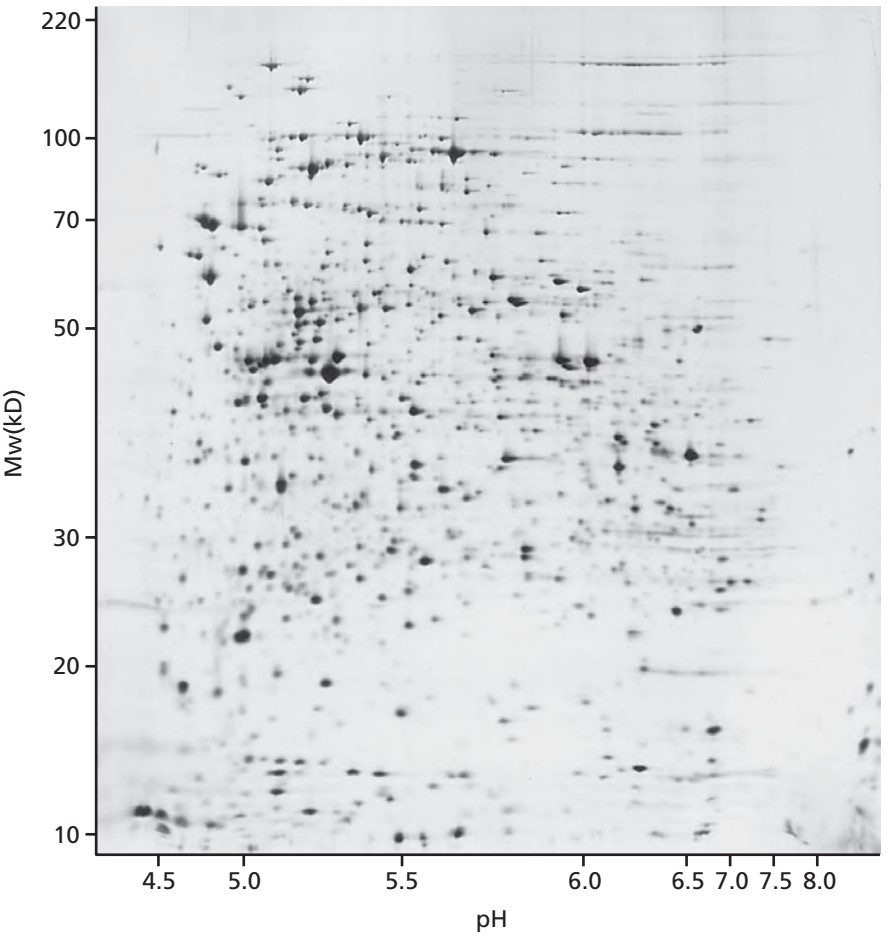
In some species, the size and sequence of every polypeptide can be determined from the sequence of the genome. There are about 4000 different polypeptides in the bacterium *Escherichia coli* with an average size of about 300 amino acid residues (M_r = 33,000). The fruit fly *Drosophila melanogaster* contains about 14,000 different polypeptides with an average size about the same as that in bacteria. Humans and other mammals have about 20,000 different polypeptides. The study of large sets of proteins, such as the entire complement of proteins produced by a cell, is part of a field of study called **proteomics**.

Classes of proteins are described in the introduction to Chapter 3, and the various classes of enzymes are described in Section 5.1.

Proteins come in a variety of shapes. Many are water-soluble, compact, roughly spherical macromolecules whose polypeptide chains are tightly folded. Such proteins—traditionally called **globular proteins**—characteristically have a hydrophobic interior and a hydrophilic surface. They possess indentations or clefts that specifically recognize and transiently bind other compounds. By selectively binding other molecules these proteins serve as dynamic agents of biological action. Many globular proteins are enzymes—the biochemical catalysts of cells. About 31% of the polypeptides in *E. coli* are classical metabolic enzymes such as those described in the next few chapters. Other proteins include various factors, carrier proteins, and regulatory proteins; 12% of the known proteins in *E. coli* fall into these categories.

The terms *globular proteins* and *fibrous proteins* are rarely used in modern scientific publications. There are many proteins that don't fit into either category.

Polypeptides can also be components of large subcellular or extracellular structures such as ribosomes, flagella and cilia, muscle, and chromatin. **Fibrous proteins** are a particular class of structural proteins that provide mechanical support to cells or organisms. Fibrous proteins are typically assembled into large cables or threads. Examples of fibrous proteins are α-keratin, the major component of hair and nails, and collagen, the major protein component of tendons, skin, bones, and teeth. Other examples of structural proteins include the protein components of viruses, bacteriophages, spores, and pollen.



▸ ***Escherichia* coli proteins.** Proteins from *E. coli* cells are separated by two-dimensional gel electrophoresis. In the first dimension, the proteins are separated by a pH gradient where each protein migrates to its isoelectric point. The second dimension separates proteins by size on an SDS–polyacrylamide gel. Each spot corresponds to a single polypeptide. There are about 4000 different proteins in *E. coli*, but some of them are present in very small quantities and can't be seen on this 2-D gel. This figure is from the Swiss-2D PAGE database. You can visit this site and click on any one of the spots to find out more about a particular protein.

Many proteins are either integral components of membranes or membrane-associated proteins. Membrane proteins account for at least 16% of the polypeptides in *E. coli* and a much higher percentage in eukaryotic cells.

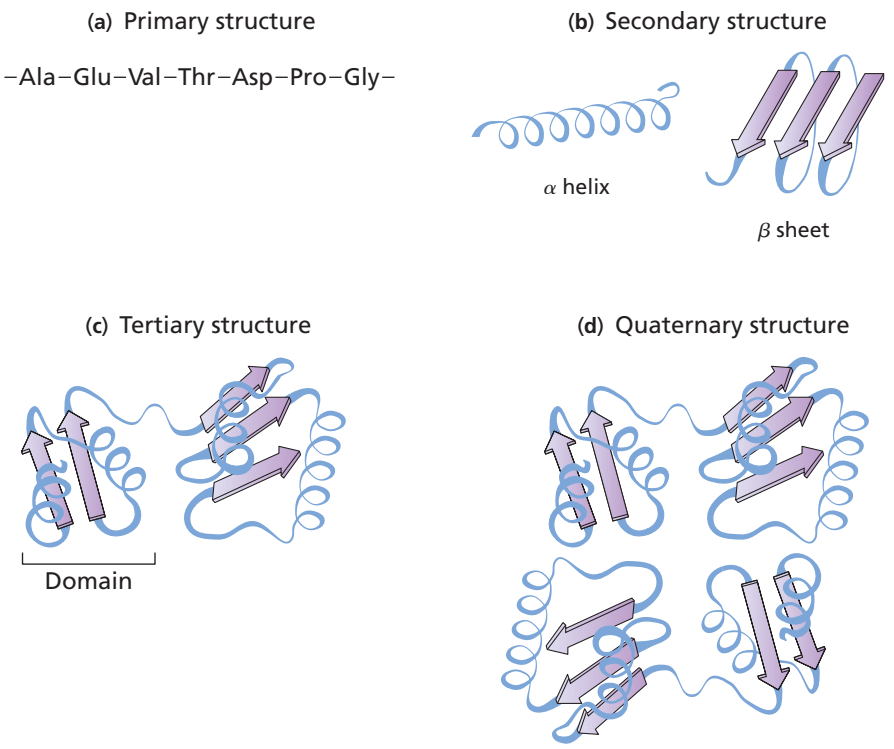
This chapter describes the molecular architecture of proteins. We will explore the conformation of the peptide bond and see that two simple shapes, the α helix and the β sheet, are common structural elements in all classes of proteins. We will describe higher levels of protein structure and discuss protein folding and stabilization. Finally, we will examine how protein structure is related to function using collagen, hemoglobin, and antibodies as examples. Above all, we will learn that proteins have properties beyond those of free amino acids. Chapters 5 and 6 describe the role of proteins as enzymes. The structures of membrane proteins are examined in more detail in Chapter 9 and proteins that bind nucleic acids are covered in Chapters 20 to 22.

4.1 There Are Four Levels of Protein Structure

Individual protein molecules have up to four levels of structure (Figure 4.1). As noted in Chapter 3, **primary structure** describes the linear sequence of amino acid residues in a protein. The three-dimensional structure of a protein is described by three additional levels: secondary structure, tertiary structure, and quaternary structure. The forces responsible for maintaining, or stabilizing, these three levels are primarily noncovalent.

Secondary structure refers to regularities in local conformations maintained by hydrogen bonds between amide hydrogens and carbonyl oxygens of the peptide backbone. The major secondary structures are α helices, β strands, and turns. Cartoons showing the structures of folded proteins usually represent α-helical regions by helices and β strands by broad arrows pointing in the N-terminal to C-terminal direction.

Tertiary structure describes the completely folded and compacted polypeptide chain. Many folded polypeptides consist of several distinct globular units linked by a short stretch of amino acid residues as shown in Figure 4.1c. Such units are called **domains**. Tertiary structures are stabilized by the interactions of amino acid side chains in nonneighboring regions of the polypeptide chain. The formation of tertiary structure brings distant portions of the primary and secondary structures close together.



◀ **Figure 4.1**
Levels of protein structure. (a) The linear sequence of amino acid residues defines the primary structure. **(b)** Secondary structure consists of regions of regularly repeating conformations of the peptide chain such as α helices and β sheets. **(c)** Tertiary structure describes the shape of the fully folded polypeptide chain. The example shown has two domains. **(d)** Quaternary structure refers to the arrangement of two or more polypeptide chains into a multisubunit molecule.

Some proteins possess **quaternary structure**—the association of two or more polypeptide chains into a multisubunit, or oligomeric, protein. The polypeptide chains of an oligomeric protein may be identical or different.

4.2 Methods for Determining Protein Structure

As we saw in Chapter 3, the amino acid sequence of polypeptides (i.e., primary structure) can be determined directly by sequencing the protein or indirectly by sequencing the gene. The usual technique for determining the three-dimensional conformation of a protein is X-ray crystallography. In this technique, a beam of collimated (parallel) X rays is aimed at a crystal of protein molecules. Electrons in the crystal diffract the X rays that are then recorded on film or by an electronic detector (Figure 4.2). Mathematical analysis of the diffraction pattern produces an image of the electron clouds surrounding atoms in the crystal. This electron density map reveals the overall shape of the molecule and the positions of each of the atoms in three-dimensional space. By combining these data with the principles of chemical bonding it is possible to deduce the location of all the bonds in a molecule and hence its overall structure. The technique of X-ray crystallography has developed to the point where it is possible to determine the structure of a protein without precise knowledge of the amino acid sequence. In practice, knowledge of the primary structure makes fitting of the electron density map much easier at the stage where chemical bonds between atoms are determined.

Initially, X-ray crystallography was used to study the simple repeating units of fibrous proteins and the structures of small biological molecules. Dorothy Crowfoot Hodgkin was one of the early pioneers in the application of X-ray crystallography to biological molecules. She solved the structure of penicillin in 1947 and developed many of the techniques used in the study of large proteins. Hodgkin received the Nobel Prize in 1964 for determining the structure of vitamin B_{12} and she later published the structure of insulin.

The chief impediment to determining the three-dimensional structure of an entire protein was the difficulty of calculating atomic positions from the positions and intensities of diffracted X-ray beams. Not surprisingly, the development of X-ray crystallography of macromolecules closely followed the development of computers. By 1962, John C. Kendrew and Max Perutz had elucidated the structures of the proteins myoglobin and hemoglobin, respectively, using large and very expensive computers at Cambridge University in the United Kingdom. Their results provided the first insights into the nature of the tertiary structures of proteins and earned them a Nobel Prize in 1962. Since then, the structures of many proteins have been revealed by X-ray crystallography. In recent years, there have been significant advances in the technology due to the availability of inexpensive high-speed computers and improvements in producing focused beams of X rays. The determination of protein structures is now limited mainly

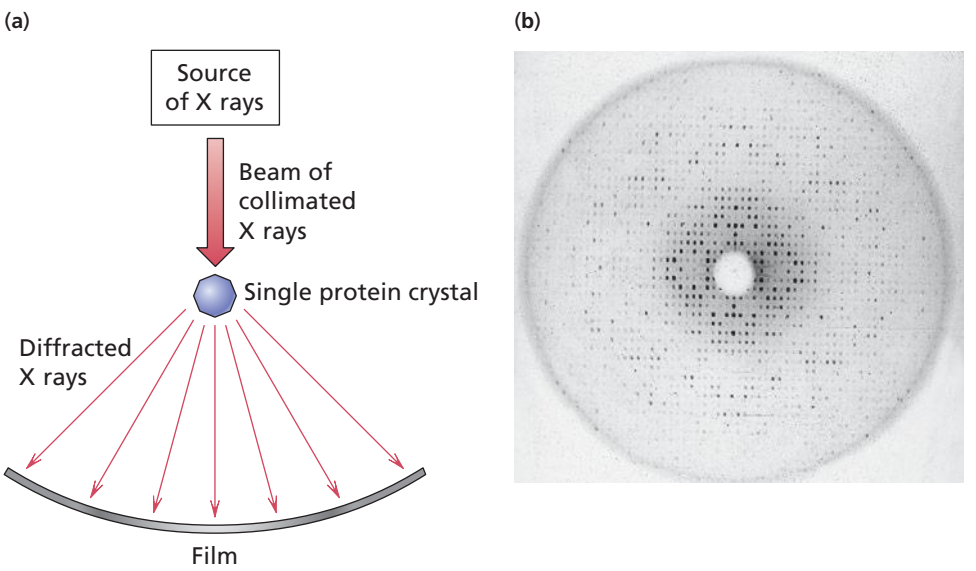


Figure 4.2 ►
X-ray crystallography. (a) Diagram of X rays diffracted by a protein crystal. **(b)** X-ray diffraction pattern of a crystal of adult human deoxyhemoglobin. The location and intensity of the spots are used to determine the three-dimensional structure of the protein.

◀ **Bioinformatics in the 1950s.** Bror Strandberg (left) and Dick Dickerson (right) carrying computer tapes from the EDSAC II computer center in Cambridge, UK. The tapes contain X-ray diffraction data from crystals of myoglobin.

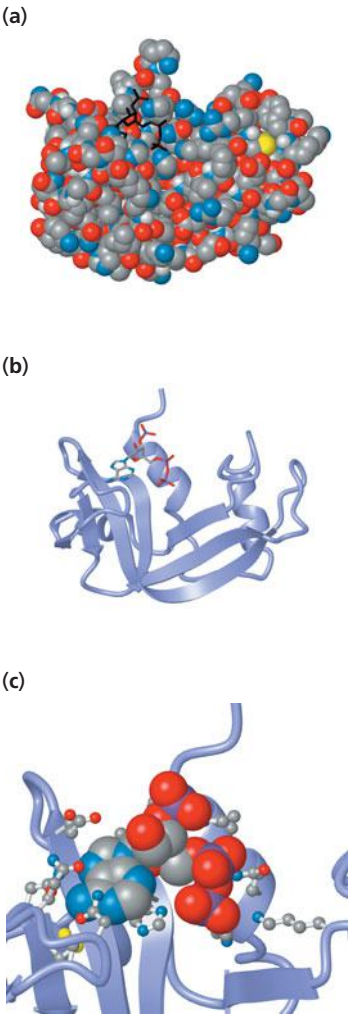
by the difficulty of preparing crystals of a quality suitable for X-ray diffraction and even that step is mostly carried out by computer-driven robots.

A protein crystal contains a large number of water molecules and it is often possible to diffuse small ligands such as substrate or inhibitor molecules into the crystal. In many cases, the proteins within the crystal retain their ability to bind these ligands and they often exhibit catalytic activity. The catalytic activity of enzymes in the crystalline state demonstrates that the proteins crystallize in their *in vivo* native conformations. Thus, the protein structures solved by X-ray crystallography are accurate representations of the structures that exist inside cells.

Once the three-dimensional coordinates of the atoms of a macromolecule have been determined, they are deposited in a data bank where they are available to other scientists. Biochemists were among the early pioneers in exploiting the Internet to share data with researchers around the world—the first public domain databases of biomolecular structures and sequences were established in the late 1970s. Many of the images in this text were created using data files from the Protein Data Bank (PDB).

Visit the website for information on how to view three-dimensional structures and retrieve data files.

◀ Max Perutz (1914–2002) (left) and John C. Kendrew (1917–1997) (right). Kendrew determined the structure of myoglobin and Perutz determined the structure of hemoglobin. They shared the Nobel Prize in 1962.



▲ **Figure 4.3**
Bovine (*Bos taurus*) ribonuclease A. Ribonuclease A is a secreted enzyme that hydrolyzes RNA during digestion. **(a)** Space-filling model showing a bound substrate analog in black. **(b)** Cartoon ribbon model of the polypeptide chain showing secondary structure. **(c)** View of the substrate-binding site. The substrate analog (5′-diphosphoadenine-3′-phosphate) is depicted as a space-filling model, and the side chains of amino acid residues are shown as ball-and-stick models. [PDB 1AFK]

We will list the PDB filename, or accession number, for every protein structure shown in this text so that you can view the three-dimensional structure on your own computer.

There are many ways of depicting the three-dimensional structure of proteins. Space-filling models (Figure 4.3a) depict each atom as a solid sphere. Such images reveal the dense, closely packed nature of folded polypeptide chains. Space-filling models of structures are used to illustrate the overall shape of a protein and the surface exposed to aqueous solvent. One can easily appreciate that the interior of folded proteins is nearly impenetrable, even by small molecules such as water.

The structure of a protein can also be depicted as a simplified cartoon that emphasizes the backbone of the polypeptide chain (Figure 4.3b). In these models, the amino acid side chains have been eliminated, making it easier to see how the polypeptide folds into a three-dimensional shape. Such models have the advantage of allowing us to see into the interior of the protein, and they also reveal elements of secondary structure such as α helices and β strands. By comparing the structures of different proteins, it is possible to recognize common folds and patterns that can't be seen in space-filling models.

The most detailed models are those that emphasize the structures of the amino acid side chains and the various covalent bonds and weak interactions between atoms (Figure 4.3c). Such detailed models are especially important in understanding how a substrate binds in the active site of an enzyme. In Figure 4.3c, the backbone is shown in the same orientation as in Figure 4.3b.

Another technique for analyzing the macromolecular structure of proteins is nuclear magnetic resonance (NMR) spectroscopy. This method permits the study of proteins in solution and therefore does not require the painstaking preparation of crystals. In NMR spectroscopy, a sample of protein is placed in a magnetic field. Certain atomic nuclei absorb electromagnetic radiation as the applied magnetic field is varied. Because absorbance is influenced by neighboring atoms, interactions between atoms that are close together can be recorded. By combining these results with the amino acid sequence and known structural constraints it is possible to calculate a number of structures that satisfy the observed interactions.

Figure 4.4 depicts the complete set of structures for bovine ribonuclease A—the same protein whose X-ray crystal structure is shown in Figure 4.3. Note that the possible structures are very similar and the overall shape of the molecule is easily seen. In some cases, the set of NMR structures may represent fluctuations, or "breathing," of the protein in solution. The similarity of the NMR and X-ray crystal structures indicates that the protein structures found in crystals accurately represent the structure of the protein in solution but in some cases the structures do not agree. Often this is due to disordered regions that do not show up in the X-ray crystal structure (Section 4.7D). On very rare occasions the protein crystallyzes in a conformation that is not the true native form. The NMR structure is thought to be more accurate.

In general, the NMR spectra for small proteins such as ribonuclease A can be easily solved but the spectrum of a large molecule can be extremely complex. For this reason, it is very difficult to determine the structure of larger proteins but the technique is very powerful for smaller proteins.

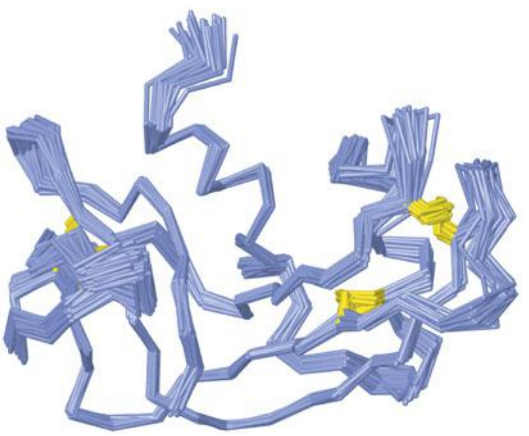


Figure 4.4 ▶
Bovine ribonuclease A NMR structure. The figure combines a set of very similar structures that satisfy the data on atomic interactions. Only the backbone of the polypeptide chain is shown. Compare this structure with that in Figure 4.3b. Note the presence of disulfide bridges (yellow), which are not shown in the images derived from the X-ray crystal structure. [PDB 2AAS].

4.3 The Conformation of the Peptide Group

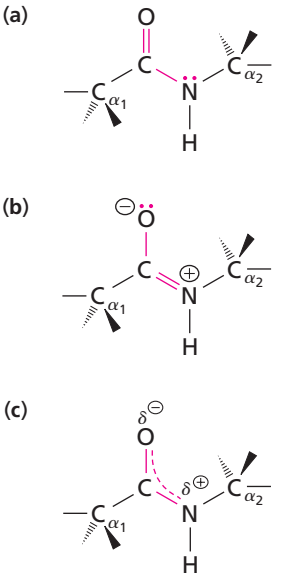
Our detailed study of protein structure begins with the structure of the peptide bonds that link amino acids in a polypeptide chain. The two atoms involved in the peptide bond, along with their four substituents (the carbonyl oxygen atom, the amide hydrogen atom, and the two adjacent α-carbon atoms), constitute the peptide group. X-ray crystallographic analyses of small peptides reveal that the bond between the carbonyl carbon and the nitrogen is shorter than typical C—N single bonds but longer than typical C═N double bonds. In addition, the bond between the carbonyl carbon and the oxygen is slightly longer than typical C═O double bonds. These measurements reveal that peptide bonds have some double-bond properties and can best be represented as a resonance hybrid (Figure 4.5).

Note that the peptide group is polar. The carbonyl oxygen has a partial negative charge and can serve as a hydrogen acceptor in hydrogen bonds. The nitrogen has a partial positive charge, and the —NH group can serve as a hydrogen donor in hydrogen bonds. Electron delocalization and the partial double-bond character of the peptide bond prevent unrestricted free rotation around the C—N bond. As a result, the atoms of the peptide group lie in the same plane (Figure 4.6). Rotation is still possible around each N—Cα bond and each Cα—C bond in the repeating N—Cα—C backbone of proteins. As we will see, restrictions on free rotation around these two additional bonds ultimately determine the three-dimensional conformation of a protein.

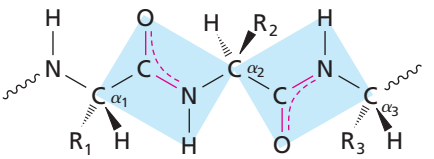
Because of the double-bond nature of the peptide bond, the conformation of the peptide group is restricted to one of two possible conformations, either *trans* or *cis* (Figure 4.7). In the *trans* conformation, the two α-carbons of adjacent amino acid residues are on opposite sides of the peptide bond and at opposite corners of the rectangle formed by the planar peptide group. In the *cis* conformation, the two α-carbons are on the same side of the peptide bond and are closer together. The *cis* and *trans* conformations arise during protein synthesis when the peptide bond is formed by joining amino acids to the growing polypeptide chain. The two conformations are not easily interconverted by free rotation around the peptide bond once it has formed.

The *cis* conformation is less favorable than the extended *trans* conformation because of steric interference between the side chains attached to the two α-carbon atoms. Consequently, nearly all peptide groups in proteins are in the *trans* conformation. Rare exceptions occur, usually at bonds involving the amide nitrogen of proline. Because of the unusual ring structure of proline, the *cis* conformation creates only slightly more steric interference than the *trans* conformation.

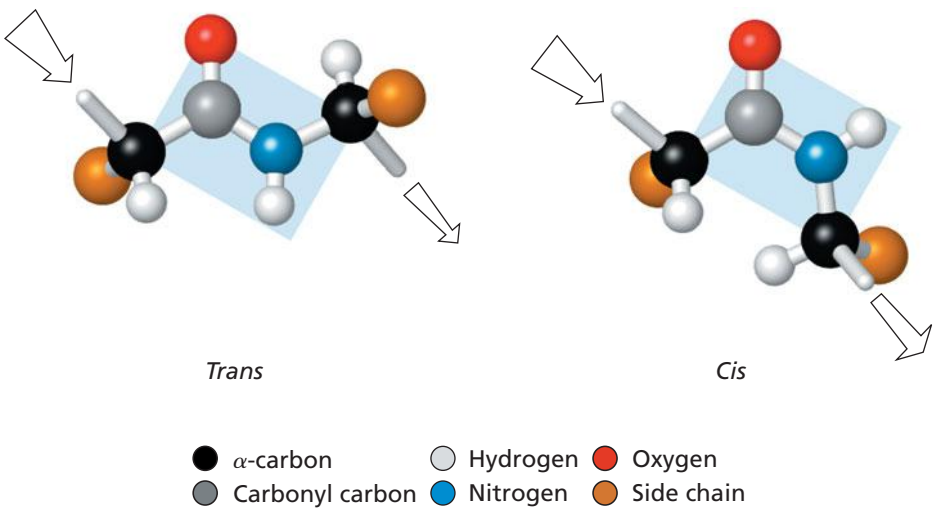
Remember that even though the atoms of the peptide group lie in a plane, rotation is still possible about the N—Cα and Cα—C bonds in the repeating N—Cα—C backbone. This rotation is restricted by steric interference between main-chain and side-chain atoms of adjacent residues. One of the most important restrictions on free rotation is steric interference between carbonyl oxygens on adjacent amino acid residues in the polypeptide



▲ **Figure 4.5**
Resonance structure of the peptide bond. **(a)** In this resonance form, the peptide bond is shown as a single C—N bond. **(b)** In this resonance form, the peptide bond is shown as a double bond. **(c)** The actual structure is best represented as a hybrid of the two resonance forms in which electrons are delocalized over the carbonyl oxygen, the carbonyl carbon, and the amide nitrogen. Rotation around the C—N bond is restricted due to the double-bond nature of the resonance hybrid form.



▲ **Figure 4.6**
Planar peptide groups in a polypeptide chain. A peptide group consists of the N—H and C═O groups involved in formation of the peptide bond, as well as the α-carbons on each side of the peptide bond. Two peptide groups are highlighted in this diagram.



◀ **Figure 4.7**
***Trans* and *cis* conformations of a peptide group.** Nearly all peptide groups in proteins are in the *trans* conformation, which minimizes steric interference between adjacent side chains. The arrows indicate the direction from the N- to the C-terminus.

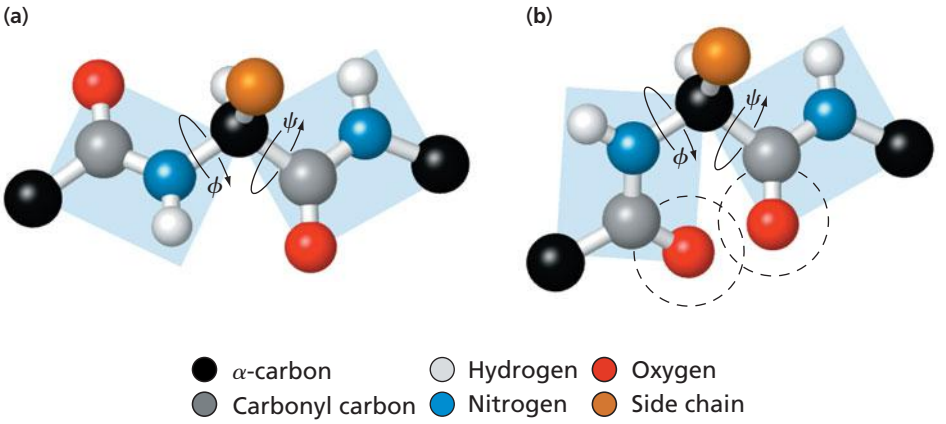
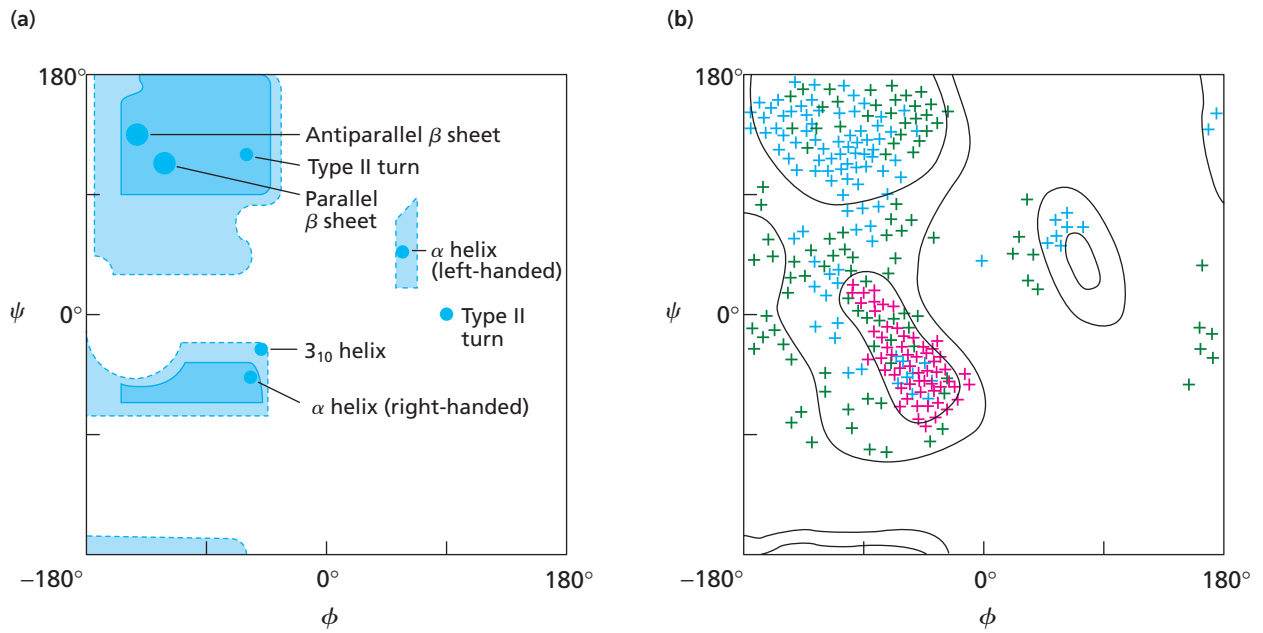


Figure 4.8 ▸
Rotation around the N—Cα and Cα—C bonds that link peptide groups in a polypeptide chain. **(a)** Peptide groups in an extended conformation. **(b)** Peptide groups in an unstable conformation caused by steric interference between carbonyl oxygens of adjacent residues. The van der Waals radii of the carbonyl oxygen atoms are shown by the dashed lines. The rotation angle around the N—Cα bond is called φ (phi), and that around the Cα—C bond is called ψ (psi). The substituents of the outer α-carbons have been omitted for clarity.

chain (Figure 4.8). The presence of bulky side chains also restricts free rotation around the N—Cα and Cα—C bonds. Proline is a special case—rotation around the N—Cα bond is constrained because it is part of the pyrrolidine ring structure of proline.

The rotation angle around the N—Cα bond of a peptide group is designated φ (phi), and that around the Cα—C bond is designated ψ (psi). The peptide bond angle is ω (omega). Because rotation around peptide bonds is hindered by their double-bond character, most of the conformation of the backbone of a polypeptide can be described by φ and ψ. Each of these angles is defined by the relative positions of four atoms of the backbone. Clockwise angles are positive, and counterclockwise angles are negative, with each having a 180° sweep. Thus, each of the rotation angles can range from −180° to +180°.

The biophysicist G. N. Ramachandran and his colleagues constructed space-filling models of peptides and made calculations to determine which values of φ and ψ are sterically permitted in a polypeptide chain. Permissible angles are shown as shaded regions in **Ramachandran plots** of φ versus ψ. Figure 4.9a shows the results of theoretical calculations—the dark, shaded regions represent permissible angles for most residues, and the lighter areas cover the φ and ψ values for smaller amino acid residues where the



▲ **Figure 4.9**
Ramachandran plot. **(a)** Solid lines indicate the range of permissible φ and ψ values based on molecular models. Dashed lines give the outer limits for an alanine residue. Large blue dots correspond to values of φ and ψ that produce recognizable conformations such as the α helix and β sheets. The positions shown for the type II turn are for the second and third residues. The white portions of the plot correspond to values of φ and ψ that were predicted to occur rarely. **(b)** Observed φ and ψ values in known structures. Crosses indicate values for typical residues in a single protein. Residues in an α helix are shown in red, β-strand residues are blue, and others are green.

R groups don't restrict rotation. Blank areas on a Ramachandran plot are nonpermissible areas, due largely to steric hindrance. The conformations of several types of ideal secondary structure fall within the shaded areas, as expected.

Another version of a Ramachandran plot is shown in Figure 4.9b. This plot is based on the observed φ and ψ angles of hundreds of proteins whose structures are known. The enclosed inner regions represent angles that are found very frequently, and the outer enclosed regions represent angles that are less frequent. Typical observed angles for α helices, β sheets, and other structures in a protein are plotted. The most important difference between the theoretical and observed Ramachandran plots is in the region around 0°φ and −90°ψ. This region should not be permitted according to the modeling studies but there are many examples of residues with these angles. It turns out that steric clashes are prevented in these regions by allowing a small amount of rotation around the peptide bond. The peptide group does not have to be exactly planar—a little bit of wiggle is permitted!

Some bulky amino acid residues have smaller permitted areas. Proline is restricted to a φ value of about −60° to −77° because its N—Cα bond is constrained by inclusion in the pyrrolidine ring of the side chain. In contrast, glycine is exempt from many steric restrictions because it lacks a β-carbon. Thus, glycine residues have greater conformational freedom than other residues and have φ and ψ values that often fall outside the shaded regions of the Ramachandran plot.

KEY CONCEPT

The three-dimensional conformation of a polypeptide backbone is defined by the φ (phi) and ψ (psi) angles of rotation around each peptide group.

BOX 4.1 FLOWERING IS CONTROLLED BY *CIS/TRANS* SWITCHES

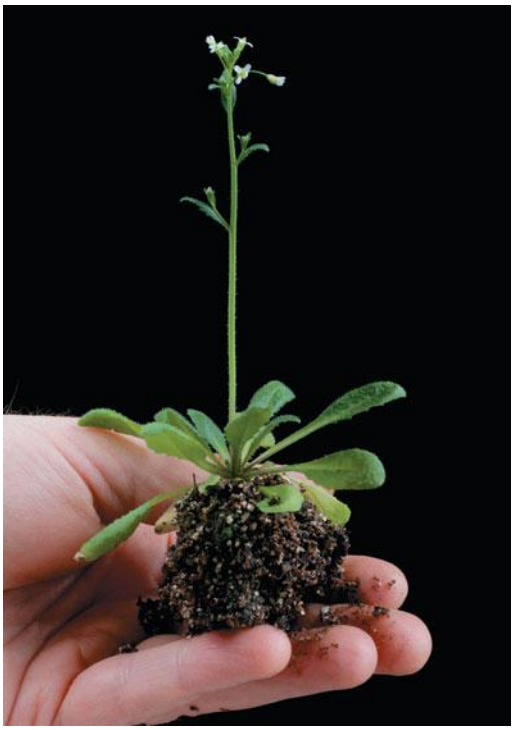
Almost all peptide groups adopt the *trans* conformation since that is the one favored during protein synthesis. It is much more stable than the *cis* conformation (with one exception). Spontaneous switching to the *cis* conformation is very rare and it is almost always accompanied by loss of function since the structure of the protein is severely affected.

However, the activity of some proteins is actually *regulated* by conformation changes due to *cis/trans* isomerization. The change in peptide group conformation invariably takes place at proline residues because the *cis* conformation is almost as stable as the *trans* conformation. This is the one exception to the rule.

Specific enzymes, called peptidyl prolyl *cis/trans* isomerases, catalyze the interconversion of *cis* and *trans* conformation at proline residues by transiently destabilizing the resonance hybrid structure of the peptide bond and allowing rotation. One important class of these enzymes recognizes Ser-Pro and Thr-Pro bonds whenever the serine and threonine residues are phosphorylated. Phosphorylation of amino acid residues is an important mechanism of regulation by covalent modification (see Section 5.9D). The gene for this type of peptidyl prolyl *cis/trans* isomerase is called *Pin1* and it is present in all eukaryotes.

In the small flowering plant, *Arabidopsis thalianna*, *Pin1* protein acts on some transcription factors that control the timing of flowering. When threonine residues are phosphorylated, the transcription factors are recognized by *Pin1* and the conformation of the Thr-Pro bond is switched from *trans* to *cis*. The resulting conformational change in the structure of the protein leads to activation of the transcription factors and transcription of the genes required for producing flowers. Flowering is considerably delayed when the synthesis of peptidyl prolyl *cis/trans* isomerase is inhibited by mutations in the *Pin1* gene.

In humans the *cis/trans* isomerase encoded by *Pin1* plays a role in regulating gene expression by modifying RNA polymerase, transcription factors, and other proteins. Mutations in this gene have been implicated in several hereditary diseases. The structure of human peptidyl prolyl *cis/trans* isomerase is shown in Figure 4.23e.



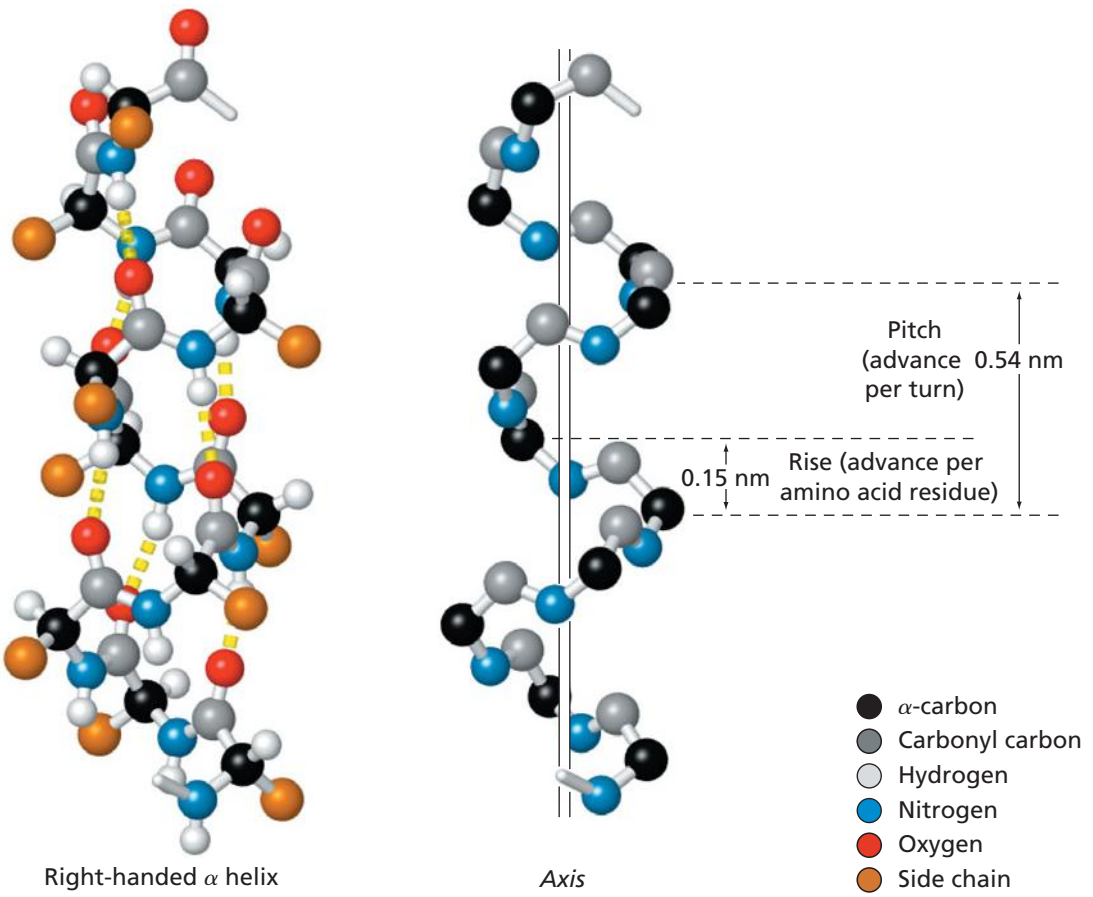
▲ *Arabidopsis thalianna*, also known as thale cress or mouse-ear cress, is a relative of mustard. It is a favorite model organism in plant biology because it is easy to grow in the laboratory.

▲ Linus Pauling (1901–1994), winner of the Nobel Prize in Chemistry in 1954 and the Nobel Peace Prize in 1962.

4.4 The α Helix

The α-helical conformation was proposed in 1950 by Linus Pauling and Robert Corey. They considered the dimensions of peptide groups, possible steric constraints, and opportunities for stabilization by formation of hydrogen bonds. Their model accounted for the major repeat observed in the structure of the fibrous protein α-keratin. This repeat of 0.50 to 0.55 nm turned out to be the pitch (the axial distance per turn) of the α helix. Max Perutz added additional support for the structure when he observed a secondary repeating unit of 0.15 nm in the X-ray diffraction pattern of α-keratin. The 0.15 nm repeat corresponds to the rise of the α helix (the distance each residue advances the helix along its axis). Perutz also showed that the α helix was present in hemoglobin, confirming that this conformation was present in more complex globular proteins.

In theory, an α helix can be either a right- or a left-handed screw. The α helices found in proteins are almost always right-handed, as shown in Figure 4.10. In an ideal α helix, the pitch is 0.54 nm, the rise is 0.15 nm, and the number of amino acid residues required for one complete turn is 3.6 (i.e., approximately 3 2/3 residues: one carbonyl group, three N—Cα—C units, and one nitrogen). Most α helices are slightly distorted in proteins but they generally have between 3.5 and 3.7 residues per turn.



▲ **Figure 4.10**
α Helix. A region of α-helical secondary structure is shown with the N-terminus at the bottom and the C-terminus at the top of the figure. Each carbonyl oxygen forms a hydrogen bond with the amide hydrogen of the fourth residue further toward the C-terminus of the polypeptide chain. The hydrogen bonds are approximately parallel to the long axis of the helix. Note that all the carbonyl groups point toward the C-terminus. In an ideal α helix, equivalent positions recur every 0.54 nm (the pitch of the helix), each amino acid residue advances the helix by 0.15 nm along the long axis of the helix (the rise), and there are 3.6 amino acid residues per turn. In a right-handed helix the backbone turns in a clockwise direction when viewed along the axis from its N-terminus. If you imagine that the right-handed helix is a spiral staircase, you will be turning to the right as you walk *down* the staircase.

Within an α helix, each carbonyl oxygen (residue *n*) of the polypeptide backbone is hydrogen-bonded to the backbone amide hydrogen of the fourth residue further toward the C-terminus (residue *n* + 4). (The three amino groups at one end of the helix and the three carbonyl groups at the other end lack hydrogen-bonding partners within the helix.) Each hydrogen bond closes a loop containing 13 atoms—the carbonyl oxygen, 11 backbone atoms, and the amide hydrogen. Thus, an α helix can also be called a 3.6_{13} helix based on its pitch and hydrogen-bonded loop size. The hydrogen bonds that stabilize the helix are nearly parallel to the long axis of the helix.

The φ and ψ angles of each residue in an α helix are similar. They cluster around a stable region of the Ramachandran plot centered at a φ value of −57° and a ψ value of −47° (Figure 4.9). The similarity of these values is what gives the α helix a regular, repeating structure. The intramolecular hydrogen bonds between residues *n* and *n* + 4 tend to "lock in" rotation around the N—Cα and Cα—C bonds restricting the φ and ψ angles to a relatively narrow range.

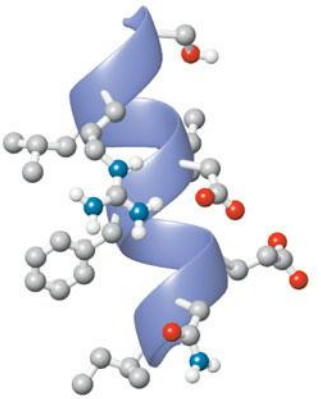
A single intrahelical hydrogen bond would not provide appreciable structural stability but the cumulative effect of many hydrogen bonds within an α helix stabilizes this conformation. Hydrogen bonds between amino acid residues are especially stable in the hydrophobic interior of a protein where water molecules do not enter and therefore cannot compete for hydrogen bonding. In an α helix, all the carbonyl groups point toward the C-terminus. The entire helix is a dipole with a positive N-terminus and a negative C-terminus since each peptide group is polar and all the hydrogen bonds point in the same direction.

The side chains of the amino acids in an α helix point outward from the cylinder of the helix and they are not involved in the hydrogen bonds that stabilize the α helix (Figure 4.11). However, the identity of the side chains affects the stability in other ways. Because of this, some amino acid residues are found in α-helical conformations more often than others. For example, alanine has a small, uncharged side chain and fits well into the α-helical conformation. Alanine residues are prevalent in the α helices of all classes of proteins. In contrast, tyrosine and asparagine with their bulky side chains are less common in α helices. Glycine, whose side chain is a single hydrogen atom, destabilizes α-helical structures since rotation around its α-carbon is so unconstrained. For this reason, many α helices begin or end with glycine residues. Proline is the least common residue in an α helix because its rigid cyclic side chain disrupts the right-handed helical conformation by occupying space that a neighboring residue of the helix would otherwise occupy. In addition, because it lacks a hydrogen atom on its amide nitrogen, proline cannot fully participate in intrahelical hydrogen bonding. For these reasons, proline residues are found more often at the ends of α helices than in the interior.

Proteins vary in their α-helical content. In some proteins most of the residues are in α helices, whereas other proteins contain very little α-helical structure. The average content of α helix in the proteins that have been examined is 26%. The length of a helix in a protein can range from about 4 or 5 residues to more than 40—the average is about 12.

Many α helices have hydrophilic amino acids on one face of the helix cylinder and hydrophobic amino acids on the opposite face. The amphipathic nature of the helix is easy to see when the amino acid sequence is drawn as a spiral called a helical wheel. The α helix shown in Figure 4.11 can be drawn as a helical wheel representing the helix viewed along its axis. Because there are 3.6 residues per turn of the helix, the residues are plotted every 100° along the spiral (Figure 4.12). Note that the helix is a right-handed screw and it is terminated by a glycine residue at the C-terminal end. The hydrophilic residues (asparagine, glutamate, aspartate, and arginine) tend to cluster on one side of the helical wheel.

Amphipathic helices are often located on the surface of a protein with the hydrophilic side chains facing outward (toward the aqueous solvent) and the hydrophobic side chains facing inward (toward the hydrophobic interior). For example, the helix shown in Figures 4.11 and 4.12 is on the surface of the water-soluble liver enzyme alcohol dehydrogenase with the side chains of the first, fifth, and eighth residues



▲ **Figure 4.11**
View of a right-handed α helix. The blue ribbon indicates the shape of the polypeptide backbone. All the side chains, shown as ball-and-stick models, project outward from the helix axis. This example is from residues Ile-355 (bottom) to Gly-365 (top) of horse liver alcohol dehydrogenase. Some hydrogen atoms are not shown. [PDB 1ADF].

▲ **A right-handed α helix.** This helix was created by Julian Voss-Andreae. It stands outside Linus Panling's childhood home in Portland, Oregon, United States.

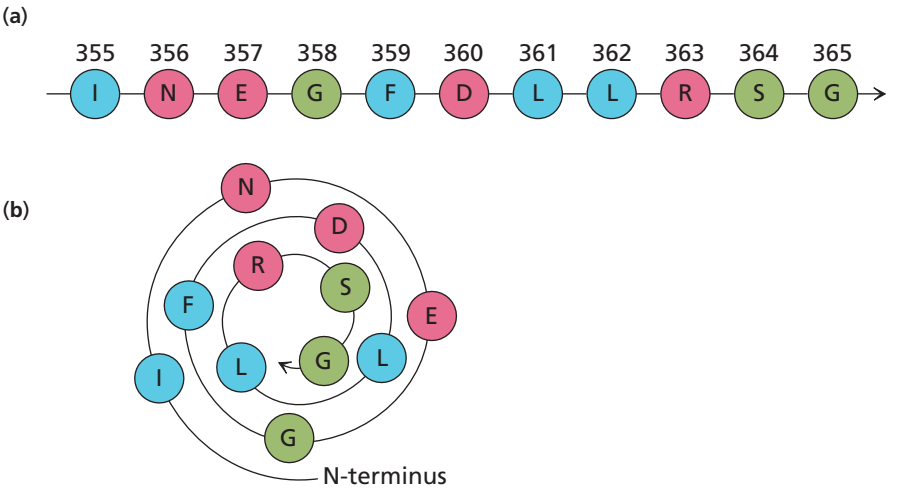


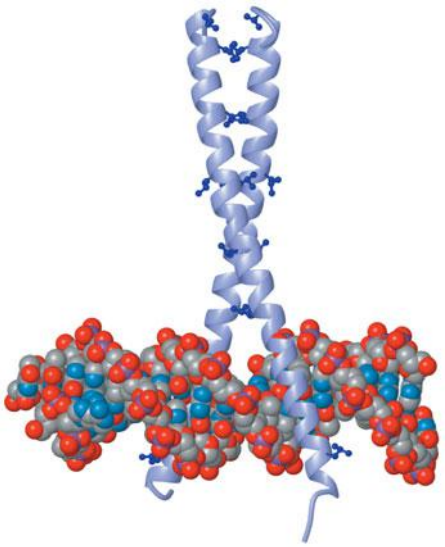
Figure 4.12 ►
α helix in horse liver alcohol dehydrogenase. Highly hydrophobic residues are blue, less hydrophobic residues are green, and highly hydrophilic residues are red. **(a)** Sequence of amino acids. **(b)** Helical wheel diagram.

The known frequencies of various amino acid residues in α helices are used to predict the secondary structure based on the primary sequence alone.

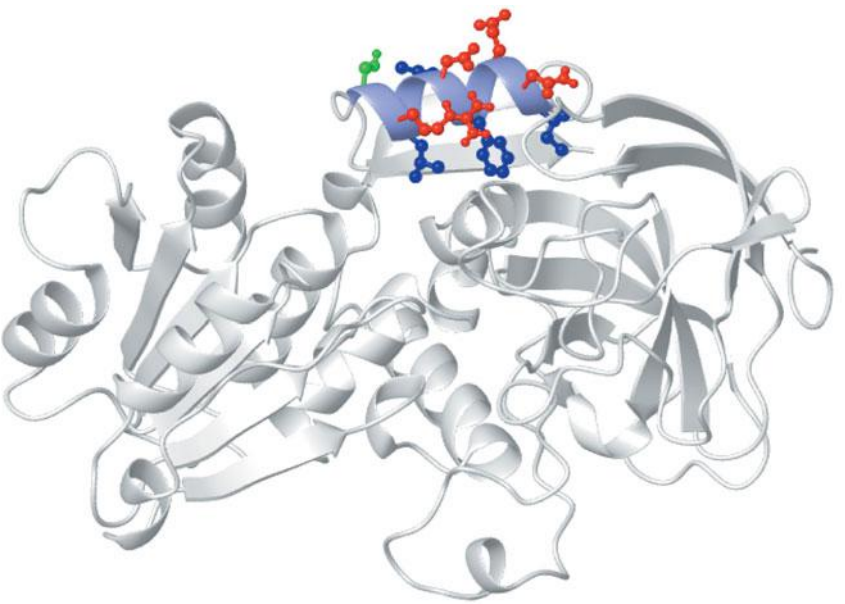
(isoleucine, phenylalanine, and leucine, respectively) buried in the protein interior (Figure 4.13).

There are many examples of two amphipathic α helices that interact to produce an extended coiled-coil structure where the two α helices wrap around each other with their hydrophobic faces in contact and their hydrophilic faces exposed to solvent. A common structure in DNA-binding proteins is called a leucine zipper (Figure 4.14). The name refers to the fact that two α helices are "zippered" together by the hydrophobic interactions of leucine residues (and other hydrophobic residues) on one side of an amphipathic helix. The ends of the helices form the DNA-binding region of the protein.

Some proteins contain a few short regions of a 3_{10} helix. Like the α helix, the 3_{10} helix is right-handed. The carbonyl oxygen of a 3_{10} helix forms a hydrogen bond with the amide hydrogen of residue $n + 3$ (as opposed to residue $n + 4$ in an α helix) so the 3_{10} helix has a tighter hydrogen-bonded ring structure than the α helix—10 atoms rather than 13—and has fewer residues per turn (3.0) and a longer pitch (0.60 nm) (Figure 4.15).

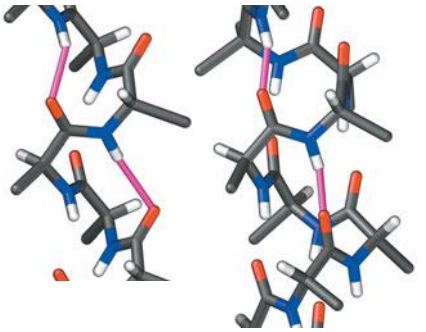


▲ Figure 4.14
Leucine zipper region of yeast (*Saccharomyces cerevisiae*). GCN4 protein bound to DNA. GCN4 is a transcription regulatory protein that binds to specific DNA sequences. The DNA-binding region consists of two amphipathic α helices, one from each of the two subunits of the protein. The side chains of leucine residues are shown in a darker blue than the ribbon. Only the leucine zipper region of the protein is shown in the figure. [PDB 1YSA].



▲ Figure 4.13
Horse (*Equus ferus*) liver alcohol dehydrogenase. The amphipathic α helix is highlighted. The side chains of highly hydrophobic residues are shown in blue, less hydrophobic residues are green, and charged residues are shown in red. Note that the side chains of the hydrophobic residues are directed toward the interior of the protein and that the side chains of charged residues are exposed to the surface. [PDB 1ADF].

The 3_{10} helix is slightly less stable than the α helix because of steric hindrances and the awkward geometry of its hydrogen bonds. When a 3_{10} helix occurs, it is usually only a few residues in length and often is the last turn at the C-terminal end of an α helix. Because of its different geometry, the φ and ψ angles of residues in a 3_{10} helix occupy a different region of the Ramachandran plot than the residues of an α helix (Figure 4.9).



▲ **Figure 4.15**
The 3_{10} helix. In the 3_{10} helix (left) hydrogen bonds (pink) form between the amide group of one residue and the carbonyl oxygen of a residue three positions away. In an α helix (right) the carbonyl group bonds to an amino acid residue four positions away.

4.5 β Strands and β Sheets

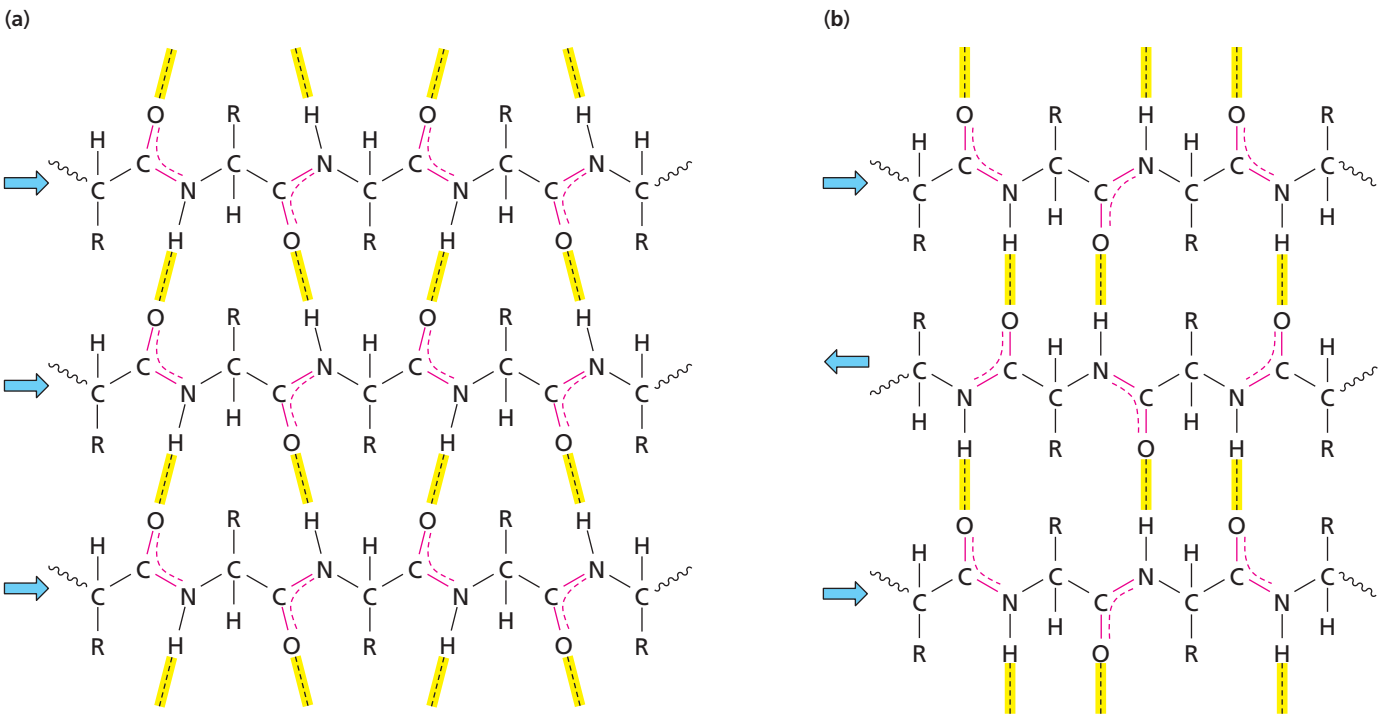
The other common secondary structure is called β structure, a class that includes β strands and β sheets. **β Strands** are portions of the polypeptide chain that are almost fully extended. Each residue in a β strand accounts for about 0.32 to 0.34 nm of the overall length in contrast to the compact coil of an α helix where each residue corresponds to 0.15 nm of the overall length. When multiple β strands are arranged side-by-side they form **β sheets**, a structure originally proposed by Pauling and Corey at the same time they developed a theoretical model of the α helix.

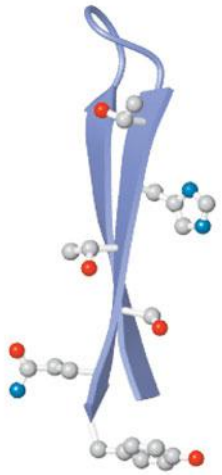
Proteins rarely contain isolated β strands because the structure by itself is not significantly more stable than other conformations. However, β sheets are stabilized by hydrogen bonds between carbonyl oxygens and amide hydrogens on adjacent β strands. Thus, in proteins, the regions of β structure are almost always found in sheets.

The hydrogen-bonded β strands can be on separate polypeptide chains or on different segments of the same chain. The β strands in a sheet can be either parallel (running in the same N- to C-terminal direction) (Figure 4.16a) or antiparallel (running in opposite N- to C-terminal directions) (Figure 4.16b). When the β strands are antiparallel, the hydrogen bonds are nearly perpendicular to the extended polypeptide chains. Note that in the antiparallel β sheet, the carbonyl oxygen and the amide hydrogen atoms of one residue form hydrogen bonds with the amide hydrogen and carbonyl oxygen of a single residue in the other strand. In the parallel arrangement, the hydrogen bonds are not perpendicular to the extended chains and each residue forms hydrogen bonds with the carbonyl and amide groups of two different residues on the adjacent strand.

Parallel sheets are less stable than antiparallel sheets, possibly because the hydrogen bonds are distorted in the parallel arrangement. The β sheet is sometimes called a **β pleated sheet** since the planar peptide groups meet each other at angles, like the folds of an accordion. As a result of the bond angles between peptide groups, the amino acid

▼ **Figure 4.16**
β Sheets. Arrows indicate the N- to C-terminal direction of the peptide chain. **(a)** Parallel β sheet. The hydrogen bonds are evenly spaced but slanted. **(b)** Antiparallel β sheet. The hydrogen bonds are essentially perpendicular to the β strands, and the space between hydrogen-bonded pairs is alternately wide and narrow.





▲ **Figure 4.17**
View of two strands of an antiparallel β sheet from influenza virus A neuraminidase. Only the side chains of the front β strand are shown. The side chains alternate from one side of the β strand to the other side. Both strands have a right-handed twist. [PDB 1BJI]

side chains point alternately above and below the plane of the sheet. A typical β sheet contains from two to as many as 15 individual β strands. Each strand has an average of six amino acid residues.

The β strands that make up β sheets are often twisted and the sheet is usually distorted and buckled. The three-dimensional view of the β sheet of ribonuclease A (Figure 4.3) shows a more realistic view of β sheets than the idealized structures in Figure 4.16.

A view of two strands of a small β sheet is shown in Figure 4.17. The side chains of the amino acid residues in the front strand alternately project to the left and to the right of (i.e., above and below) the β strand, as described above. Typically, β strands twist slightly in a right-hand direction; that is, they twist clockwise as you look along one strand.

The φ and ψ angles of the bonds in a β strand are restricted to a broad range of values occupying a large, stable region in the upper left-hand corner of the Ramachandran plot. The typical angles for residues in parallel and antiparallel strands are not identical (see Figure 4.9). Because most β strands are twisted, the φ and ψ angles exhibit a broader range of values than those seen in the more regular α helix.

Although we usually think of β sheets as examples of secondary structure this is not, strictly speaking, correct. In many cases, the individual β strands are located in different regions of the protein and only come together to form the β sheet when the protein adopts its final tertiary conformation. Sometimes the quaternary structure of a protein gives rise to a large β sheet. Some proteins are almost entirely β sheets but most proteins have a much lower β-strand content.

In the previous section we noted that amphipathic α helices have hydrophobic side chains that project outward on one side of the helix. This is the side that interacts with the rest of the protein creating a series of hydrophobic interactions that help stabilize the tertiary structure. The side chains of β sheets project alternately above and below the plane of the β strands. One surface may consist of hydrophobic side chains that allow the β sheet to lie on top of other hydrophobic residues in the interior of the protein.

An example of such hydrophobic interactions between two β sheets is seen in the structure of the coat protein of grass pollen grains (Figure 4.18a). This protein is the major allergen affecting people who are allergic to grass pollen. One surface of each β sheet contains hydrophobic side chains and the opposite surface has hydrophilic side chains. The two hydrophobic surfaces interact to form the hydrophobic core of the protein and the hydrophilic surfaces are exposed to solvent as shown in Figure 4.18b. This is an example of a β sandwich, one of several arrangements of secondary structural elements that are covered in more detail in the section on tertiary structure (Section 4.7).

KEY CONCEPT

There are only three different kinds of common secondary structure: α helix, β strand, and turns.

4.6 Loops and Turns

In both an α helix and a β strand there are consecutive residues with a similar conformation that is repeated throughout the structure. Proteins also contain stretches of nonrepeating three-dimensional structure. Most of these non-repeating regions of secondary structure can be characterized as loops or turns since they cause directional changes in the polypeptide backbone. The conformations of peptide groups in nonrepetitive regions are constrained just as they are in repetitive regions. They have φ and ψ values that are usually well within the permitted regions of the Ramachandran plot and often close to the values of residues that form α helices or β strands.

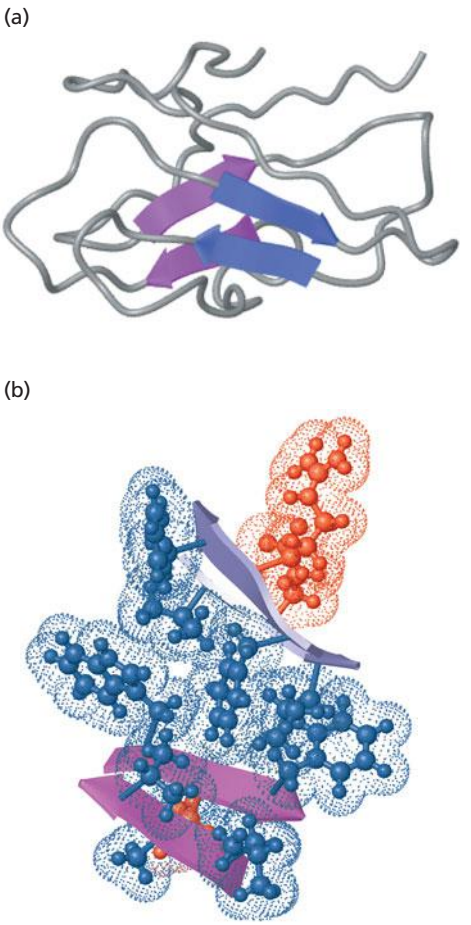
Loops and turns connect α helices and β strands and allow the polypeptide chain to fold back on itself producing the compact three-dimensional shape seen in the native structure. As much as one-third of the amino acid residues in a typical protein are found in such nonrepetitive structures. **Loops** often contain hydrophilic residues and are usually found on the surfaces of proteins where they are exposed to solvent and form hydrogen bonds with water. Some loops consist of many residues of extended nonrepetitive structure. About 10% of the residues can be found in such regions.

▲ **U-turns are allowed in proteins.**

Loops containing only a few (up to five) residues are referred to as **turns** if they cause an abrupt change in the direction of a polypeptide chain. The most common types of tight turns are called **reverse turns**. They are also called **β turns** because they often connect different antiparallel β strands. (Recall that in order to create a β sheet the polypeptide must fold so that two or more regions of β strand are adjacent to one another as shown in Figure 4.17.) This terminology is misleading since β turns can also connect α helices or an α helix and a β strand.

There are two common types of β turn, designated type I and type II. Both types of turn contain four amino acid residues and are stabilized by hydrogen bonding between the carbonyl oxygen of the first residue and the amide hydrogen of the fourth residue (Figure 4.19). Both type I and type II turns produce an abrupt (usually about 180°) change in the direction of the polypeptide chain. In type II turns, the third residue is glycine about 60% of the time. Proline is often the second residue in both types of turns.

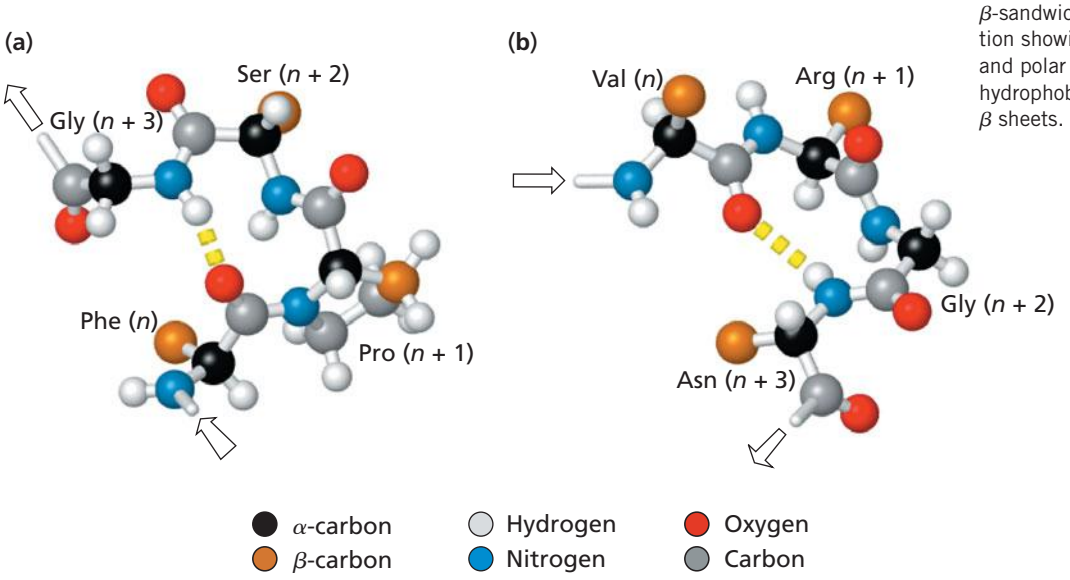
Proteins contain many turn structures. They all have internal hydrogen bonds that stabilize the structure and that's why they can be considered a form of secondary structure. Turns make up a significant proportion of the structure in many proteins. Some of the bonds in turn residues have φ and ψ angles that lie outside the "permitted" regions of a typical Ramachandran plot (Figure 4.9). This is especially true of residues in the third position of type II turns where there is an abrupt change in the direction of the backbone. This residue is often glycine so the bond angles can adopt a wider range of values without causing steric clashes between the side-chain atoms and the backbone atoms. Ramachandran plots usually show only the permitted regions for all residues except glycine—this is why the rotation angles of type II turns appear to lie in a restricted area.



▲ **Figure 4.18**
Structure of PHL P2 from Timothy grass (*Phleum pratense*) pollen. (a) The two short, two-stranded, antiparallel β sheets are highlighted in blue and purple to show their orientation within the protein. **(b)** View of the β-sandwich structure in a different orientation showing hydrophobic residues (blue) and polar residues (red). A number of hydrophobic interactions connect the two β sheets. [PDB 1BMW].

4.7 Tertiary Structure of Proteins

Tertiary structure results from the folding of a polypeptide (which may already possess some regions of α helix and β structure) into a closely packed three-dimensional structure. An important feature of tertiary structure is that amino acid residues that are far apart in the primary structure are brought together permitting interactions among their side chains. Whereas secondary structure is stabilized by hydrogen bonding between amide hydrogens and carbonyl oxygens of the polypeptide backbone, tertiary



▲ **Figure 4.19**
Reverse turns. (a) Type I β turn. The structure is stabilized by a hydrogen bond between the carbonyl oxygen of the first N-terminal residue (Phe) and the amide hydrogen of the fourth residue (Gly). Note the proline residue at position $n+1$. **(b)** Type II β turn. This turn is also stabilized by a hydrogen bond between the carbonyl oxygen of the first N-terminal residue (Val) and the amide hydrogen of the fourth residue (Asn). Note the glycine residue at position $n+2$. [PDB 1AHL (giant sea anemone neurotoxin)].

structure is stabilized primarily by noncovalent interactions (mostly the hydrophobic effect) between the side chains of amino acid residues. Disulfide bridges, though covalent, are also elements of tertiary structure they are not part of the primary structure since they form only after the protein folds.

A. Supersecondary Structures

Supersecondary structures, or **motifs**, are recognizable combinations of α helices, β strands, and loops that appear in a number of different proteins. Sometimes motifs are associated with a particular function although structurally similar motifs may have different functions in different proteins. Some common motifs are shown in Figure 4.20.

One of the simplest motifs is the helix–loop–helix (Figure 4.20a). This structure occurs in a number of calcium-binding proteins. Glutamate and aspartate residues in the loop of these proteins form part of the calcium-binding site. In certain DNA-binding proteins a version of this supersecondary structure is called a helix–turn–helix motif since the residues that connect the helices form a reverse turn. In these proteins, the residues of the α helices bind DNA.

The coiled-coil motif consists of two amphipathic α helices that interact through their hydrophobic edges (Figure 4.20b) as in the leucine zipper example (Figure 4.14). Several α helices can associate to form a helix bundle (Figure 4.20c). In this case, the individual α helices have opposite orientations, whereas they are parallel in the coiled-coil motif.

The $\beta\alpha\beta$ unit consists of two parallel β strands linked to an intervening α helix by two loops (Figure 4.20d). The helix connects the C-terminal end of one β strand to the N-terminal end of the next and often runs parallel to the two strands. A hairpin consists of two adjacent antiparallel β strands connected by a β turn (Figure 4.20e). (One example of a hairpin motif is shown in Figure 4.16.)

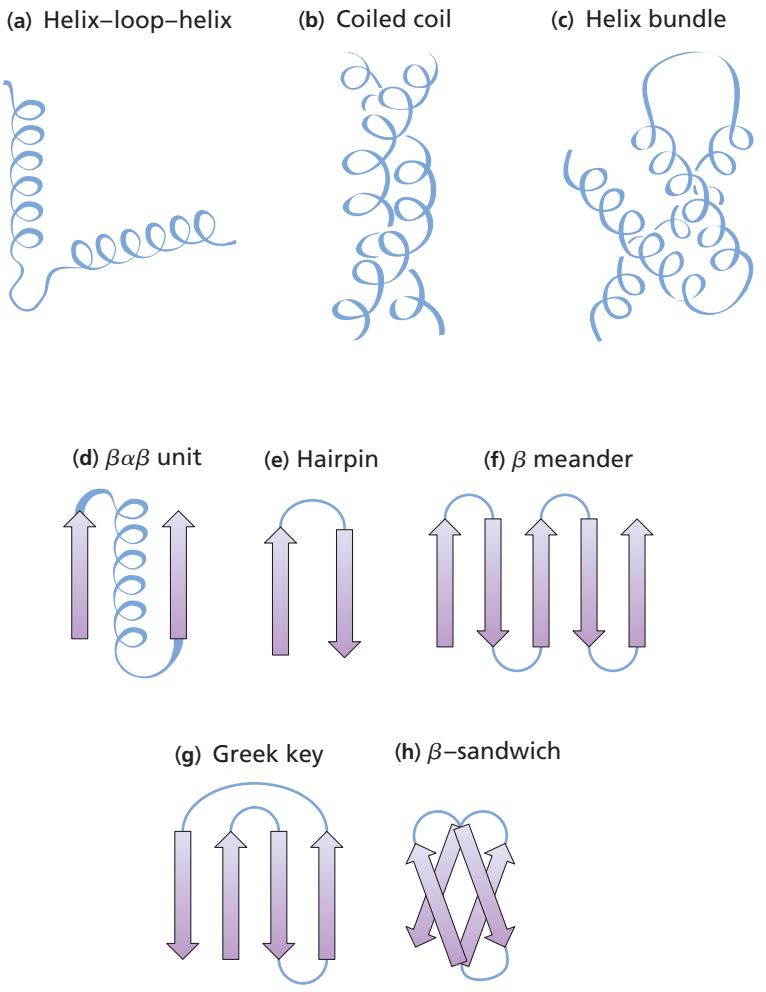
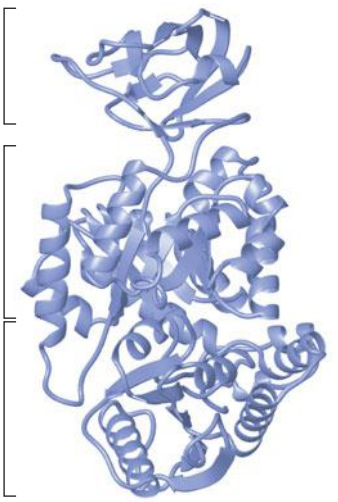


Figure 4.20 ▸
Common motifs. In folded proteins α helices and strands are commonly connected by loops and turns to form supersecondary structures, shown here as two-dimensional representations. Arrows indicate the N- to C-terminal direction of the peptide chain.

The β meander motif (Figure 4.20f) is an antiparallel β sheet composed of sequential β strands connected by loops or turns. The order of strands in the β sheet is the same as their order in the sequence of the polypeptide chain. The β meander sheet may contain one or more hairpins but, more typically, the strands are joined by larger loops. The Greek key motif takes its name from a design found on classical Greek pottery. This is a β sheet motif linking four antiparallel β strands such that strands 3 and 4 form the outer edges of the sheet and strands 1 and 2 are in the middle of the sheet. The β sandwich motif is formed when β strands or sheets stack on top of one another (Figure 4.20h). The figure shows an example of a β sandwich where the β strands are connected by short loops and turns, but β sandwiches can also be formed by the interaction of two β sheets in different regions of the polypeptide chain, as seen in Figure 4.18.

B. Domains

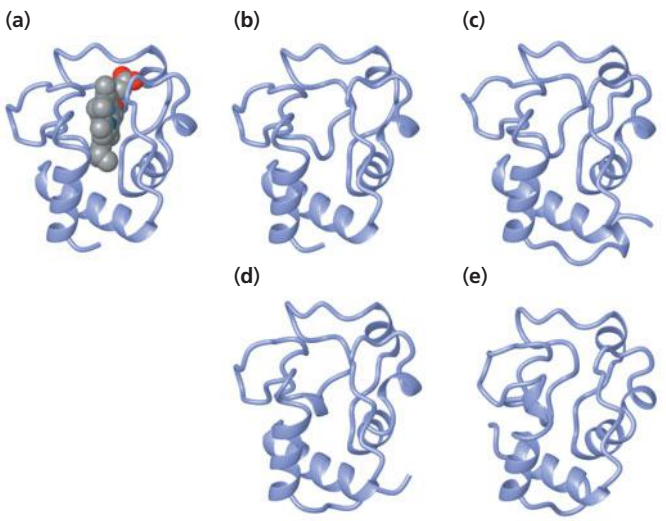
Many proteins are composed of several discrete, independently folded, compact units called **domains**. Domains may consist of combinations of motifs. The size of a domain varies from as few as 25 to 30 amino acid residues to more than 300. An example of a protein with multiple domains is shown in Figure 4.21. Note that each domain is a distinct compact unit consisting of various elements of secondary structure. Domains are usually connected by loops but they are also bound to each other through weak interactions formed by the amino acid side chains on the surface of each domain. The top domain of pyruvate kinase in Figure 4.21 contains residues 116 to 219, the central domain contains residues 1 to 115 plus 220 to 388, and the bottom domain contains residues 389 to 530. In general, domains consist of a contiguous stretch of amino acid residues as in the top and bottom domains of pyruvate kinase but in some cases a single domain may contain two or more different regions of the polypeptide chain as in the middle domain.



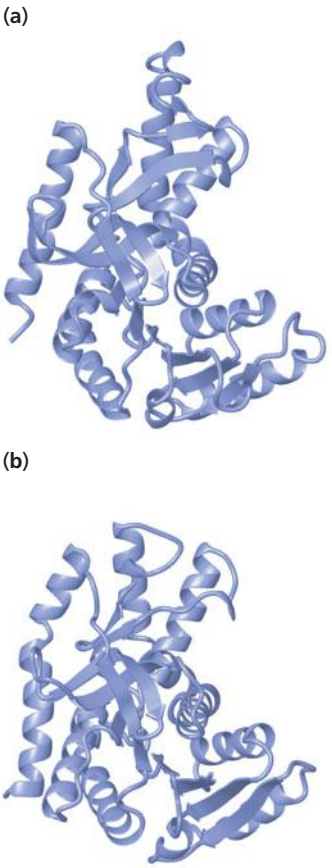
▲ **Figure 4.21**
Pyruvate kinase from cat (*Felis domesticus*). The main polypeptide chain of this common enzyme folds into three distinct domains as indicated by brackets. [PDB 1PKM].

The evolutionary conservation of protein structure is one of the most important observations that has emerged from the study of proteins in the past few decades. This conservation is most easily seen in the case of single-domain homologous proteins from different species. For example, in Chapter 3 we examined the sequence similarity of cytochrome *c* and showed that the similarities in primary structure could be used to construct a phylogenetic tree that reveals the evolutionary relationships of the proteins from different species (Section 3.11). As you might expect, the tertiary structures of cytochrome *c* proteins are also highly conserved (Figure 4.22). Cytochrome *c* is an example of a protein that contains a heme prosthetic group. The conservation of protein structure is a reflection of its interaction with heme and its conserved function as an electron transport protein in diverse species.

Some domain structures occur in many different proteins whereas others are unique. In general, proteins can be grouped into families according to similarities in domain structures and amino acid sequence. All of the members of a family have descended from a common ancestral protein. Some biochemists believe that there may be only a few thousand families



◀ **Figure 4.22**
Conservation of cytochrome *c* structure. **(a)** Tuna (*Thunnus alalunga*) cytochrome *c* bound to heme [PDB 5CYT]. **(b)** Tuna cytochrome *c* polypeptide chain. **(c)** Rice (*Oryza sativa*) cytochrome *c* [PDB 1CCR]. **(d)** Yeast (*Saccharomyces cerevisiae*) cytochrome *c* [PDB 1YCC]. **(e)** Bacterial (*Rhodopila globiformis*) cytochrome *c* [PDB 1HRO].



▲ **Figure 4.23**
Structural similarity of lactate and malate dehydrogenase. (a) *Bacillus stereothermophilus* lactate dehydrogenase [PDB 1LDN]. **(b)** *Escherichia coli* malate dehydrogenase [PDB 1EMD].

The enzymatic activities of lactate dehydrogenase and malate dehydrogenase are compared in Box 7.1.

suggesting that all modern proteins are descended from only a few thousand proteins that were present in the most primitive organisms living 3 billion years ago.

Lactate dehydrogenase and malate dehydrogenase are different enzymes that belong to the same family of proteins. Their structures are very similar as shown in Figure 4.23. The sequences of the proteins are only 23% identical. In spite of the obvious similarity in structure, Nevertheless, this level of sequence similarity is significant enough to conclude that the two proteins are homologous. They descend from a common ancestral gene that duplicated billions of years ago before the last common ancestor of all extant species of bacteria. Both lactate dehydrogenase and malate dehydrogenase are present in the same species which is why they are members of a family of related proteins. Protein families contain related proteins that are present in the same species. The cytochrome *c* proteins shown in Figure 4.22 are evolutionarily related but strictly speaking they are not members of a protein family because there is only one of them in each species. Protein familes arise from gene duplication events.

Protein domains can be classified by their structures. One commonly used classification scheme groups these domains into four categories. The "all-α" category contains domains that consist almost entirely of α helices and loops. "All-β" domains contain only β sheets and nonrepetitive structures that link β strands. The other two categories contain domains that have a mixture of α helices and β strands. Domains in the "α/β" class have supersecondary structures such as the $\beta\alpha\beta$ motif and others in which regions of α helix and β strand alternate in the polypeptide chain. In the "$\alpha + \beta$" category, the domains consist of local clusters of α helices and β sheet where each type of secondary structure arises from separate contiguous regions of the polypeptide chain.

Protein domains can be further classified by the presence of characteristic folds within each of the four main structural categories. A fold is a combination of secondary structures that form the core of a domain. Figure 4.24 on pages 103–104 shows selected examples of proteins from each of the main categories and illustrates a number of common domain folds. Some domains have easily recognizable folds, such as the β meander that contains antiparallel β strands connected by hairpin loops (Figure 4.20f), or helix bundles (Figure 4.19c). Other folds are more complex (Figure 4.25).

The important point about Figure 4.24 is not to memorize the structures of common proteins and folds. The key concept is that proteins can adopt an amazing variety of different sizes and shapes (tertiary structure) even though they contain only three basic forms of secondary structure.

C. Domain Structure, Function, and Evolution

The relationship between domain structure and function is complex. Often a single domain has a particular function such as binding small molecules or catalyzing a single reaction. In multifunctional enzymes, each catalytic activity can be associated with one of several domains found in a single polypeptide chain (Figure 4.24j). However, in many cases the binding of small molecules and the formation of the active site of an enzyme take place at the interface between two separate domains. These interfaces often form crevices, grooves, and pockets that are accessible on the surface of the protein. The extent of contact between domains varies from protein to protein.

The unique shapes of proteins, with their indentations, interdomain interfaces, and other crevices, allow them to fulfill dynamic functions by selectively and transiently binding other molecules. This property is best illustrated by the highly specific binding of reactants (substrates) to substrate-binding sites, or active sites, of enzymes. Because many binding sites are positioned toward the interior of a protein, they are relatively free of water. When substrates bind, they fit so well that some of the few remaining water molecules in the binding site are displaced.

D. Intrinsically Disordered Proteins

This section on tertiary structure wouldn't be complete without mentioning those proteins and domains that have no stable three-dimensional structure. These intrinsically disordered proteins (and domains) are quite common and the lack of secondary and tertiary structure is encoded in the amino acid sequences. There has been selection for

clusters of charged residues (positive or negative) and proline residues that maintain the polypeptide chain in a disordered state.

Many of these proteins interact with other proteins. They contain short amino acid sequences that serve as binding sites and these binding sites are within the intrinsically disordered regions. This allows easy access to the binding site. If a protein contains two different binding sites for other proteins then the disordered polypeptide chain acts as a tether to bring the two binding proteins closer together. Several transcription factors also contain disordered regions when they are not bound to DNA. These regions become ordered when the proteins interact with DNA.

KEY CONCEPT

There are only three basic types of secondary structure but thousands of tertiary folds and domains.

Speculations on the possible relationship between protein domains and gene organization will be presented in Chapter 21.

4.8 Quaternary Structure

Many proteins exhibit an additional level of organization called *quaternary structure*. Quaternary structure refers to the organization and arrangement of subunits in a protein with multiple subunits. Each subunit is a separate polypeptide chain. A multisubunit protein is referred to as an oligomer (proteins with only one polypeptide chain are monomers). The subunits of a multisubunit protein may be identical or different. When the subunits are identical, dimers and tetramers predominate. When the subunits differ, each type often has a different function. A common shorthand method for describing oligomeric proteins uses Greek letters to identify types of subunits and subscript numerals to indicate numbers of subunits. For example, an $\alpha_2\beta\gamma$ protein contains two subunits designated α and one each of subunits designated β and γ.

The subunits within an oligomeric protein always have a defined stoichiometry and the arrangement of the subunits gives rise to a stable structure where subunits are usually held together by weak noncovalent interactions. Hydrophobic interactions are the principal forces involved although electrostatic forces may contribute to the proper alignment of the subunits. Because intersubunit forces are usually rather weak, the subunits of an oligomeric protein can often be separated in the laboratory. *In vivo,* however, the subunits usually remain tightly associated.

Examples of several multisubunit proteins are shown in Figure 4.26. In the case of triose phosphate isomerase (Figure 4.26a) and HIV protease (Figure 4.26b), the identical subunits associate through weak interactions between the side chains found mainly in loop regions. Similar interactions are responsible for the formation of the MS2 capsid protein that consists of a trimer of identical subunits (Figure 4.26d). In this case, the trimer units assemble into a more complex structure—the bacteriophage particle. The enzyme HGPRT (Figure 4.26e) is a tetramer formed from the association of two pairs of nonidentical subunits. Each of the subunits is a recognizable domain.

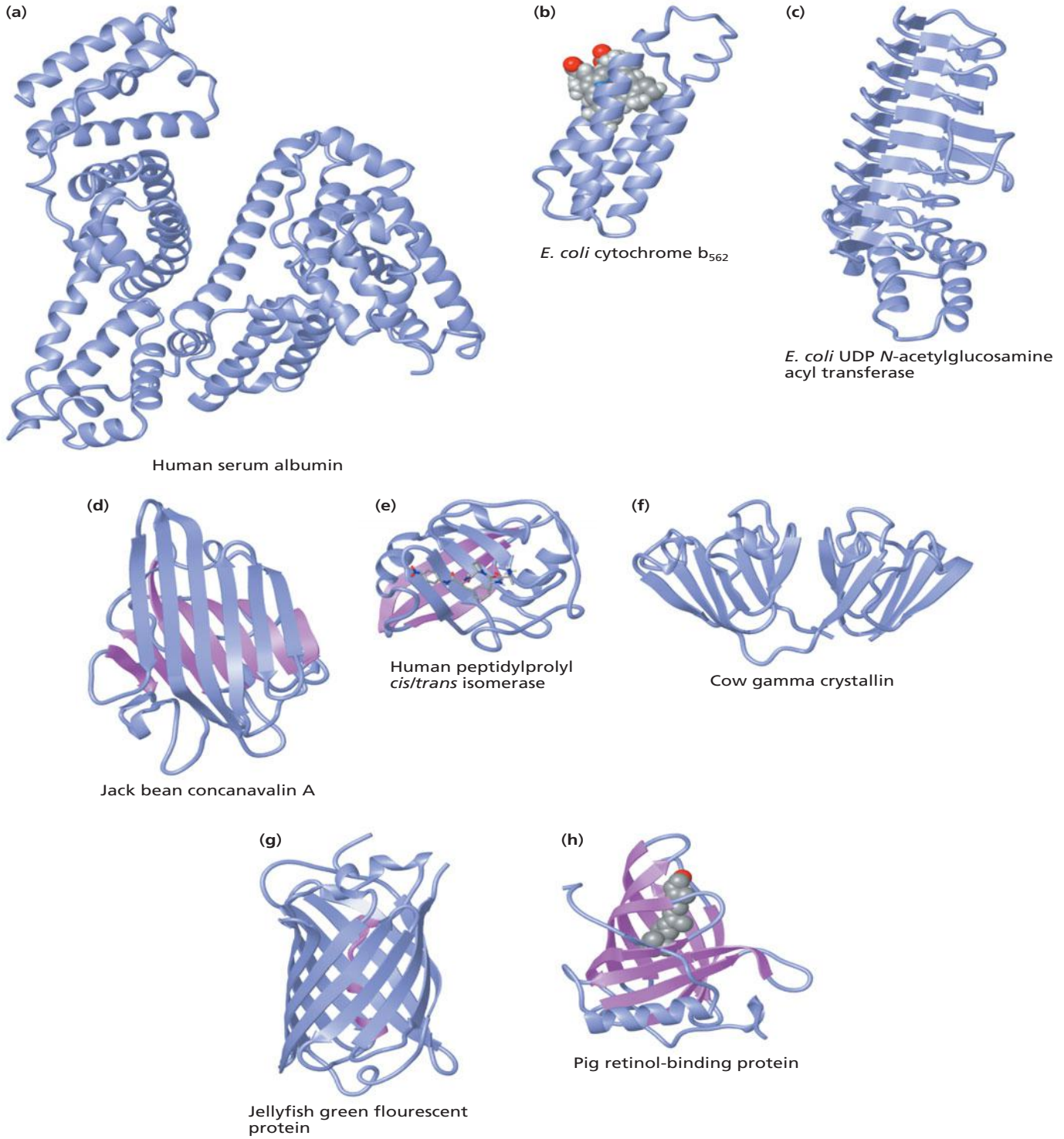
The potassium channel protein (Figure 4.26c) is an example of a tetramer of identical subunits where the subunits interact to form a membrane-spanning region consisting of an eight-helix bundle. The subunits do not form separate domains within the protein but instead come together to form a single channel. The bacterial photosystem shown in Figure 4.26f is a complex example of quaternary structure. Three of the subunits contribute to a large membrane-bound helix bundle while a fourth subunit (a cytochrome) sits on the exterior surface of the membrane.

The structures and functions of bacterial and plant photosystems are described in Chapter 15.

Determination of the subunit composition of an oligomeric protein is an essential step in the physical description of a protein. Typically, the molecular weight of the native oligomer is estimated by gel-filtration chromatography and then the molecular weight of each chain is determined by SDS–polyacrylamide gel electrophoresis (Section 3.6). For a protein having only one type of chain, the ratio of the two values provides the number of chains per oligomer.

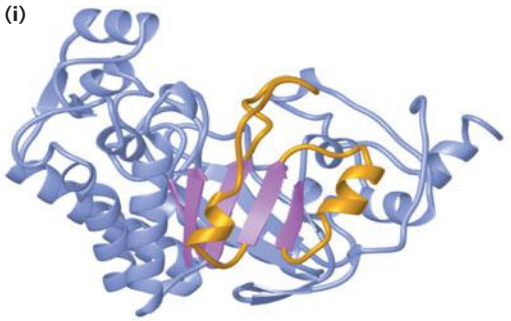
The fact that a large proportion of proteins consist of multiple subunits is probably related to several factors:

1. Oligomers are usually more stable than their dissociated subunits suggesting that quaternary structure prolongs the life of a protein *in vivo.*
2. The active sites of some oligomeric enzymes are formed by residues from adjacent polypeptide chains.

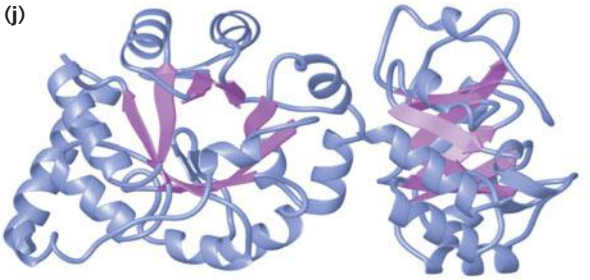


▲ Figure 4.24

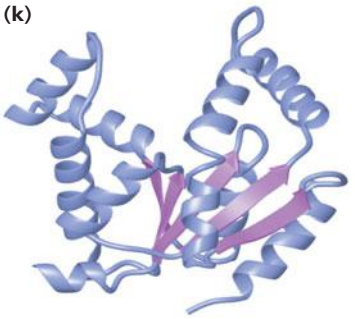
Examples of tertiary structure in selected proteins. (a) Human (*Homo sapiens*) serum albumin [PDB 1BJ5] (class: all-α). This protein has several domains consisting of layered α helices and helix bundles. **(b)** *Escherichia coli* cytochrome $\mathbf{b}_{562}$ [PDB 1QPU] (class: all-α). This is a heme-binding protein consisting of a single four-helix bundle domain. **(c)** *Escherichia coli* UDP N-acetylglucosamine acyl transferase [PDB 1LXA] (class: all-β). The structure of this enzyme shows a classic example of a β helix domain. **(d)** Jack bean (*Canavalia ensiformis*) concanavalin A [PDB 1CON] (class: all-β). This carbohydrate-binding protein (lectin) is a single-domain protein made up of a large β sandwich fold. **(e)** Human (*Homo sapiens*) peptidylprolyl *cis/trans* isomerase [PDB 1VBS] (class: all-β). The dominant feature of the structure is a β sandwich fold. **(f)** Cow (*Bos taurus*) γ-crystallin [PDB 1A45] (class: all-β) This protein contains two β barrel domains. **(g)** Jellyfish (*Aequorea victoria*) green fluorescent protein [PDB 1GFL] (class: all-β). This is a β barrel structure with a central α helix. The strands of the sheet are antiparallel. **(h)** Pig (*Sus scrofa*) retinol-binding protein [PDB 1AQB] (class: all-β). Retinol binds in the interior of a β barrel fold. **(I)** Brewer's yeast (*Saccharomyces carlsburgensis*) old yellow enzyme (FMN oxidoreductase) [PDB 1OYA] (class: α/β). The central fold is an α/β barrel with parallel β strands connected by α helices. Two of the connecting α helical regions are highlighted in yellow. **(j)** *Escherichia coli* enzyme required for tryptophan biosynthesis [PDB 1PII] (class: α/β). This is a bifunctional enzyme containing two distinct domains. Each domain is an example of an α/β barrel. The left-hand domain contains the indolglycerol phosphate



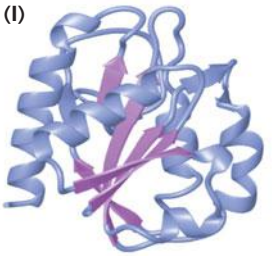
Yeast FMN oxidoreductase
(old yellow enzyme)



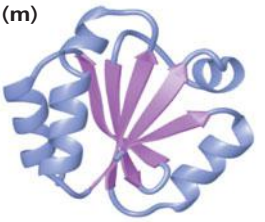
E. coli tryptophan biosynthesis enzyme



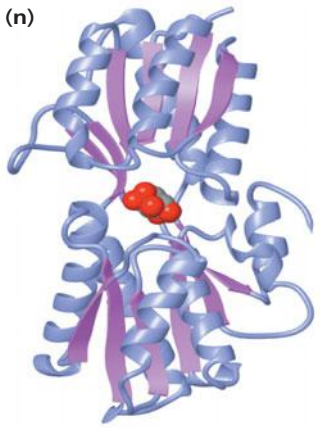
Pig adenylyl kinase



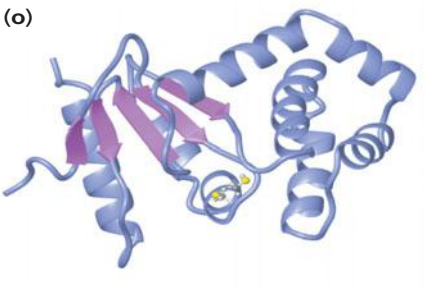
E. coli flavodoxin



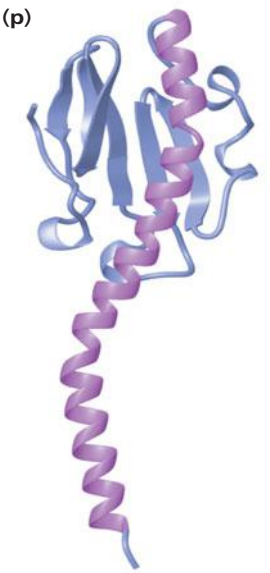
Human thioredoxin



E. coli L-arabinose-binding
protein



E. coli thiol-disulfide
oxidoreductase



Neisseria gonorrhea pilin

▲ Figure 4.24 (*continued*)
synthetase activity, and the right-hand domain contains the phosphoribosylanthranilate isomerase activity. **(k)** Pig (*Sus scrofa*) adenylyl kinase [PDB 3ADK] (class: α/β). This single-domain protein consists of a five-stranded parallel β sheet with layers of α helices above and below the sheet. The substrate binds in the prominent groove between α helices. **(l)** *Escherichia coli* flavodoxin [PDB 1AHN] (class: α/β). The fold is a five-stranded parallel twisted sheet surrounded by α helices. **(m)** Human (*Homo sapiens*) thioredoxin [PDB 1ERU] (class: α/β). The structure of this protein is very similar to that of *E. coli* flavodoxin except that the five-stranded twisted sheet in the thioredoxin fold contains a single antiparallel strand. **(n)** *Escherichia coli* L-arabinose-binding protein [PDB 1ABE] (class: α/β). This is a two-domain protein where each domain is similar to that in *E. coli* flavodoxin. The sugar L-arabinose binds in the cavity between the two domains. **(o)** *Escherichia coli* DsbA (thiol-disulfide oxidoreductase/disulfide isomerase) [PDB 1A23] (class: α/β). The predominant feature of this structure is a (mostly) antiparallel β sheet sandwiched between α helices. Cysteine side chains at the end of one of the α helices are shown (sulfur atoms are yellow). **(p)** *Neisseria gonorrhea* pilin [PDB 2PIL] (class: $\alpha + \beta$). This polypeptide is one of the subunits of the pili on the surface of the bacteria responsible for gonorrhea. There are two distinct regions of the structure: a β sheet and a long α helix.

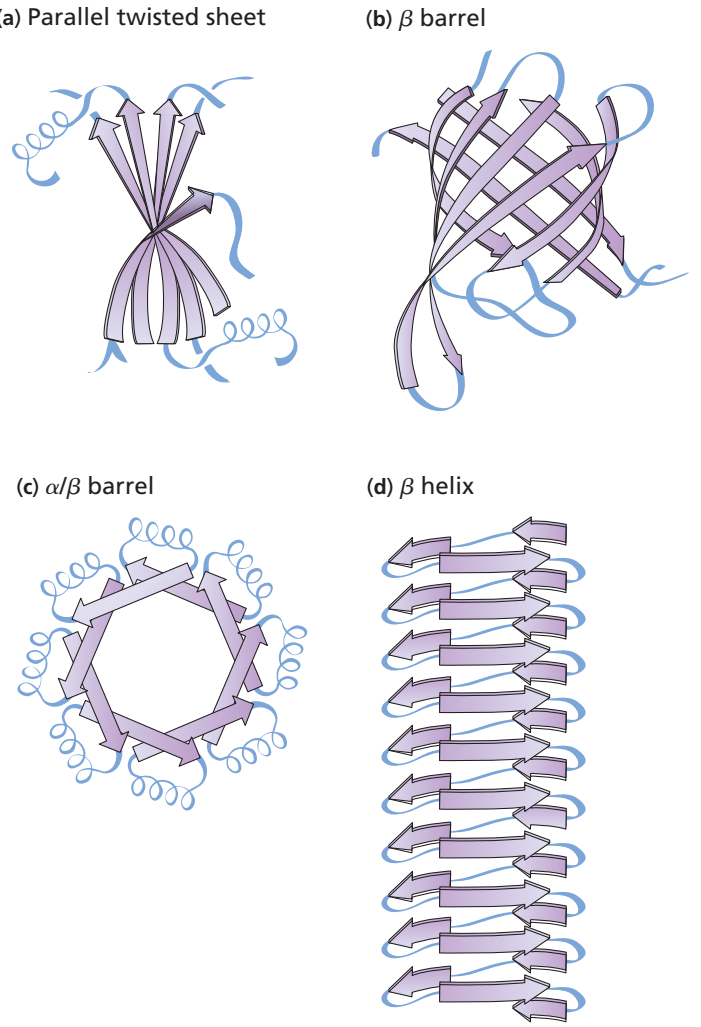
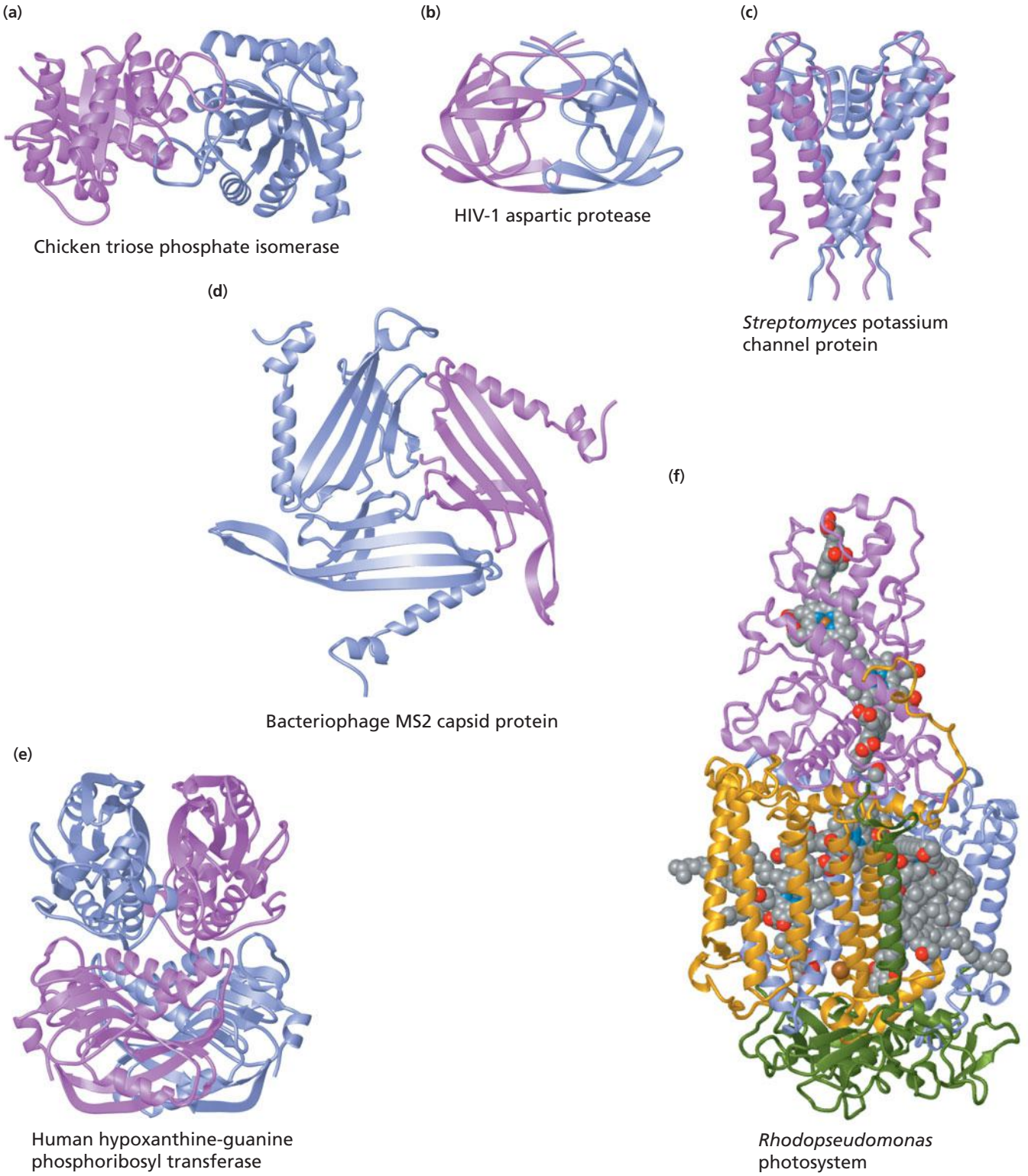


Figure 4.25 ►
Common domain folds.

3. The three-dimensional structures of many oligomeric proteins change when the proteins bind ligands. Both the tertiary structures of the subunits and the quaternary structures (i.e., the contacts between subunits) may be altered. Such changes are key elements in the regulation of the biological activity of certain oligomeric proteins.
4. Different proteins can share the same subunits. Since many subunits have a defined function (e.g., ligand binding), evolution has favored selection for different combinations of subunits to carry out related functions. This is more efficient than selection for an entirely new monomeric protein that duplicates part of the function.
5. A multisubunit protein may bring together two sequential enzymatic steps where the product of the first reaction becomes the substrate of the second reaction. This gives rise to an effect known as channeling (Section 5.11).

As shown in Figure 4.26, the variety of multisubunit proteins ranges from simple homodimers such as triose phosphate isomerase to large complexes such as the photosystems in bacteria and plants. We would like to know how many proteins are monomers and how many are oligomers but studies of cell proteomes—the complete complement of proteins —have only begun.

Table 4.1 on page 108 shows the results of a survey of *E. coli* proteins in the SWISS-PROT database. Of those polypeptides that have been analyzed, only about 19% are in monomers. Dimers are the largest class among the oligomers, and homodimers—where the two subunits are identical—represent 31% of all proteins. The next largest class is tetramers of identical subunits. Note that trimers are relatively rare. Most proteins exhibit dyad symmetry meaning that you can usually draw a line through a protein dividing it into two halves that are symmetrical about this axis. This dyad symmetry is seen even in



▲ **Figure 4.26**

Quaternary structure. (a) Chicken (*Gallus gallus*) triose phosphate isomerase [PDB 1TIM]. This protein has two identical subunits with α/β barrel folds. **(b)** HIV-1 aspartic protease [PDB 1DIF]. This protein has two identical all-β subunits that bind symmetrically. HIV protease is the target of many new drugs designed to treat AIDS patients. **(c)** *Streptomyces lividans* potassium channel protein [PDB 1BL8]. This membrane-bound protein has four identical subunits, each of which contributes to a membrane-spanning eight-helix bundle. **(d)** Bacteriophage MS2 capsid protein [PDB 2MS2]. The basic unit of the MS2 capsid is a trimer of identical subunits with a large β sheet. **(e)** Human (*Homo sapiens*) hypoxanthine-guanine phosphoribosyl transferase (HGPRT) [PDB 1BZY]. HGPRT is a tetrameric protein containing two different types of subunit. **(f)** *Rhodopseudomonas viridis* photosystem [PDB 1PRC]. This complex, membrane-bound protein has two identical subunits (orange, blue) and two other subunits (purple, green) bound to several molecules of photosynthetic pigments.

Table 4.1 Natural occurrence of oligomeric proteins in *Escherichia coli*

Oligomeric state	Number of homooligomers	Number of heterooligomers	Percent
Monomer	72		19.4
Dimer	115	27	38.2
Trimer	15	5	5.4
Tetramer	62	16	21.0
Pentamer	1	1	0.1
Hexamer	20	1	5.6
Heptamer	1	1	0.1
Octamer	3	6	2.4
Nonamer	0	0	0.0
Decamer	1	0	0.0
Undecamer	0	1	0.0
Dodecamer	4	2	1.6
Higher oligomers	8		2.2
Polymers	10		2.7

heterooligomers such as hypoxanthine-guanine phosphoribosyl transferase (HGPRT, Figure 4.26e) and hemoglobin (Section 4.14). Of course, there are many exceptions, especially when the oligomers are large complexes.

We will encounter many other examples of multisubunit proteins throughout this textbook, especially in the chapters on information flow (Chapters 20–22). DNA polymerase, RNA polymerase, and the ribosome are excellent examples. Other examples include GroEL (Section 4.11D) and pyruvate dehydrogenase (Section 13.1). Many of these large proteins are easily seen in electron micrographs, as illustrated in Figure 4.27.

Large complexes are referred to, metaphorically, as *protein machines* since the various polypeptide components work together to carry out a complex reaction. The term

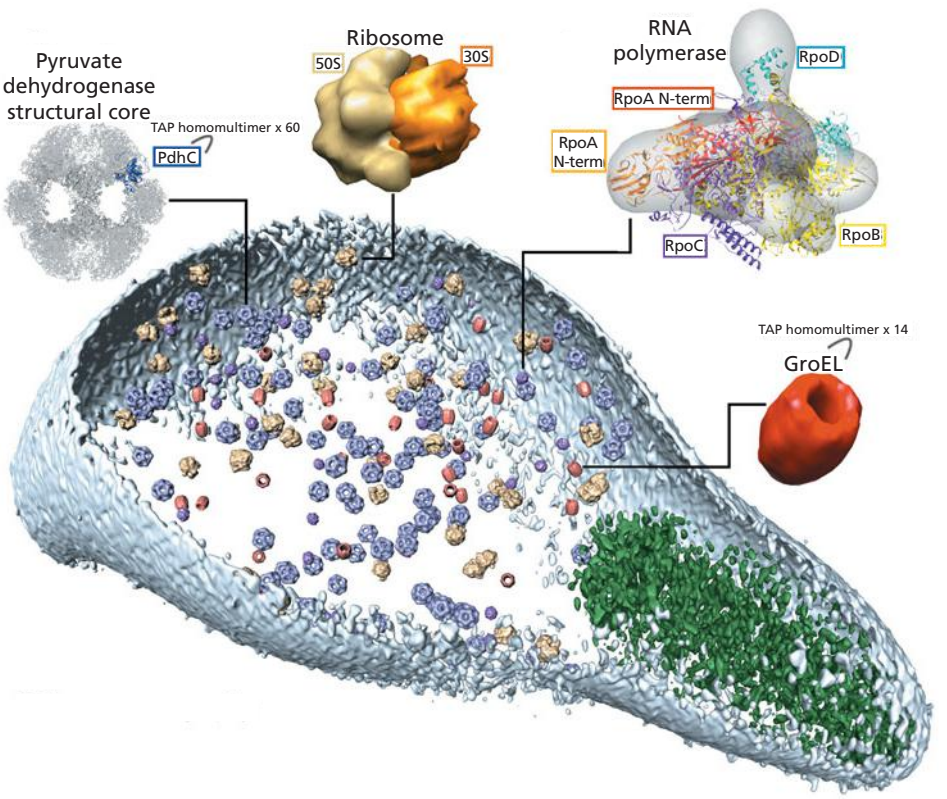
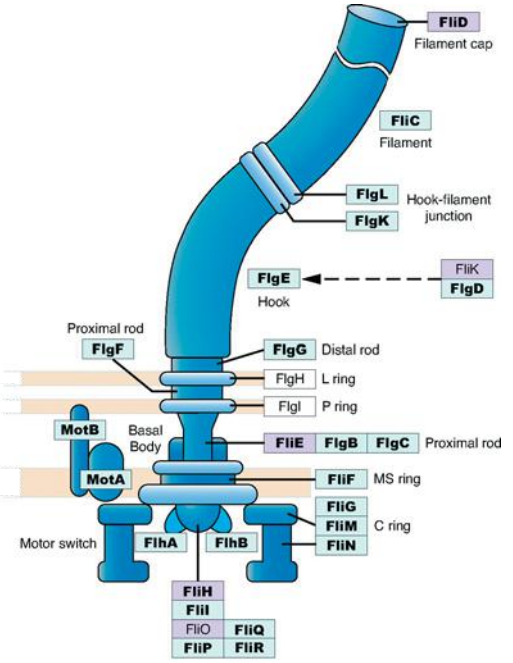


Figure 4.27 ►
Large protein complexes in the bacterium *Mycoplasma pneumoniae*. *M. pneumoniae* causes some forms of pneumonia in humans. This species has one of the smallest genomes known (689 protein-encoding genes). Most of those genes are likely to represent the minimum proteome of a living cell. The cell contains several large complexes found in all cells: pyruvate dehygrogenase (purple), ribosome (yellow), GroEL (red), and RNA polymerase (orange). It also contains a rod (green) found only in some bacteria. [Adapted from Kühner et al. (2009). Proteome organization in a genome-reduced bacterium. *Science* 326:1235–1240]

was originally coined to describe complexes such as the replisome (Figure 20.15) but there are many other examples, including those shown in Figure 4.27.

The bacterial flagellum (Figure 4.28) is a spectacular example of a protein machine. The complex drives the rotation of a long flagellum using protonmotive force as an energy source (Section 14.3). More than 50 genes are required to build the flagellum in *E. coli* but surveys of other bacteria reveal that there are only about 21 core proteins required to build a functional flagellum. The evolutionary history of this protein machine is being actively investigated and it appears that it was built up by combining simpler components involved in ATP synthesis and membrane secretion.



▲ **Figure 4.28**
Bacterial flagellum. The bacterial flagellum is a protein machine composed of 21 core subunits found in all species (blue boxes). Two additional subunits are missing in Firmicutes (white boxes) and five others are sporadically distributed. The flagellum (hook + filament + cap) spins as the motor complex rotates. The three layers represent the outer membrane (top), the peptidoglycan layer (middle), and the cytoplasmic membrane (bottom). (Courtesy of Howard Ochman.)

4.9 Protein–Protein Interactions

The various subunits in multisubunit proteins bind to each other so strongly that they rarely dissociate inside the cell. These protein–protein contacts are characterized by a number of weak interactions. We have already become familiar with the type of interactions involved: hydrogen bonds, charge–charge interactions, van der Waals forces, and hydrophobic interactions (Section 2.5). In some cases the contact areas between two subunits are localized to small patches on the surface of the polypeptides but while in other cases there can be extensive contact spread over large portions of the polypeptides. The distinguishing feature of subunit contacts is the cumulative effect of a large number of individual weak interactions giving a binding strength that is sufficient to keep the subunits together.

In addition to subunit–subunit contacts, there are many other types of protein–protein interactions that are less stable. These range from transient contacts between external proteins and receptors on the cell surface to weak interactions between various enzymes in metabolic pathways. These weak interactions are much more difficult to detect but they are essential components of many biochemical reactions.

Consider a simple interaction between two proteins, P1 and P2, to give a complex P1:P2. The equilibrium between the free and bound molecules can be described by either an association constant (K_a) or a dissociation constant (K_d) ($K_a = 1/K_d$).

$$\text{P1} + \text{P2} \rightleftharpoons \text{P1:P2} \qquad K_a = \frac{[\text{P1:P2}]}{[\text{P1}][\text{P2}]} \tag{4.1}$$

$$\text{P1:P2} \rightleftharpoons \text{P1} + \text{P2} \qquad K_d = \frac{[\text{P1}][\text{P2}]}{[\text{P1:P2}]} \tag{4.2}$$

Typical association constants for the binding of subunits in a multimeric protein are greater than $10^8\ M^{-1}$ ($K_a > 10^8\ M^{-1}$) and can range as high as $10^{14}\ M^{-1}$ for very tight interactions. At the other extreme are protein–protein interactions that are so weak they have no biological significance. These can be fortuitous interactions that arise from time to time because any two polypeptides will almost always form some kind of weak contact. The lower limit of relevant association constants is about $10^4\ M^{-1}$ ($K_a < 10^4\ M^{-1}$). The really interesting cases are those with association constants between these two values.

The binding of transcription factors to RNA polymerase is one example of weak protein–protein interactions that are very important. The association constants range from about $10^5\ M^{-1}$ to $10^7\ M^{-1}$. The interactions between proteins in signaling pathways also fall into this range as do the interactions between enzymes in metabolic pathways.

Let's look at what these association constants mean in terms of protein concentrations. As the concentrations of P1 and P2 increase it becomes more and more likely that they will interact and bind to each other. At some concentration, the rate of binding (a second-order reaction) becomes comparable to the rate of dissociation (a first-order reaction) and complexes will be present in appreciable amounts. Using the association constant, we can calculate the ratio of free polypeptide (P1 or P2) as a fraction of the total concentration of either one ($P1_T$ or $P2_T$). This ratio [free]/[total] tells us how much of the complex will be present at a given protein concentration.

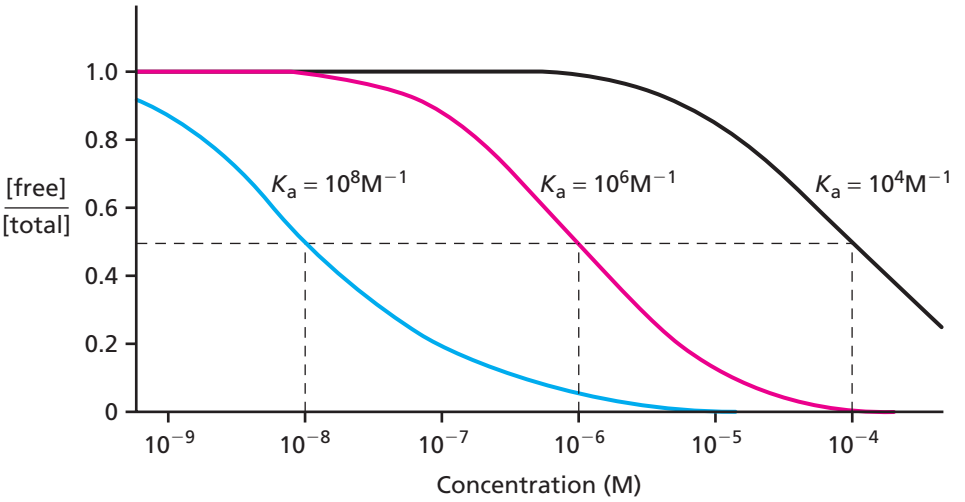


Figure 4.29 ►
Association constants and protein concentration. The ratio of free unbound protein to total protein is shown for a protein–protein interaction at three different association constants. Assuming that the concentration of the other component is in excess, the concentrations at which half the molecules are in complex and half are free corresponds to the reciprocal of the association constant. [Adapted from van Holde, Johnson, and Ho, *Principles of Physical Biochemistry*, Prentice Hall.]

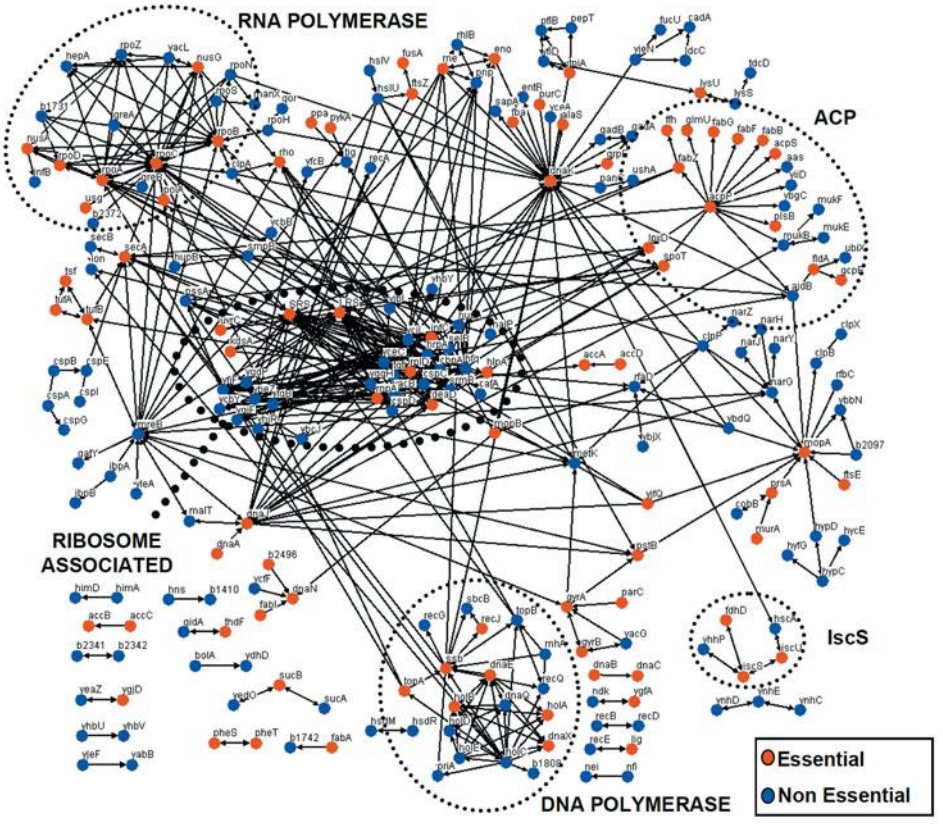
The curves in Figure 4.29 show these ratios for three different association constants corresponding to very weak ($K_a = 10^4\ M^{-1}$), moderate ($K_a = 10^6\ M^{-1}$), and very strong ($K_a = 10^8\ M^{-1}$) protein–protein interactions. If we assume that one of the components is present in excess, then the curves represent the concentrations of only the rate-limiting polypeptide. One can demonstrate mathematically that for simple systems the point at which half of the polypeptide is free and half is in a complex corresponds to the reciprocal of the association constant. For example, if $K_a = 10^8\ M^{-1}$ then most of the polypeptide will be bound at any concentration over 10^{-8} M.

What does this mean in terms of molecules per cell? For an *E. coli* cell whose volume is about 2×10^{-15} l it means that as long as there are more than a dozen molecules per cell the complex will be stable if $K_a > 10^8\ M^{-1}$. This is why large oligomeric complexes can exist in *E. coli* even if there are only a few dozen per cell. Most eukaryotic cells are 1000 times larger and there must be 12,000 molecules in order to achieve a concentration of 10^{-8} M. Figure 4.29 also shows why it is impossible for weak interactions to produce significant numbers of P1:P2 complexes. The protein concentration has to be greater than 10^{-4} M in order for the complex to be present in significant quantity and this concentration corresponds to 120,000 molecules in an *E. coli* cell or 120 million molecules in a eukaryotic cell. There are no free polypeptides present at such concentrations so weak interactions of this magnitude are biologically meaningless.

There are many techniques for detecting moderate binding. These include direct techniques such as affinity chromatopraphy, immunoprecipitation, and chemical cross-linking. Newer techniques rely on more sophisticated manipulations such as phage display, two-hybrid analysis, and genetic methods. Many workers are attempting to map the interactions of every protein in the cell using these techniques. An example of such an "interactome" for many *E. coli* proteins is shown in Figure 4.30. Note that strong interactions between the subunits of oligomers are easily detected as shown by lines connecting the subunits of RNA polymerase, the ribosome, and DNA polymerase. Other lines connect RNA polymerase to various transcription factors—these represent moderate interactions. Further studies of the "interactome" in various species should give us a much better picture of the complex protein–protein interactions in living cells.

4.10 Protein Denaturation and Renaturation

Environmental changes or chemical treatments may disrupt the native conformation of a protein causing loss of biological activity. Such a disruption is called **denaturation**. The amount of energy needed to cause denaturation is often small, perhaps equivalent to that needed for the disruption of three or four hydrogen bonds. Some proteins may unfold completely when denatured to form a random coil (a fluctuating chain considered to be totally disordered) but most denatured proteins retain considerable internal structure. It is sometimes possible to find conditions under which small denatured proteins can spontaneously renature, or refold, following denaturation.



◄ **Figure 4.30**
***E. coli* interactome.** Each point on the diagram represents a single *E. coli* protein. Red dots are essential proteins and blue dots are nonessential proteins. Lines joining the points indicate experimentally determined protein–protein interactions. Five large complexes are shown: RNA polymerase, DNA polymerase, ribosome and associated proteins, proteins interacting with cysteine desulfurase (IscS), and proteins associated with acyl carrier protein (ACP). (The role of ACP is described in Section 16.1.) [Adapted from Butland et al. (2005)]

Proteins are commonly denatured by heating. Under the appropriate conditions, a modest increase in temperature will result in unfolding and loss of secondary and tertiary structure. An example of thermal denaturation is shown in Figure 4.31. In this experiment, a solution containing bovine ribonuclease A is heated slowly and the structure of the protein is monitored by various techniques that measure changes in conformation. All of these techniques detect a change when denaturation occurs. In the case of bovine ribonuclease A, thermal denaturation also requires a reducing agent that disrupts internal disulfide bridges allowing the protein to unfold.

Denaturation takes place over a relatively small range of temperature. This indicates that unfolding is a cooperative process where the destabilization of just a few weak interactions leads to almost complete loss of native conformation. Most proteins have a characteristic "melting" temperature (T_m) that corresponds to the temperature at the midpoint of the transition between the native and denatured forms. The T_m depends on pH and the ionic strength of the solution.

Most proteins are stable at temperatures up to 50°C to 60°C under physiological conditions. Some species of bacteria, such as those that inhabit hot springs and the vicinity of deep ocean thermal vents, thrive at temperatures well above this range. Proteins in these species denature at much higher temperatures as expected. Biochemists are actively studying these proteins in order to determine how they resist denaturation.

Proteins can also be denatured by two types of chemicals—chaotropic agents and detergents (Section 2.4). High concentrations of chaotropic agents, such as urea and guanidinium salts (Figure 4.32), denature proteins by allowing water molecules to solvate nonpolar groups in the interior of proteins. The water molecules disrupt the hydrophobic interactions that normally stabilize the native conformation. The hydrophobic tails of

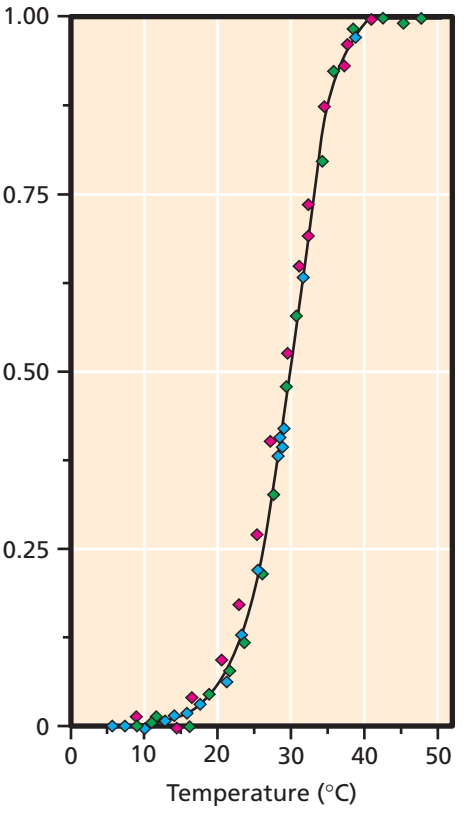
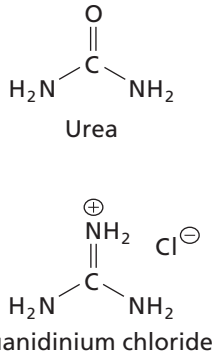
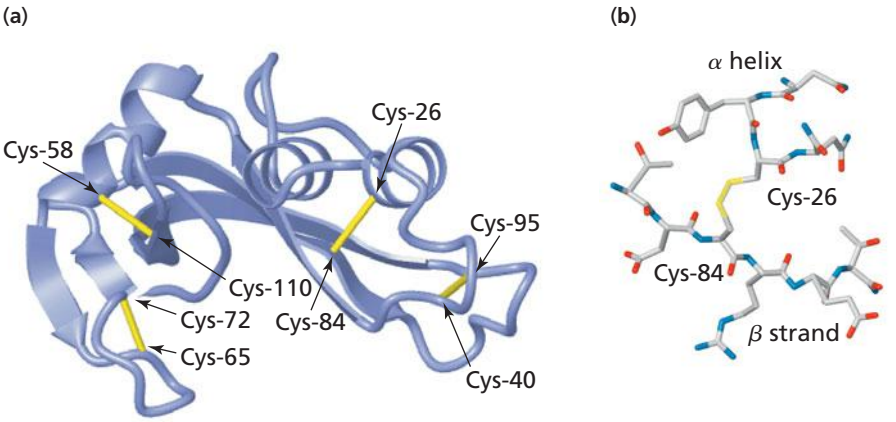


Figure 4.31 ►
Heat denaturation of ribonuclease A. A solution of ribonuclease A in 0.02 M KCl at pH 2.1 was heated. Unfolding was monitored by changes in ultraviolet absorbance (blue), viscosity (red), and optical rotation (green). The y-axis is the fraction of the molecule unfolded at each temperature. [Adapted from Ginsburg, A., and Carroll, W. R. (1965). Some specific ion effects on the conformation and thermal stability of ribonuclease. *Biochemistry* 4:2159–2174.



▲ **Figure 4.32**
Urea and guanidinium chloride.



▲ **Figure 4.33**
Disulfide bridges in bovine ribonuclease A. (a) Location of disulfide bridges in the native protein. **(b)** View of the disulfide bridge between Cys-26 and Cys-84 [PDB 2AAS].

detergents, such as sodium dodecyl sulfate (Figure 2.8), also denature proteins by penetrating the protein interior and disrupting hydrophobic interactions.

The native conformation of some proteins (e.g., ribonuclease A) is stabilized by disulfide bonds. Disulfide bonds are not generally found in intracellular proteins but are sometimes found in proteins that are secreted from cells. The presence of disulfide bonds stabilizes proteins by making them less susceptible to unfolding and subsequent degradation when they are exposed to the external environment. Disulfide bond formation does not drive protein folding; instead, the bonds form where two cysteine residues are appropriately located once the protein has folded. Formation of a disulfide bond requires oxidation of the thiol groups of the cysteine residues (Figure 3.4), probably by disulfide-exchange reactions involving oxidized glutathione, a cysteine-containing tripeptide.

Figure 4.33a shows the locations of the disulfide bridges in ribonuclease A. (Compare this orientation of the protein with that shown in Figure 4.3.) There are four disulfide bridges. They can link adjacent β strands, β strands to α helices, or β strands to loops. Figure 4.33b is a view of the disulfide bridge between a cysteine residue in an α helix (Cys-26) and a cysteine residue in a β strand (Cys-84). Note that the S—S bond does not align with the cysteine side chains. Disulfide bridges will form whenever the two cysteine sulfhydryl groups are in close proximity in the native conformation.

The numbering convention for amino acid residues in a polypeptide starts at the N-terminal end (Section 3.5). Cys-26 is the 26th residue from the N-terminus.

Complete denaturation of proteins containing disulfide bonds requires cleavage of these bonds in addition to disruption of hydrophobic interactions and hydrogen bonds. 2-Mercaptoethanol or other thiol reagents can be added to a denaturing medium in order to reduce any disulfide bonds to sulfhydryl groups (Figure 4.34). Reduction of the disulfide bonds of a protein is accompanied by oxidation of the thiol reagent.

In a series of classic experiments, Christian B. Anfinsen and his coworkers studied the renaturation pathway of ribonuclease A that had been denatured in the presence of thiol reducing agents. Since ribonuclease A is a relatively small protein (124 amino acid

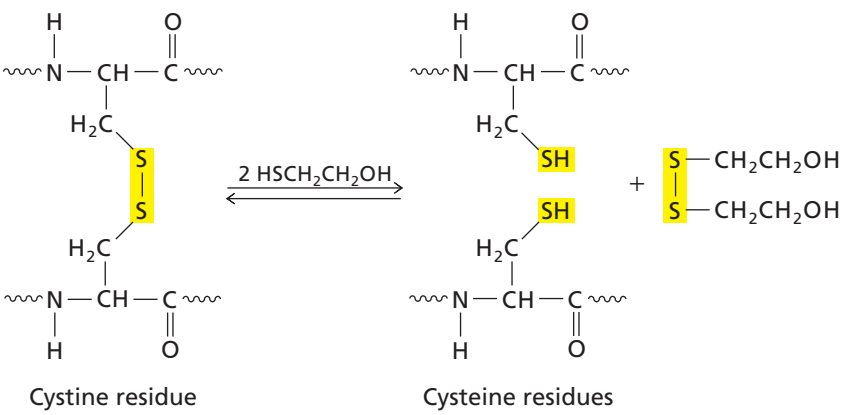


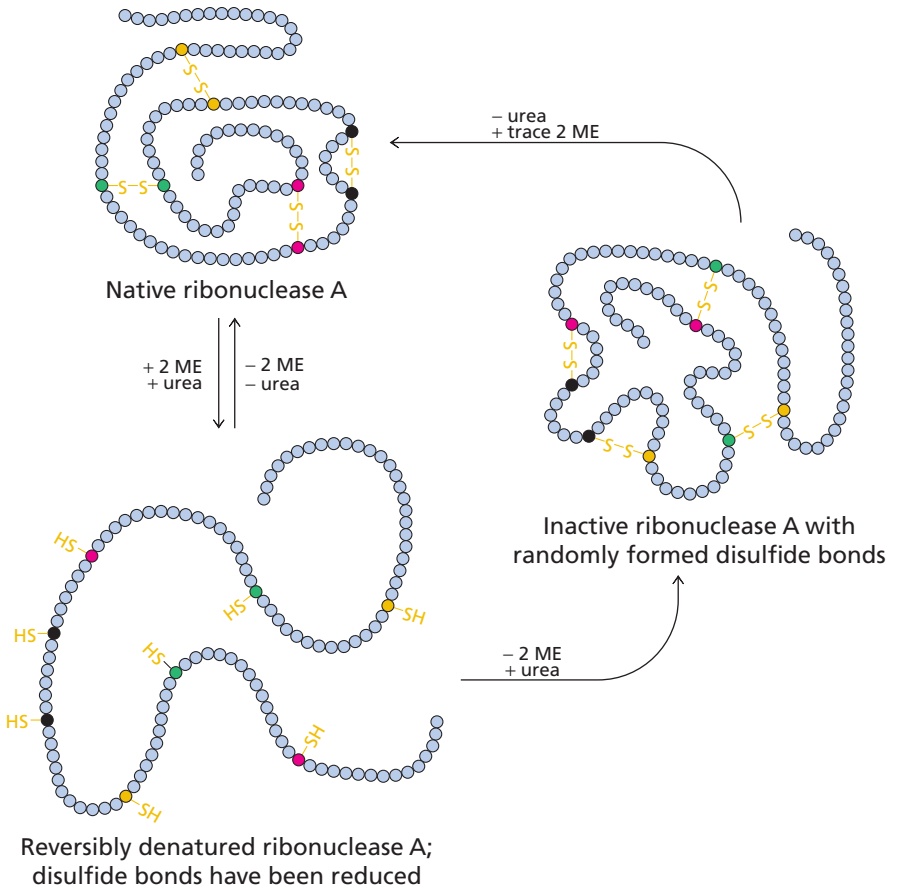
Figure 4.34 ►
Cleaving disulfide bonds. When a protein is treated with excess 2-mercaptoethanol ($HSCH_2CH_2OH$), a disulfide-exchange reaction occurs in which each cystine residue is reduced to two cysteine residues and 2-mercaptoethanol is oxidized to a disulfide.

residues), it refolds (renatures) quickly once it is returned to conditions where the native form is stable (e.g., cooled below the melting temperature or removed from a solution containing chaotropic agents). Anfinsen was among the first to show that denatured proteins can refold spontaneously to their native conformation indicating that the information required for the native three-dimensional conformation is contained in the amino acid sequence of the polypeptide chain. In other words, the primary structure determines the tertiary structure.

Denaturation of ribonuclease A with 8 M urea containing 2-mercaptoethanol results in complete loss of tertiary structure and enzymatic activity and yields a polypeptide chain containing eight sulfhydryl groups (Figure 4.35). When 2-mercaptoethanol is removed and oxidation is allowed to occur in the presence of urea, the sulfhydryl groups pair randomly so that only about 1% of the protein population forms the correct four disulfide bonds recovering original enzymatic activity. (If the eight sulfhydryl groups pair randomly, 105 disulfide-bonded structures are possible—7 possible pairings for the first bond, 5 for the second, 3 for the third, and 1 for the fourth ($7 \times 5 \times 3 \times 1 = 105$)—but only one of these structures is correct.) However, when urea and 2-mercaptoethanol are removed simultaneously and dilute solutions of the reduced protein are then exposed to air, ribonuclease A spontaneously regains its native conformation, its correct set of disulfide bonds, and its full enzymatic activity. The inactive proteins containing randomly formed disulfide bonds can be renatured if urea is removed, a small amount of 2-mercaptoethanol is added, and the solution gently warmed. Anfinsen's experiments demonstrate that the correct disulfide bonds can form only after the protein folds into its native conformation. Anfinsen concluded that the renaturation of ribonuclease A is spontaneous, driven entirely by the free energy gained in changing to the stable physiological conformation. This conformation is determined by the primary structure.

Proteins occasionally adopt a nonnative conformation and form inappropriate disulfide bridges when they fold inside a cell. Anfinsen discovered an enzyme, called protein disulfide isomerase (PDI), that catalyzes reduction of these incorrect bonds. All

▲ **Christian B. Anfinsen (1916–1995).** Anfinsen was awarded the Nobel Prize in Chemistry in 1972 for his work on the refolding of proteins.



◄ **Figure 4.35**
Denaturation and renaturation of ribonuclease A. Treatment of native ribonuclease A (top) with urea in the presence of 2-mercaptoethanol unfolds the protein and disrupts disulfide bonds to produce reduced, reversibly denatured ribonuclease A (bottom). When the denatured protein is returned to physiological conditions in the absence of 2-mercaptoethanol, it refolds into its native conformation and the correct disulfide bonds form. However, when 2-mercaptoethanol alone is removed, ribonuclease A reoxidizes in the presence of air, but the disulfide bonds form randomly, producing inactive protein (such as the form shown on the right). When urea is removed, a trace of 2-mercaptoethanol is added to the randomly reoxidized protein, and the solution is warmed gently, the disulfide bonds break and re-form correctly to produce native ribonuclease A.

living cells contain such an activity. The enzyme contains two reduced cysteine residues positioned in the active site. When the misfolded protein binds, the enzyme catalyzes a disulfide-exchange reaction whereby the disulfide in the misfolded protein is reduced and a new disulfide bridge is created between the two cysteine residues in the enzyme. The misfolded protein is then released and it can refold into the low-energy native conformation. The structure of the reduced form of *E. coli* disulfide isomerase (DsbA) is shown in Figure 4.24o.

4.11 Protein Folding and Stability

New polypeptides are synthesized in the cell by a translation complex that includes ribosomes, mRNA, and various factors (Chapter 21). As the newly synthesized polypeptide emerges from the ribosome, it folds into its characteristic three-dimensional shape. Folded proteins occupy a low-energy well that makes the native structure much more stable than alternative conformations (Figure 4.36). The *in vitro* experiments of Anfinsen and many other biochemists demonstrate that many proteins can fold spontaneously to reach this low-energy conformation. In this section we discuss the characteristics of those proteins that fold into a stable three-dimensional structure.

It is thought that as a protein folds the first few interactions trigger subsequent interactions. This is an example of cooperative effects in protein folding—the phenomenon whereby the formation of one part of a structure leads to the formation of the remaining parts of the structure. As the protein begins to fold, it adopts lower and lower energies and begins to fall into the energy well shown in Figure 4.36. The protein may become temporarily trapped in a local energy well (shown as small dips in the energy diagram) but eventually it reaches the energy minimum at the bottom of the well. In its final, stable, conformation, the native protein is much less sensitive to degradation than an extended, unfolded polypeptide chain. Thus, native proteins can have half-lives of many cell generations and some molecules may last for decades.

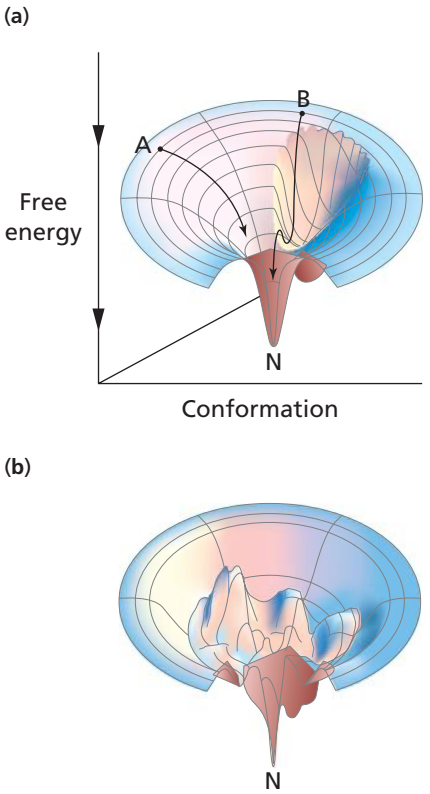
Folding is extremely rapid—in most cases the native conformation is reached in less than a second. Protein folding and stabilization depend on several noncovalent forces including the hydrophobic effect, hydrogen bonding, van der Waals interactions, and charge–charge interactions. Although noncovalent interactions are weak individually, collectively they account for the stability of the native conformations of proteins. The weakness of each noncovalent interaction gives proteins the resilience and flexibility to undergo small conformational changes. (Covalent disulfide bonds also contribute to the stability of certain proteins.)

In multidomain proteins the different domains fold independently of one another as much as possible. One of the reasons for limitations on the size of a domain (usually < 200 residues) is that large domains would fold too slowly if domains were larger than 300 residues. The rate of spontaneous folding would be too slow to be useful.

No actual protein-folding pathway has yet been described in detail but current research is focused on intermediates in the folding pathways of a number of proteins. Several hypothetical folding pathways are shown in Figure 4.37. During protein folding, the polypeptide collapses upon itself due to the hydrophobic effect and elements of secondary structure begin to form. This intermediate is called a molten globule. Subsequent steps involve rearrangement of the backbone chain to form characteristic motifs and, finally, the stable native conformation.

The mechanism of protein folding is one of the most challenging problems in biochemistry. The process is spontaneous and must be largely determined by the primary structure (sequence) of the polypeptide. It should be possible, therefore, to predict the structure of a protein from knowledge of its amino acid sequence. Much progress has been made in recent years by modeling the folding process using fast computers.

In the remainder of this section, we examine the forces that stabilize protein structure in more detail. We will also describe the role of chaperones in protein folding.



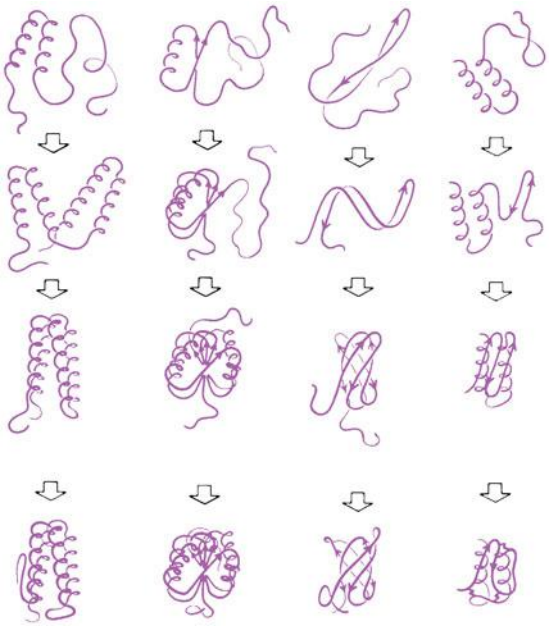
▲ **Figure 4.36**
Energy well of protein folding. The funnels represent the free-energy potential of folding proteins. **(a)** A simplified funnel showing two possible pathways to the low-energy native protein. In path B, the polypeptide enters a local low-energy minimum as it folds. **(b)** A more realistic version of the possible free-energy forms of a folding protein with many local peaks and dips.

KEY CONCEPT

Most proteins fold spontaneously into a conformation with the lowest energy.

A. The Hydrophobic Effect

Proteins are more stable in water when their hydrophobic side chains are aggregated in the protein interior rather than exposed on the surface to the aqueous medium. Because



◀ **Figure 4.37**
Hypothetical protein-folding pathways. The initially extended polypeptide chains form partial secondary structures, then approximate tertiary structures, and finally the unique native conformations. The arrows within the structures indicate the direction from the N- to the C-terminus.

water molecules interact more strongly with each other than with the nonpolar side chains of a protein, the side chains are forced to associate with one another causing the polypeptide chain to collapse into a more compact molten globule. The entropy of the polypeptide decreases as it becomes more ordered. This decrease is more than offset by the increase in solvent entropy as water molecules that were previously bound to the protein are released. (Folding also disrupts extended cages of water molecules surrounding hydrophobic groups.) This overall increase in the entropy of the system provides the major driving force for protein folding.

Whereas nonpolar side chains are driven into the interior of the protein, most polar side chains remain in contact with water on the surface of the protein. The sections of the polar backbone that are forced into the interior of a protein neutralize their polarity by hydrogen bonding to each other, often generating secondary structures. Thus, the hydrophobic nature of the interior not only accounts for the association of hydrophobic residues but also contributes to the stability of helices and sheets. Studies of folding pathways indicate that hydrophobic collapse and formation of secondary structures occur simultaneously

Localized examples of this hydrophobic effect are the interactions of the hydrophobic side of an amphipathic α helix with the protein core (Section 4.4) and the hydrophobic region between β sheets in the β-sandwich structure (Section 4.5). Most of the examples shown in Figures 4.25 and 4.26 contain juxtaposed regions of secondary structure that are stabilized by hydrophobic interactions between the side chains of hydrophobic amino acid residues.

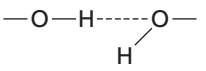
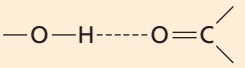
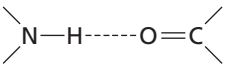
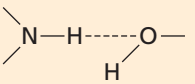
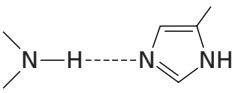
B. Hydrogen Bonding

Hydrogen bonds contribute to the cooperativity of folding and help stabilize the native conformations of proteins. The hydrogen bonds in α helices, β sheets, and turns are the first to form, giving rise to defined regions of secondary structure. The final native structure also contains hydrogen bonds between the polypeptide backbone and water, between the polypeptide backbone and polar side chains, between two polar side chains, and between polar side chains and water. Table 4.2 shows some of the many types of hydrogen bonds found in proteins along with their typical bond lengths. Most hydrogen bonds in proteins are of the N—H—O type. The distance between the donor and acceptor atoms varies from 0.26 to 0.34 nm and the bonds may deviate from linearity by up to 40°. Recall that hydrogen bonds within the hydrophobic core of a protein are much more stable than those that form near the surface because the internal hydrogen bonds don't compete with water molecules.

KEY CONCEPT

Entropically driven reactions are reactions where the most important thermodynamic change is an increase in entropy of the system. We can say that the system is much more disordered at the end of the reaction than at the beginning. In the case of hydrophobic interactions, the change in entropy is mostly due to the release of ordered water molecules that shield hydrophobic groups (Section 2.5D).

Table 4.2 Examples of hydrogen bonds in proteins

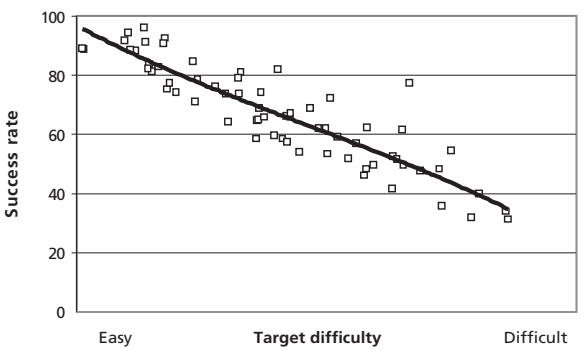
Type of hydrogen bond		Typical distance between donor and acceptor atom (nm)
Hydroxyl-hydroxyl		0.28
Hydroxyl-carbonyl		0.28
Amide-carbonyl		0.29
Amide-hydroxyl		0.30
Amide-imidazole nitrogen		0.31

BOX 4.2 CASP: THE PROTEIN FOLDING GAME

The basic principles of protein folding are reasonably well understood and it seems certain that if a protein has a stable three-dimensional structure it will be determined largely by the primary structure (sequence). This has led to efforts to predict tertiary structure from knowing the amino acid sequence. Biochemists have made huge advances in this theoretical work in the last 30 years.

The value of such work has to be assessed by making predictions of the structure of unknown proteins. This led in 1996 to the beginning of CASP–Critical Assessment of Methods of Protein Structure Prediction. This is a sort of game with no prizes other than the honor of being successful. Protein folding groups are given the amino acid sequences of a number of targets and asked to predict the three-dimensional structure. The targets are drawn from those proteins whose structures have just been determined but the data haven't yet been published. Contestants have only a few weeks to send in their predictions before the actual structures become known.

The results of the 2008 CASP round are shown in the figure. There were 121 targets and thousands of predictions were submitted. Success ranged from nearly 100% for easy proteins to only about 30% for difficult ones. ("Easy" targets are those where the Protein Data Bank (PDB) already contains the structures of several homologous proteins. "Difficult" targets are proteins with new folds that have never been solved.) The success rate for moderately difficult targets has climbed over the years as the prediction methods improved, but there's plenty of opportunity to make winning predictions at the very difficult end of the scale.



C. Van der Waals Interactions and Charge–Charge Interactions

Van der Waals contacts between nonpolar side chains also contribute to the stability of proteins. The extent of stabilization due to optimized van der Waals interactions is difficult to determine. The cumulative effect of many van der Waals interactions probably makes a significant contribution to stability because nonpolar side chains in the interior of a protein are densely packed.

Charge–charge interactions between oppositely charged side chains may make a small contribution to the stability of proteins but most ionic side chains are found on the surfaces where they are solvated and can contribute only minimally to the overall stabilization of the protein. Nevertheless, two oppositely charged ions occasionally form an ion pair in the interior of a protein. Such ion pairs are much stronger than those exposed to water.

D. Protein Folding Is Assisted by Molecular Chaperones

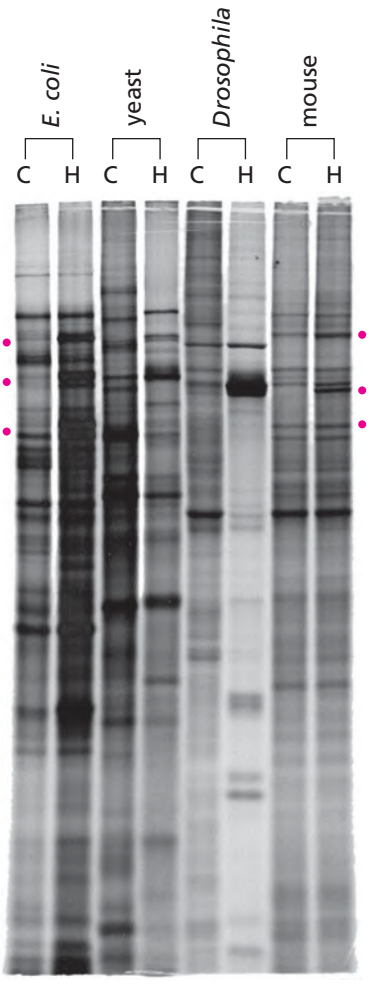
Studies of protein folding have led to two general observations regarding the folding of polypeptide chains into biologically active proteins. First, protein folding does not involve a random search in three-dimensional space for the native conformation. Instead, protein folding appears to be a cooperative, sequential process in which formation of the first few structural elements assists in the alignment of subsequent structural features. [The need for cooperativity is illustrated by a calculation made by Cyrus Levinthal. Consider a polypeptide of 100 residues. If each residue had three possible conformations that could interconvert on a picosecond time scale then a random search of all possible conformations for the complete polypeptide would take 10^{87} seconds—many times the estimated age of the universe (6×10^{17} seconds)!]

Second, to a first approximation the folding pattern and the final conformation of a protein depend on its primary structure. (Many proteins bind metal ions and coenzymes as described in Chapter 7. These external ligands are also required for proper folding.) As we saw in the case of ribonuclease A, simple proteins may fold spontaneously into their native conformations in a test tube without any energy input or assistance. Larger proteins will also fold spontaneously into their native structures since the final conformation represents the minimal free energy form. However, larger proteins are more likely to become temporarily trapped in a local energy well of the type illustrated in Figure 4.36b. The presence of such metastable incorrect conformations at best slows the rate of protein folding and at worst causes the folding intermediates to aggregate and fall out of solution. In order to overcome this problem inside the cell, the rate of correct protein folding is enhanced by a group of ubiquitous special proteins called **molecular chaperones**.

Chaperones increase the *rate* of correct folding of some proteins by binding newly synthesized polypeptides before they are completely folded. They prevent the formation of incorrectly folded intermediates that may trap the polypeptide in an aberrant form. Chaperones can also bind to unassembled protein subunits to prevent them from aggregating incorrectly and precipitating before they are assembled into a complete multisubunit protein.

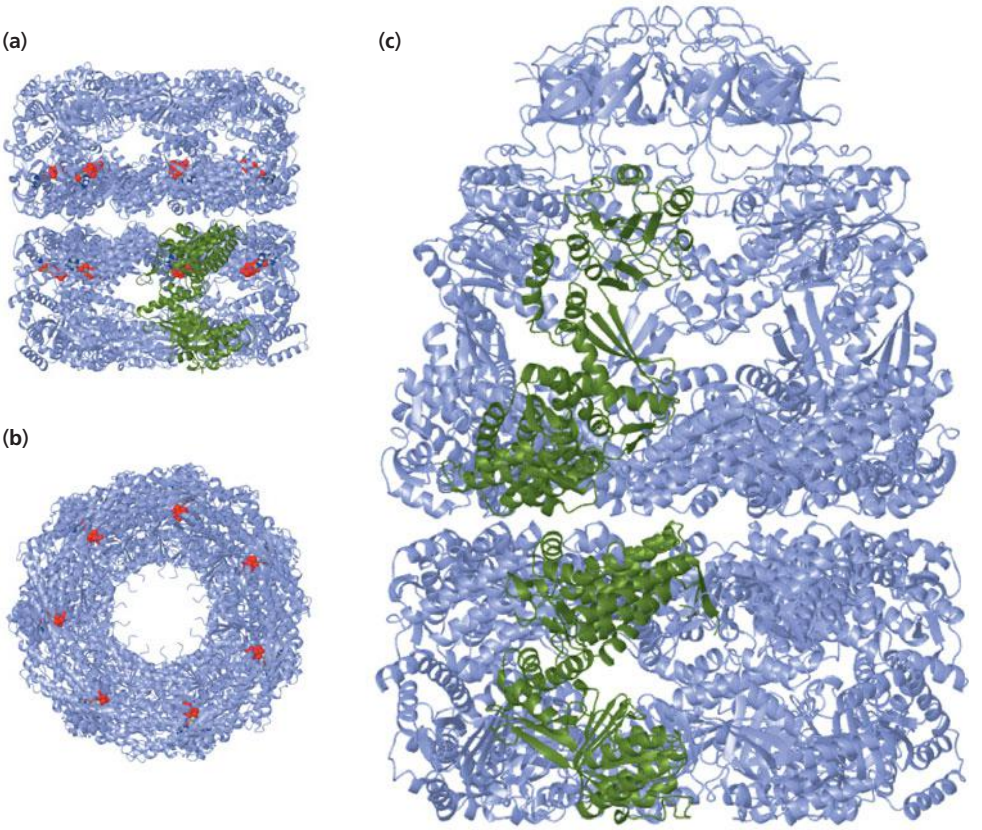
There are many different chaperones. Most of them are heat shock proteins—proteins that are synthesized in response to temperature increases (heat shock) or other changes that cause protein denaturation *in vivo*. The role of heat shock proteins—now recognized as chaperones—is to repair the damage caused by temperature increases by binding to denatured proteins and helping them to refold rapidly into their native conformation.

The major heat shock protein is Hsp70 (heat shock protein, M_r = 70,000). This protein is present in all species except for some species of archaebacteria. In bacteria, it is also called DnaK. The normal role of the chaperone Hsp70 is to bind to nascent



▶ **Heat shock proteins.** Proteins were synthesized for a short time in the presence of radioactive amino acids then run on an SDS–polyacrylamide gel. The gel was exposed to film to detect radioactive proteins. The resulting autoradiograph shows only those proteins that were labeled during the time of exposure to radioactive amino acids. Lanes "C" are proteins synthesized at normal growth temperatures, and lanes "H" are proteins synthesized during a short heat shock where cells are shifted to a temperature a few degrees above their normal growth temperature. The induction of heat shock proteins (chaperones) in four different species is shown. Red dots indicate major heat shock proteins: top = Hsp90, middle = Hsp70, bottom = Hsp60(GroEL).

► **Figure 4.38**
***Escherichia coli* chaperonin (GroE).** The core structure consists of two identical rings composed of seven GroEL subunits. Unfolded proteins bind to the central cavity. Bound ATP molecules can be identified by their red oxygen atoms. **(a)** Side view. **(b)** Top view showing the central cavity. [PDB 1DER]. **(c)** During folding the size of the central cavity of one of the rings increases and the end is capped by a protein containing seven GroES subunits. [PDB 1AON].

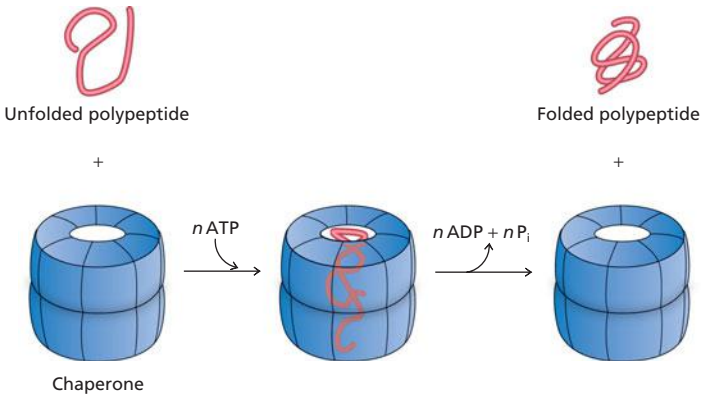


proteins while they are being synthesized in order to prevent aggregation or entrapment in a local low-energy well. The binding and release of nascent polypeptides is coupled to the hydrolysis of ATP and usually requires additional accessory proteins. Hsp70/DnaK is one of the most highly conserved proteins known in all of biology. This indicates that chaperone-assisted protein folding is an ancient and essential requirement for efficient synthesis of proteins with the correct three-dimensional structure.

Another important and ubiquitous chaperone is called chaperonin (also called GroE in bacteria). Chaperonin is also a heat shock protein (Hsp60) that plays an important and essential role in assisting normal protein folding inside the cell.

E. coli chaperonin is a complex multisubunit protein. The core structure consists of two rings containing seven identical GroEL subunits. Each subunit can bind a molecule of ATP (Figure 4.38a). A simplified version of chaperonin-assisted folding is shown in Figure 4.39 . Unfolded proteins bind to the hydrophobic central cavity enclosed by the rings. When folding is complete, the protein is released by hydrolysis of the bound ATP molecules. The actual pathway is more complicated and requires an additional component that serves as a cap sealing one end of the central cavity while the folding process takes place.

Figure 4.39 ►
Chaperonin-assisted protein folding. The unfolded polypeptide enters the central cavity of chaperonin, where it folds. The hydrolysis of several ATP molecules is required for chaperonin function.



The cap contains seven GroES subunits forming an additional ring (Figure 4.38c). The conformation of the GroEL ring can be altered during folding to increase the size of the cavity and the role of the cap is to prevent the unfolded protein from being released prematurely.

As mentioned earlier, some proteins tend to aggregate during folding in the absence of chaperones. Aggregation is probably due to temporary formation of hydrophobic surfaces on folding intermediates. The intermediates bind to each other and the result is that they are taken out of solution and are no longer able to explore the conformations represented by the energy funnel shown in Figure 4.36. Chaperonins isolate polypeptide chains in the folding cavity and thus prevent folding intermediates from aggregating. The folding cavity serves as an "Anfinsen cage" that allows the chain to reach the correct low-energy conformation without interference from other folding intermediates.

The central cavity of chaperonin is large enough to accommodate a polypeptide chain of about 630 amino acid residues (M_r = 70,000). Thus, the folding of most small and medium-sized proteins can be assisted by chaperonin. However, only about 5% to 10% of *E. coli* proteins (i.e., about 300 different proteins) appear to interact with chaperonin during protein synthesis. Medium-sized proteins and those of the α/β structural class are more likely to require chaperonin-assisted folding. Smaller proteins are able to fold quickly on their own. Many of the remaining proteins in the cell require other chaperones, such as HSP70/DnaK.

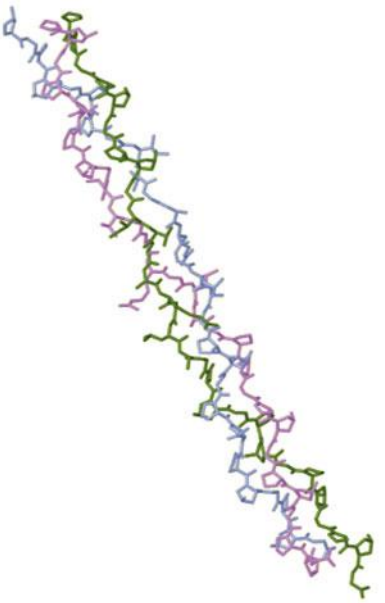
Chaperones appear to inhibit incorrect folding and assembly pathways by forming stable complexes with surfaces on polypeptide chains that are exposed only during synthesis, folding, and assembly. Even in the presence of chaperones, protein folding is spontaneous; for this reason, chaperone-assisted protein folding has been described as assisted self-assembly.

4.12 Collagen, a Fibrous Protein

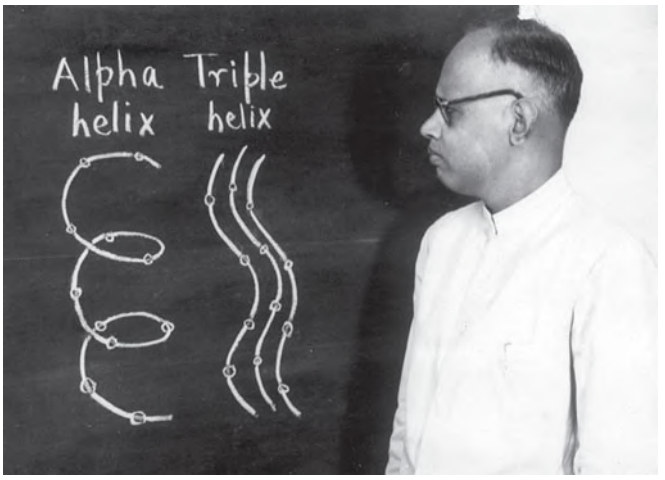
To conclude our examination of the three-dimensional structure of proteins, we examine several proteins to see how their structures are related to their biological functions. The proteins selected for more detailed study are the structural protein collagen, the oxygen-binding proteins myoglobin and hemoglobin (Sections 4.12 to 4.13), and antibodies (Section 4.14).

Collagen is the major protein component of the connective tissue of vertebrates. It makes up about 30% of the total protein in mammals. Collagen molecules have remarkably diverse forms and functions. For example, collagen in tendons forms stiff, ropelike fibers of tremendous tensile strength whereas in skin, collagen takes the form of loosely woven fibers permitting expansion in all directions.

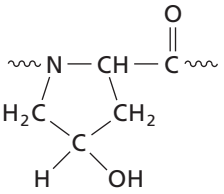
The structure of collagen was worked out by G. N. Ramachandran (famous for his Ramachandran plots, Section 4.3). The molecule consists of three left-handed helical chains coiled around each other to form a right-handed supercoil (Figure 4.40).



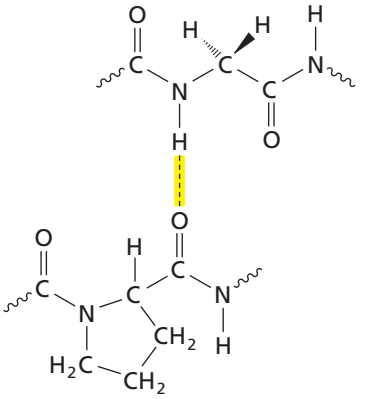
▲ **Figure 4.40**
The human type III collagen triple helix. The extended region of collagen contains three identical subunits (purple, light blue, and green). Three left-handed collagen helices are coiled around one another to form a right-handed supercoil. [PDB 1BKV]



◄ **G.N. Ramachandran (1922–2001).** In this photograph he is illustrating the difference between an α helix and the left-handed triple helix of collagen. Note that he has deliberately drawn the α helix as a left-handed helix and not the standard right-handed form found in most proteins.



▲ **Figure 4.41**
4-Hydroxyproline residue. 4-Hydroxyproline residues are formed by enzyme-catalyzed hydroxylation of proline residues.



◄ **Figure 4.42**
Interchain hydrogen bonding in collagen. The amide hydrogen of a glycine residue in one chain is hydrogen-bonded to the carbonyl oxygen of a residue, often proline, in an adjacent chain.

Each left-handed helix in collagen has 3.0 amino acid residues per turn and a pitch of 0.94 nm giving a rise of 0.31 nm per residue. Consequently, a collagen helix is more extended than an α helix and the coiled-coil structure of collagen is not the same as the coiled-coil motif discussed in Section 4.7. (Several proteins unrelated to collagen also form similar three-chain supercoils.)

The collagen triple helix is stabilized by interchain hydrogen bonds. The sequence of the protein in the helical region consists of multiple repeats of the form –Gly–X–Y–, where X is often proline and Y is often a modified proline called 4-hydroxyproline (Figure 4.41). The glycine residues are located along the central axis of the triple helix, where tight packing of the protein strands can accommodate no other residue. For each –Gly–X–Y– triplet, one hydrogen bond forms between the amide hydrogen atom of glycine in one chain and the carbonyl oxygen atom of residue X in an adjacent chain (Figure 4.42). Hydrogen bonds involving the hydroxyl group of hydroxyproline may also stabilize the collagen triple helix. Unlike the more common α helix, the collagen helix has no intrachain hydrogen bonds.

In addition to hydroxyproline, collagen contains an additional modified amino acid residue called 5-hydroxylysine (Figure 4.43). Some hydroxylysine residues are covalently bonded to carbohydrate residues, making collagen a glycoprotein. The role of this glycosylation is not known.

The requirement for vitamin C is explained in Section 7.9.

Hydroxyproline and hydroxylysine residues are formed when specific proline and lysine residues are hydroxylated after incorporation into the polypeptide chains of collagen. The hydroxylation reactions are catalyzed by enzymes and require ascorbic acid (vitamin C). Hydroxylation is impaired in the absence of vitamin C, and the triple helix of collagen is not assembled properly.

The limited conformational flexibility of proline and hydroxyproline residues prevents the formation of α helices in collagen chains and also makes collagen somewhat rigid. (Recall that proline is almost never found in α helices.) The presence of glycine residues at every third position allows collagen chains to form a tightly wound left-handed helix that accommodates the proline residues. (Recall that the flexibility of glycine residues tends to disrupt the right-handed α helix.)

Collagen triple helices aggregate in a staggered fashion to form strong, insoluble fibers. The strength and rigidity of collagen fibers result in part from covalent

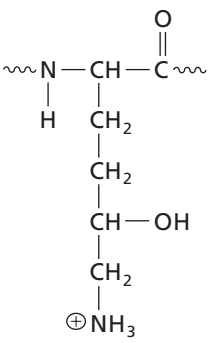
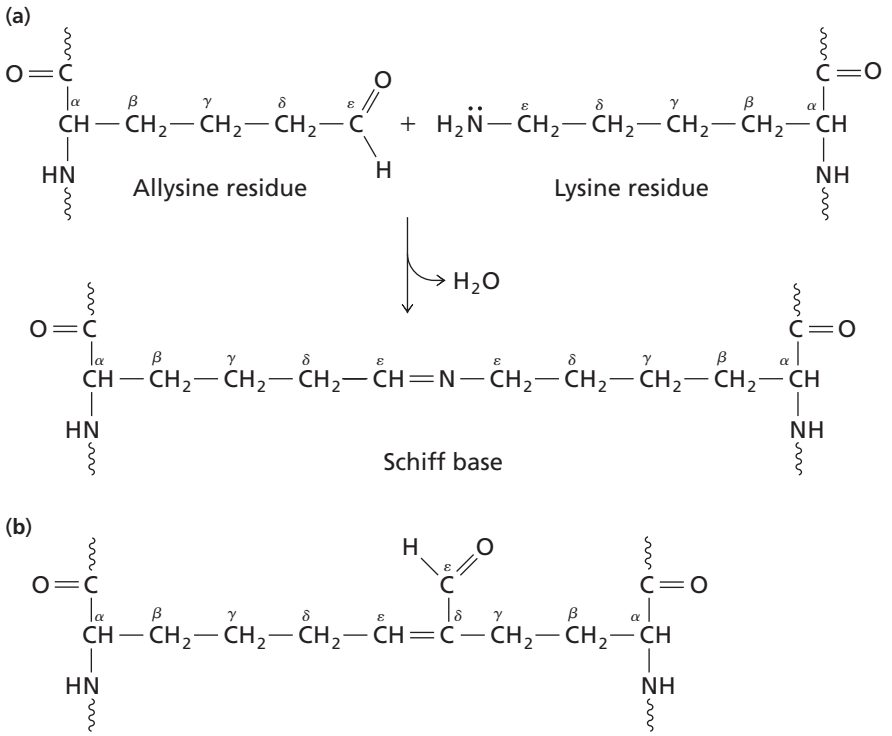


Figure 4.43 ►
5-Hydroxylysine residue. 5-Hydroxylysine residues are formed by enzyme-catalyzed hydroxylation of lysine residues.



◄ **Figure 4.44**
Covalent cross-links in collagen. (a) An allysine residue condenses with a lysine residue to form an intermolecular Schiff-base cross-link. **(b)** Two allysine residues condense to form an intramolecular cross-link.

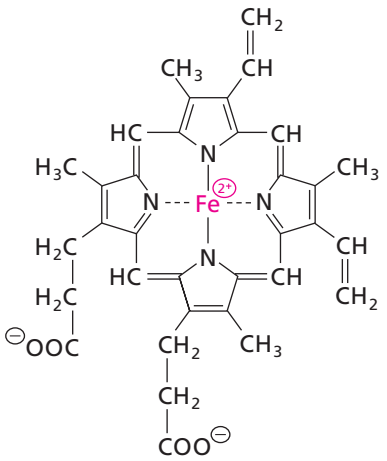
cross-links. The $—CH_2NH_3^+$ groups of the side chains of some lysine and hydroxylysine residues are converted enzymatically to aldehyde groups (—CHO), producing allysine and hydroxyallysine residues. Allysine residues (and their hydroxy derivatives) react with the side chains of lysine and hydroxylysine residues to form **Schiff bases**, complexes formed between carbonyl groups and amines (Figure 4.44a). These Schiff bases usually form between collagen molecules. Allysine residues also react with other allysine residues by aldol condensation to form cross-links, usually between the individual strands of the triple helix (Figure 4.44b). Both types of cross-links are converted to more stable bonds during the maturation of tissues, but the chemistry of these conversions is unknown.

BOX 4.3 STRONGER THAN STEEL

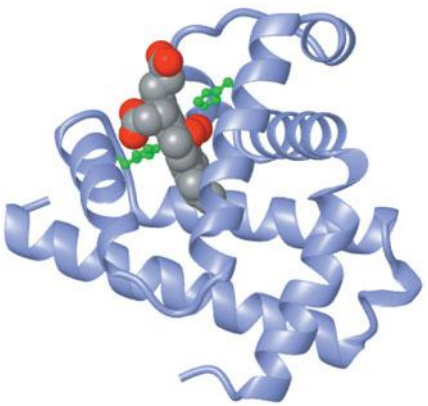
Not all fibrous proteins are composed of α helices. Silk is composed of a number of proteins that are predominantly β strands. The dragline silk of the spider, *Nephila clavipes*, for example, contains two proteins called spidroin 1 and spidroin 2. Both proteins contain multiple stretches of alanine residues separated by residues that are mostly glycine. The structure of this silk is not known in spite of major efforts by many laboratories. However, it is known that the proteins contain extensive regions of β strands.

There are many different kinds of spider silk and spiders have specialized glands for each type. The silk fiber produced by the major ampulate gland is called dragline silk; it is the fiber that spiders use to drop out of danger or anchor their webs. This silk fiber is quite literally stronger than steel cable. Materials manufactured from dragline silk would be very useful in a number of applications, one of which would be personal armor because dragline silk is stronger than Kevlar. So far it has not been possible to make significant amounts of silk in the laboratory without relying on spiders.

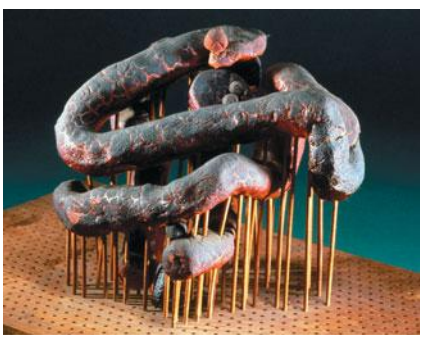
***Nephila clavipes*, the golden silk spider.** ►



▲ **Figure 4.45**
Chemical structure of the Fe(II)-protoporphyrin IX heme group in myoglobin and hemoglobin. The porphyrin ring provides four of the six ligands that surround the iron atom.



▲ **Figure 4.46**
Sperm whale (*Physeter catodon*) oxymyoglobin. Myoglobin consists of eight α helices. The heme prosthetic group binds oxygen (red). His-64 (green) forms a hydrogen bond with oxygen, and His-93 (green) is complexed to the iron atom of the heme. [PDB 1A6M].



▲ John Kendrew's original model of myoglobin determined from his X-ray diffraction data in the 1950s. The model is made of plasticine. It was the first three-dimensional model of a protein.

4.13 Structures of Myoglobin and Hemoglobin

Like most proteins, myoglobin (Mb) and the related protein hemoglobin (Hb) carry out their biological functions by selectively and reversibly binding other molecules—in this case, molecular oxygen (O_2). Myoglobin is a relatively small monomeric protein that facilitates the diffusion of oxygen in vertebrates. It is responsible for supplying oxygen to muscle tissue in reptiles, birds, and mammals. Hemoglobin is a larger tetrameric protein that carries oxygen in blood.

The red color associated with the oxygenated forms of myoglobin and hemoglobin (e.g., the red color of oxygenated blood) is due to a heme prosthetic group (Figure 4.45). (A prosthetic group is a protein-bound organic molecule essential for the activity of the protein.) Heme consists of a tetrapyrrole ring system (protoporphyrin IX) complexed with iron. The four pyrrole rings of this system are linked by methene (—CH═) bridges so that the unsaturated porphyrin is highly conjugated and planar. The bound iron is in the ferrous, or Fe^{2+}, oxidation state; it forms a complex with six ligands, four of which are the nitrogen atoms of protoporphyrin IX. (Other proteins, such as cytochrome *a* and cytochrome *c*, contain different porphyrin/heme groups.)

Myoglobin is a member of a family of proteins called globins. The tertiary structure of sperm whale myoglobin shows that the protein consists of a bundle of eight α helices (Figure 4.46). It is a member of the all-α structural category. The globin fold has several groups of α helices that form a layered structure. Adjacent helices in each layer are tilted at an angle that allows the side chains of the amino acid residues to interdigitate.

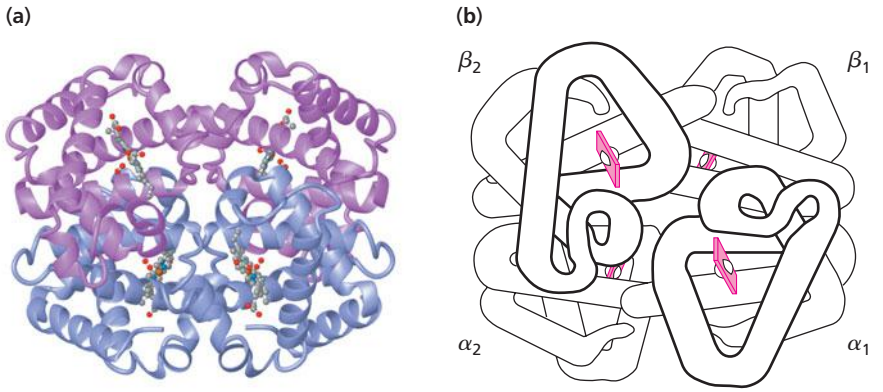
The interior of myoglobin is made up almost exclusively of hydrophobic amino acid residues, particularly those that are highly hydrophobic—valine, leucine, isoleucine, phenylalanine, and methionine. The surface of the protein contains both hydrophilic and hydrophobic residues. As is the case with most proteins, the tertiary structure of myoglobin is stabilized by hydrophobic interactions within the core. Folding of the polypeptide chain is driven by the energy minimization that results from formation of this hydrophobic core.

The heme prosthetic group of myoglobin occupies a hydrophobic cleft formed by three α helices and two loops. The binding of the porphyrin moiety to the polypeptide is due to a number of weak interactions including hydrophobic interactions, van der Waals contacts, and hydrogen bonds. There are no covalent bonds between the porphyrin and the amino acid side chains of myoglobin. The iron atom of heme is the site of oxygen binding as shown in Figure 4.46. Two histidine residues interact with the iron atom and the bound oxygen. Accessibility of the heme group to molecular oxygen depends on slight movement of nearby amino acid side chains. We will see later that the hydrophobic crevices of myoglobin and hemoglobin are essential for the reversible binding of oxygen.

In vertebrates, O_2 is bound to molecules of hemoglobin for transport in red blood cells, or erythrocytes. Viewed under a microscope, a mature mammalian erythrocyte is a biconcave disk that lacks a nucleus or other internal membrane-enclosed compartments (Figure 4.47). A typical human erythrocyte is filled with approximately 3×10^8 hemoglobin molecules.

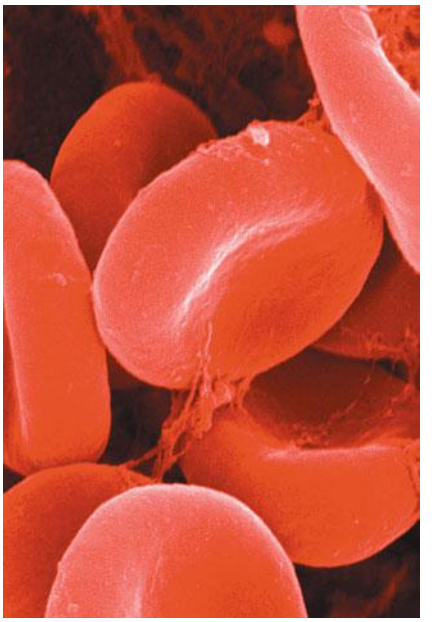
Hemoglobin is more complex than myoglobin because it is a multisubunit protein. In adult mammals, hemoglobin contains two different globin subunits called *α-globin* and *β-globin*. Hemoglobin is an $\alpha_2\beta_2$ tetramer—it contains two α chains and two β chains. Each of these globin subunits is similar in structure and sequence to myoglobin reflecting their evolution from a common ancestral globin gene in primitive chordates.

Each of the four globin subunits contains a heme prosthetic group identical to that found in myoglobin. The α and β subunits face each other across a central cavity (Figure 4.48). The tertiary structure of each of the four chains is almost identical to that of myoglobin (Figure 4.49). The α chain has seven α helices, and the β chain has eight. (Two short α helices found in β-globin and myoglobin are fused into one larger one in α-globin) Hemoglobin, however, is not simply a tetramer of myoglobin molecules. Each α chain interacts extensively with a β chain so hemoglobin is actually a dimer of $\alpha\beta$ subunits. We will see in the following section that the presence of multiple subunits is responsible for oxygen-binding properties that are not possible with single-chain myoglobin.



▲ **Figure 4.48**
Human (*Homo sapiens*) oxyhemoglobin. (a) Structure of human oxyhemoglobin showing two α and two β subunits. Heme groups are shown as stick models. [PDB 1HND]. **(b)** Schematic diagram of the hemoglobin tetramer. The heme groups are red.

▼ **Figure 4.47**
Scanning electron micrograph of mammalian erythrocytes. Each cell contains approximately 300 million hemoglobin molecules. The cells have been artificially colored.



4.14 Oxygen Binding to Myoglobin and Hemoglobin

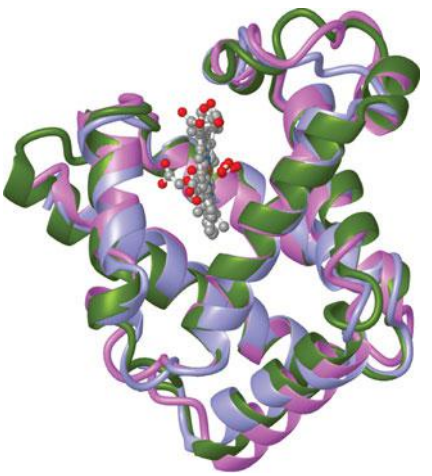
The oxygen-binding activities of myoglobin and hemoglobin provide an excellent example of how protein structure relates to physiological function. These proteins are among the most intensely studied proteins in biochemistry. They were the first complex proteins whose structure was determined by X-ray crystallography (Section 4.2). A number of the principles described here for oxygen-binding proteins also hold true for the enzymes that we will study in Chapters 5 and 6. In this section we examine the chemistry of oxygen binding to heme, the physiology of oxygen binding to myoglobin and hemoglobin, and the regulatory properties of hemoglobin.

A. Oxygen Binds Reversibly to Heme

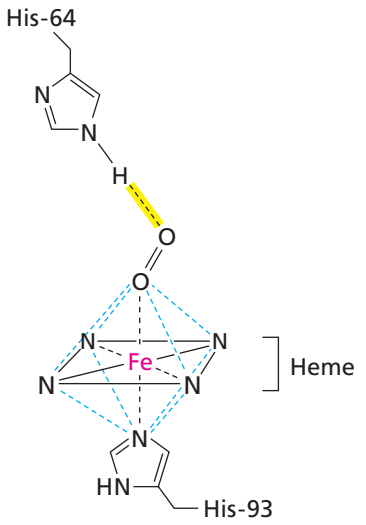
We will use myoglobin as an example of oxygen binding to the heme prosthetic group but the same principles apply to hemoglobin. The reversible binding of oxygen is called **oxygenation**. Oxygen-free myoglobin is called *deoxymyoglobin* and the oxygen-bearing molecule is called *oxymyoglobin*. (The two forms of hemoglobin are called *deoxyhemoglobin* and *oxyhemoglobin*.)

Some substituents of the heme prosthetic group are hydrophobic—this feature allows the prosthetic group to be partially buried in the hydrophobic interior of the myoglobin molecule. Recall from Figure 4.46 that there are two polar residues, His-64 and His-93, situated near the heme group. In oxymyoglobin, six ligands are coordinated to the ferrous iron, with the ligands in octahedral geometry around the metal cation (Figures 4.50 and 4.51). Four of the ligands are the nitrogen atoms of the tetrapyrrole ring system; the fifth ligand is an imidazole nitrogen from His-93 (called the proximal histidine); and the sixth ligand is molecular oxygen bound between the iron and the imidazole side chain of His-64 (called the distal histidine). In deoxymyoglobin, the iron is coordinated to only five ligands because oxygen is not present. The nonpolar side chains of Val-68 and Phe-43, shown in Figure 4.51, contribute to the hydrophobicity of the oxygen-binding pocket and help hold the heme group in place. Several side chains block the entrance to the heme-containing pocket in both oxymyoglobin and deoxymyoglobin. The protein structure in this region must vibrate, or breathe, rapidly to allow oxygen to bind and dissociate.

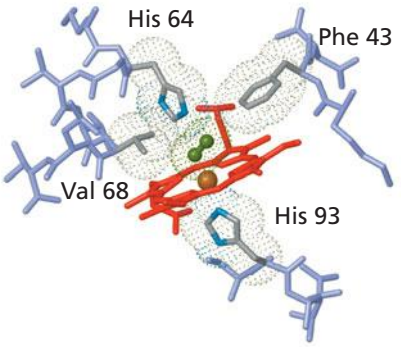
The hydrophobic crevice of the globin polypeptide holds the key to the ability of myoglobin and hemoglobin to suitably bind and release oxygen. Free heme does not reversibly bind oxygen in aqueous solution; instead, the Fe^{2+} of the heme is almost instantly oxidized to Fe^{3+}. (Oxidation is equivalent to the loss of an electron, as described in Section 6.1C. Reduction is the gain of an electron. Oxidation and reduction refer to the transfer of electrons and not to the presence or absence of oxygen molecules.)



▲ **Figure 4.49**
Tertiary structure of myoglobin, α-globin, and β-globin. The orientations of the individual α-globin and β-globin subunits of hemoglobin have been shifted in order to reveal the similarities in tertiary structure. The three structures have been superimposed. All of the structures are from the oxygenated forms shown in Figures 4.46 and 4.48. Color code: α-globin (blue), β-globin (purple), myoglobin (green).



▲ **Figure 4.50**
Oxygen-binding site of sperm whale oxymyoglobin. The heme prosthetic group is represented by a parallelogram with a nitrogen atom at each corner. The blue dashed lines illustrate the octahedral geometry of the coordination complex.



▲ **Figure 4.51**
The oxygen-binding site in sperm whale myoglobin. Fe(II) (orange) lies in the plane of the heme group. Oxygen (green) is bound to the iron atom and the amino acid side chain of His-64. Val-68 and Phe-43 contribute to the hydrophobic environment of the oxygen-binding site. [PDB 1AGM].

The structure of myoglobin and hemoglobin prevents the permanent transfer of an electron or irreversible oxidation thereby ensuring the reversible binding of molecular oxygen for transport. The ferrous iron atom of heme in hemoglobin is partially oxidized when O_2 is bound. An electron is temporarily transferred toward the oxygen atom that is attached to the iron so that the molecule of dioxygen is partially reduced. If the electron were transferred completely to the oxygen, the complex would be $Fe^{3+}—O_2^{\ominus}$ (a superoxide anion attached to ferric iron). The globin crevice prevents complete electron transfer and enforces return of the electron to the iron atom when O_2 dissociates.

B. Oxygen-Binding Curves of Myoglobin and Hemoglobin

Oxygen binds reversibly to myoglobin and hemoglobin. The extent of binding at equilibrium depends on the concentration of the protein and the concentration of oxygen. This relationship is depicted in oxygen-binding curves (Figure 4.52). In these figures, the fractional saturation (Y) of a fixed amount of protein is plotted against the concentration of oxygen (measured as the partial pressure of gaseous oxygen, pO_2). The fractional saturation of myoglobin or hemoglobin is the fraction of the total number of molecules that are oxygenated.

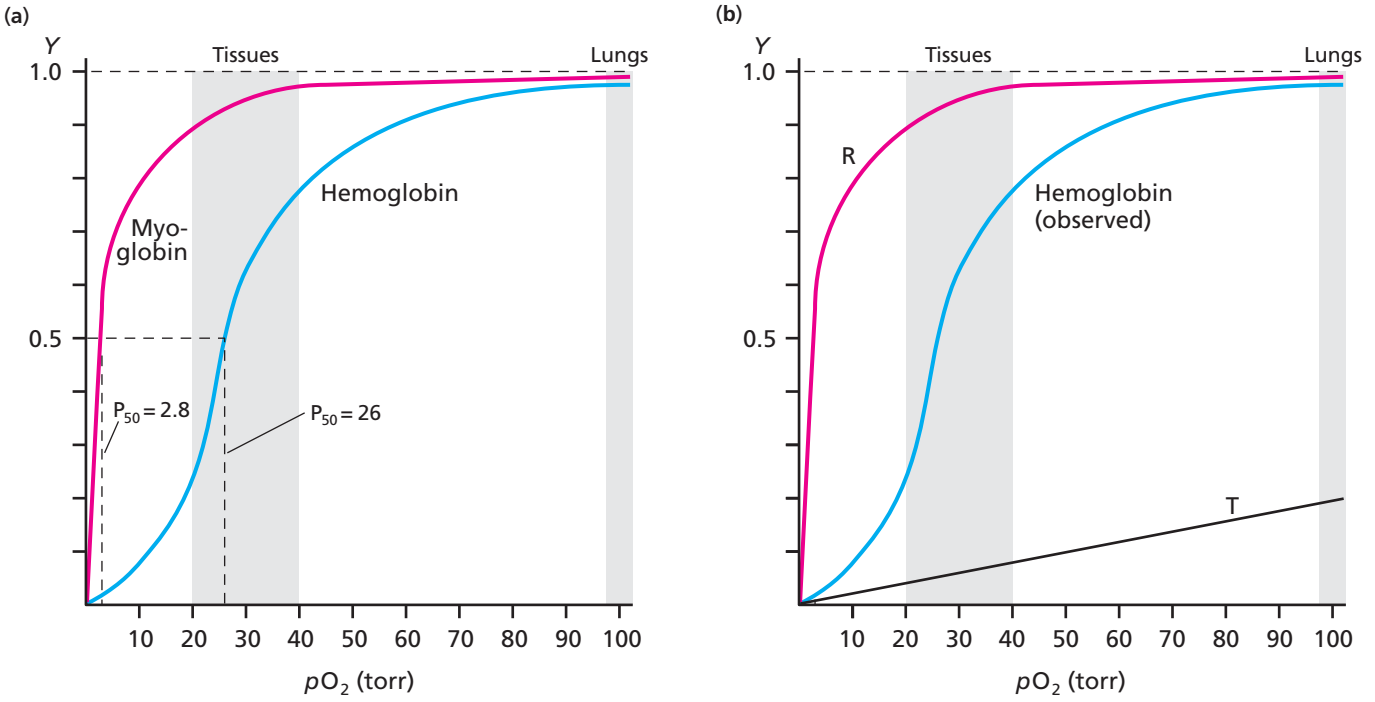
$$Y = \frac{[MbO_2]}{[MbO_2] + [Mb]} \qquad \textbf{(4.3)}$$

The oxygen-binding curve of myoglobin is hyperbolic (Figure 4.52), indicating that there is a single equilibrium constant for the binding of O_2 to the macromolecule. In contrast, the curve depicting the relationship between oxygen concentrations and binding to hemoglobin is sigmoidal. Sigmoidal (S-shaped) binding curves indicate that more than one molecule of ligand is binding to each protein. In this case, up to four molecules of O_2 bind to hemoglobin, one per heme group of the tetrameric protein. The shape of the curve indicates that the oxygen-binding sites of hemoglobin interact such that the binding of one molecule of oxygen to one heme group facilitates binding of oxygen molecules to the other hemes. The oxygen affinity of hemoglobin increases as each oxygen molecule is bound. This interactive binding phenomenon is termed *positive cooperativity* of binding.

The partial pressure at half-saturation (P_{50}) is a measure of the affinity of the protein for O_2. A low P_{50} indicates a high affinity for oxygen since the protein is half-saturated with oxygen at a low oxygen concentration; similarly, a high P_{50} signifies a low affinity. Myoglobin molecules are half-saturated at a pO_2 of 2.8 torr (1 atmosphere = 760 torr). The P_{50} for hemoglobin is much higher (26 torr) reflecting its lower affinity for oxygen. The heme prosthetic groups of myoglobin and hemoglobin are identical but the affinities of these groups for oxygen differ because the microenvironments provided by the proteins are slightly different. Oxygen affinity is an intrinsic property of the protein. It is similar to the equilibrium binding/dissociation constants that are commonly used to describe the binding of ligands to other proteins and enzymes (Section 4.9).

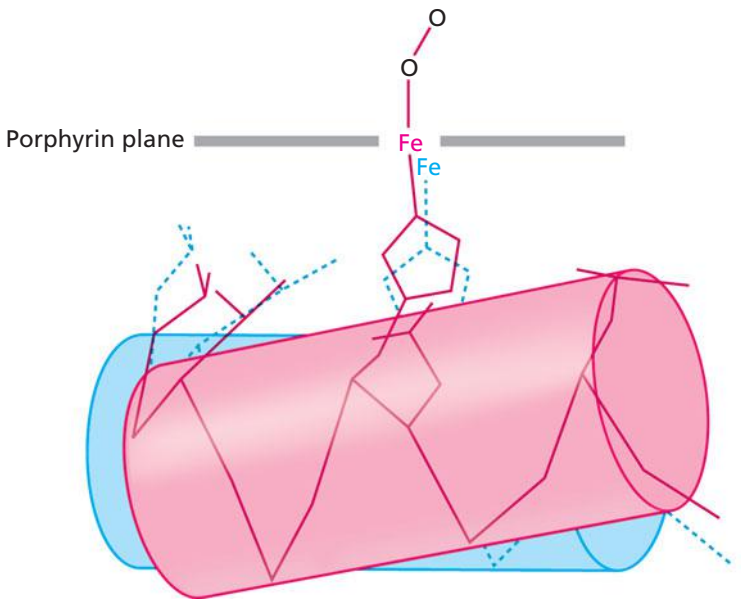
As Figure 4.52 shows, at the high pO_2 found in the lungs (about 100 torr) both myoglobin and hemoglobin are nearly saturated. However, at pO_2 values below about 50 torr, myoglobin is still almost fully saturated whereas hemoglobin is only partially saturated. Much of the oxygen carried by hemoglobin in erythrocytes is released within the capillaries of tissues where pO_2 is low (20 to 40 torr). Myoglobin in muscle tissue then binds oxygen released from hemoglobin. The differential affinities of myoglobin and hemoglobin for oxygen thus lead to an efficient system for oxygen delivery from the lungs to muscle.

The cooperative binding of oxygen by hemoglobin can be related to changes in the protein conformation that occur on oxygenation. Deoxyhemoglobin is stabilized by several intra- and intersubunit ion pairs. When oxygen binds to one of the subunits, it causes a movement that disrupts these ion pairs and favors a slightly different conformation. The movement is triggered by the reactivity of the heme iron atom (Figure 4.53). In deoxyhemoglobin, the iron atom is bound to only five ligands (as in myoglobin). It is slightly larger than the cavity within the porphyrin ring and lies below the plane of the ring. When O_2—the sixth ligand—binds to the iron atom, the electronic structure of the iron



▲ **Figure 4.52**

Oxygen-binding curves of myoglobin and hemoglobin. (a) Comparison of myoglobin and hemoglobin. The fractional saturation (Y) of each protein is plotted against the partial pressure of oxygen (pO_2). The oxygen-binding curve of myoglobin is hyperbolic, with half-saturation ($Y = 0.5$) at an oxygen pressure of 2.8 torr. The oxygen-binding curve of hemoglobin in whole blood is sigmoidal, with half-saturation at an oxygen pressure of 26 torr. Myoglobin has a greater affinity than hemoglobin for oxygen at all oxygen pressures. In the lungs, where the partial pressure of oxygen is high, hemoglobin is nearly saturated with oxygen. In tissues, where the partial pressure of oxygen is low, oxygen is released from oxygenated hemoglobin and transferred to myoglobin. **(b)** O_2 binding by the different states of hemoglobin. The oxy (R, or high-affinity) state of hemoglobin has a hyperbolic binding curve. The deoxy (T, or low-affinity) state of hemoglobin would also have a hyperbolic binding curve but with a much higher concentration for half-saturation. Solutions of hemoglobin containing mixtures of low- and high-affinity forms show sigmoidal binding curves with intermediate oxygen affinities.



▲ **Figure 4.53**

Conformational changes in a hemoglobin chain induced by oxygenation. When the heme iron of a hemoglobin subunit is oxygenated (red), the proximal histidine residue is pulled toward the porphyrin ring. The helix containing the histidine also shifts position, disrupting ion pairs that cross-link the subunits of deoxyhemoglobin (blue).

changes, its diameter decreases, and it moves into the plane of the porphyrin ring pulling the helix that contains the proximal histidine. The change in tertiary structure results in a slight change in quaternary structure and this allows the remaining subunits to bind oxygen more readily. The entire tetramer appears to shift from the deoxy to the oxy conformation only after at least one oxygen molecule binds to each $\alpha\beta$ dimer. (For further discussion, see Section 5.9C.)

The conformational change of hemoglobin is responsible for the positive cooperativity of binding seen in the binding curve (Figure 4.52a). The shape of the curve is due to the combined effect of the two conformations (Figure 4.52b). The completely deoxygenated form of hemoglobin has a low affinity for oxygen and thus exhibits a hyperbolic binding curve with a very high concentration of half-saturation. Only a small amount of hemoglobin is saturated at low oxygen concentrations. As the concentration of oxygen increases, some of the hemoglobin molecules bind a molecule of oxygen and this increases their affinity for oxygen so that they are more likely to bind additional oxygen. This causes the sigmoidal curve and also a sharp rise in binding. More molecules of hemoglobin are in the oxy conformation. If all of the hemoglobin molecules were in the oxy conformation, a solution would exhibit a hyperbolic binding curve. Release of the oxygen molecules allows the hemoglobin molecule to re-form the ion pairs and resume the deoxy conformation.

The two conformations of hemoglobin are called the T (tense) and R (relaxed) states, using the standard terminology for such conformational changes. In hemoglobin, the deoxy conformation, which resists oxygen binding, is considered the inactive (T) state, and the oxy conformation, which facilitates oxygen binding, is considered the active (R) state. The R and T states are in dynamic equilibrium.

BOX 4.4 EMBRYONIC AND FETAL HEMOGLOBINS

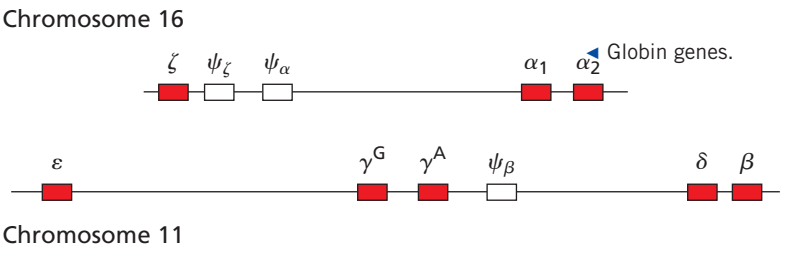
The human α globin genes are located on chromosome 16 in a cluster of related members of the globin gene family. There are two different genes encoding α globin: α_1 and α_2. Upstream of these genes there is another functional gene called ζ (zeta). The locus includes two nonfunctional pseudogenes, one related to ζ ($\psi\zeta$) and the other derived from a duplicated α globin gene (ψ_α).

The β globin gene is on chromosome 11 and it is also located at a locus where there are other members of the globin gene family. The functional genes are δ, two related γ globin genes (γ^A and γ^G), and an ε (epsilon) gene. This locus also contains a pseudogene related to β (ψ_β).

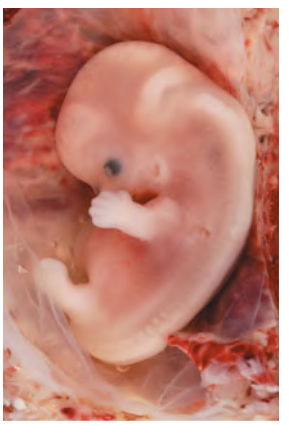
The other globin genes encode hemoglobin subunits that are expressed in the early embryo and in the fetus. The embryonic hemoglobins are called Gower 1 ($\zeta_2\varepsilon_2$), Gower 2 ($\alpha_2\varepsilon_2$), and Portland ($\zeta_2\gamma_2$). The fetal hemoglobin has the subunit composition $\alpha_2\gamma_2$. The adult hemoglobins are $\alpha_2\beta_2$ and $\alpha_2\delta_2$.

During early embryogenesis, the growing embryo gets oxygen from the mother's blood through the placenta. The concentration of oxygen in the embryo is much lower than the concentration of oxygen in adult blood. The embryonic hemoglobins compensate by binding oxygen much more tightly, their P_{50} values range from 4 to 12 torr—much lower than the value of adult hemoglobin (26 torr). The fetal hemoglobins bind oxygen less tightly than the embryonic hemoglobin but tighter than the adult hemoglobins (P_{50} = 20 torr).

Expression of the various globin genes is carefully regulated so that the right genes are transcribed at the right time. Sometimes mutations arise where the fetal γ globin genes are inappropriately expressed in adults. The result is a phenotype known as Hereditary Persistence of Fetal Hemoglobin (HPFH). This is just one of hundreds of hemoglobin variants that have been detected in humans. You can read about them on a database called Online Mendelian Inheritance in Man (OMIM), the most complete and accurate database of human genetic diseases (ncbi.nlm.nih.gov/omim).



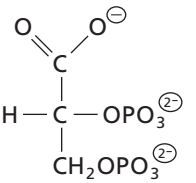
► Human fetus.



▲ Julian Voss-Andreae created a sculpture called "Heart of Steel (Hemoglobin)" in 2005 in the City of Lake Oswego, Oregon. The sculpture is a depiction of a hemoglobin molecule with a bound oxygen atom. The original sculpture was shiny steel (left). After 10 days (middle) it had started to rust as the iron in the steel reacted with oxygen in the atmosphere. After several months (right) the sculpture was completely rust colored.

C. Hemoglobin Is an Allosteric Protein

The binding and release of oxygen by hemoglobin are regulated by **allosteric interactions** (from the Greek *allos*, "other"). In this respect, hemoglobin—a carrier protein, not an enzyme—resembles certain regulatory enzymes (Section 5.9). Allosteric interactions occur when a specific small molecule, called an **allosteric modulator**, or **allosteric effector**, binds to a protein (usually an enzyme) and modulates its activity. The allosteric modulator binds reversibly at a site separate from the functional binding site of the protein. An effector molecule may be an activator or an inhibitor. A protein whose activity is modulated by allosteric effectors is called an **allosteric protein**.



▲ 2,3-Bisphospho-D-glycerate (2,3BPG).

Allosteric modulation is accomplished by small but significant changes in the conformations of allosteric proteins. It involves cooperativity of binding that is regulated by binding of the allosteric effector to a distinct site that doesn't overlap the normal binding site of a substrate, product, or transported molecule such as oxygen. An allosteric protein is in an equilibrium in which its active shape (R state) and its inactive shape (T state) are rapidly interconverting. A substrate, which obviously binds at the active site (to heme in hemoglobin), binds most avidly when the protein is in the R state. An allosteric inhibitor, which binds at an allosteric or regulatory site, binds most avidly to the T state. The binding of an allosteric inhibitor to its own site causes the allosteric protein to change rapidly from the R state to the T state. The binding of a substrate to the active site (or an allosteric activator to the allosteric site) causes the reverse change. The change in conformation of an allosteric protein caused by binding or release of an effector extends from the allosteric site to the functional binding site (the active site). The activity level of an allosteric protein depends on the relative proportions of molecules in the R and T forms and these, in turn, depend on the relative concentrations of the substrates and modulators that bind to each form.

The synthesis of 2,3BPG in red blood cells is described in Box 11.2 (Chapter 11).

The molecule 2,3-*bis*phospho-D-glycerate (2,3BPG) is an allosteric effector of mammalian hemoglobin. The presence of 2,3BPG in erythrocytes raises the P_{50} for binding of oxygen to adult hemoglobin to about 26 torr—much higher than the P_{50} for oxygen binding to purified hemoglobin in aqueous solution (about 12 torr). In other words, 2,3BPG in erythrocytes substantially lowers the affinity of deoxyhemoglobin for oxygen. The concentrations of 2,3BPG and hemoglobin within erythrocytes are nearly equal (about 4.7 mM).

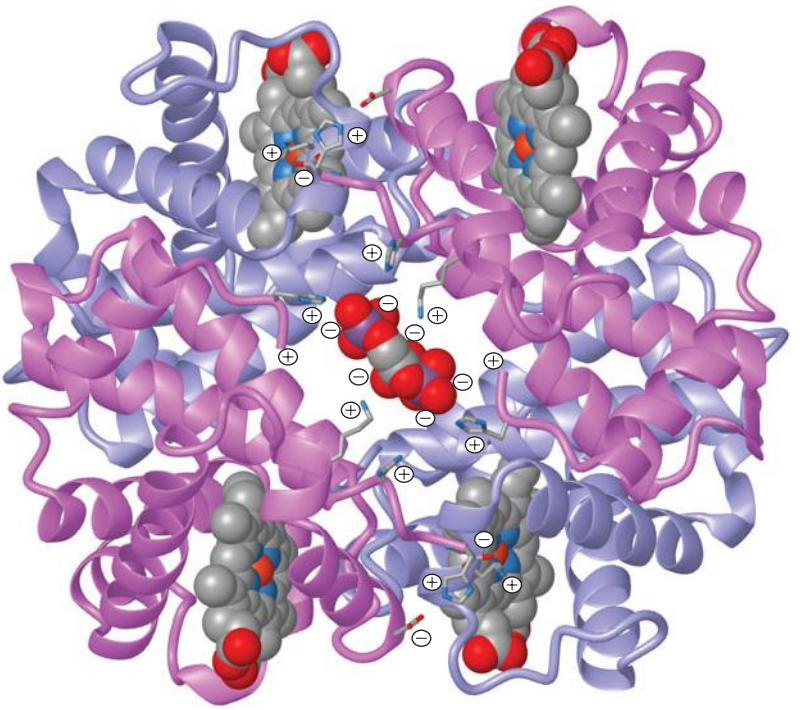


Figure 4.54 ▶
Binding of 2,3BPG to deoxyhemoglobin. The central cavity of deoxyhemoglobin is lined with positively charged groups that are complementary to the carboxylate and phosphate groups of 2,3BPG. Both 2,3BPG and the ion pairs shown help stabilize the deoxy conformation. The α subunits are shown in pink, the β subunits in blue, and the heme prosthetic groups in red.

R and T conformations are explained more thoroughly in Section 5.10, "Theory of Allostery."

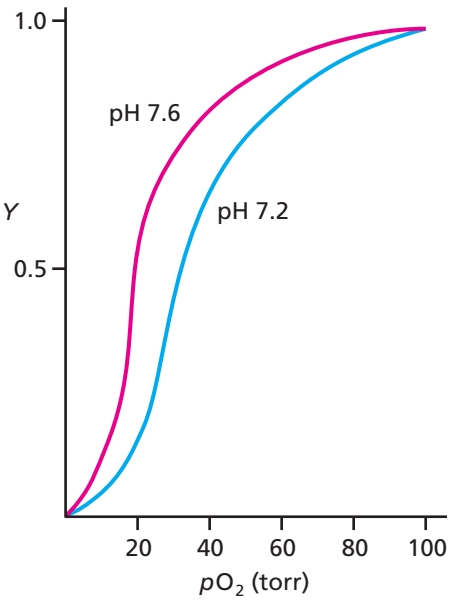
The effector 2,3BPG binds in the central cavity of hemoglobin between the two β subunits. In this binding pocket there are six positively charged side chains and the N-terminal α-amino group of each β chain forming a cationic binding site (Figure 4.54). In deoxyhemoglobin, these positively charged groups can interact electrostatically with the five negative charges of 2,3BPG. When 2,3BPG is bound, the deoxy conformation (the T state, which has a low affinity for O_2) is stabilized and conversion to the oxy conformation (the R or high-affinity state) is inhibited. In oxyhemoglobin, the β chains are closer together and the allosteric binding site is too small to accommodate 2,3BPG. The reversibly bound ligands O_2 and 2,3BPG have opposite effects on the R $\rightleftharpoons$ T equilibrium. Oxygen binding increases the proportion of hemoglobin molecules in the oxy (R) conformation and 2,3BPG binding increases the proportion of hemoglobin molecules in the deoxy (T) conformation. Because oxygen and 2,3BPG have different binding sites, 2,3BPG is a true allosteric effector.

In the absence of 2,3BPG, hemoglobin is nearly saturated at an oxygen pressure of about 20 torr. Thus, at the low partial pressure of oxygen that prevails in the tissues (20 to 40 torr), hemoglobin without 2,3BPG would not unload its oxygen. In the presence of equimolar 2,3BPG, however, hemoglobin is only about one-third saturated at 20 torr. The allosteric effect of 2,3BPG causes hemoglobin to release oxygen at the low partial pressures of oxygen in the tissues. In muscle, myoglobin can bind some of the oxygen that is released.

Additional regulation of the binding of oxygen to hemoglobin involves carbon dioxide and protons, both of which are products of aerobic metabolism. CO_2 decreases the affinity of hemoglobin for O_2 by lowering the pH inside red blood cells. Enzyme-catalyzed hydration of CO_2 in erythrocytes produces carbonic acid, H_2CO_3, which dissociates to form bicarbonate and a proton thereby lowering the pH.

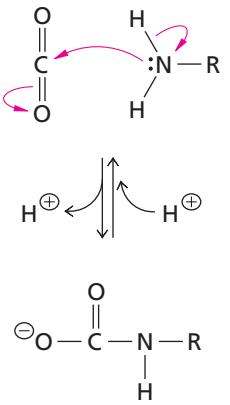
$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^{\oplus} + HCO_3^{\ominus} \qquad \textbf{(4.4)}$$

The lower pH leads to protonation of several groups in hemoglobin. These groups then form ion pairs that help stabilize the deoxy conformation. The increase in the concentration of CO_2 and the concomitant decrease in pH raise the P_{50} of hemoglobin (Figure 4.55). This phenomenon, called the Bohr effect, increases the efficiency of the oxygen delivery system. In inhaling lungs, where the CO_2 level is low, O_2 is readily picked up by hemoglobin; in metabolizing tissues, where the CO_2 level is relatively high and the pH is relatively low, O_2 is readily unloaded from oxyhemoglobin.



▲ Figure 4.55
Bohr effect. Lowering the pH decreases the affinity of hemoglobin for oxygen.

Carbon dioxide is transported from the tissues to the lungs in two ways. Most CO_2 produced by metabolism is transported as dissolved bicarbonate ions but some carbon dioxide is carried by hemoglobin itself the form of carbamate adducts (Figure 4.56). At the pH of red blood cells (7.2) and at high concentrations of CO_2, the unprotonated amino groups of the four N-terminal residues of deoxyhemoglobin (pK_a values between 7 and 8) can react reversibly with CO_2 to form carbamate adducts. The carbamates of oxyhemoglobin are less stable than those of deoxyhemoglobin. When hemoglobin reaches the lungs, where the partial pressure of CO_2 is low and the partial pressure of O_2 is high, hemoglobin is converted to its oxygenated state and the CO_2 that was bound is released.



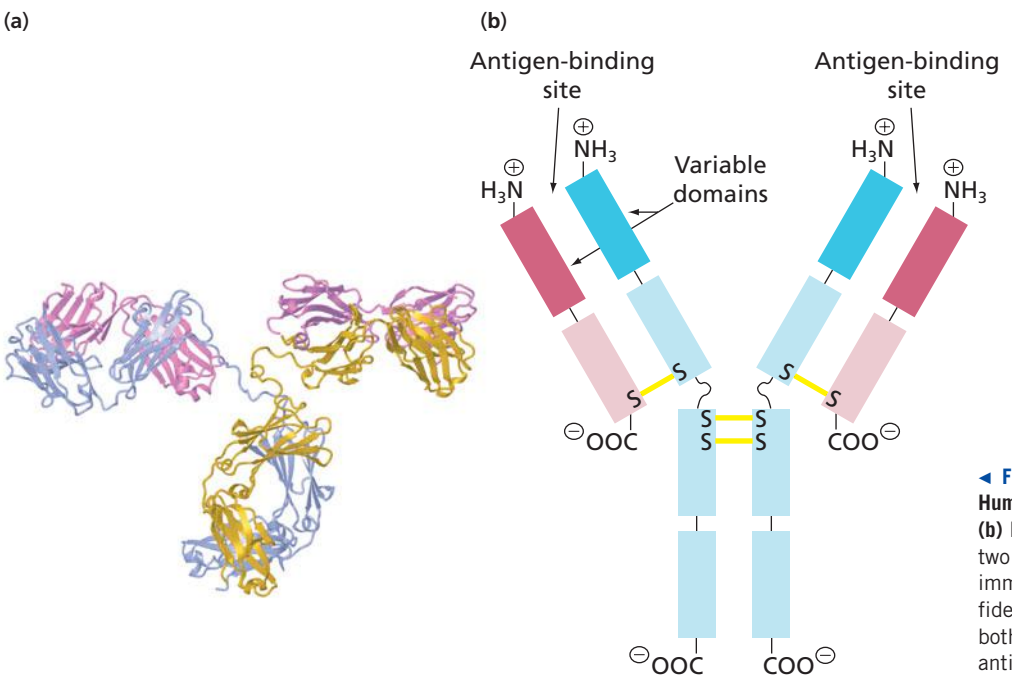
▲ **Figure 4.56**
Carbamate adduct. Carbon dioxide produced by metabolizing tissues can react reversibly with the N-terminal residues of the globin chains of hemoglobin, converting them to carbamate adducts.

4.15 Antibodies Bind Specific Antigens

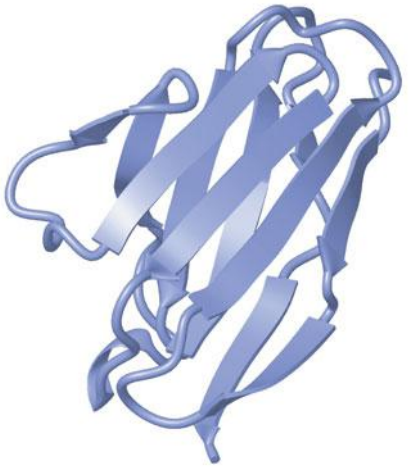
Vertebrates possess a complex immune system that eliminates foreign substances including infectious bacteria and viruses. As part of this defense system, vertebrates synthesize proteins called **antibodies** (also known as immunoglobulins) that specifically recognize and bind **antigens**. Many different types of foreign compounds can serve as antigens that produce an immune response. Antibodies are synthesized by white blood cells called lymphocytes—each lymphocyte and its descendants synthesize the same antibody. Because animals are exposed to many foreign substances over their lifetimes, they develop a huge array of antibody-producing lymphocytes that persist at low levels for many years and can later respond to the antigen during reinfection. The memory of the immune system is the reason certain infections do not recur in an individual despite repeated exposure. Vaccines (inactivated pathogens or analogs of toxins) administered to children are effective because immunity established in childhood lasts through adulthood.

When an antigen—either novel or previously encountered—binds to the surface of lymphocytes, these cells are stimulated to proliferate and produce soluble antibodies for secretion into the bloodstream. The soluble antibodies bind to the foreign organism or substance forming antibody–antigen complexes that precipitate and mark the antigen for destruction by a series of interacting proteases or by lymphocytes that engulf the antigen and digest it intracellularly.

The most abundant antibodies in the bloodstream are of the immunoglobulin G class (IgG). These are Y-shaped oligomers composed of two identical light chains and two identical heavy chains connected by disulfide bonds (Figure 4.57). Immunoglobulins are glycoproteins containing covalently bound carbohydrates attached to the heavy chains. The N-termini of pairs of light and heavy chains are close together. Light chains contain two domains and heavy chains contain four domains. Each of the domains consists of



◀ **Figure 4.57**
Human antibody structure. (a) Structure. **(b)** Diagram. Two heavy chains (blue) and two light chains (red) of antibodies of the immunoglobulin G class are joined by disulfide bonds (yellow). The variable domains of both the light and heavy chains (where antigen binds) are colored more darkly.

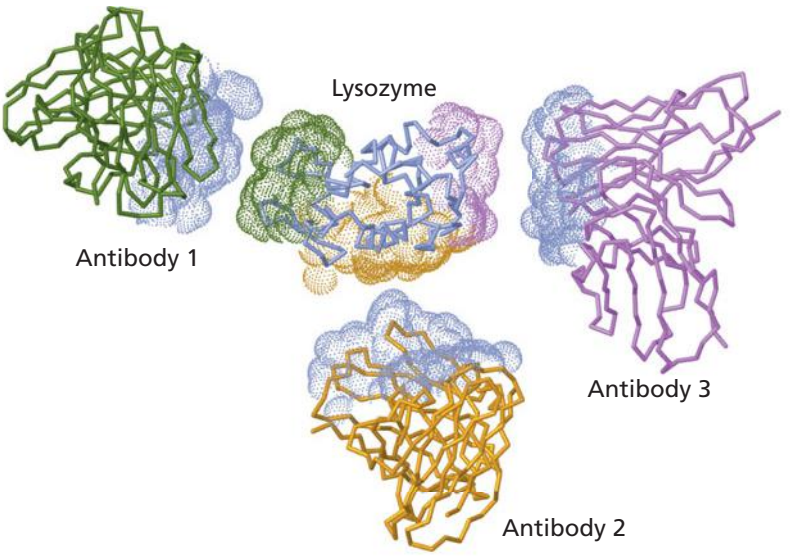


▲ **Figure 4.58**
The immunoglobulin fold. The domain consists of a sandwich of two antiparallel β sheets. [PDB 1REI].

about 110 residues assembled into a common motif called the immunoglobulin fold whose characteristic feature is a sandwich composed of two antiparallel β sheets (Figure 4.58). This domain structure is found in many other proteins of the immune system.

The N-terminal domains of antibodies are called the variable domains because of their sequence diversity. They determine the specificity of antigen binding. X-ray crystallographic studies have shown that the antigen-binding site of a variable domain consists of three loops, called hypervariable regions, that differ widely in size and sequence. The loops from a light chain and a heavy chain combine to form a barrel, the upper surface of which is complementary to the shape and polarity of a specific antigen. The match between the antigen and antibody is so close that there is no space for water molecules between the two. The forces that stabilize the interaction of antigen with antibody are primarily hydrogen bonds and electrostatic interactions. An example of the interaction of antibodies with a protein antigen is shown in Figure 4.59.

Antibodies are used in the laboratory for the detection of small quantities of various substances because of their remarkable antigen-binding specificity. In a common type of immunoassay, fluid containing an unknown amount of antigen is mixed with a solution of labeled antibody and the amount of antibody–antigen complex formed is measured. The sensitivity of these assays can be enhanced in a variety of ways to make them suitable for diagnostic tests.



▲ **Figure 4.59**
Binding of three different antibodies to an antigen (the protein lysozyme). The structures of the three antigen–antibody complexes have been determined by X-ray crystallography. This composite view, in which the antigen and antibodies have been separated, shows the surfaces of the antigen and antibodies that interact. Only parts of the three antibodies are shown.

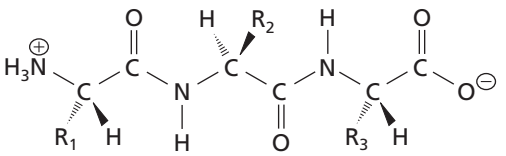
Summary

1. Proteins fold into many different shapes, or conformations. Many proteins are water-soluble, roughly spherical, and tightly folded. Others form long filaments that provide mechanical support to cells and tissues. Membrane proteins are integral components of membranes or are associated with membranes.
2. There are four levels of protein structure: primary (sequence of amino acid residues), secondary (regular local conformation, stabilized by hydrogen bonds), tertiary (compacted shape of the entire polypeptide chain), and quaternary (assembly of two or more polypeptide chains into a multisubunit protein).
3. The three-dimensional structures of biopolymers, such as proteins can be determined by X-ray crystallography and NMR spectroscopy.
4. The peptide group is polar and planar. Rotation around the N—Cα and Cα—C bonds is described by φ and ψ.
5. The α helix, a common secondary structure, is a coil containing approximately 3.6 amino acid residues per turn. Hydrogen bonds between amide hydrogens and carbonyl oxygens are roughly parallel to the helix axis.

6. The other common type of secondary structure, β structure, often consists of either parallel or antiparallel β strands that are hydrogen-bonded to each other to form β sheets.
7. Most proteins include stretches of nonrepeating conformation, including turns and loops that connect α helices and β strands.
8. Recognizable combinations of secondary structural elements are called motifs.
9. The tertiary structure of proteins consists of one or more domains, which may have recognizable structures and may be associated with particular functions.
10. In proteins that possess quaternary structure, subunits are usually held together by noncovalent interactions.
11. The native conformation of a protein can be disrupted by the addition of denaturing agents. Renaturation may be possible under certain conditions.
12. Folding of a protein into its biologically active state is a sequential, cooperative process driven primarily by the hydrophobic effect. Folding can be assisted by chaperones.
13. Collagen is the major fibrous protein of connective tissues. The three left-handed helical chains of collagen form a right-handed supercoil.
14. The compact, folded structures of proteins allow them to selectively bind other molecules. The heme-containing proteins myoglobin and hemoglobin bind and release oxygen. Oxygen binding to hemoglobin is characterized by positive cooperativity and allosteric regulation.
15. Antibodies are multidomain proteins that bind foreign substances, or antigens, marking them for destruction. The variable domains at the ends of the heavy and light chains interact with the antigen.

Problems

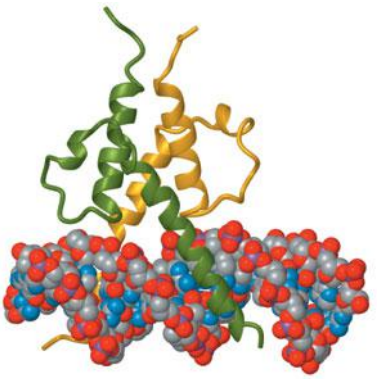
1. Examine the following tripeptide:



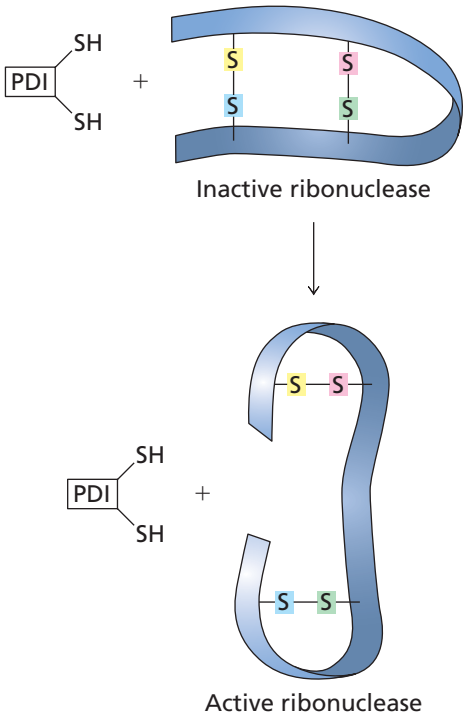
   (a) Label the α-carbon atoms and draw boxes around the atoms of each peptide group.
   (b) What do the R groups represent?
   (c) Why is there limited free rotation around the carbonyl C═O to N amide bonds?
   (d) Assuming that the chemical structure represents the correct conformation of the peptide linkage, are the peptide groups in the *cis* or the *trans* conformation?
   (e) Which bonds allow rotation of peptide groups with respect to each other?
2. (a) Characterize the hydrogen-bonding pattern of (1) an α helix and (2) a collagen triple helix.
   (b) Explain how the amino acid side chains are arranged in each of these helices.
3. Explain why (1) glycine and (2) proline residues are not commonly found in α helices.
4. A synthetic 20 amino acid polypeptide named Betanova was designed as a small soluble molecule that would theoretically form stable β-sheet structures in the absence of disulfide bonds. NMR of Betanova in solution indicates that it does, in fact, form a three-stranded antiparallel β sheet. Given the sequence of Betanova below:
   (a) Draw a ribbon diagram for Betanova indicating likely residues for each hairpin turn between the β strands.
   (b) Show the interactions that are expected to stabilize this β-sheet structure.

   Betanova RGWSVQNGKYTNNGKTTEGR

5. Each member of an important family of 250 different DNA-binding proteins is composed of a dimer with a common protein motif. This motif permits each DNA-binding protein to recognize and bind to specific DNA sequences. What is the common protein motif in the structure below?



6. Refer to Figure 4.21 to answer the following questions.
   (a) To which of the four major domain categories does the middle domain of pyruvate kinase (PK) belong (all α all β, α/β, $\alpha + \beta$)?
   (b) Describe any characteristic domain "fold" that is prominent in this middle domain of PK.
   (c) Identify two other proteins that have the same fold as the middle domain of pyruvate kinase.
7. Protein disulfide isomerase (PDI) markedly increases the rate of correct refolding of the inactive ribonuclease form with random disulfide bonds (Figure 4.35). Show the mechanism for the PDI-catalyzed rearrangement of a nonnative (inactive) protein with incorrect disulfide bonds to the native (active) protein with correct disulfide bonds.



8. Myoglobin contains eight α helices, one of which has the following sequence:

–Gln–Gly–Ala–Met–Asn–Lys–Ala–Leu–Glu–His–Phe–Arg–Lys–Asp–Ile–Ala–Ala–

Which side chains are likely to be on the side of the helix that faces the interior of the protein? Which are likely to be facing the aqueous solvent? Account for the spacing of the residues facing the interior.

9. Homocysteine is an α-amino acid containing one more methylene group in its side chain than cysteine (side chain = $—CH_2CH_2SH$). Homocysteinuria is a genetic disease characterized by elevated levels of homocysteine in plasma and urine, as well as skeletal deformities due to defects in collagen structure. Homocysteine reacts readily with allysine under physiological conditions. Show this reaction and suggest how it might lead to defective cross-linking in collagen.

10. The larval form of the parasite *Schistosoma mansoni* infects humans by penetrating the skin. The larva secretes enzymes that catalyze the cleavage of peptide bonds between residues X and Y in the sequence –Gly–Pro–X–Y– (X and Y can be any of several amino acids). Why is this enzyme activity important for the parasite?

11. (a) How does the reaction of carbon dioxide with water help explain the Bohr effect? Include the equation for the formation of bicarbonate ion from CO_2 and water, and explain the effects of $H^{\oplus}$ and CO_2 on hemoglobin oxygenation.

(b) Explain the physiological basis for the intravenous administration of bicarbonate to shock victims.

12. Fetal hemoglobin (Hb F) contains serine in place of the cationic histidine at position 143 of the β chains of adult hemoglobin (Hb A). Residue 143 faces the central cavity between the β chains.

(a) Why does 2,3BPG bind more tightly to deoxy Hb A than to deoxy Hb F?

(b) How does the decreased affinity of Hb F for 2,3BPG affect the affinity of Hb F for O_2?

(c) The P_{50} for Hb F is 18 torr, and the P_{50} for Hb A is 26 torr. How do these values explain the efficient transfer of oxygen from maternal blood to the fetus?

13. Amino acid substitutions at the $\alpha\beta$ subunit interfaces of hemoglobin may interfere with the $R \rightleftharpoons T$ quaternary structural changes that take place on oxygen binding. In the hemoglobin variant Hb_{Yakima}, the R form is stabilized relative to the T form, and P_{50} = 12 torr. Explain why the mutant hemoglobin is less efficient than normal hemoglobin (P_{50} = 26 torr) in delivering oxygen to working muscle, where O_2 may be as low as 10 to 20 torr.

14. The spider venom from the Chilean Rose Tarantula (*Grammostola spatulata*) contains a toxin that is a 34-amino acid protein. It is thought to be a globular protein that partitions into the lipid membrane to exert its effect. The sequence of the protein is:

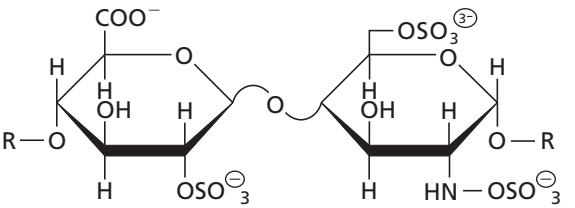
ECGKFMWKCKNSNDCCKDLVCSSRWKWCVLASPF

(a) Identify the hydrophobic and highly hydrophilic amino acids in the protein.

(b) The protein is thought to have a hydrophobic face that interacts with the lipid membrane. How can the hydrophobic amino acids far apart in sequence interact to form a hydrophobic face?

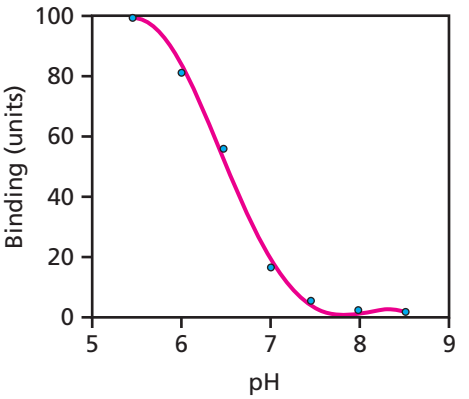
[Adapted from Lee, S. and MacKinnon, R. (2004). *Nature* 430: 232–235.]

15. Selenoprotein P is an unusual extracellular protein that contains 8–10 selenocysteine residues and has a high content of cysteine and histidine residues. Selenoprotein P is found both as a plasma protein and as a protein strongly associated with the surface of cells. The association of selenoprotein P with cells is proposed to occur through the interaction of selenoprotein P with high-molecular-weight carbohydrate compounds classified as glycosaminoglycans. One such compound is heparin (see structure on next page). Binding studies of selenoprotein P to heparin were carried out under different pH conditions. The results are shown in the graph on next page.



(a) How is the binding of selenoprotein P to heparin dependent upon pH?

(b) Give possible structural reasons for the binding dependence.

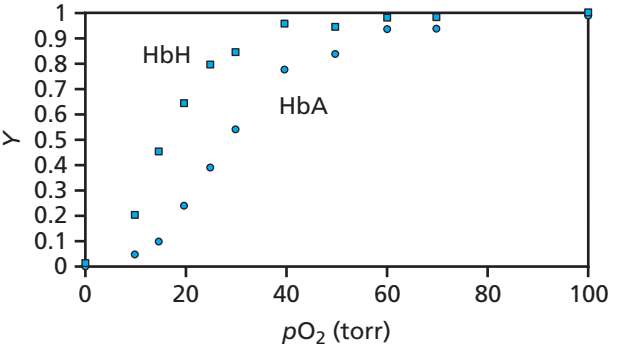


(*Hint:* Use the information about which amino acids are abundant in selenoprotein P in your answer).

[Adapted from Arteel, G. E., Franken, S., Kappler, J., and Sies, H. (2000). *Biol. Chem.* 381:265–268.]

16. Gelatin is processed collagen that comes from the joints of animals. When gelatin is mixed with hot water, the triple helix structure unwinds and the chains separate, becoming random coils that dissolve in the water. As the dissolved gelatin mixture cools, the collagen forms a matrix that traps water; as a result, the mixture turns into the jiggling semisolid mass that is recognizable as Jell-O™. The directions on a box of gelatin include the following: "Chill until slightly thickened, then add 1 to 2 cups cooked or raw fruits or vegetables. Fresh or frozen pineapple must be cooked before adding." If the pineapple is not cooked, the gelatin will not set properly. Pineapple belongs to a group of plants called Bromeliads and contains a protease called bromelain. Explain why pineapple must be cooked before adding to gelatin.

17. Hb Helsinki (HbH) is a hemoglobin mutant in which the lysine residue at position 82 has been replaced with methionine. The mutation is in the beta chain, and residue 82 is found in the central cavity of hemoglobin. The oxygen binding curves for normal adult hemoglobin (HbA, ●) and HbH (■) at pH 7.4 in the presence of a physiological concentration of 2,3BPG are shown in the graph.



[Adapted from Ikkala, E., Koskela, J., Pikkarainen, P., Rahiala, E.L., El-Hazmi, M. A., Nagai, K., Lang, A., and Lehmann, H. *Acta Haematol.* (1976). 56:257–275.]

Explain why the curve for HbH is shifted from the curve for HbA. Does this mutation stabilize the R or T state? What result does this mutation have on oxygen affinity?

Selected Readings

General

Clothia, C., and Gough, J. (2009). Genomic and structural aspects of protein evolution. *Biochem. J.* 419:15–28. doi: 10,1042/BJ20090122.

Creighton, T. E. (1993). *Proteins: Structures and Molecular Properties*, 2nd ed. (New York: W. H. Freeman), Chapters 4–7.

Fersht, A. (1998). *Structure and Mechanism in Protein Structure* (New York: W. H. Freeman).

Goodsell, D., and Olson, A. J. (1993). Soluble proteins: size, shape, and function. *Trends Biochem. Sci.* 18:65–68.

Goodsell, D. S., and Olson, A. J. (2000). Structural symmetry and protein function. *Annu. Rev. Biophys, Biomolec. Struct.* 29:105–153.

Kyte, J. (1995). *Structure in Protein Chemistry* (New York: Garland).

Protein Structure

Branden, C., and Tooze, J. (1991). *Introduction to Protein Structure* 2nd ed. (New York: Garland).

Chothia, C., Hubbard, T., Brenner, S., Barns, H., and Murzin, A. (1997). Protein folds in the all-β and all-α classes. *Annu. Rev. Biophys. Biomol. Struct.* 26:597–627.

Edison, A. S. (2001). Linus Pauling and the planar peptide bond. *Nat. Struct. Biol.* 8:201–202.

Harper, E. T., and Rose, G. D. (1993). Helix stop signals in proteins and peptides: the capping box. *Biochemistry* 32:7605–7609.

Phizicky, E., and Fields, S. (1995). Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* 59:94–123.

Rhodes, G. (1993). *Crystallography Made Crystal Clear* (San Diego: Academic Press).

Richardson, J. S., and Richardson, D. C. (1989). Principles and patterns of protein conformation. In *Prediction of Protein Structure and the Principles of Protein Conformation*, G. D. Fasman, ed. (New York: Plenum), pp. 1–98.

Wang, Y., Liu, C., Yang, D., and Yu, H. (2010). *Pin1At* encoding a peptidyl-prolyl *cis/trans* isomerase regulates flowering time in arabidopsis. *Molec. Cell.* 37:112–122.

Uversky, V. N., and Dunker, A. K. (2010). Understanding protein non-folding. *Biochim. Biophys. Acta.* 1804:1231–1264.

Protein Folding and Stability

Daggett, V., and Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* 28:18–25.

Dill, K. A. Ozkan, S. B., Shell, M. S., and Weik, T. R. (2008). The protein folding problem. *Annu. Rev. Biophys.* 37:289–316.

Feldman, D. E., and Frydman, J. (2000). Protein folding *in vivo*: the importance of molecular chaperones. *Curr. Opin. Struct. Biol.* 10:26–33.

Kryshtafovych, A., Fidelis, K., and Moult, J. (2009). CASP8 results in context of previous experiments. *Proteins.* 77(suppl 9):217–228.

Matthews, B. W. (1993). Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.* 62:139–160.

Saibil, H. R. and Ranson, N. A. (2002). The chaperonin folding machine. *Trends Biochem. Sci.* 27:627–632.

Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A., and Horwich, A. L. (1998). Structure and function in GroEL-mediated protein folding. *Annu. Rev. Biochem.* 67:581–608.

Smith, C. A. (2000). How do proteins fold? *Biochem. Ed.* 28:76–79.

Specific Proteins

Ackers, G. K., Doyle, M. L., Myers, D., and Daugherty, M. A. (1992). Molecular code for cooperativity in hemoglobin. *Science* 255:54–63.

Brittain, T. (2002). Molecular aspects of embryonic hemogloin function. *Molec. Aspects Med.* 23:293–342.

Davies, D. R., Padlan, E. A., and Sheriff, S. (1990). Antibody-antigen complexes. *Annu. Rev. Biochem.* 59:439–473.

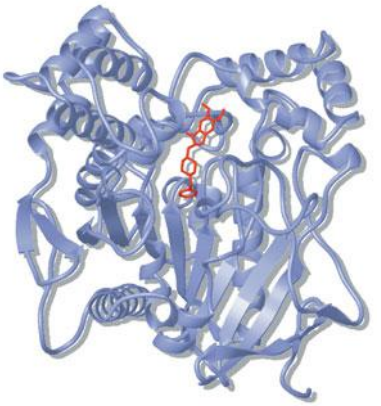
Eaton, W. A., Henry, E. R., Hofrichter, J., and Mozzarelli, A. (1999). Is cooperative binding by hemoglobin really understood? *Nature Struct. Biol.* 6(4):351–357.

Kadler, K. (1994). Extracellular matrix 1: fibril-forming collagens. *Protein Profile* 1:519–549.

Liu, R., and Ochman, H. (2007). Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci.* (USA). 104:7116–7121.

Perutz, M. F. (1978). Hemoglobin structure and respiratory transport. *Sci. Am.* 239(6):92–125.

Perutz, M. F., Wilkinson, A. J., Paoli, M., and Dodson, G. G. (1998). The stereochemical mechanism of the cooperative effects in hemoglobin revisited. *Annu. Rev. Biophys. Biomol. Struct.* 27:1–34.



Properties of Enzymes

We have seen how the three-dimensional shapes of proteins allow them to serve structural and transport roles. We now discuss their functions as enzymes. Enzymes are extraordinarily efficient, selective, biological catalysts. Every living cell has hundreds of different enzymes catalyzing the reactions essential for life—even the simplest living organisms contain hundreds of different enzymes. In multicellular organisms, the complement of enzymes differentiates one cell type from another but most of the enzymes we discuss in this book are among the several hundred common to all cells. These enzymes catalyze the reactions of the central metabolic pathways necessary for the maintenance of life.

In the absence of the enzymes, metabolic reactions will not proceed at significant rates under physiological conditions. The primary role of enzymes is to enhance the rates of these reactions to make life possible. Enzyme-catalyzed reactions are 10^3 to 10^{20} times faster than the corresponding uncatalyzed reactions. A catalyst is defined as a substance that speeds up the attainment of equilibrium. It may be temporarily changed during the reaction but it is unchanged in the overall process since it recycles to participate in multiple reactions. Reactants bind to a catalyst and products dissociate from it. Note that a catalyst does not change the *position* of the reaction's equilibrium (i.e., it does not make an unfavorable reaction favorable). Instead, it lowers the amount of energy needed in order for the reaction to proceed. Catalysts speed up both the forward and reverse reactions by converting a one- or two-step process into several smaller steps each needing less energy than the uncatalyzed reaction.

Enzymes are highly specific for the reactants, or **substrates**, they act on, but the degree of substrate specificity varies. Some enzymes act on a group of related substrates, and others on only a single compound. Many enzymes exhibit **stereospecificity** meaning

I was awed by enzymes and fell instantly in love with them. I have since had love affairs with many enzymes (none as enduring as with DNA polymerase), but I have never met a dull or disappointing one.

—Arthur Kornberg (2001)

KEY CONCEPT

Catalysts speed up the rate of forward and reverse reactions but they don't change the equilibrium concentrations.

Top: The enzyme acetylcholinesterase with the reversible inhibitor donepezil hydrochloride (Aricept; shown in red) occupying the active site. Aricept is used to improve mental functioning in patients with Alzheimer's disease. It is thought to act by inhibiting the breakdown of the neurotransmitter acetylcholine in the brain, thus prolonging the neurotransmitter effects. (It does not, however, affect the course of the disease.) [PDB 1EVE]

▲ **Enzyme reaction.** This is a large-scale enzyme reaction where milk is being curdled to make Appenzeller cheese. The reaction is catalyzed by rennet (rennin), which was originally derived from cow stomach. Rennet contains the enzyme chymosin, a protease that cleaves the milk protein casein between phenylalanine and methionine residues. The reaction releases a hydrophobic fragment of casein that aggregates and precipitates forming curd.

that they act on only a single stereoisomer of the substrate. Perhaps the most important aspect of enzyme specificity is **reaction specificity**—that is, the lack of formation of wasteful by-products. Reaction specificity is reflected in the exceptional purity of product (essentially 100%)—much higher than the purity of products of typical catalyzed reactions in organic chemistry. The specificity of enzymes not only saves energy for cells but also precludes the buildup of potentially toxic metabolic by-products.

Enzymes can do more than simply increase the rate of a single, highly specific reaction. Some can also combine, or couple, two reactions that would normally occur separately. This property allows the energy gained from one reaction to be used in a second reaction. Coupled reactions are a common feature of many enzymes—the hydrolysis of ATP, for example, is often coupled to less favorable metabolic reactions.

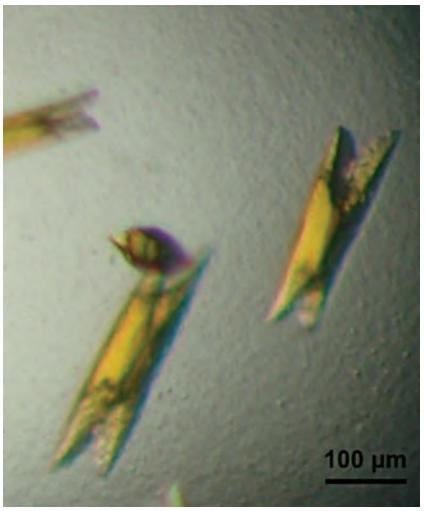
Some enzymatic reactions function as control points in metabolism. As we will see, metabolism is regulated in a variety of ways including alterations in the concentrations of enzymes, substrates, and enzyme inhibitors and modulation of the activity levels of certain enzymes. Enzymes whose activity is regulated generally have a more complex structure than unregulated enzymes. With few exceptions, regulated enzymes are oligomeric molecules that have separate binding sites for substrates and effectors, the compounds that act as regulatory signals. The fact that enzyme activity can be regulated is an important property that distinguishes biological catalysts from those encountered in a chemistry lab.

Some of the first biochemistry departments in universities were called Departments of Zymology.

The word *enzyme* is derived from a Greek word meaning "in yeast." It indicates that these catalysts are present inside cells. In the late 1800s, scientists studied the fermentation of sugars by yeast cells. Vitalists (who maintained that organic compounds could be made only by living cells) said that intact cells were needed for fermentation. Mechanists claimed that enzymes in yeast cells catalyze the reactions of fermentation. The latter conclusion was supported by the observation that cell-free extracts of yeast can catalyze fermentation. This finding was soon followed by the identification of individual reactions and the enzymes that catalyze them.

Catalytic RNA molecules are discussed in Chapters 21 and 22.

A generation later, in 1926, James B. Sumner crystallized the first enzyme (urease) and proved that it is a protein. Five more enzymes were purified in the next decade and also found to be proteins: pepsin, trypsin, chymotrypsin, carboxypeptidase, and Old Yellow Enzyme (a flavoprotein NADPH oxidase). Since then, almost all enzymes have been shown to be proteins or proteins plus cofactors. Certain RNA molecules also exhibit catalytic activity but they are not usually referred to as enzymes.



▲ **Crystals of a bacterial (*Shewanella oneidensis*) homologue of Old Yellow Enzyme.** (Courtesy of J. Elegheert and S. N. Savvides)

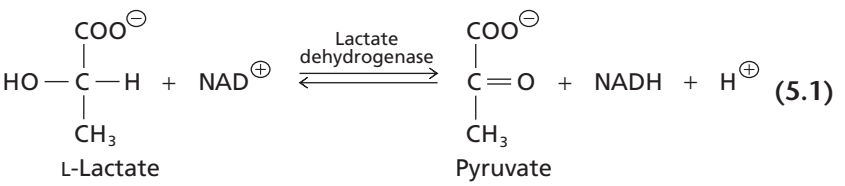
We begin this chapter with a description of enzyme classification and nomenclature. Next, we discuss kinetic analysis (measurements of reaction rates) emphasizing how kinetic experiments can reveal the properties of an enzyme and the nature of the complexes it forms with substrates and inhibitors. Finally, we describe the principles of inhibition and activation of regulatory enzymes. Chapter 6 explains how enzymes work at the chemical level and uses serine proteases to illustrate the relationship between protein structure and enzymatic function. Chapter 7 is devoted to the biochemistry of coenzymes, the organic molecules that assist some enzymes in their catalytic roles by providing reactive groups not found on amino acid side chains. In the remaining chapters we will present many other examples illustrating the four main properties of enzymes: (1) they function as catalysts, (2) they catalyze highly specific reactions, (3) they can couple reactions, and (4) their activity can be regulated.

5.1 The Six Classes of Enzymes

Most of the classical metabolic enzymes are named by adding the suffix *-ase* to the name of their substrates or to a descriptive term for the reactions they catalyze. For example, urease has urea as a substrate. Alcohol dehydrogenase catalyzes the removal of hydrogen from alcohols (i.e., the oxidation of alcohols). A few enzymes, such as trypsin and amylase, are known by their historic names. Many newly discovered enzymes are named after their genes or for some nondescriptive characteristic. For example, RecA is named after the *rec*A gene and HSP70 is a heat shock protein—both enzymes catalyze the hydrolysis of ATP.

A committee of the International Union of Biochemistry and Molecular Biology (IUBMB) maintains a classification scheme that categorizes enzymes according to the general class of organic chemical reaction that is catalyzed. The six categories—oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases—are defined below with an example of each type of enzyme. The IUBMB classification scheme assigns a unique number, called the enzyme classification number, or EC number, to each enzyme. IUBMB also assigns a unique systematic name to each enzyme; it may be different from the common name of an enzyme. This book usually refers to enzymes by their common names.

1. **Oxidoreductases** catalyze oxidation–reduction reactions. Most of these enzymes are commonly referred to as **dehydrogenases**. Other enzymes in this class are called oxidases, peroxidases, oxygenases, or reductases. There is a trend in biochemistry to refer to more and more of these enzymes by their systematic name, oxidoreductases, rather than the more common names in the older biochemical literature. One example of an oxidoreductase is lactate dehydrogenase (EC 1.1.1.27) also called lactate:NAD oxidoreductase. This enzyme catalyzes the reversible conversion of L-lactate to pyruvate. The oxidation of L-lactate is coupled to the reduction of the coenzyme nicotinamide adenine dinucleotide ($NAD^{\oplus}$).



(5.1)

2. **Transferases** catalyze group transfer reactions and many require the presence of coenzymes. In group transfer reactions a portion of the substrate molecule usually binds covalently to the enzyme or its coenzyme. This group includes kinases, enzymes that catalyze the transfer of a phosphoryl group from ATP. Alanine transaminase, whose systematic name is L-alanine:2-oxyglutarate aminotransferase

BOX 5.1 ENZYME CLASSIFICATION NUMBERS

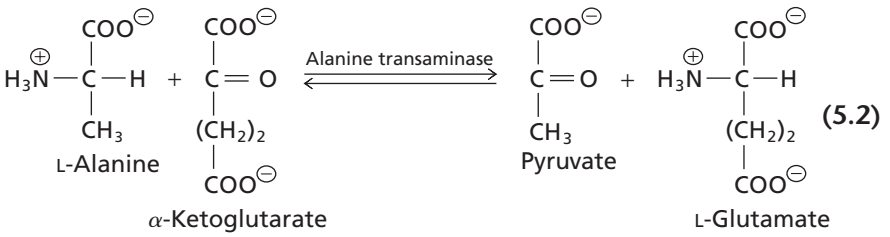
The enzyme classification number for malate dehydrogenase is EC 1.1.1.37. This enzyme has an activity similar to that of lactate dehydrogenase described under oxidoreductases (see Figure 4.23, Box 13.3).

The first number identifies this enzyme as a member of the first class of enzymes (oxidoreductases). The second number identifies the substrate group that malate dehydrogenase recognizes. Subclass 1.1 means that the substrate is a HC—OH group. The third number specifies the electron acceptor for this class of enzymes. Subclass 1.1.1 is for enzymes that use NAD^+ or $NADP^+$ as an acceptor. The final number means that malate dehydrogenase is the 37th enzyme in this category.

Compare the EC number of malate dehydrogenase with that of lactate dehydrogenase to see how similar enzymes have similar classification numbers.

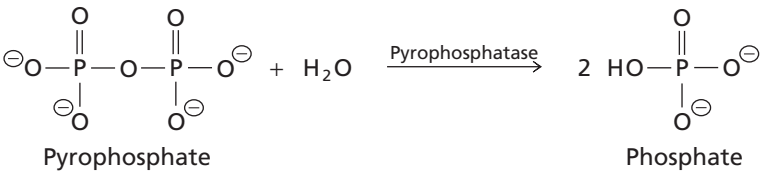
Accurate enzyme identification and classification is an important and essential part of modern biological databases. The entire classification database can be seen at **www.chem.qmul.ac.uk/iubmb/enzyme/.**

(EC 2.6.1.2), is a typical transferase. It transfers an amino group from L-alanine to α-ketoglutarate (2-oxoglutarate).



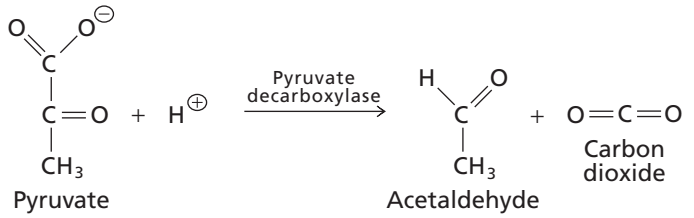
(5.2)

3. **Hydrolases** catalyze hydrolysis. They are a special class of transferases with water serving as the acceptor of the group transferred. Pyrophosphatase is a simple example of a hydrolase. The systematic name of this enzyme is diphosphate phosphohydrolase (EC 3.6.1.1).



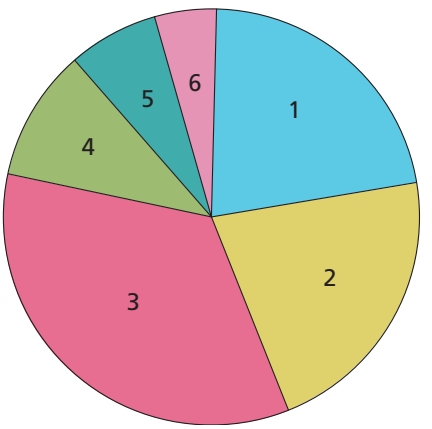
(5.3)

4. **Lyases** catalyze lysis of a substrate generating a double bond in nonhydrolytic, nonoxidative, elimination reactions. In the reverse direction, lyases catalyze the addition of one substrate to the double bond of a second substrate. Pyruvate decarboxylase belongs to this class of enzymes since it splits pyruvate into acetaldehyde and carbon dioxide. The systematic name for pyruvate decarboxylase, 2-oxo-acid carboxy-lyase (EC 4.1.1.1), is rarely used.

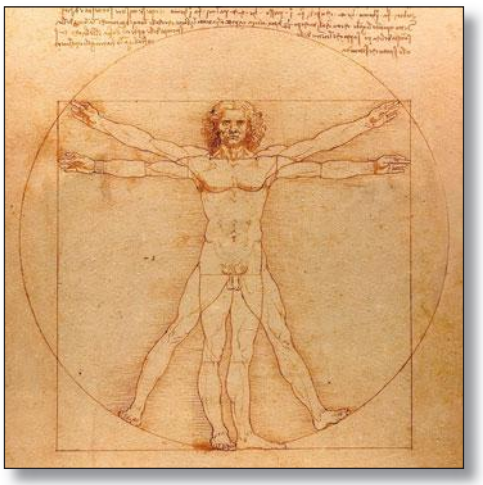


(5.4)

5. **Isomerases** catalyze structural change within a single molecule (isomerization reactions). Because these reactions have only one substrate and one product, they are among the simplest enzymatic reactions. Alanine racemase (EC 5.1.1.1) is an



▲ **Distribution of all known enzymes by EC classification number.** 1. oxidoreductases; 2. transferases; 3. hydrolases; 4. lyases; 5. isomerases; 6. ligases.

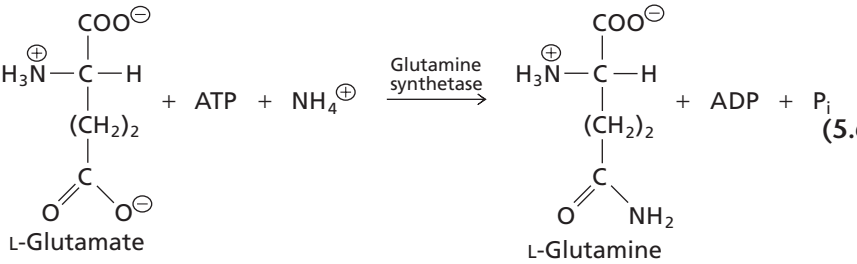


The human genome contains genes for about 1000 different enzymes catalyzing reactions in several hundred metabolic pathways (humancyc.org/). Since many enzymes have multiple subunits there are about 3000 different genes devoted to making enzymes. We have about 20,000 genes so most of the genes in our genome do not encode enzymes or enzyme subunits.

isomerase that catalyzes the interconversion of L-alanine and D-alanine. The common name is the same as the systematic name.

L-Alanine D-Alanine (5.5)

6. **Ligases** catalyze ligation, or joining, of two substrates. These reactions require the input of chemical potential energy in the form of a nucleoside triphosphate such as ATP. Ligases are usually referred to as **synthetases**. Glutamine synthetase, or L-glutamate:ammonia ligase (ADP-forming) (EC 6.3.1.2), uses the energy of ATP hydrolysis to join glutamate and ammonia to produce glutamine.



L-Glutamate L-Glutamine (5.6)

From the examples given above we see that most enzymes have more than one substrate although the second substrate may be only a molecule of water or a proton. Although enzymes catalyze both forward and reverse reactions, one-way arrows are often used when the equilibrium favors a great excess of product over substrate. Remember that when a reaction reaches equilibrium the enzyme must be catalyzing both the forward and reverse reactions at the same rate.

5.2 Kinetic Experiments Reveal Enzyme Properties

We begin our study of enzyme properties by examining the *rates* of enzyme-catalyzed reactions. Such studies fall under the category of enzyme kinetics (from the Greek *kinetikos*, "moving"). This is an appropriate place to begin since the most important property of enzymes is that they act as catalysts, speeding up the rates of reactions. Enzyme kinetics provides indirect information about the specificities and catalytic mechanisms of enzymes. Kinetic experiments also reveal whether an enzyme is regulated.

Recall that concentrations are indicated by square brackets: [P] signifies the concentration of product, [E] the concentration of enzyme, and [S] the concentration of the substrate.

Most enzyme research in the first half of the 20th century was limited to kinetic experiments. This research revealed how the rates of reactions are affected by variations in experimental conditions or changes in the concentration of enzyme or substrate. Before discussing enzyme kinetics in depth, let's review the principles of kinetics for nonenzymatic chemical systems. These principles are then applied to enzymatic reactions.

A. Chemical Kinetics

Kinetic experiments examine the relationship between the amount of product (P) formed in a unit of time ($\Delta[P]/\Delta t$) and the experimental conditions under which the reaction takes place. The basis of most kinetic measurements is the observation that the rate, or **velocity** (v), of a reaction varies directly with the concentration of each reactant (Section 1.4). This observation is expressed in a **rate equation**. For example, the rate equation for the nonenzymatic conversion of substrate (S) to product in an isomerization reaction is written as

$$\frac{\Delta[P]}{\Delta t} = v = k[S] \qquad (5.7)$$

The rate equation reflects the fact that the velocity of a reaction depends on the concentration of the substrate ([S]). The symbol k is the rate constant and indicates the speed or efficiency of a reaction. Each reaction has a different rate constant. The units of the rate constant for a simple reaction are s^{-1}.

As a reaction proceeds, the amount of product ([P]) increases and the amount of substrate ([S]) decreases. An example of the progress of several reactions is shown in Figure 5.1a. The velocity is the slope of the progress curve over a particular interval of time. The shape of the curves indicates that the velocity is decreasing over time as expected since the substrate is being depleted.

In this hypothetical example, the velocity of the reaction might eventually become zero when the substrate is used up. This would explain why the curve flattens out at extended time points. (See below for another explanation.) We are interested in the relationship between substrate concentration and the velocity of a reaction since if we know these two values we can use Equation 5.7 to calculate the rate constant. The only accurate substrate concentration is the one we prepare at the beginning of the experiment because the concentration changes during the experiment. The velocity of the reaction at the very beginning is the value that we want to know. This value represents the rate of the reaction at a known substrate concentration before it changes.

The initial velocity (v_0) can be determined from the slope of the progress curves (Figure 5.1a) or from the derivatives of the curves. A graph of initial velocity versus substrate concentration at the beginning of the experiment gives a straight line as shown in Figure 5.1b. The slope of the curve in Figure 5.1b is the rate constant.

The experiment shown in Figure 5.1 will only determine the forward rate constant since the data were collected under conditions where there was no reverse reaction. This is another important reason for calculating initial velocity (v_0) rather than the rate at later time points. In a reversible reaction, the flattening of the progress curves does not represent zero velocity. Instead, it simply indicates that there is no net increase in product over time because the reaction has reached equilibrium.

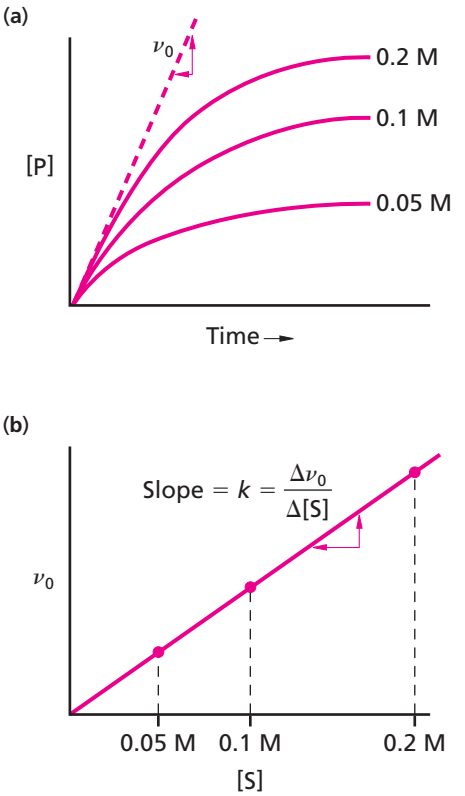
A better description of our simple reaction would be

$$S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} P \qquad \textbf{(5.8)}$$

For a more complicated single-step reaction, such as the reaction $S_1 + S_2 \rightarrow P_1 + P_2$, the rate is determined by the concentrations of both substrates. If both substrates are present at similar concentrations, the rate equation is

$$v = k[S_1][S_2] \qquad \textbf{(5.9)}$$

The rate constant for reactions involving two substrates has the units $M^{-1}\ s^{-1}$. These rate constants can be easily determined by setting up conditions where the concentration of one substrate is very high and the other is varied. The rate of the reaction will depend on the concentration of the rate-limiting substrate.



▲ **Figure 5.1**
Rate of a simple chemical reaction. (a) The amount of product produced over time is plotted for several different initial substrate concentrations. The initial velocity v_0 is the slope of the progress curve at the beginning of the reaction. **(b)** The initial velocity as a function of initial substrate concentration. The slope of the curve is the rate constant.

KEY CONCEPT

The rate or velocity of a reaction depends on the concentration of substrate.

B. Enzyme Kinetics

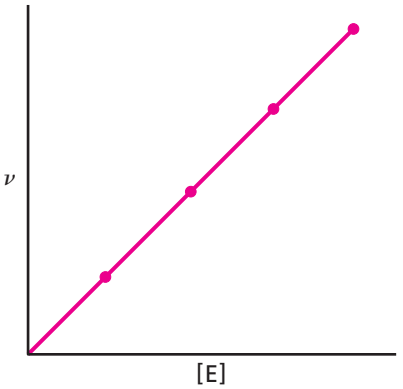
One of the first great advances in biochemistry was the discovery that enzymes bind substrates transiently. In 1894, Emil Fischer proposed that an enzyme is a rigid template, or lock, and that the substrate is a matching key. Only specific substrates can fit into a given enzyme. Early studies of enzyme kinetics confirmed that an enzyme (E) binds a substrate to form an **enzyme–substrate complex (ES)**. ES complexes are formed when ligands bind noncovalently in their proper places in the active site. The substrate interacts transiently with the protein catalyst (and with other substrates in a multisubstrate reaction) on its way to forming the product of the reaction.

Let's consider a simple enzymatic reaction; namely, the conversion of a single substrate to a product. Although most enzymatic reactions have two or more substrates, the general principles of enzyme kinetics can be described by assuming the simple case of one substrate and one product.

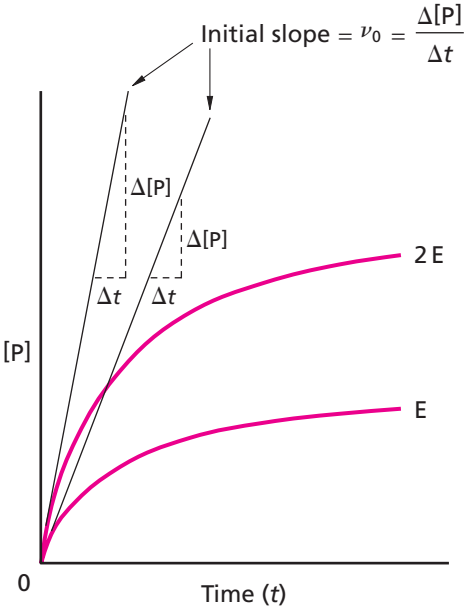
$$E + S \longrightarrow ES \longrightarrow E + P \qquad \textbf{(5.10)}$$

KEY CONCEPT

The enzyme–substrate complex (ES) is a transient intermediate in an enzyme catalyzed reaction.



▲ **Figure 5.2**
Effect of enzyme concentration ([E]), on the initial velocity (v) of an enzyme-catalyzed reaction at a fixed, saturating [S]. The reaction rate is affected by the concentration of enzyme but not by the concentration of the other reactant, S.



▲ **Figure 5.3 Progress curve for an enzyme-catalyzed reaction.** [P], the concentration of product, increases as the reaction proceeds. The initial velocity of the reaction, v_0, is the slope of the initial linear portion of the curve. Note that the rate of the reaction doubles when twice as much enzyme (2E, upper curve) is added to an otherwise identical reaction mixture.

This reaction takes place in two distinct steps—the formation of the enzyme–substrate complex and the actual chemical reaction accompanied by the dissociation of the enzyme and product. Each step has a characteristic rate. The overall rate of an enzymatic reaction depends on the concentrations of both the substrate and the catalyst (enzyme). When the amount of enzyme is much less than the amount of substrate the reaction will depend on the amount of enzyme.

The straight line in Figure 5.2 illustrates the effect of enzyme concentration on the reaction velocity in a pseudo first-order reaction. The more enzyme present, the faster the reaction. These conditions are used in enzyme assays to determine the concentrations of enzymes. The concentration of enzyme in a test sample can be easily determined by comparing its activity to a reference curve similar to the model curve in Figure 5.2. Under these experimental conditions, there are sufficient numbers of substrate molecules so that every enzyme molecule binds a molecule of substrate to form an ES complex, a condition called saturation of E with S. Enzyme assays measure the amount of product formed in a given time period. In some assay methods, a recording spectrophotometer can be used to record data continuously; in other methods, samples are removed and analyzed at intervals. The assay is performed at a constant pH and temperature, generally chosen for optimal enzyme activity or for approximation to physiological conditions.

If we begin an enzyme-catalyzed reaction by mixing substrate and enzyme then there is no product present during the initial stages of the reaction. Under these conditions we can ignore the reverse reaction where P binds to E and is converted to S. The reaction can be described by

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \qquad \textbf{(5.11)}$$

The rate constants k_1 and k_{-1} in Reaction 5.11 govern the rates of association of S with E and dissociation of S from ES, respectively. This first step is an equilibrium binding interaction similar to the binding of oxygen to hemoglobin. The rate constant for the second step is k_2, the rate of formation of product from ES. Note that conversion of the ES complex to free enzyme and product is shown by a one-way arrow because the rate of the reverse reaction (E + P → EP) is negligible at the start of a reaction. The velocity measured during this short period is the **initial velocity** (v_0) described in the previous section. The formation and dissociation of ES complexes are usually very rapid reactions because only noncovalent bonds are being formed and broken. In contrast, the conversion of substrate to product is usually rate limiting. It is during this step that the substrate is chemically altered.

Enzyme kinetics differs from simple chemical kinetics because the rates of enzyme-catalyzed reactions depend on the concentration of enzyme and the enzyme is neither a product nor a substrate of the reaction. The rates also differ because substrate has to bind to enzyme before it can be converted to product. In an enzyme-catalyzed reaction, the initial velocities are obtained from progress curves, just as they are in chemical reactions. Figure 5.3 shows the progress curves at two different enzyme concentrations in the presence of a high initial concentration of substrate ([S] >> [E]). In this case, the rate of product formation depends on enzyme concentration and not on the substrate concentration. Data from experiments such as those shown in Figure 5.3 can be used to plot the curve shown in Figure 5.2.

5.3 The Michaelis–Menten Equation

Enzyme-catalyzed reactions, like any chemical reaction, can be described mathematically by rate equations. Several constants in the equations indicate the efficiency and specificity of an enzyme and are therefore useful for comparing the activities of several enzymes or for assessing the physiological importance of a given enzyme. The first rate equations were derived in the early 1900s by examining the effects of variations in substrate concentration. Figure 5.4 a shows a typical result where the initial velocity (v_0) of a reaction is plotted against the substrate concentration ([S]).

The data can be explained by the reaction shown in Reaction 5.11. The first step is a bimolecular interaction between the enzyme and substrate to form an ES complex. At high substrate concentrations (right-hand side of the curve in Figure 5.4) the initial velocity doesn't change very much as more S is added. This indicates that the amount of enzyme has become rate-limiting in the reaction. The concentration of enzyme is an important component of the overall reaction as expected for formation of an ES complex. At low substrate concentrations (left-hand side of the curve in Figure 5.4), the initial velocity is very sensitive to changes in the substrate concentration. Under these conditions most enzyme molecules have not yet bound substrate and the formation of the ES complex depends on the substrate concentration.

The shape of the v_0 vs. [S] curve is that of a rectangular hyperbola. Hyperbolic curves indicate processes involving simple dissociation as we saw for the dissociation of oxygen from oxymyoglobin (Section 4.13B). This is further evidence that the simple reaction under study is bimolecular involving the association of E and S to form an ES complex. The equation for a rectangular hyperbola is

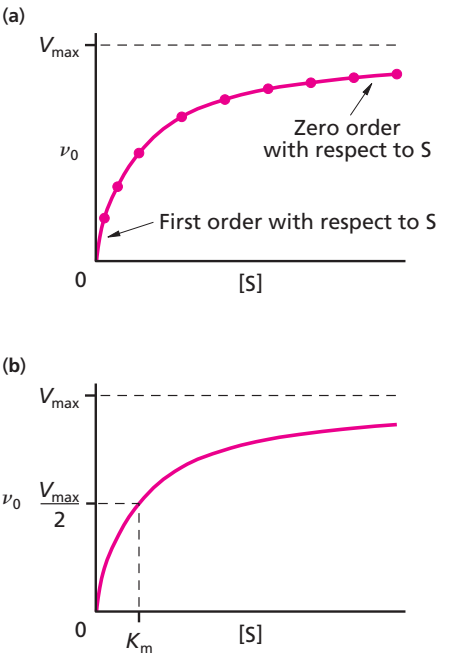
$$y = \frac{ax}{b + x} \tag{5.12}$$

where a is the asymptote of the curve (the value of y at an infinite value of x) and b is the point on the x axis corresponding to a value of $a/2$. In enzyme kinetic experiments, $y = v_0$ and $x = [S]$. The asymptote value (a) is called V_{max}. It's the maximum velocity of the reaction at infinitely large substrate concentrations. We often show the V_{max} value on v_0 vs. [S] plots but if you look at the figure it's not obvious why this particular asymptote was chosen. One of the characteristics of hyperbolic curves is that the curve seems to flatten out at moderate substrate concentrations at a level that seems far less than the V_{max} value. The true V_{max} is not determined by trying to estimate the position of the asymptote from the shape of the curve; instead, it is precisely and correctly determined by fitting the data to the general equation for a rectangular hyperbola.

The b term in the general equation for a rectangular hyperbola is called the **Michaelis constant** (K_m) defined as the concentration of substrate when v_0 is equal to one-half V_{max} (Figure 5.4b). The complete rate equation is

$$v_0 = \frac{V_{max}[S]}{K_m + [S]} \tag{5.13}$$

This is called the **Michaelis–Menten equation**, named after Leonor Michaelis and Maud Menten. Note how the general form of the equation compares to Equation 5.12. The Michaelis–Menten equation describes the relationship between the initial velocity of a reaction and the substrate concentration. In the following section we derive the Michaelis–Menten equation by a kinetic approach and then consider the meaning of the various constants.



▲ **Figure 5.4**
Plots of initial velocity (v_0) versus substrate concentration ([S]) for an enzyme-catalyzed reaction. (a) Each experimental point is obtained from a separate progress curve using the same concentration of enzyme. The shape of the curve is hyperbolic. At low substrate concentrations, the curve approximates a straight line that rises steeply. In this region of the curve, the reaction is highly dependent on the concentration of substrate. At high concentrations of substrate, the enzyme is almost saturated, and the initial rate of the reaction does not change much when substrate concentration is further increased. **(b)** The concentration of substrate that corresponds to half-maximum velocity is called the Michaelis constant (K_m). The enzyme is half-saturated when S = K_m.

A. Derivation of the Michaelis–Menten Equation

One common derivation of the Michaelis–Menten equation is termed the *steady state derivation*. It was proposed by George E. Briggs and J. B. S. Haldane. This derivation postulates a period of time (called the steady state) during which the ES complex is formed at the same rate that it decomposes so that the concentration of ES is constant. The initial velocity is used in the steady state derivation because we assume that the concentration of product ([P]) is negligible. The steady state is a common condition for metabolic reactions in cells.

If we assume a constant steady state concentration of ES then the rate of formation of product depends on the rate of the chemical reaction and the rate of dissociation of P from the enzyme. The rate limiting step is the right-hand side of Reaction 5.11 and the velocity depends on the rate constant k_2 and the concentration of ES.

$$ES \xrightarrow{k_2} E + P \qquad v_0 = k_2[ES] \tag{5.14}$$

▲ Leonor Michaelis (1875–1949).

The steady-state derivation solves Equation 5.14 for [ES] using terms that can be measured such as the rate constant, the total enzyme concentration ($[E]_{total}$), and the substrate concentration ([S]). [S] is assumed to be greater than $[E]_{total}$ but not necessarily saturating. For example, soon after a small amount of enzyme is mixed with substrate [ES] becomes constant because the overall rate of decomposition of ES (the sum of the rates of conversion of ES to E + S and to E + P) is equal to the rate of formation of the ES complex from E + S. The rate of formation of ES from E + S depends on the concentration of free enzyme (enzyme molecules not in the form of ES) which is $[E]_{total} - [ES]$. The concentration of the ES complex remains constant until consumption of S causes [S] to approach $[E]_{total}$. We can express these statements as a mathematical equation.

$$\underset{\text{Rate of ES formation}}{k_1([E]_{total} - [ES])[S]} = \underset{\text{Rate of ES decomposition}}{(k_{-1} + k_2)[ES]} \quad \textbf{(5.15)}$$

Equation 5.15 is rearranged to collect the rate constants.

$$\frac{k_{-1} + k_2}{k_1} = K_m = \frac{1[E]_{total} - [ES]2[S]}{[ES]} \quad \textbf{(5.16)}$$

The ratio of rate constants on the left-hand side of Equation 5.16 is the Michaelis constant, K_m. Next, this equation is solved for [ES] in several steps.

$$[ES]K_m = ([E]_{total} - [ES])[S] \quad \textbf{(5.17)}$$

Expanding,

$$[ES]K_m = ([E]_{total}[S]) - ([ES][S]) \quad \textbf{(5.18)}$$

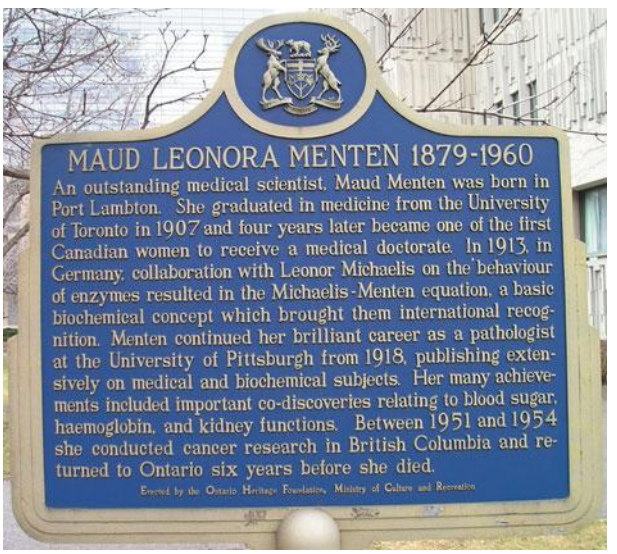
Collecting [ES] terms,

$$[ES](K_m + [S]) = [E]_{total}[S] \quad \textbf{(5.19)}$$

and

$$[ES] = \frac{[E]_{total}[S]}{K_m + [S]} \quad \textbf{(5.20)}$$

▼ Maud Menten (1879–1960).



Equation 5.20 describes the steady-state ES concentration using terms that can be measured in an experiment. Substituting the value of [ES] into the velocity equation (Equation 5.14) gives

$$v_0 = k_2[\text{ES}] = \frac{k_2[\text{E}]_{\text{total}}[\text{S}]}{K_m + [\text{S}]} \tag{5.21}$$

As indicated by Figure 5.4a, when the concentration of S is very high the enzyme is saturated and essentially all the molecules of E are present as ES. Adding more S has almost no effect on the reaction velocity. The only way to increase the velocity is to add more enzyme. Under these conditions the velocity is at its maximum rate (V_{max}) and this velocity is determined by the total enzyme concentration and the rate constant k_2. Thus, by definition,

$$V_{max} = k_2[\text{E}]_{\text{total}} \tag{5.22}$$

Substituting this in Equation 5.21 gives the most familiar form of the Michaelis–Menten equation.

$$v_0 = \frac{V_{max}[\text{S}]}{K_m + [\text{S}]} \tag{5.23}$$

We've already seen that this form of the Michaelis–Menten equation adequately describes the data from kinetic experiments. In this section we've shown that the same equation can be derived from a theoretical consideration of the implications of Reaction 5.11, the equation for an enzyme-catalyzed reaction. The agreement between theory and data gives us confidence that the theoretical basis of enzyme kinetics is sound.

KEY CONCEPT

The constant k_{cat} is the number of moles of substrate converted to product per second per mole of enzyme.

B. The Catalytic Constant k_{cat}

At high substrate concentration, the overall velocity of the reaction is V_{max} and the rate is determined by the enzyme concentration. The rate constant observed under these conditions is called the **catalytic constant, k_{cat}**, defined as

$$V_{max} = k_{cat}[\text{E}]_{\text{total}} \qquad k_{cat} = \frac{V_{max}}{[\text{E}]_{\text{total}}} \tag{5.24}$$

where k_{cat} represents the number of moles of substrate converted to product per second per mole of enzyme (or per mole of active site for a multisubunit enzyme) under saturating conditions. In other words, k_{cat} indicates the maximum number of substrate molecules converted to product each second by each active site. This is often called the **turnover number**. The catalytic constant measures how quickly a given enzyme can catalyze a specific reaction—it's a very useful way of describing the effectiveness of an enzyme. The unit for k_{cat} is s^{-1} and the reciprocal of k_{cat} is the time required for one catalytic event. Note that the enzyme concentration must be known in order to calculate k_{cat}.

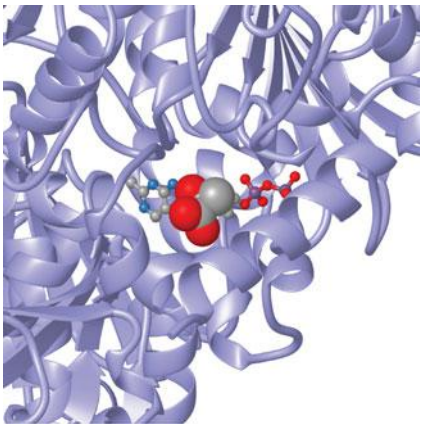
For a simple reaction, such as Reaction 5.11, the rate-limiting step is the conversion of substrate to product and the dissociation of product from the enzyme (ES → E + P). Under these conditions k_{cat} is equal to k_2 (Equation 5.14). Many enzyme reactions are more complex. If one step is clearly rate-limiting then its rate constant is the k_{cat} for that reaction. If the mechanism is more complex then k_{cat} may be a combination of several different rate constants. This is why we need a different rate constant (k_{cat}) to describe the overall rate of the enzyme-catalyzed reaction. In most cases you can assume that k_{cat} is a good approximation of k_2.

Representative values of k_{cat} are listed in Table 5.1. Most enzymes are potent catalysts with k_{cat} values of 10^2 to 10^3 s^{-1}. This means that at high substrate concentrations a single

Table 5.1 Examples of catalytic constants

Enzyme	$k_{cat}(s^{-1})$*
Papain	10
Ribonuclease	10^2
Carboxypeptidase	10^2
Trypsin	10^2 (to 10^3)
Acetylcholinesterase	10^3
Kinases	10^3
Dehydrogenases	10^3
Transaminases	10^3
Carbonic anhydrase	10^6
Superoxide dismutase	10^6
Catalase	10^7

*The catalytic constants are given only as orders of magnitude.



▲ **Substrate binding.** Pyruvate carboxylase binds pyruvate, HCO_3^-, and ATP. The structure of the active site of the yeast (*Saccharomyces cerevisiae*) enzyme is shown here with a bound molecule of pyruvate (space-filling representation) and the cofactor biotin (ball-and-stick). The K_m value for pyruvate binding is 4×10^{-4} M. The K_m values for HCO_3^-, and ATP binding are 1×10^{-3} M and 6×10^{-5} M. [PDB 2VK1]

enzyme molecule will convert 100–1000 molecules of substrate to product every second. This rate is limited by a number of factors that will be discussed in the next chapter (Chapter 6: *Mechanisms of Enzymes*).

Some enzymes are extremely rapid catalysts with k_{cat} values of 10^6 s^{-1} or greater. Mammalian carbonic anhydrase, for example, must act very rapidly in order to maintain equilibrium between aqueous CO_2 and bicarbonate (Section 2.10). As we will see in Section 6.4B, superoxide dismutase and catalase are responsible for rapid decomposition of the toxic oxygen metabolites superoxide anion and hydrogen peroxide, respectively. Enzymes that catalyze a million reactions per second often act on small substrate molecules that diffuse rapidly inside the cell.

C. The Meanings of K_m

The Michaelis constant has a number of meanings. Equation 5.16 defined K_m as the ratio of the combined rate constants for the breakdown of ES divided by the constant for its formation. If the rate constant for product formation (k_2) is much smaller than either k_1 or k_{-1}, as is often the case, k_2 can be neglected and K_m is equivalent to k_{-1}/k_1. In this case K_m is the same as the equilibrium constant for dissociation of the ES complex to E +S. Thus, K_m becomes a measure of the affinity of E for S. The lower the value of K_m, the more tightly the substrate is bound. K_m is also one of the parameters that determines the shape of the v_0 vs. [S] curve shown in Figure 5.4b. It is the substrate concentration when the initial velocity is one-half the V_{max} value. This meaning follows directly from the general equation for a rectangular hyperbola.

K_m values are sometimes used to distinguish between different enzymes that catalyze the same reaction. For example, mammals have several different forms of lactate dehydrogenase, each with a distinct K_m value. Although it is useful to think of K_m as representing the equilibrium dissociation constant for ES, this is not always valid. For many enzymes K_m is a more complex function of the rate constants. This is especially true when the reaction occurs in more than two steps.

Typical K_m values for enzymes range from 10^{-2} to 10^{-5} M. Since these values often represent apparent dissociation constants their reciprocal is an apparent association (binding) constant. You can see by comparison with protein–protein interactions (Section 4.9) that the binding of enzymes to substrates is much weaker.

5.4 Kinetic Constants Indicate Enzyme Activity and Catalytic Proficiency

KEY CONCEPT

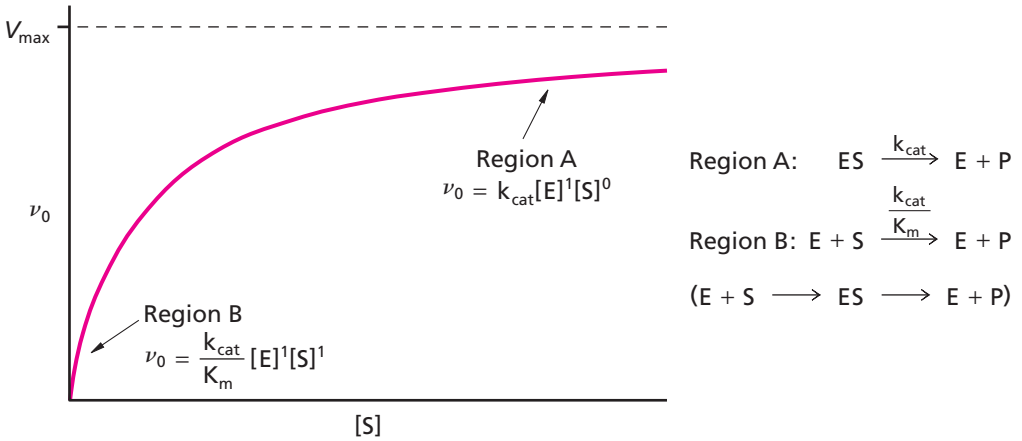
K_m is the substrate concentration when the rate of the reaction is one-half the V_{max} value. It is often an approximation of the equilibrium dissociation constant of the reaction ES $\rightleftharpoons$ E + S.

We've seen that the kinetic constants K_m and k_{cat} can be used to gauge the relative activities of enzymes and substrates. In most cases, K_m is a measure of the stability of the ES complex and k_{cat} is similar to the rate constant for the conversion of ES to E + P when the substrate is not limiting (region A in Figure 5.5). Recall that k_{cat} is a measure of the catalytic activity of an enzyme indicating how many reactions a molecule of enzyme can catalyze per second.

Examine region B of the hyperbolic curve in Figure 5.5. The concentration of S is very low and the curve approximates a straight line. Under these conditions, the reaction rate depends on the concentrations of both substrate and enzyme. In chemical terms, this is a second-order reaction and the velocity depends on a second-order rate constant defined by

$$v_0 = k[E][S] \qquad (5.25)$$

We are interested in knowing how to determine this second-order rate constant since it tells us the rate of the enzyme-catalyzed reaction under physiological conditions. When Michaelis and Menten first wrote the full rate equation they used the form that included $k_{cat}[E]_{total}$ rather than V_{max} (Equation 5.24). Now that we understand the meaning of k_{cat}



▲ **Figure 5.5 Meanings of k_{cat} and k_{cat}/K_m.** The catalytic constant (k_{cat}) is the rate constant for conversion of the ES complex to E + P. It is measured most easily when the enzyme is saturated with substrate (region A on the Michaelis–Menten curve shown). The ratio k_{cat}/K_m is the rate constant for the conversion of E + S to E + P at very low concentrations of substrate (region B). The reactions measured by these rate constants are summarized below the graph.

we can substitute $k_{cat}[E]_{total}$ in the Michaelis–Menten equation (Equation 5.23) in place of V_{max}. If we consider only the region of the Michaelis–Menten curve at a very low [S] then this equation can be simplified by neglecting the [S] in the denominator since [S] is much less than K_m.

$$v_0 = \frac{k_{cat}[E][S]}{K_m + [S]} = \frac{k_{cat}}{K_m}[E][S] \qquad \textbf{(5.26)}$$

Comparing Equations 5.25 and 5.26 reveals that the second-order rate constant is closely approximated by k_{cat}/K_m. Thus, the ratio k_{cat}/K_m is an apparent second-order rate constant for the formation of E + P from E + S when the overall reaction is limited by the encounter of S with E. This ratio approaches 10^8 to 10^9 M^{-1} s^{-1}, the fastest rate at which two uncharged solutes can approach each other by diffusion at physiological temperature. Enzymes that can catalyze reactions at this extremely rapid rate are discussed in Section 6.4.

The k_{cat}/K_m ratio is useful for comparing the activities of different enzymes. It is also possible to assess the efficiency of an enzyme by measuring its catalytic proficiency. This value is equal to the rate constants for a reaction in the presence of the enzyme (k_{cat}/K_m) divided by the rate constant for the same reaction in the absence of the enzyme (k_n). Surprisingly few catalytic proficiency values are known because most chemical reactions occur extremely slowly in the absence of enzymes—so slowly that their nonenzymatic rates are very difficult to measure. The reaction rates are often measured in special steel-enclosed glass vessels at temperatures in excess of 300°C.

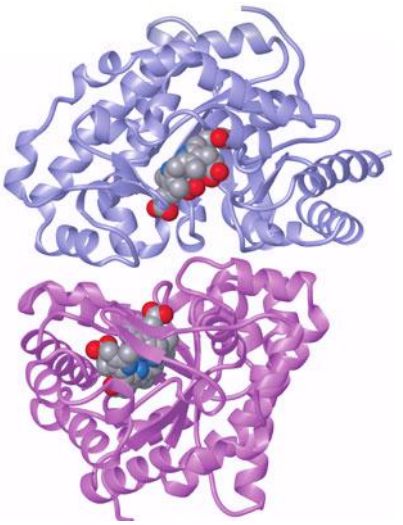
Table 5.2 lists several examples of known catalytic proficiencies. Typical values range from 10^{14} to 10^{20} but some are quite a bit higher (up to 10^{24}). The current record holder is uroporphyrinogen decarboxylase, an enzyme required for a step in the porphyrin synthesis pathway. The difficulty in obtaining rate constants for nonenzymatic reactions is illustrated by the half-life for the uncatalyzed reaction—about 2 billion years! The catalytic proficiency values in Table 5.2 emphasize one of the main properties of enzymes, namely, their ability to increase the rates of reactions that would normally occur too slowly to be useful.

5.5 Measurement of K_m and V_{max}

The kinetic parameters of an enzymatic reaction can provide valuable information about the specificity and mechanism of the reaction. The key parameters are K_m and V_{max} because k_{cat} can be calculated if V_{max} is known.

Table 5.2 Catalytic proficiencies of some enzymes

	Nonenzymatic rate constant (k_n in s^{-1})	Enzymatic rate constant (k_{cat}/K_m in $M^{-1}s^{-1}$)	Catalytic proficiency
Carbonic anhydrase	10^{-1}	7×10^6	7×10^7
Chymotrypsin	4×10^{-9}	9×10^7	2×10^{16}
Chorismate mutase	10^{-5}	2×10^6	2×10^{11}
Triose phosphate isomerase	4×10^{-6}	4×10^8	10^{14}
Cytidine deaminase	10^{-10}	3×10^6	3×10^{16}
Adenosine deaminase	2×10^{-10}	10^7	5×10^{16}
Mandelate racemase	3×10^{-13}	10^6	3×10^{18}
β-Amylase	7×10^{-14}	10^7	10^{20}
Fumarase	10^{-13}	10^9	10^{21}
Arginine decarboxylase	9×10^{-16}	10^6	10^{21}
Alkaline phosphatase	10^{-15}	3×10^7	3×10^{22}
Orotidine 5′-phosphate decarboxylase	3×10^{-16}	6×10^7	2×10^{23}
Uroporphyrinogen decarboxylase	10^{-17}	2×10^7	2×10^{24}



▲ **Maximum catalytic proficiency.** Uroporphyrinogen decarboxylase is the current record holder for maximum catalytic proficiency. It catalyzes a step in the heme synthesis pathway. The enzyme shown here is a human *(Homo sapiens)* variant with a bound porphoryrin molecule at the active site of each monomer. [PDB 2Q71]

K_m and V_{max} for an enzyme-catalyzed reaction can be determined in several ways. Both values can be obtained by the analysis of initial velocities at a series of substrate concentrations and a fixed concentration of enzyme. In order to obtain reliable values for the kinetic constants the [S] points must be spread out both below and above K_m to produce a hyperbola. It is difficult to determine either K_m or V_{max} directly from a graph

BOX 5.2 HYPERBOLAS VERSUS STRAIGHT LINES

We have seen that a plot of substrate concentration ([S]) versus the initial velocity of a reaction (v_0) produces a hyperbolic curve as shown in Figures 5.4 and 5.5. The general equation for a rectangular hyperbola (Equation 5.12) and the Michaelis–Menten equation have the same form (Equation 5.13).

It's very difficult to determine V_{max} from a plot of enzyme kinetic data since the hyperbolic curve that shows the relationship between substrate concentration and initial velocity is asymptotic to V_{max} and it is experimentally difficult to achieve the concentration of substrate required to estimate V_{max}. For these reasons, it is often easier to convert the hyperbolic curve to a linear form that matches the general formula $y = mx + b$, where m is the slope of the line and b is the y-axis intercept. The first step in transforming the original Michaelis–Menten equation to this general form of a linear equation is to invert the terms so that the K_m + [S] term is on top of the right-hand side. This is done by taking the reciprocal of each side—a transformation that will be familiar to many who are familiar with hyperbolic curves.

The next two steps involve separating terms and canceling [S] in the second term on the right-hand side of the equation. This form of the Michaelis–Menten equation is called the Lineweaver–Burk equation and it resembles the general form of a linear equation, $y = mx + b$, where y is the reciprocal of v_0 and x values are the reciprocal of [S]. A plot of data in this form is referred to as a double-reciprocal plot. The slope of the line will be K_m/V_{max} and the y-axis intercept will be $1/V_{max}$.

The original reason for this sort of transformation was to calculate K_m and V_{max} from experimental data. It was easier to plot the reciprocal values of v_0 and [S] and draw a straight line through the points in order to calculate the kinetic constants. Nowadays, there are computer programs that can accurately fit the data to a hyperbolic curve and calculate the constants so the Lineweaver–Burk plot is no longer necessary for this type of analysis. In this book we will still use the Lineweaver–Burk plots to illustrate some general features of enzyme kinetics but they are rarely used for their original purpose of data analysis.

$$\frac{1}{v_0} = \frac{K_m + [S]}{V_{max}[S]} \qquad \frac{1}{v_0} = \frac{K_m}{V_{max}[S]} + \frac{[S]}{V_{max}[S]} \qquad \frac{1}{v_0} = \left(\frac{K_m}{V_{max}}\right)\frac{1}{[S]} + \frac{1}{V_{max}}$$

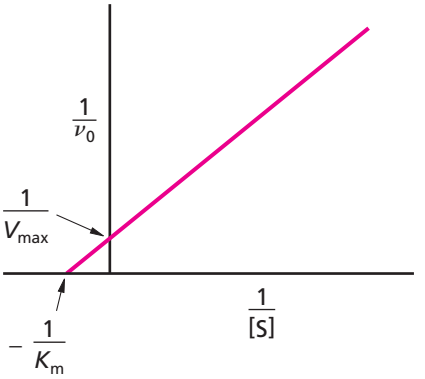
of initial velocity versus concentration because the curve approaches V_{max} asymptotically. However, accurate values can be determined by using a suitable computer program to fit the experimental results to the equation for the hyperbola.

The Michaelis–Menten equation can be rewritten in order to obtain values for V_{max} and K_m from straight lines on graphs. The most commonly used transformation is the double-reciprocal, or Lineweaver–Burk, plot in which the values of $1/v_0$ are plotted against 1/[S] (Figure 5.6). The absolute value of $1/K_m$ is obtained from the intercept of the line at the x axis, and the value of $1/V_{max}$ is obtained from the y intercept. Although double-reciprocal plots are not the most accurate methods for determining kinetic constants, they are easily understood and provide recognizable patterns for the study of enzyme inhibition, an extremely important aspect of enzymology that we will examine shortly.

Values of k_{cat} can be obtained from measurements of V_{max} only when the absolute concentration of the enzyme is known. Values of K_m can be determined even when enzymes have not been purified provided that only one enzyme in the impure preparation can catalyze the observed reaction.

Lineweaver–Burk equation:

$$\frac{1}{v_0} = \left(\frac{K_m}{V_{max}}\right)\frac{1}{[S]} + \frac{1}{V_{max}}$$



▲ **Figure 5.6**
Double-reciprocal (Lineweaver–Burk) plot. This plot is derived from a linear transformation of the Michaelis–Menten equation. Values of $1/v_0$ are plotted as a function of 1/[S] values.

5.6 Kinetics of Multisubstrate Reactions

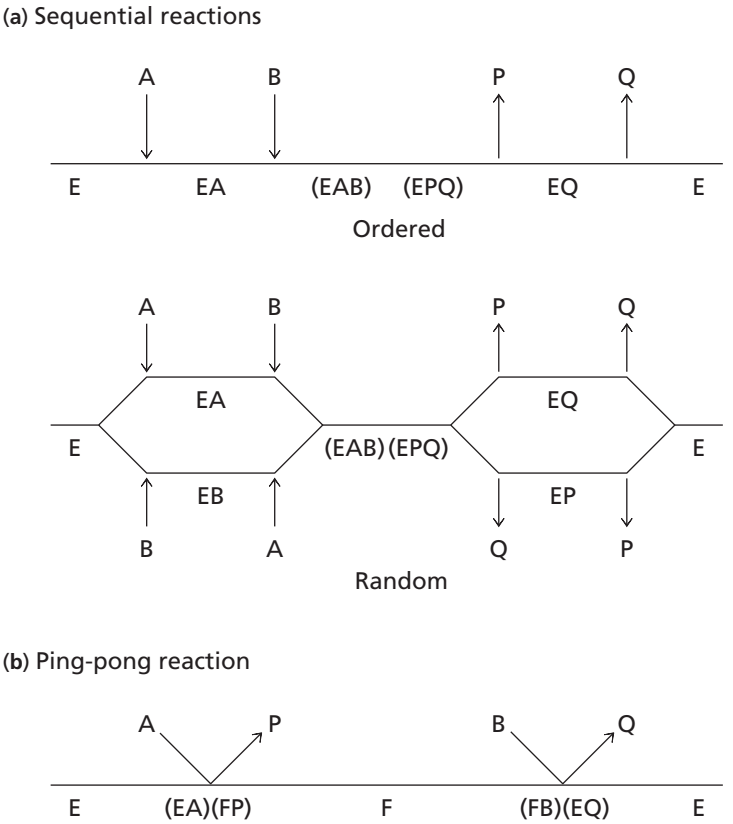
Until now, we have only been considering reactions where a single substrate is converted to a single product. Let's consider a reaction in which two substrates, A and B, are converted to products P and Q.

$$E + A + B \rightleftharpoons (EAB) \rightarrow E + P + Q \qquad \textbf{(5.27)}$$

Kinetic measurements for such multisubstrate reactions are a little more complicated than simple one-substrate enzyme kinetics. For many purposes, such as designing an enzyme assay, it's sufficient simply to determine the K_m for each substrate in the presence of saturating amounts of each of the other substrates as we described for chemical reactions (Section 5.2A). The simple enzyme kinetics discussed in this chapter can be extended to distinguish among several mechanistic possibilities for multisubstrate reactions, such as group transfer reactions. This is done by measuring the effect of variations in the concentration of one substrate on the kinetic results obtained for the other.

Multisubstrate reactions can occur by several different kinetic schemes. These schemes are called **kinetic mechanisms** because they are derived entirely from kinetic experiments. Kinetic mechanisms are commonly represented using the notation introduced by W. W. Cleland. The sequence of steps proceeds from left to right (Figure 5.7). The addition of substrate molecules (A, B, C, . . .) to the enzyme and the release of products (P, Q, R, . . .) from the enzyme are indicated by arrows pointing toward (substrate binding) or from (product release) the line. The various forms of the enzyme (free E, ES complexes, or EP complexes) are written under a horizontal line. The ES complexes that undergo chemical transformation when the active site is filled are shown in parentheses.

Sequential reactions (Figure 5.7a) require all the substrates to be present before any product is released. Sequential reactions can be either **ordered**, with an obligatory order for the addition of substrates and release of products, or **random**. In **ping-pong reactions** (Figure 5.7b), a product is released before all the substrates are bound. In a bisubstrate ping-pong reaction, the first substrate is bound, the enzyme is altered by substitution, and the first product is released. Then the second substrate is bound, the altered enzyme is restored to its original form, and the second product is released. A ping-pong mechanism is sometimes called a substituted-enzyme mechanism because of the covalent binding of a portion of a substrate to the enzyme. The binding and release of ligands in a ping-pong mechanism are usually indicated by slanted lines. The two forms of the enzyme are represented by E (unsubstituted) and F (substituted).



▲ **Figure 5.7**
Notation for bisubstrate reactions. (a) In sequential reactions, all substrates are bound before a product is released. The binding of substrates may be either ordered or random. **(b)** In ping-pong reactions, one substrate is bound and a product is released, leaving a substituted enzyme. A second substrate is then bound and a second product released, restoring the enzyme to its original form.

5.7 Reversible Enzyme Inhibition

An enzyme **inhibitor** (I) is a compound that binds to an enzyme and interferes with its activity. Inhibitors can act by preventing the formation of the ES complex or by blocking the chemical reaction that leads to the formation of product. As a general rule, inhibitors are small molecules that bind reversibly to the enzyme they inhibit. Cells contain many natural enzyme inhibitors that play important roles in regulating metabolism. Artificial inhibitors are used experimentally to investigate enzyme mechanisms and decipher metabolic pathways. Some drugs, and many poisons, are enzyme inhibitors.

Some inhibitors bind covalently to enzymes causing irreversible inhibition but most biologically relevant inhibition is reversible. Reversible inhibitors are bound to enzymes by the same weak, noncovalent forces that bind substrates and products. The equilibrium between free enzyme (E) plus inhibitor (I) and the EI complex is characterized by a dissociation constant. In this case, the constant is called the **inhibition constant**, K_i.

Irreversible inhibitors are described in Section 5.8.

$$E + I \rightleftharpoons EI \qquad K_d = K_i = \frac{[E][I]}{[EI]} \tag{5.28}$$

KEY CONCEPT

Reversible inhibitors bind to enzymes and either prevent substrate binding or block the reaction leading to formation of product.

The basic types of reversible inhibition are competitive, uncompetitive, noncompetitive and mixed. These can be distinguished experimentally by their effects on the kinetic behavior of enzymes (Table 5.3). Figure 5.8 shows diagrams representing modes of reversible enzyme inhibition.

Table 5.3 Effects of reversible inhibitors on kinetic constants

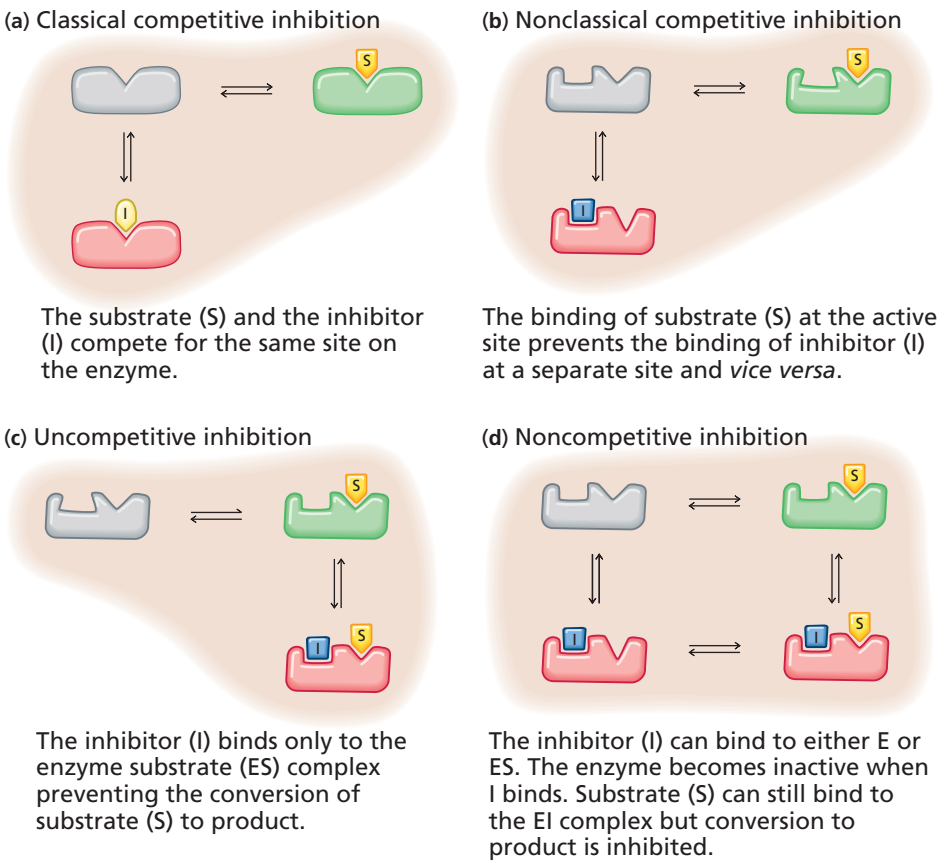
Type of inhibitor	Effect
Competitive (I binds to E only)	Raises K_m V_{max} remains unchanged
Uncompetitive (I binds to ES only)	Lowers V_{max} and K_m Ratio of V_{max}/K_m remains unchanged
Noncompetitive (I binds to E or ES)	Lowers V_{max} K_m remains unchanged

A. Competitive Inhibition

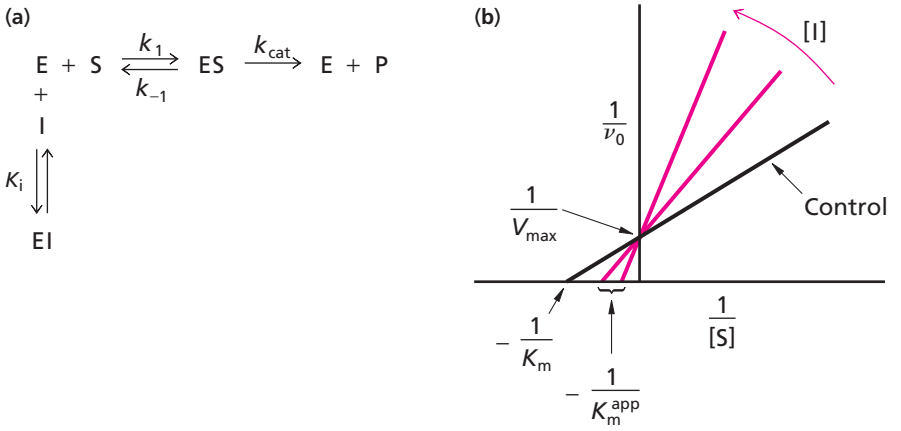
Competitive inhibitors are the most commonly encountered inhibitors in biochemistry. In competitive inhibition, the inhibitor can bind only to free enzyme molecules that have not bound any substrate. **Competitive inhibition** is illustrated in Figure 5.8 and by the kinetic scheme in Figure 5.9a. In this scheme only ES can lead to the formation of product. The formation of an EI complex removes enzyme from the normal pathway.

Once a competitive inhibitor is bound to an enzyme molecule, a substrate molecule cannot bind to that enzyme molecule. Conversely, the binding of substrate to an enzyme molecule prevents the binding of an inhibitor. In other words, S and I compete for binding to the enzyme molecule. Most commonly, S and I bind at the same site on the enzyme, the active site. This type of inhibition is termed classical competitive inhibition (Figure 5.8). This is not the only kind of competitive inhibition (see Figure 5.8). In some cases, such as allosteric enzymes (Section 5.10), the inhibitor binds at a different site and this alters the substrate binding site preventing substrate binding. This type of inhibition is called nonclassical competitive inhibition. When both I and S are

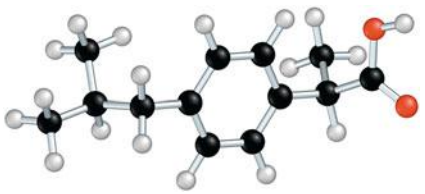
▲ **Competitive inhibition.** The active ingredient in the weed killer Roundup© is glyphosate, a competitive inhibitor of the plant enzyme 5-enolpyruvylshikimate-3-phosphate synthase. (See Box 17.2 in Chapter 17.)



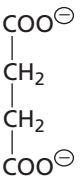
◄ **Figure 5.8**
Diagrams of reversible enzyme inhibition. In this scheme, catalytically competent enzymes are green and inactive enzymes are red.



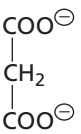
▲ **Figure 5.9**
Competitive inhibition. (a) Kinetic scheme illustrating the binding of I to E. Note that this is an expansion of Equation 5.11 that includes formation of the EI complex. **(b)** Double-reciprocal plot. In competitive inhibition, V_{max} remains unchanged and K_m increases. The black line labeled "Control" is the result in the absence of inhibitor. The red lines are the results in the presence of inhibitor, with the arrow showing the direction of increasing [I].



▲ Ibuprofen, the active ingredient in many over-the-counter painkillers, is a competitive inhibitor of the enzyme cyclooxygenase (COX). (See Box 16.1 Chapter 16.)



Succinate



Malonate

present in a solution, the proportion of the enzyme that is able to form ES complexes depends on the concentrations of substrate and inhibitor and their relative affinities for the enzyme.

The amount of EI can be reduced by increasing the concentration of S. At sufficiently high concentrations the enzyme can still be saturated with substrate. Therefore, the maximum velocity is the same in the presence or in the absence of an inhibitor. The more competitive inhibitor present, the more substrate needed for half-saturation. We have shown that the concentration of substrate at half-saturation is K_m. In the presence of increasing concentrations of a competitive inhibitor, K_m increases. The new value is usually referred to as the apparent K_m (K_m^{app}). On a double-reciprocal plot, adding a competitive inhibitor shows as a decrease in the absolute value of the intercept at the x axis $1/K_m$, whereas the y intercept $1/V_{max}$ remains the same (Figure 5.9b).

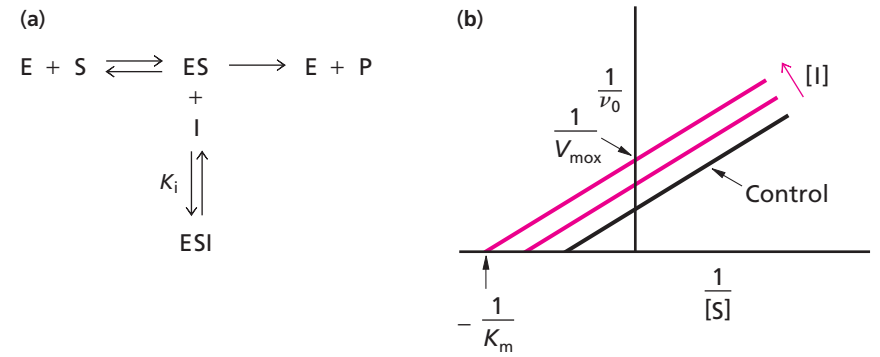
Many classical competitive inhibitors are substrate analogs—compounds that are structurally similar to substrates. The analogs bind to the enzyme but do not react. For example, the enzyme succinate dehydrogenase converts succinate to fumarate (Section 13.3#6). Malonate resembles succinate and acts as a competitive inhibitor of the enzyme.

B. Uncompetitive Inhibition

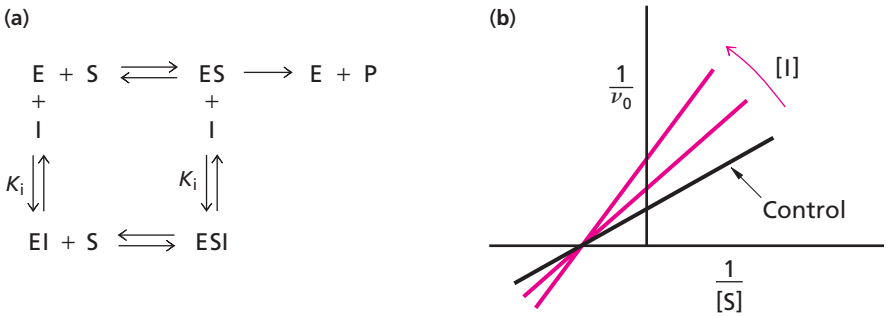
Uncompetitive inhibitors bind only to ES and not to free enzyme (Figure 5.10a). In **uncompetitive inhibition**, V_{max} is decreased ($1/V_{max}$ is increased) by the conversion of some molecules of E to the inactive form ESI. Since it is the ES complex that binds I, the decrease in V_{max} is not reversed by the addition of more substrate. Uncompetitive inhibitors also decrease the K_m (seen as an increase in the absolute value of $1/K_m$ on a double-reciprocal plot) because the equilibria for the formation of both ES and ESI are shifted toward the complexes by the binding of I. Experimentally, the lines on a double-reciprocal plot representing varying concentrations of an uncompetitive inhibitor all have the same slope indicating proportionally decreased values for K_m and V_{max} (Figure 5.10b). This type of inhibition usually occurs only with multisubstrate reactions.

C. Noncompetitive Inhibition

Noncompetitive inhibitors can bind to E or ES forming inactive EI or ESI complexes, respectively (Figure 5.11a). These inhibitors are not substrate analogs and do not bind at the same site as S. The classic case of noncompetitive inhibition is characterized by an



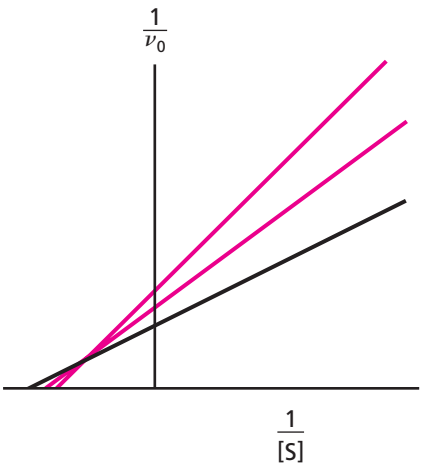
◀ **Figure 5.10**
Uncompetitive inhibition. (a) Kinetic scheme illustrating the binding of I to ES. **(b)** Double-reciprocal plot. In uncompetitive inhibition, both V_{max} and K_m decrease (i.e., the absolute values of both $1/V_{max}$ and $1/K_m$ obtained from the y and x intercepts, respectively, increase). The ratio K_m/V_{max}, the slope of the lines, remains unchanged.



◀ **Figure 5.11**
Classic noncompetitive inhibition. (a) Kinetic scheme illustrating the binding of I to E or ES. **(b)** Double-reciprocal plot. V_{max} decreases, but K_m remains the same.

apparent decrease in V_{max} ($1/V_{max}$ appears to increase) with no change in K_m. On a double-reciprocal plot, the lines for classic noncompetitive inhibition intersect at the point on the x axis corresponding to $1/K_m$ (Figure 5.11b). The common x-axis intercept indicates that K_m isn't affected. The effect of noncompetitive inhibition is to reversibly titrate E and ES with I removing active enzyme molecules from solution. This inhibition cannot be overcome by the addition of S. Classic noncompetitive inhibition is rare but examples are known among allosteric enzymes. In these cases, the noncompetitive inhibitor probably alters the conformation of the enzyme to a shape that can still bind S but cannot catalyze any reaction.

Most enzymes do not conform to the classic form of noncompetitive inhibition where K_m is unchanged. In most cases, both K_m and V_{max} are affected because the affinity of the inhibitor for E is different than its affinity for ES. These cases are often referred to as **mixed inhibition** (Figure 5.12).



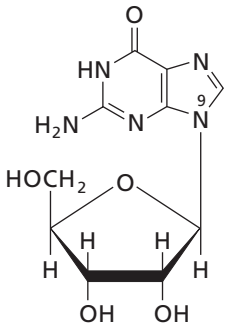
▲ **Figure 5.12**
Double-reciprocal plot showing mixed Inhibition. Both V_{max} and K_m are affected when the inhibitor binds with different affinities to E and ES.

D. Uses of Enzyme Inhibition

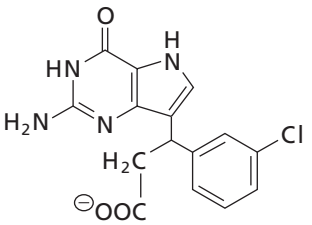
Reversible enzyme inhibition provides a powerful tool for probing enzyme activity. Information about the shape and chemical reactivity of the active site of an enzyme can be obtained from experiments involving a series of competitive inhibitors with systematically altered structures.

The pharmaceutical industry uses enzyme inhibition studies to design clinically useful drugs. In many cases, a naturally occurring enzyme inhibitor is used as the starting point for drug design. Instead of using random synthesis and testing of potential inhibitors, some investigators are turning to a more efficient approach known as *rational drug design*. Theoretically, with the greatly expanded bank of knowledge about enzyme structure, inhibitors can now be rationally designed to fit the active site of a target enzyme. The effects of a synthetic compound are tested first on isolated enzymes and then in biological systems. However, even if a compound has suitable inhibitory activity, other problems may be encountered. For example, the drug may not enter the target cells, may be rapidly metabolized to an inactive compound, may be toxic to the host organism, or the target cell may develop resistance to the drug.

(a)



(b)



▲ **Figure 5.13**
Comparison of a substrate and a designed inhibitor of purine nucleoside phosphorylase. The two substrates of this enzyme are guanosine and inorganic phosphate. **(a)** Guanosine. **(b)** A potent inhibitor of the enzyme. N-9 of guanosine has been replaced by a carbon atom. The chlorinated benzene ring binds to the sugar-binding site of the enzyme, and the acetate side chain binds to the phosphate-binding site.

The advances made in drug synthesis are exemplified by the design of a series of inhibitors of the enzyme purine nucleoside phosphorylase. This enzyme catalyzes a degradative reaction between phosphate and the nucleoside guanosine whose structure is shown in Figure 5.13a. With computer modeling, the structures of potential inhibitors were designed and fit into the active site of the enzyme. One such compound (Figure 5.13b) was synthesized and found to be 100 times more inhibitory than any compound made by the traditional trial-and-error approach. Researchers hope that the rational design approach will produce a drug suitable for treating autoimmune disorders such as rheumatoid arthritis and multiple sclerosis.

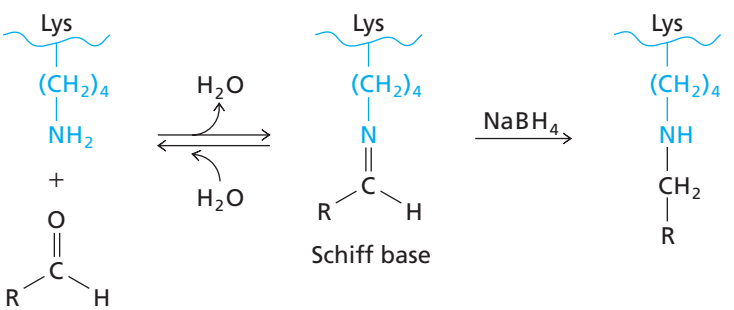
5.8 Irreversible Enzyme Inhibition

In contrast to a reversible enzyme inhibitor, an **irreversible enzyme inhibitor** forms a stable covalent bond with an enzyme molecule thus removing active molecules from the enzyme population. Irreversible inhibition typically occurs by alkylation or acylation of the side chain of an active-site amino acid residue. There are many naturally occurring irreversible inhibitors as well as the synthetic examples described here.

An important use of irreversible inhibitors is the identification of amino acid residues at the active site by specific substitution of their reactive side chains. In this process, an irreversible inhibitor that reacts with only one type of amino acid is incubated with a solution of enzyme that is then tested for loss of activity. Ionizable side chains are modified by acylation or alkylation reactions. For example, free amino groups such as the ϵ-amino group of lysine react with an aldehyde to form a Schiff base that can be stabilized by reduction with sodium borohydride ($NaBH_4$) (Figure 5.14).

The nerve gas diisopropyl fluorophosphate (DFP) is one of a group of organic phosphorus compounds that inactivate hydrolases with a reactive serine as part of the active site. These enzymes are called serine proteases or serine esterases, depending on their reaction specificity. The serine protease chymotrypsin, an important digestive enzyme, is inhibited irreversibly by DFP (Figure 5.15). DFP reacts with the serine residue at chymotrypsin's active site (Ser-195) to produce diisopropylphosphoryl-chymotrypsin.

Some organophosphorus inhibitors are used in agriculture as insecticides; others, such as DFP, are useful reagents for enzyme research. The original organophosphorus nerve gases are extremely toxic poisons developed for military use. The major biological action of these poisons is irreversible inhibition of the serine esterase acetylcholinesterase that catalyzes hydrolysis of the neurotransmitter acetylcholine. When acetylcholine released from an activated nerve cell binds to its receptor on a second nerve cell, it triggers a nerve impulse. The action of acetylcholinesterase restores the cell to its resting state. Inhibition of this enzyme can cause paralysis.



▲ **Figure 5.14**
Reaction of the ϵ-amino group of a lysine residue with an aldehyde. Reduction of the Schiff base with sodium borohydride ($NaBH_4$) forms a stable substituted enzyme.

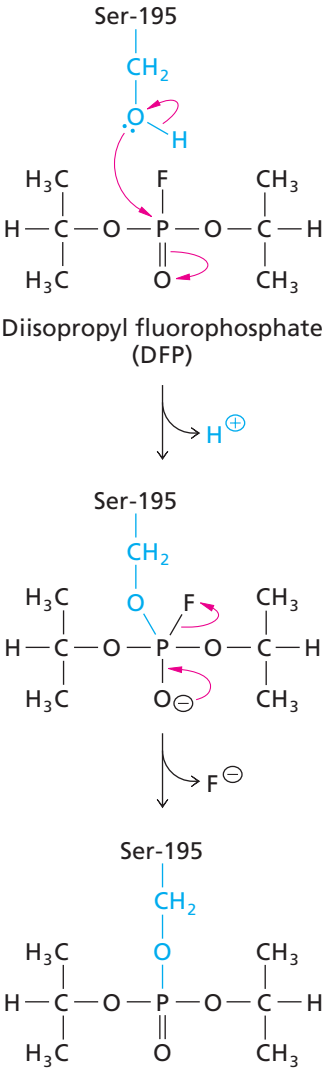


Figure 5.15 ►
Irreversible Inhibition by DFP. Diisopropyl fluorophosphate (DFP) reacts with a single, highly nucleophilic serine residue (Ser-195) at the active site of chymotrypsin, producing inactive diisopropylphosphoryl-chymotrypsin. DFP inactivates serine proteases and serine esterases.

5.9 Regulation of Enzyme Activity

At the beginning of this chapter, we listed several advantages to using enzymes as catalysts in biochemical reactions. Clearly, the most important advantage is to speed up reactions that would otherwise take place too slowly to sustain life. One of the other advantages of enzymes is that their catalytic activity can be regulated in various ways. The *amount* of an enzyme can be controlled by regulating the rate of its synthesis or degradation. This mode of control occurs in all species but it often takes many minutes or hours to synthesize new enzymes or to degrade existing enzymes.

In all organisms, rapid control—on the scale of seconds or less—can be accomplished through reversible modulation of the activity of **regulated enzymes**. In this context, we define regulated enzymes as those enzymes whose activity can be modified in a manner that affects the rate of an enzyme-catalyzed reaction. In many cases, these regulated enzymes control a key step in a metabolic pathway. The activity of a regulated enzyme changes in response to environmental signals, allowing the cell to respond to changing conditions by adjusting the rates of its metabolic processes.

In general, regulated enzymes become more active catalysts when the concentrations of their substrates increase or when the concentrations of the products of their metabolic pathways decrease. They become less active when the concentrations of their substrates decrease or when the products of their metabolic pathways accumulate. Inhibition of the first enzyme unique to a pathway conserves both material and energy by preventing the accumulation of intermediates and the ultimate end product. The activity of regulated enzymes can be controlled by noncovalent allosteric modulation or covalent modification.

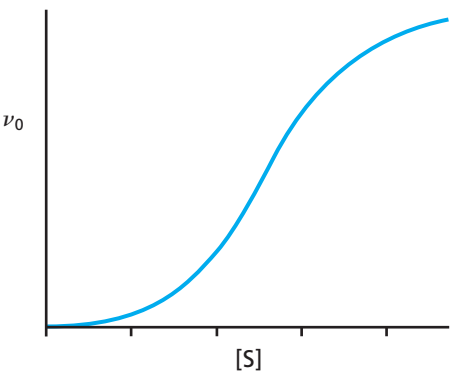
Allosteric enzymes are enzymes whose properties are affected by changes in structure. The structural changes are mediated by interaction with small molecules. We saw an example of allostery in the previous chapter when we examined the binding of oxygen to hemoglobin. Allosteric enzymes often do not exhibit typical Michaelis–Menten kinetics due to cooperative binding of substrate, as is the case with hemoglobin.

Figure 5.16 shows a v_0 versus [S] curve for an allosteric enzyme with cooperative binding of substrate. Sigmoidal curves result from the transition between two states of the enzyme. In the absence of substrate, the enzyme is in the T state. The conformation of each subunit is in a shape that binds substrate inefficiently and the rate of the reaction is slow. As substrate concentration is increased, enzyme molecules begin to bind substrate even though the affinity of the enzyme in the T state is low. When a subunit binds substrate, the enzyme undergoes a conformational change that converts the enzyme to the R state and the reaction takes place. The kinetic properties of the enzyme subunit in the T state and the R state are quite different—each conformation by itself could exhibit standard Michaelis–Menten kinetics.

The conformational change in the subunit that initially binds a substrate molecule affects the other subunits in the multisubunit enzyme. The conformations of these other subunits are shifted toward the R state where their affinity for substrate is much higher. They can now bind substrate at a much lower concentration than when they were in the T state.

Allosteric phenomena are responsible for the reversible control of many regulated enzymes. In Section 4.13C, we saw how the conformation of hemoglobin and its affinity for oxygen change when 2,3-*bis*phosphoglycerate is bound. Many regulated enzymes also undergo allosteric transitions between active (R) states and inactive (T) states. These enzymes have a second ligand-binding site away from their catalytic centers called the **regulatory site** or **allosteric site**. An allosteric inhibitor or activator, also called an allosteric modulator or **allosteric effector**, binds to the regulatory site and causes a conformational change in the regulated enzyme. This conformational change is transmitted

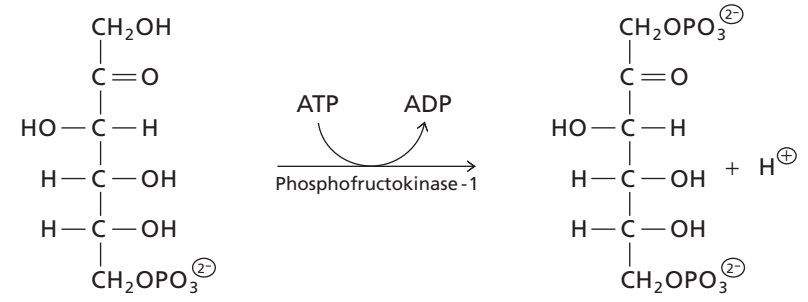
Aspartate transcarbamoylase (ATCase), another well-characterized allosteric enzyme, is described in Chapter 18.



▲ **Figure 5.16**
Cooperativity. Plot of initial velocity as a function of substrate concentration for an allosteric enzyme exhibiting cooperative binding of substrate.

KEY CONCEPT

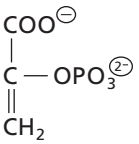
Allosteric enzymes often have multiple subunits and substrate binding is cooperative. This produces a sigmoidal curve when velocity is plotted against substrate concentration.



▲ **Figure 5.17**
Reaction catalyzed by phosphofructokinase-1.

to the active site of the enzyme, which changes shape sufficiently to alter its activity. The regulatory and catalytic sites are physically distinct regions of the protein—usually located on separate domains and sometimes on separate subunits. Allosterically regulated enzymes are often larger than other enzymes.

First, we examine an enzyme that undergoes allosteric (noncovalent) regulation and then we list some general properties of such enzymes. Next, we describe two models that explain allosteric regulation in terms of changes in the conformation of regulated enzymes. Finally, we discuss a closely related group of regulatory enzymes—those subject to covalent modification.



▲ **Figure 5.18**
Phosphoenolpyruvate. This intermediate of glycolysis is an allosteric inhibitor of phosphofructokinase-1 from *Escherichia coli.*

A. Phosphofructokinase Is an Allosteric Enzyme

Bacterial phosphofructokinase-1 (*Escherichia coli*) provides a good example of allosteric inhibition and activation. Phosphofructokinase-1 catalyzes the ATP-dependent phosphorylation of fructose 6-phosphate to produce fructose 1,6-*bis*phosphate and ADP (Figure 5.17). This reaction is one of the first steps of glycolysis, an ATP-generating pathway for glucose degradation described in detail in Chapter 11. Phosphoenolpyruvate (Figure 5.18), an intermediate near the end of the glycolytic pathway, is an allosteric *inhibitor* of *E. coli* phosphofructokinase-1. When the concentration of phosphoenolpyruvate rises, it indicates that the pathway is blocked beyond that point. Further production of phosphoenolpyruvate is prevented by inhibiting phosphofructokinase-1 (see feedback inhibition, Section 10.2C).

ADP is an allosteric activator of phosphofructokinase-1. This may seem strange from looking at Figure 5.17 but keep in mind that the overall pathway of glycolysis results in net synthesis of ATP from ADP. Rising ADP levels indicate a deficiency of ATP and glycolysis needs to be stimulated. Thus, ADP activates phosphofructokinase-1 in spite of the fact that ADP is a product in this particular reaction.

Phosphoenolpyruvate and ADP affect the binding of the substrate fructose 6-phosphate to phosphofructokinase-1. Kinetic experiments have shown that there are four binding sites on phosphofructokinase-1 for fructose 6-phosphate and structural experiments have confirmed that *E. coli* phosphofructokinase-1 (M_r 140,000) is a tetramer consisting of four identical subunits. Figure 5.19 shows the structure of the enzyme complexed with its products, fructose 1,6-*bis*phosphate and ADP, and a second molecule of ADP, an allosteric activator. Two of the subunits shown in Figure 5.19a associate to form a dimer. The two products are bound in the active site located between two domains of each chain—ADP is bound to the large domain and fructose 1,6-*bis*phosphate is bound mostly to the small domain. Two of these dimers interact to form the complete tetrameric enzyme.

A notable feature of the structure of phosphofructokinase-1 (and a general feature of regulated enzymes) is the physical separation of the active site and the regulatory

site on each subunit. (In some regulated enzymes the active sites and regulatory sites are on different subunits.) The activator ADP binds at a distance from the active site in a deep hole between the subunits. When ADP is bound to the regulatory site, phosphofructokinase-1 assumes the R conformation, which has a high affinity for fructose 6-phosphate. When the smaller compound phosphoenolpyruvate is bound to the same regulatory site the enzyme assumes a different conformation, the T conformation, which has a lower affinity for fructose 6-phosphate. The transition between conformations is accomplished by a slight rotation of one rigid dimer relative to the other. The cooperativity of substrate binding is tied to the concerted movement of an arginine residue in each of the four fructose 6-phosphate binding sites located near the interface between the dimers. Movement of the side chain of this arginine from the active site lowers the affinity for fructose 6-phosphate. In many organisms, phosphofructokinase-1 is larger and is subject to more complex allosteric regulation than in *E. coli* as you will see in Chapter 11.

Activators can affect either V_{max} or K_m or both. It's important to recognize that the binding of an activator alters the structure of an enzyme and this alteration converts it to a different form that may have quite different kinetic properties. In most cases, the differences between the kinetic properties of the R and T forms are more complex than the differences we saw with enzyme inhibitors in Section 5.7.

B. General Properties of Allosteric Enzymes

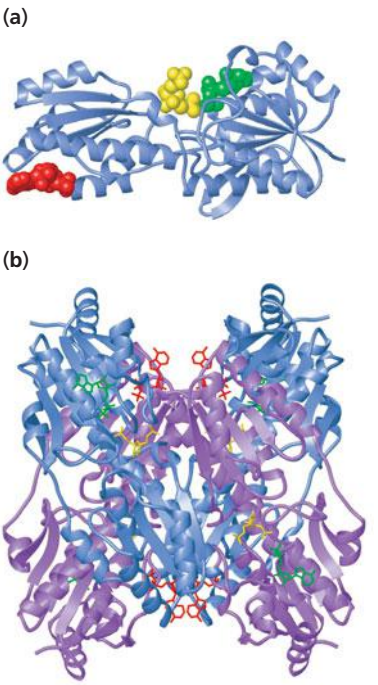
Examination of the kinetic and physical properties of allosteric enzymes has shown that they have the following general features:

1. The activities of allosteric enzymes are changed by metabolic inhibitors and activators. Often these allosteric effectors do not resemble the substrates or products of the enzyme. For example, phosphoenolpyruvate (Figure 5.18) resembles neither the substrate nor the product (Figure 5.17) of phosphofructokinase. Consideration of the structural differences between substrates and metabolic inhibitors originally led to the conclusion that allosteric effectors are bound to regulatory sites separate from catalytic sites.
2. Allosteric effectors bind noncovalently to the enzymes they regulate. (There is a special group of regulated enzymes whose activities are controlled by covalent modification, described in Section 5.10D.) Many effectors alter the K_m of the enzyme for a substrate; but some alter the V_{max}. Allosteric effectors themselves are not altered chemically by the enzyme.
3. With few exceptions, regulated enzymes are multisubunit proteins. (But not all multisubunit enzymes are regulated.) The individual polypeptide chains of a regulated enzyme may be identical or different. For those with identical subunits (such as phosphofructokinase-1 from *E. coli*), each polypeptide chain can contain both the catalytic and regulatory sites and the oligomer is a symmetric complex, most often possessing two or four protein chains. Regulated enzymes composed of nonidentical subunits have more complex, but usually symmetric, arrangements.
4. An allosterically regulated enzyme usually has at least one substrate for which the v_0 versus [S] curve is sigmoidal rather than hyperbolic (Section 5.9). Phosphofructokinase-1 exhibits Michaelis–Menten (hyperbolic) kinetics with respect to one substrate, ATP, but sigmoidal kinetics with respect to its other substrate, fructose 6-phosphate. A sigmoidal curve is caused by positive cooperativity of substrate binding and this is made possible by the presence of multiple substrate binding sites in the enzyme—four binding sites in the case of tetrameric phosphofructokinase-1.

The allosteric R $\rightleftharpoons$ T transition between the active and the inactive conformations of a regulatory enzyme is rapid. The ratio of R to T is controlled by the concentrations of the various ligands and the relative affinities of each conformation for these ligands. In the simplest cases, substrate and activator molecules bind only to enzyme in the R state (E_R) and inhibitor molecules bind only to enzyme in the T state (E_T).

KEY CONCEPT

Allosteric effectors shift the concentrations of the R and T forms of an allosteric enzyme.



▲ **Figure 5.19**
The R conformation of phosphofructokinase-1 from *E. coli*. The enzyme is a tetramer of identical chains. **(a)** Single subunit, shown as a ribbon. The products, fructose 1,6-*bis*phosphate (yellow) and ADP (green), are bound in the active site. The allosteric activator ADP (red) is bound in the regulatory site. **(b)** Tetramer. Two are blue, and two are purple. The products, fructose 1,6-*bis*phosphate (yellow) and ADP (green), are bound in the four active sites. The allosteric activator ADP (red) is bound in the four regulatory sites, at the interface of the subunits. [PDB 1PFK].

The relationship between the regulation of an individual enzyme and a pathway is discussed in Section 10.2B, where we encounter terms such as feedback inhibition and feedforward activation.

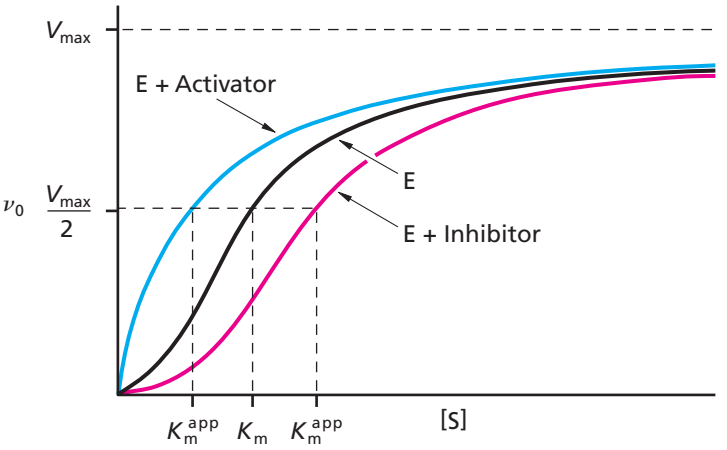
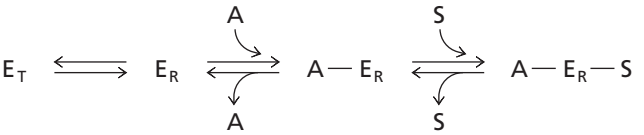


Figure 5.20 ►
Role of cooperativity of binding in regulation. The activity of an allosteric enzyme with a sigmoidal binding curve can be altered markedly when either an activator or an inhibitor is bound to the enzyme. Addition of an activator can lower the apparent K_m raising the activity at a given [S]. Conversely, addition of an inhibitor can raise the apparent K_m producing less activity at a given [S].

(5.29)



(5.30)

These simplified examples illustrate the main property of allosteric effectors—they shift the steady-state concentrations of free E_T and E_R.

Figure 5.20 illustrates the regulatory role that cooperative binding can play. Addition of an activator can shift the sigmoidal curve toward a hyperbolic shape, lowering the apparent K_m (the concentration of substrate required for half-saturation) and raising the activity at a given [S]. The addition of an inhibitor can raise the apparent K_m of the enzyme and lower its activity at any particular concentration of substrate.

The addition of S leads to an increase in the concentration of enzyme in the R conformation. Conversely, the addition of inhibitor increases the proportion of the T species. Activator molecules bind preferentially to the R conformation leading to an increase in the R/T ratio. Note that this simplified scheme does not show that there are multiple interacting binding sites for both S and I.

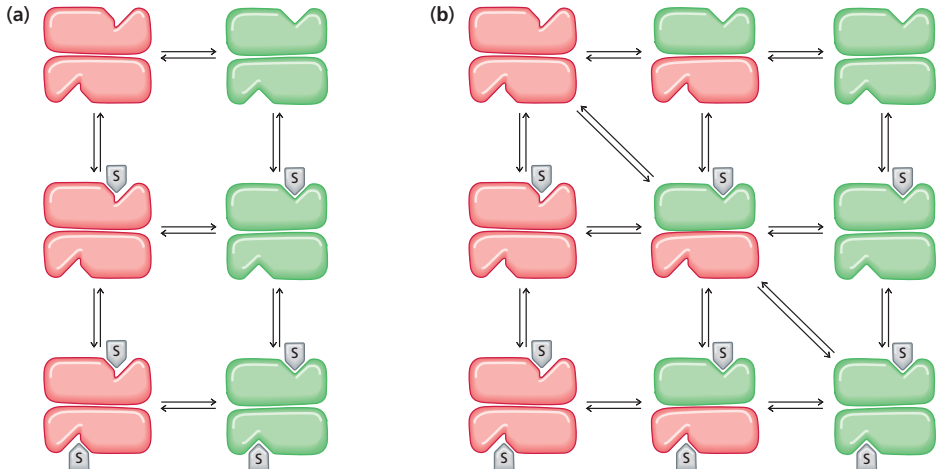
Some allosteric inhibitors are nonclassical competitive inhibitors (Figure 5.8). For example, Figure 5.20 describes an enzyme that has a higher apparent K_m for its substrate in the presence of the allosteric inhibitor but an unaltered V_{max}. Therefore, the allosteric modulator is a competitive inhibitor.

Some regulatory enzymes exhibit noncompetitive inhibition patterns where binding of a modulator at the regulatory site does not *prevent* substrate from binding but appears to distort the conformation of the active site sufficiently to *decrease* the activity of the enzyme.

C. Two Theories of Allosteric Regulation

Recall that most proteins are made up of two or more polypeptide chains (Section 4.8). Enzymes are typical proteins—most of them have multiple subunits. This complicates our understanding of regulation. There are two general models that explain the cooperative binding of ligands to multimeric proteins. Both models describe the cooperative transitions in simple quantitative terms.

The **concerted model**, or symmetry model, was devised to explain the cooperative binding of identical ligands, such as substrates. It was first proposed in 1965 by



▲ **Figure 5.21**
Two models for cooperativity of binding of substrate (S) to a tetrameric protein. A two-subunit protein is shown for simplicity. In all cases, the enzymatically active subunit (R) is colored green and the inactive conformation (T) is colored red. **(a)** In the simplified concerted model, both subunits are either in the R conformation or the T conformation. Substrate (S) can bind to subunits in either conformation but binding to T is assumed to be weaker than binding to R. Cooperativity is explained by postulating that when substrate binds to a subunit in the T conformation (red), it shifts the protein into a conformation where both subunits are in the R conformation. **(b)** In the sequential model, one subunit may be in the R conformation while another is in the T conformation. As in the concerted model, both conformations can bind substrate. Cooperativity is achieved by postulating that substrate binding causes the subunit to shift to the R conformation and that when one subunit has adopted the R conformation, the other one is more likely to bind substrate and undergo a conformation change (diagonal lines).

Jacques Monod, Jeffries Wyman, and Jean-Pierre Changeux and it's sometimes known as the MWC model. The concerted model assumes there is one substrate binding site on each subunit. According to the concerted model, the conformation of each subunit is constrained by its association with other subunits and when the protein changes conformation it retains its molecular symmetry (Figure 5.21a). Thus, there are two conformations in equilibrium, R and T. When a subunit is in the R conformation it has a high affinity for the substrate. Subunits in the T conformation have a low affinity for the substrate. The binding of substrate to one subunit shifts the equilibrium since it "locks" the other subunits in the R conformation making it more likely that the other subunits will bind substrate. This explains the cooperativity of substrate binding.

When the conformation of the protein changes, the affinity of its substrate binding sites also changes. The concerted model was extended to include the binding of allosteric effectors and it can be simplified by assuming that the substrate binds only to the R conformation and the allosteric effectors bind preferentially to one of the conformations—inhibitors bind only to subunits in the T conformation and activators bind only to subunits in the R conformation. The concerted model is based on the observed structural symmetry of regulatory enzymes. It suggests that all subunits of a given protein molecule have the same conformation, either all R or all T.

When the enzyme shifts from one conformation to the other, all subunits change conformation in a concerted manner. Experimental data obtained with a number of enzymes can be explained by this simple theory. For example, many of the properties of phosphofructokinase-1 from *E. coli* fit the concerted theory. In most cases, however, the concerted theory does not adequately account for all of the observations concerning a particular enzyme. Their behavior is more complex than that suggested by this simple all-or-nothing model.

The **sequential model** was first proposed by Daniel Koshland, George Némethy, and David Filmer (KNF model). It is a more general model because it allows for both subunits to exist in two different conformations within the same multimeric protein. The specific induced-fit version or the model is based on the idea that a ligand may induce a change in the tertiary structure of each subunit to which it binds. This subunit–ligand

complex may change the conformations of neighboring subunits to varying extents. Like the concerted model, the sequential model assumes that only one shape has a high affinity for the ligand but it differs from the concerted model in allowing for the existence of both high- and low-affinity subunits in a multisubunit protein (Figure 5.21b).

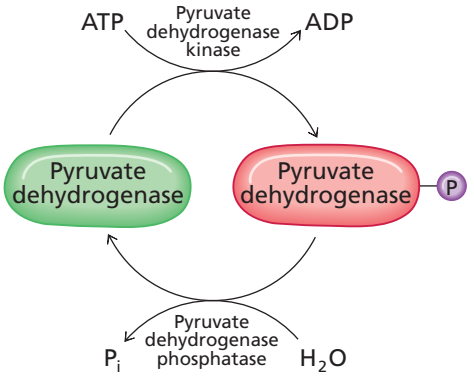
Hundreds of allosteric proteins have been studied and the majority show cooperative binding of substrates and/or effector molecules. It has proven to be very difficult to distinguish between the concerted and sequential models. Many proteins exhibit binding behavior that can best be explained as a mixture of the all-or-nothing shift of the concerted model and the stepwise shift of the sequential model.

D. Regulation by Covalent Modification

The activity of an enzyme can be modified by the covalent attachment and removal of groups on the polypeptide chain. Regulation by covalent modification is usually slower than the allosteric regulation described above. It's important to note that the covalent modification of regulated enzymes must be reversible, otherwise it wouldn't be a form of regulation. The modifications usually require additional modifying enzymes for activation and inactivation. The activities of these modifying enzymes may themselves be allosterically regulated or regulated by covalent modification. Enzymes controlled by covalent modification are believed to generally undergo R $\rightleftharpoons$ T transitions but they may be frozen in one conformation or the other by a covalent substitution.

The most common type of covalent modification is phosphorylation of one or more specific serine residues, although in some cases threonine, tyrosine, or histidine residues are phosphorylated. An enzyme called a protein kinase catalyzes the transfer of the terminal phosphoryl group from ATP to the appropriate serine residue of the regulated enzyme. The phosphoserine of the regulated enzyme is hydrolyzed by the activity of a protein phosphatase, releasing phosphate and returning the enzyme to its dephosphorylated state. Individual enzymes differ as to whether it is their phosphorylated or dephosphorylated forms that are active.

The reactions involved in the regulation of mammalian pyruvate dehydrogenase by covalent modification are shown in Figure 5.22. Pyruvate dehydrogenase catalyzes a reaction that connects the pathway of glycolysis to the citric acid cycle. Phosphorylation of pyruvate dehydrogenase, catalyzed by the allosteric enzyme pyruvate dehydrogenase kinase, inactivates the dehydrogenase. The kinase can be activated by any of several metabolites. Phosphorylated pyruvate dehydrogenase is reactivated under different metabolic conditions by hydrolysis of its phosphoserine residue, catalyzed by pyruvate dehydrogenase phosphatase.



▲ **Figure 5.22**
Regulation of mammalian pyruvate dehydrogenase. Pyruvate dehydrogenase, an interconvertible enzyme, is inactivated by phosphorylation catalyzed by pyruvate dehydrogenase kinase. It is reactivated by hydrolysis of its phosphoserine residue, catalyzed by an allosteric hydrolase called pyruvate dehydrogenase phosphatase.

5.10 Multienzyme Complexes and Multifunctional Enzymes

In some cases, different enzymes that catalyze sequential reactions in the same pathway are bound together in a multienzyme complex. In other cases, different activities may be found on a single multifunctional polypeptide chain. The presence of multiple activities on a single polypeptide chain is usually the result of a gene fusion event.

Some multienzyme complexes are quite stable. We will encounter several of these complexes in other chapters. In other multienzyme complexes the proteins may be associated more weakly (Section 4.9). Because these complexes dissociate easily it has been difficult to demonstrate their existence and importance. Attachment to membranes or cytoskeletal components is another way that enzymes may be associated.

The metabolic advantages of multienzyme complexes and multifunctional enzymes include the possibility of **metabolite channeling**. Channeling of reactants between active sites can occur when the product of one reaction is transferred directly to the next active site without entering the bulk solvent. This can vastly increase the rate of a reaction by decreasing transit times for intermediates between enzymes and by producing local high concentrations of intermediates. Channeling can also protect chemically labile intermediates from degradation by the solvent. Metabolic channeling is one way in which enzymes can effectively couple separate reactions.

One of the best-characterized examples of channeling involves the enzyme tryptophan synthase that catalyzes the last two steps in the biosynthesis of tryptophan (Section 17.3F). Tryptophan synthase has a tunnel that conducts a reactant between its two active sites. The structure of the enzyme not only prevents the loss of the reactant to the bulk solvent but also provides allosteric control to keep the reactions occurring at the two active sites in phase.

Several other enzymes have two or three active sites connected by a molecular tunnel. Another mechanism for metabolite channeling involves guiding the reactant along a path of basic amino acid side chains on the surface of coupled enzymes. The metabolites (most of which are negatively charged) are directed between active sites by the electrostatically positive surface path. The fatty acid synthase complex catalyzes a sequence of seven reactions required for the synthesis of fatty acids. The structure of this complex is described in Chapter 16 (Section 16.1).

The search for enzyme complexes and the evaluation of their catalytic and regulatory roles is an extremely active area of research.

The regulation of pyruvate dehydrogenase activity is explained in Section 13.5. An example of a signal transduction pathway involving covalent modification is described in Section 12.6.

Summary

1. Enzymes, the catalysts of living organisms, are remarkable for their catalytic efficiency and their substrate and reaction specificity. With few exceptions, enzymes are proteins or proteins plus cofactors. Enzymes are grouped into six classes (oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases) according to the nature of the reactions they catalyze.
2. The kinetics of a chemical reaction can be described by a rate equation.
3. Enzymes and substrates form noncovalent enzyme–substrate complexes. Consequently, enzymatic reactions are characteristically first order with respect to enzyme concentration and typically show hyperbolic dependence on substrate concentration. The hyperbola is described by the Michaelis–Menten equation.
4. Maximum velocity (V_{max}) is reached when the substrate concentration is saturating. The Michaelis constant (K_m) is equal to the substrate concentration at half-maximal reaction velocity—that is, at half-saturation of E with S.
5. The catalytic constant (k_{cat}), or turnover number, for an enzyme is the maximum number of molecules of substrate that can be transformed into product per molecule of enzyme (or per active site) per second. The ratio k_{cat}/K_m is an apparent second-order rate constant that governs the reaction of an enzyme when the substrate is dilute and nonsaturating. k_{cat}/K_m provides a measure of the catalytic efficiency of an enzyme.
6. K_m and V_{max} can be obtained from plots of initial velocity at a series of substrate concentrations and at a fixed enzyme concentration.
7. Multisubstrate reactions may follow a sequential mechanism with binding and release events being ordered or random, or a ping-pong mechanism.
8. Inhibitors decrease the rates of enzyme-catalyzed reactions. Reversible inhibitors may be competitive (increasing the apparent value of K_m without changing V_{max}), uncompetitive (appearing to decrease K_m and V_{max} proportionally), noncompetitive (appearing to decrease V_{max} without changing K_m), or mixed. Irreversible enzyme inhibitors form covalent bonds with the enzyme.
9. Allosteric modulators bind to enzymes at a site other than the active site and alter enzyme activity. Two models, the concerted model and the sequential model, describe the cooperativity of allosteric enzymes. Covalent modification, usually phosphorylation, of certain regulatory enzymes can also regulate enzyme activity.
10. Multienzyme complexes and multifunctional enzymes are very common. They can channel metabolites between active sites.

Problems

1. Initial velocities have been measured for the reaction of α-chymotrypsin with tyrosine benzyl ester [S] at six different substrate concentrations. Use the data below to make a reasonable estimate of the V_{max} and K_m value for this substrate.

mM [S]	0.00125	0.01	0.04	0.10	2.0	10
(mM/min)	14	35	56	66	69	70

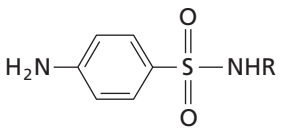
2. Why is the k_{cat}/K_m value used to measure the catalytic proficiency of an enzyme?
   (a) What are the upper limits for k_{cat}/K_m values for enzymes?
   (b) Enzymes with k_{cat}/K_m values approaching these upper limits are said to have reached "catalytic perfection." Explain.
3. Carbonic anhydrase (CA) has a 25,000-fold higher activity ($k_{cat} = 10^6\ s^{-1}$) than orotidine monophosphate decarboxylase (OMPD) ($k_{cat} = 40\ s^{-1}$). However, OMPD provides more than a 10^{10} higher "rate acceleration" than CA (Table 5.2). Explain how this is possible.
4. An enzyme that follows Michaelis–Menten kinetics has a K_m of 1 μM. The initial velocity is 0.1 μM min^{-1} at a substrate concentration of 100 μM. What is the initial velocity when [S] is equal to (a) 1 mM, (b) 1 μM, or (c) 2 μM?
5. Human immunodeficiency virus 1 (HIV-1) encodes a protease (M_r 21,500) that is essential for the assembly and maturation of the virus. The protease catalyzes the hydrolysis of a heptapeptide substrate with a k_{cat} of 1000 s^{-1} and a K_m of 0.075 M.

(a) Calculate V_{max} for substrate hydrolysis when HIV-1 protease is present at 0.2 mg ml^{-1}.

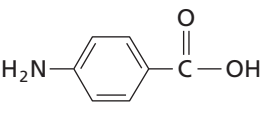
(b) When —C(O)NH— of the heptapeptide is replaced by $—CH_2NH—$, the resulting derivative cannot be cleaved by HIV-1 protease and acts as an inhibitor. Under the same experimental conditions as in part (a), but in the presence of 2.5 μM inhibitor, V_{max} is 9.3×10^{-3} M s^{-1}. What kind of inhibition is occurring? Is this type of inhibition expected for a molecule of this structure?

6. Draw a graph of v_0 versus [S] for a typical enzyme reaction (a) in the absence of an inhibitor, (b) in the presence of a competitive inhibitor, and (c) in the presence of a noncompetitive inhibitor.

7. Sulfonamides (sulfa drugs) such as sulfanilamide are antibacterial drugs that inhibit the enzyme dihydropteroate synthase (DS) that is required for the synthesis of folic acid in bacteria. There is no corresponding enzyme inhibition in animals because folic acid is a required vitamin and cannot be synthesized. If *p* aminobenzoic acid (PABA) is a substrate for DS, what type of inhibition can be predicted for the bacterial synthase enzyme in the presence of sulfonamides? Draw a double reciprocal plot for this type of inhibition with correctly labeled axes and identify the uninhibited and inhibited lines.



Sulfonamides
(R = H, sulfanilamide)



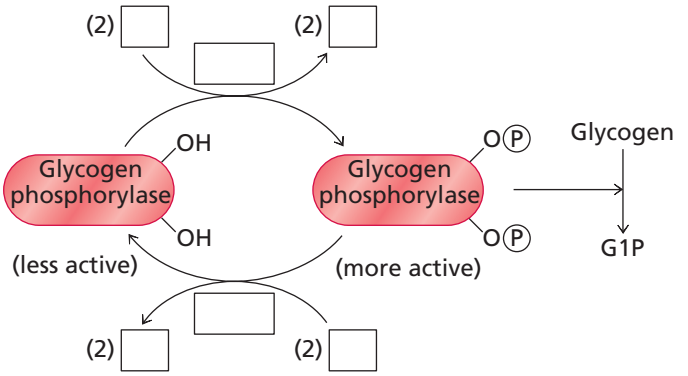
p-Aminobenzoic acid

8. (a) Fumarase is an enzyme in the citric acid cycle that catalyzes the conversion of fumarate to L-malate. Given the fumarate (substrate) concentrations and initial velocities below, construct a Lineweaver–Burk plot and determine the V_{max} and K_m values for the fumarase-catalyzed reaction.

Fumarate (mM)	Rate (mmol l^{-1} min^{-1})
02.0	2.5
03.3	3.1
05.0	3.6
10.0	4.2

(b) Fumarase has a molecular weight of 194,000 and is composed of four identical subunits, each with an active site. If the enzyme concentration is 1×10^{-2} M for the experiment in part (a), calculate the k_{cat} value for the reaction of fumarase with fumarate. *Note*: The units for k_{cat} are reciprocal seconds (s^{-1}).

9. Covalent enzyme regulation plays an important role in the metabolism of muscle glycogen, an energy storage molecule. The active phosphorylated form of glycogen phosphorylase (GP) catalyzes the degradation of glycogen to glucose 1-phosphate. Using pyruvate dehydrogenase as a model (Figure 5.23), fill in the boxes below for the activation and inactivation of muscle glycogen phosphorylase.



10. Regulatory enzymes in metabolic pathways are often found at the first step that is unique to that pathway. How does regulation at this point improve metabolic efficiency?

11. ATCase is a regulatory enzyme at the beginning of the pathway for the biosynthesis of pyrimidine nucleotides. ATCase exhibits positive cooperativity and is activated in vitro by ATP and inhibited by the pyrimidine nucleotide cytidine triphosphate (CTP). Both ATP and CTP affect the K_m for the substrate aspartate but not V_{max}. In the absence of ATP or CTP, the concentration of aspartate required for half-maximal velocity is about 5 mM at saturating concentrations of the second substrate, carbamoyl phosphate. Draw a v_0 versus [aspartate] plot for ATCase, and indicate how CTP and ATP affect v_0 when [aspartate] = 5 mM.

12. The cytochrome P450 family of monooxygenase enzymes are involved in the clearance of foreign compounds (including drugs) from our body. P450s are found in many tissues, including the liver, intestine, nasal tissues, and lung. For every drug that is approved for human use the pharmaceutical company must investigate the metabolism of the drug by cytochrome P450. Many of the adverse drug–drug interactions known to occur are a result of interactions with the cytochrome P450 enzymes. A significant portion of drugs are metabolized by one of the P450 enzymes, P450 3A4. Human intestinal P450 3A4 is known to metabolize midazolam, a sedative, to a hydroxylated product, 1′-hydroxymidazolam. The kinetic data given below are for the reaction catalyzed by P450 3A4.

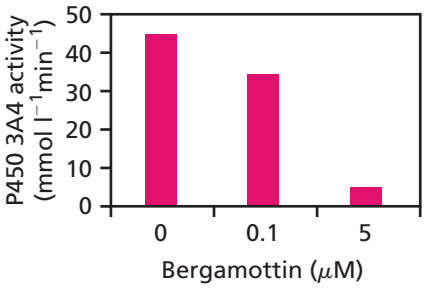
(a) Focusing on the first two columns, determine the K_m and V_{max} for the enzyme using a Lineweaver–Burk plot.

(b) Ketoconazole, an antifungal, is known to cause adverse drug-drug interactions when administered with midazolam. Using the data in the table, determine the type of inhibition that ketoconazole exerts on the P450-catalyzed hydroxylation of midazolam.

Midazolam(μM)	Rate of product formation (pmol l^{-1} min^{-1})	Rate of product formation in the presence of 0.1 μM ketoconazole (pmol l^{-1} min^{-1})
1	100	11
2	156	18
4	222	27
8	323	40

[Adapted from Gibbs, M. A., Thummel, K. E., Shen, D. D., and Kunze, K. L. *Drug Metab. Dispos.* (1999). 27:180–187]

13. Patients who are taking certain medications are warned by their physicians to avoid taking these medications with grapefruit juice, which contains many compounds including bergamottin. Cytochrome P450 3A4 is a monooxygenase that is known to metabolize drugs to their inactive forms. The following results were obtained when P450 3A4 activity was measured in the absence or presence of bergamottin.



(a) What is the effect of adding bergamottin to the P450-catalyzed reaction?

(b) Why could it be dangerous for a patient to take certain medications with grapefruit juice?

[Adapted from Wen, Y. H., Sahi, J., Urda, E., Kalkarni, S., Rose, K., Zheng, X., Sinclair, J. F., Cai, H., Strom, S. C., and Kostrubsky, V. E. *Drug Metab. Dispos.* (2002). 30:977–984.]

14. Use the Michaelis-Menten equation (Equation 5.14) to demonstrate the following:

(a) v_0 becomes independent of [S] when [S]>>K_m.

(b) The reaction is first order with respect to S when [S]<<K_m.

(c) [S]>>K_m when v_0 is one-half V_{max}.

Selected Readings

Enzyme Catalysis

Fersht, A. (1985). *Enzyme Structure and Mechanism*, 2nd ed. (New York: W. H. Freeman).

Lewis, C. A., and Wolfenden, R. (2008). Uroporphyrinogen decarboxylation as a benchmark for the catalytic proficiency of enzymes. *Proc. Natl. Acad. Sci. (USA).* 105:17328–17333.

Miller, B. G., and Wolfenden, R. (2002). Catalytic proficiency: the unusual case of OMP decarboxylase. *Annu. Rev. Biochem.* 71, 847–885.

Sigman, D. S., and Boyer, P. D., eds. (1990–1992). *The Enzymes*, Vols. 19 and 20, 3rd ed. (San Diego: Academic Press).

Webb, E. C., ed. (1992). *Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes* (San Diego; Academic Press).

Enzyme Kinetics and Inhibition

Bugg, C. E., Carson, W. M., and Montgomery, J. A. (1993). Drugs by design. *Sci. Am.* 269(6):92–98.

Chandrasekhar, S. (2002). Thermodynamic analysis of enzyme catalysed reactions: new insights into the Michaelis-Menten equation. *Res. Cehm. Intermed.* 28:265–275.

Cleland, W. W. (1970). *Steady State Kinetics. The Enzymes*, Vol. 2, 3rd ed., P. D. Boyer, ed. (New York: Academic Press), pp. 1–65.

Cornish-Bowden, A. (1999). Enzyme kinetics from a metabolic perspective. *Biochem. Soc. Trans.* 27:281–284.

Northrop, D. B. (1998). On the meaning of K_m and V/K in enzyme Kinetics. *J. Chem. Ed.* 75:1153–1157.

Radzicka, A., and Wolfenden, R. (1995). A proficient enzyme. *Science* 267:90–93.

Segel, I. H. (1975) *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady State Enzyme Systems* (New York: Wiley-Interscience).

Regulated Enzymes

Ackers, G. K., Doyle, M. L., Myers, D., and Daugherty, M. A. (1992). Molecular code for cooperativity in hemoglobin. *Science* 255:54–63.

Barford, D. (1991). Molecular mechanisms for the control of enzymic activity by protein phosphorylation. *Biochim. Biophys. Acta* 1133:55–62.

Hilser, V. J. (2010). An ensemble view of allostery. *Science* 327:653–654.

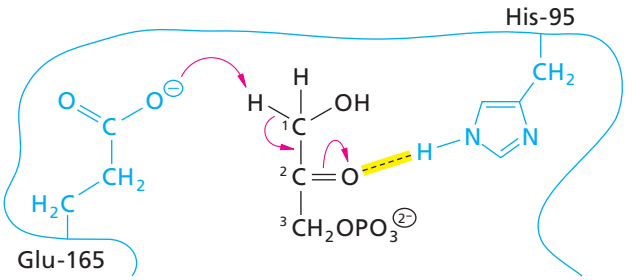
Hurley, J. H., Dean, A. M., Sohl, J. L., Koshland, D. E., Jr., and Stroud, R. M. (1990). Regulation of an enzyme by phosphorylation at the active site. *Science* 249:1012–1016.

Schirmer, T., and Evans, P. R. (1990). Structural basis of the allosteric behavior of phosphofructokinase. *Nature* 343:140–145.

Metabolite Channeling

Pan, P., Woehl, E., and Dunn, M. F. (1997). Protein architecture, dynamics and allostery in tryptophan synthase channeling. *Trends Biochem. Sci.* 22:22–27.

Vélot, C., Mixon, M. B., Teige, M., and Srere, P. A. (1997). Model of a quinary structure between Krebs TCA cycle enzymes: a model for the metabolon. *Biochemistry* 36:14271–14276.



Mechanisms of Enzymes

The previous chapter described some general properties of enzymes with an emphasis on enzyme kinetics. In this chapter, we see how enzymes catalyze reactions by studying the molecular details of catalyzed reactions. Individual enzyme mechanisms have been deduced by a variety of methods including kinetic experiments, protein structural studies, and studies of nonenzymatic model reactions. The results of such studies show that the extraordinary catalytic ability of enzymes results from simple physical and chemical properties, especially the binding and proper positioning of reactants in the active sites of enzymes. Chemistry, physics, and biochemistry have combined to take much of the mystery out of enzymes and recombinant DNA technology now allows us to test the theories proposed by enzyme chemists. Observations for which there were no explanations just a half-century ago are now thoroughly understood.

The mechanisms of many enzymes are well established and they give us a general picture of how enzymes function as catalysts. We begin this chapter with a review of simple chemical mechanisms, followed by a brief discussion of catalysis. We then examine the major modes of enzymatic catalysis: acid–base and covalent catalysis (classified as chemical effects) and substrate binding and transition state stabilization (classified as binding effects). We end the chapter with some specific examples of enzyme mechanisms.

I think that enzymes are molecules that are complementary in structure to the activated complexes of the reactions that they catalyze.

—Linus Pauling (1948)

6.1 The Terminology of Mechanistic Chemistry

The mechanism of a reaction is a detailed description of the molecular, atomic, and even subatomic events that occur during the reaction. Reactants, products, and any intermediates must be identified. A number of laboratory techniques are used to determine the mechanism of a reaction. For example, the use of isotopically labeled reactants can trace the path of individual atoms and kinetic techniques can measure the changes in chemical bonds of a reactant or solvent during the reaction. Study of the stereochemical changes that occur during the reaction can give a three-dimensional view of the process. For any proposed enzyme mechanism, the mechanistic information about the reactants and intermediates must be coordinated with the three-dimensional structure of the enzyme. This is an important part of understanding structure–function relationships—one of the main themes in biochemistry.

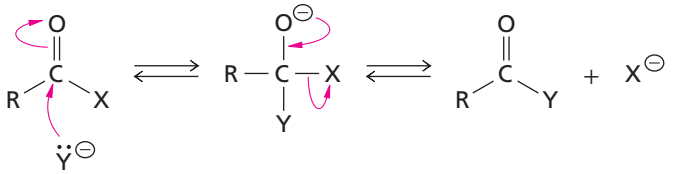
Top: A step from the mechanism of the triose phosphate isomerase reaction.

Enzymatic mechanisms are described using the same symbolism developed in organic chemistry to represent the breaking and forming of chemical bonds. The movement of electrons is the key to understanding chemical (and enzymatic) reactions. We will review chemical mechanisms in this section and in the following sections we will discuss catalysis and present several specific enzyme mechanisms. This discussion should provide sufficient background for you to understand all the enzyme-catalyzed reactions presented in this book.

A. Nucleophilic Substitutions

Many chemical reactions have ionic substrate, intermediates, or products. There are two types of ionic molecules: one species is electron rich, or **nucleophilic**, and the other species is electron poor, or **electrophilic** (Section 2.6). A nucleophile has a negative charge or an unshared electron pair. We usually think of the nucleophile as attacking the electrophile and call the mechanism a nucleophilic attack or a nucleophilic substitution. In mechanistic chemistry, the movement of a *pair* of electrons is represented by a curved arrow pointing from the available electrons of the nucleophile to the electrophilic center. These "electron pushing" diagrams depict the breaking of an existing covalent bond or the formation of a new covalent bond. The reaction mechanism usually involves an intermediate.

Many biochemical reactions are **group transfer reactions** where a group is moved from one molecule to another. Many of these reactions involve a charged intermediate. The transfer of an acyl group, for example, can be written as the general mechanism

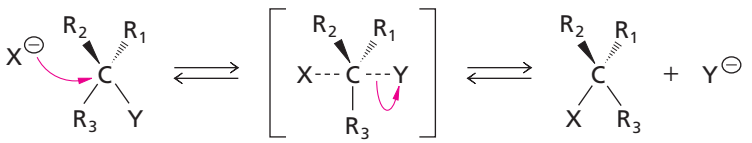


(6.1)

The nucleophile $Y^{\ominus}$ attacks the carbonyl carbon (i.e., adds to the carbonyl carbon atom) to form a tetrahedral addition intermediate from which $X^{\ominus}$ is eliminated. $X^{\ominus}$ is called the leaving group—the group displaced by the attacking nucleophile. This is an example of a **nucleophilic substitution** reaction.

Another type of nucleophilic substitution involves direct displacement. In this mechanism, the attacking group, or molecule, adds to the face of the central atom opposite the leaving group to form a transition state having five groups associated with the central atom. This **transition state** is unstable. It has a structure between that of the reactant and that of the product. (Transition states are shown in square brackets to identify them as unstable, transient entities.)

Transition states are discussed further in Section 6.2.



(6.2)

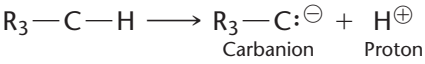
Transition state

Note that both types of nucleophilic substitution mechanisms involve a transitory state. In the first type (Reaction 6.1), the reaction proceeds in a stepwise manner forming an intermediate molecule that may be stable enough to be detected. In the second type of mechanism (Reaction 6.2), the addition of the attacking nucleophile and the displacement of the leaving group occur simultaneously. The transition state is not a stable intermediate.

B. Cleavage Reactions

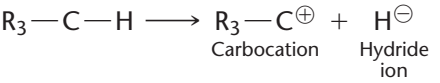
We will also encounter cleavage reactions. Covalent bonds can be cleaved in two ways: either both electrons can stay with one atom or one electron can remain with each atom.

The two electrons will stay with one atom in most reactions so that an ionic intermediate and a leaving group are formed. For example, cleavage of a C—H bond almost always produces two ions. If the carbon atom retains both electrons then the carbon-containing compound becomes a **carbanion** and the other product is a proton.



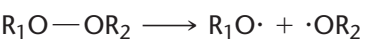
(6.3)

If the carbon atom loses both electrons, the carbon-containing compound becomes a cationic ion called a **carbocation** and the hydride ion carries a pair of electrons.



(6.4)

In the second, less common, type of bond cleavage, one electron remains with each product to form two free radicals that are usually very unstable. (A **free radical**, or radical, is a molecule or atom with an unpaired electron.)



(6.5)

C. Oxidation–Reduction Reactions

Loss of Electrons = Oxidation (LEO)
Gain of Electrons = Reduction (GER)

Remember the phrase: LEO (the lion) says GER

Oxidation Is Loss (OIL)
Reduction Is Gain (RIG)

Remember the phrase: OIL RIG

Oxidation–reduction reactions are central to the supply of biological energy. In an oxidation–reduction (redox) reaction, electrons from one molecule are transferred to another. The terminology here can be a bit confusing so it's important to master the meaning of the words *oxidation* and *reduction*—they will come up repeatedly in the rest of the book. **Oxidation** is the loss of electrons: a substance that is oxidized will have fewer electrons when the reaction is complete. **Reduction** is the gain of electrons: a substance that gains electrons in a reaction is reduced. Oxidation and reduction reactions always occur together. One substrate is oxidized and the other is reduced. An oxidizing agent is a substance that causes an oxidation—it takes electrons from the substrate that is oxidized. Thus, oxidizing agents gain electrons (i.e., they are reduced). A reducing agent is a substance that donates electrons (and is oxidized in the process).

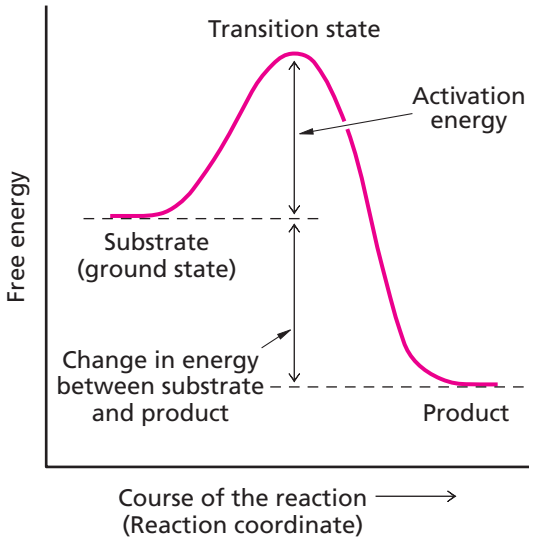
Oxidations can take several forms, such as removal of hydrogen (dehydrogenation), addition of oxygen, or removal of electrons. Dehydrogenation is the most common form of biological oxidation. Recall that oxidoreductases (enzymes that catalyze oxidation–reduction reactions) represent a large class of enzymes and dehydrogenases (enzymes that catalyze removal of hydrogen) are a major subclass of oxidoreductases (Section 5.1).

Most dehydrogenations occur by C—H bond cleavage producing a hydride ion ($H^{\ominus}$). The substrate is oxidized because it loses the electrons associated with the hydride ion. Such reactions will be accompanied by a corresponding reduction where another substrate gains electrons by reacting with the hydride ion. The dehydrogenation of lactate (Equation 5.1) is an example of the removal of hydrogen. In this case, the oxidation of lactate is coupled to the reduction of the coenzyme $NAD^{\oplus}$. The role of cofactors in oxidation–reduction reactions will be discussed in the next chapter (Section 7.3) and the free energy of these reactions is described in Section 10.9.

6.2 Catalysts Stabilize Transition States

In order to understand catalysis it's necessary to appreciate the importance of transition states and intermediates in chemical reactions. The rate of a *chemical* reaction depends on how often reacting molecules collide in such a way that a reaction is favored. The colliding substances must be in the correct orientation and must possess sufficient energy to approach the physical configuration of the atoms and bonds of the final product.

As mentioned above, the transition state is an unstable arrangement of atoms in which chemical bonds are in the process of being formed or broken. Transition states



◀ **Figure 6.1**
Energy diagram for a single-step reaction. The upper arrow shows the activation energy for the forward reaction. Molecules of substrate that have more free energy than the activation energy pass over the activation barrier and become molecules of product. For reactions with a high activation barrier, energy in the form of heat must be provided in order for the reaction to proceed.

have extremely short lifetimes of about 10^{-14} to 10^{-13} second, the time of one bond vibration. Although they are very difficult to detect, their structures can be predicted. The energy required to reach the transition state from the ground state of the reactants is called the *activation energy* of the reaction and is often referred to as the *activation barrier*.

The progress of a reaction can be represented by an energy diagram, or energy profile. Figure 6.1 is an example that shows the conversion of a substrate (reactant) to a product in a single step. The *y* axis shows the free energies of the reacting species. The *x* axis, called the *reaction coordinate*, measures the progress of the reaction, beginning with the substrate on the left and proceeding to the product on the right. This axis is not time but rather the progress of bond breaking and bond formation of a particular molecule. The transition state occurs at the peak of the activation barrier—this is the energy level that must be exceeded for the reaction to proceed. The lower the barrier the more stable the transition state and the more often the reaction proceeds.

Intermediates, unlike transition states, can be sufficiently stable to be detected or isolated. When there is an intermediate in a reaction, the energy diagram has a trough that represents the free energy of the intermediate as shown in Figure 6.2. This reaction has two transition states, one preceding formation of the intermediate and one preceding its conversion to product. The slowest step, the rate-determining or rate-limiting step, is the step with the highest energy transition state. In Figure 6.2, the rate-determining step is the formation of the intermediate. The intermediate is metastable because relatively little energy is required for the intermediate either to continue to product or to revert to the original reactant. Proposed intermediates that are too short-lived to be isolated or detected are often enclosed in square brackets like transition states, which they presumably closely resemble.

Catalysts create reaction pathways that have lower activation energies than those of uncatalyzed reactions. Catalysts participate directly in reactions by stabilizing the transition states along the reaction pathways. Enzymes are catalysts that accelerate reactions by lowering the overall activation energy. They achieve rate enhancement by providing a multistep pathway (with one or several intermediates) in which each of the steps has lower activation energy than the corresponding stages in the nonenzymatic reaction.

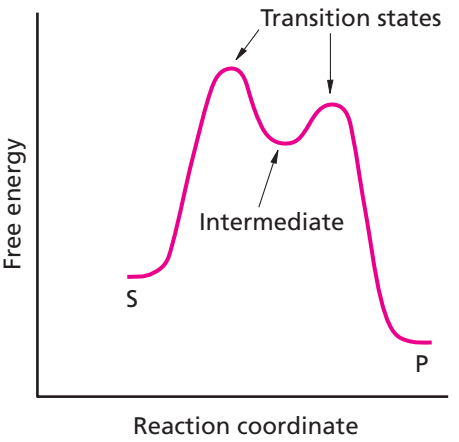
The first step in an enzymatic reaction is the formation of a noncovalent enzyme–substrate complex, ES. In a reaction between A and B, formation of the EAB complex collects and positions the reactants making the probability of reaction much higher for the enzyme-catalyzed reaction than for the uncatalyzed reaction. Figures 6.3a and 6.3b show a hypothetical case in which substrate binding is the only mode of catalysis by an enzyme. In this example, the activation energy is lowered by bringing the reactants together in the substrate binding site. Correct substrate binding accounts for a large part of the catalytic power of enzymes.

The active sites of enzymes bind substrates and products. They also bind transition states. In fact, transition states are likely to bind to active sites much more tightly than

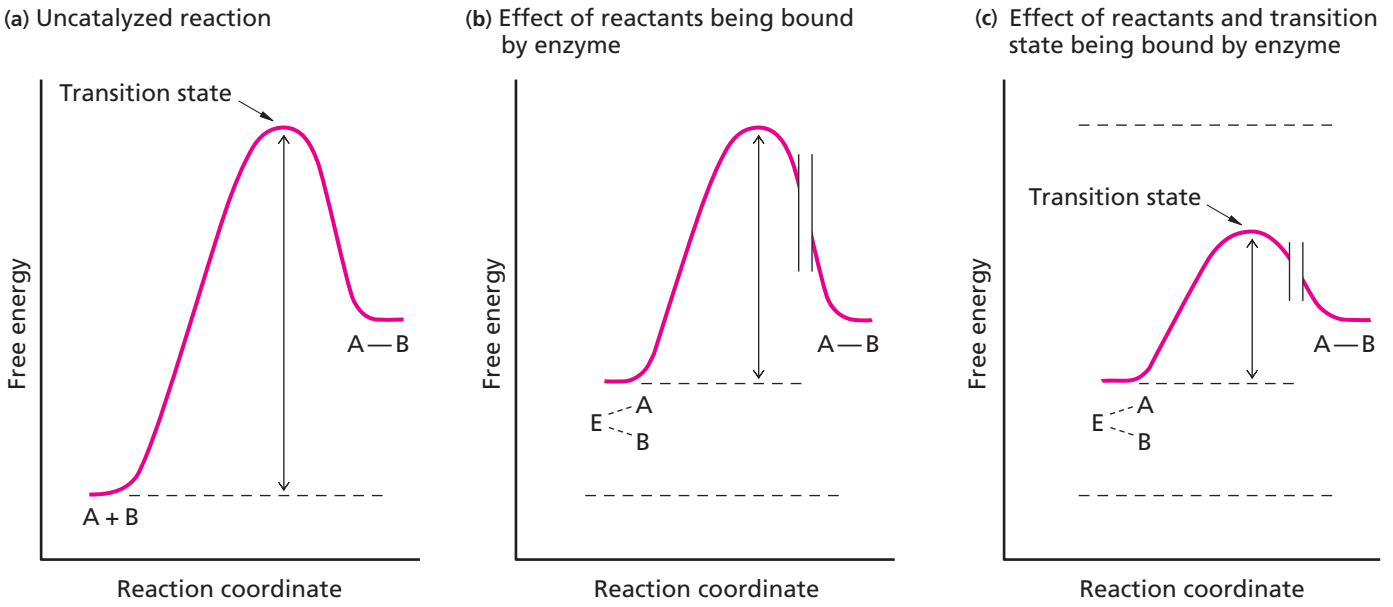
KEY CONCEPT

Transition states are unstable molecules with free energies higher than either the substrate or the product.

The meaning of activation energy is described in Section 1.4D.



▲ **Figure 6.2**
Energy diagram for a reaction with an intermediate. The intermediate occurs in the trough between the two transition states. The rate-determining step in the forward direction is formation of the first transition state, the step with the higher energy transition state. S represents the substrate, and P represents the product.



▲ **Figure 6.3**
Enzymatic catalysis of the reaction A + B → A—B. (a) Energy diagram for an uncatalyzed reaction. **(b)** Effect of reactant binding. Collection of the two reactants in the EAB complex properly positions them for reaction, makes formation of the transition state more frequent, and hence lowers the activation energy. **(c)** Effect of transition-state stabilization. An enzyme binds the transition state more tightly than it binds substrates, further lowering the activation energy. Thus, an enzymatic reaction has a much lower activation energy than an uncatalyzed reaction. (The breaks in the reaction curves indicate that the enzymes provide multistep pathways.)

substrates do. The extra binding interactions stabilize the transition state, further lowering the activation energy (Figure 6.3c). We will see that the binding of substrates followed by the binding of transition states provides the greatest rate acceleration in enzyme catalysis.

We return to binding phenomena later in this chapter after we examine the chemical processes that underlie enzyme function. (Note that enzyme-catalyzed reactions are usually reversible. The same principles apply to the reverse reaction. The activation energy is lowered by binding the "products" and stabilizing the transition state.)

6.3 Chemical Modes of Enzymatic Catalysis

In addition to reactive amino acid residues, there may be metal ions or coenzymes in the active site. The role of these cofactors in enzyme catalysis is described in Chapter 7.

The formation of an ES complex places reactants in proximity to reactive amino acid residues in the enzyme active site. Ionizable side chains participate in two kinds of chemical catalysis; acid–base catalysis and covalent catalysis. These are the two major chemical modes of catalysis.

A. Polar Amino Acid Residues in Active Sites

The active site cavity of an enzyme is generally lined with hydrophobic amino acid residues. However, a few polar, ionizable residues (and a few molecules of water) may also be present in the active site. Polar amino acid residues (or sometimes coenzymes) undergo chemical changes during enzymatic catalysis. These residues make up much of the catalytic center of the enzyme.

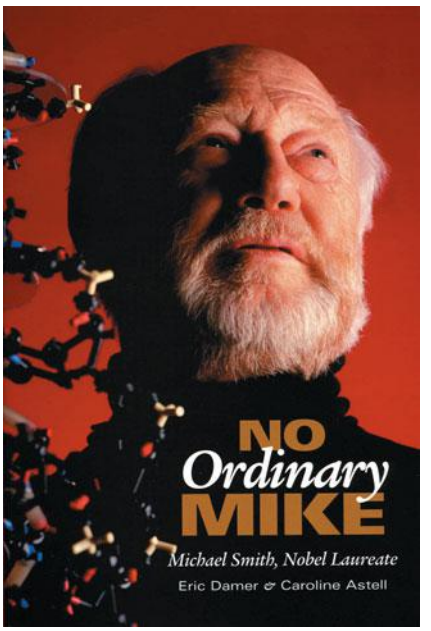
Table 6.1 lists the ionizable residues found in the active sites of enzymes. Histidine, which has a pK_a of about 6 to 7 in proteins, is often an acceptor or a donor of protons. Aspartate, glutamate, and occasionally lysine can also participate in proton transfer. Certain amino acids, such as serine and cysteine, are commonly involved in group-transfer reactions. At neutral pH, aspartate and glutamate usually have negative charges, and lysine and arginine have positive charges. These anions and cations can serve as sites for electrostatic binding of oppositely charged groups on substrates.

BOX 6.1 SITE-DIRECTED MUTAGENESIS MODIFIES ENZYMES

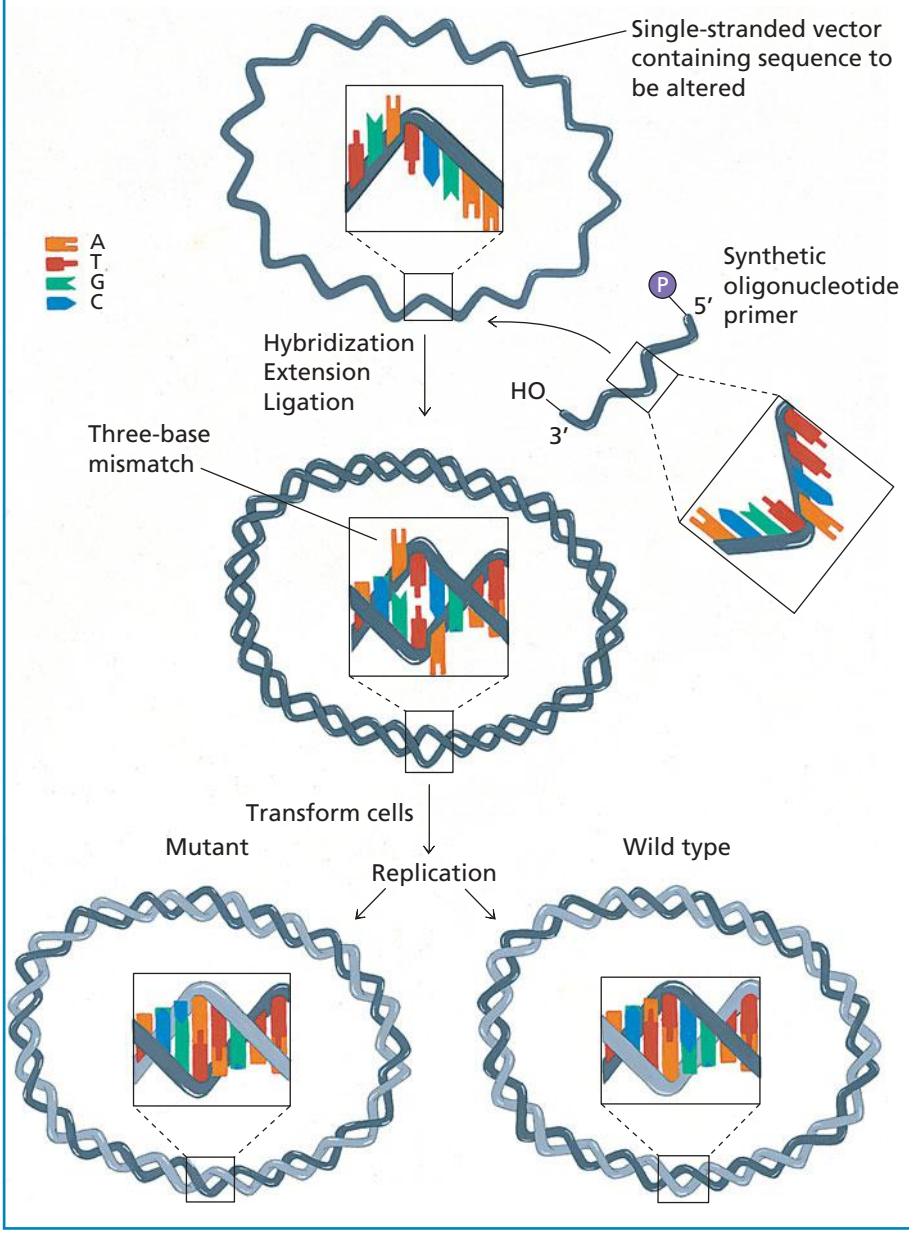
It is possible to test the functions of the amino acid side chains of an enzyme using the technique of *site-directed mutagenesis* (see Section 23.10). This technique has had a huge impact on our understanding of structure–function relationships of enzymes.

In site-directed mutagenesis, a desired mutation is engineered directly into a gene by synthesizing an oligonucleotide that contains the mutation flanked by sequences identical to the target gene. When this oligonucleotide is used as a primer for DNA replication *in vitro*, the new copy of the gene contains the desired mutation. Since alterations can be made at any position in a gene, specific changes in proteins can be engineered allowing direct testing of hypotheses about the functional role of key amino acid residues. Site-directed mutagenesis is commonly used to introduce single codon mutations into genes, resulting in single amino acid substitutions.

The mutated gene can be introduced into bacterial cells where modified enzymes are synthesized from the gene. The structure and activity of the mutant protein can then be analyzed to see the effect of changing an individual amino acid.



▲ Michael Smith (1932–2000), received the Nobel Prize in Chemistry in 1993 for inventing site-directed mutagenesis.



◄ Oligonucleotide-directed, site-specific mutagenesis. A synthetic oligonucleotide containing the desired change (3 bp) is annealed to the single-stranded vector containing the sequence to be altered. The synthetic oligonucleotide serves as a primer for the synthesis of a complementary strand. The double-stranded, circular heteroduplex is transformed into *E. coli* cells where replication produces mutant and wild-type DNA molecules.

Table 6.1 Catalytic functions of reactive groups of ionizable amino acids

Amino acid	Reactive group	Net charge at pH 7	Principal functions
Aspartate	$-COO^{\ominus}$	−1	Cation binding; proton transfer
Glutamate	$-COO^{\ominus}$	−1	Cation binding; proton transfer
Histidine	Imidazole	Near 0	Proton transfer
Cysteine	$-CH_2SH$	Near 0	Covalent binding of acyl groups
Tyrosine	Phenol	0	Hydrogen bonding to ligands
Lysine	$NH_3^{\oplus}$	+1	Anion binding; proton transfer
Arginine	Guanidinium	+1	Anion binding
Serine	$-CH_2OH$	0	Covalent binding of acyl groups

Table 6.2 Typical pK_a values of ionizable groups of amino acids in proteins

Group	pK_a
Terminal α-carboxyl	3–4
Side-chain carboxyl	4–5
Imidazole	6–7
Terminal α-amino	7.5–9
Thiol	8–9.5
Phenol	9.5–10
ε-Amino	~10
Guanidine	~12
Hydroxymethyl	~16

Table 6.3 Frequency distribution of catalytic residues in enzymes

	% of catalytic residues	% of all residues
His	18	3
Asp	15	6
Arg	11	5
Glu	11	6
Lys	9	6
Cys	6	1
Tyr	6	4
Asn	5	4
Ser	4	5
Gly	4	8

The pK_a values of the ionizable groups of amino acid residues in proteins may differ from the values of the same groups in free amino acids (Section 3.4). Table 6.2 lists the typical pK_a values of ionizable groups of amino acid residues in proteins. Compare these ranges to the exact values for free amino acids in Table 3.2. A given ionizable group can have different pK_a values within a protein because of differing microenvironments. These differences are usually small but can be significant.

Occasionally, the side chain of a catalytic amino acid residue exhibits a pK_a quite different from the one shown in Table 6.2. Bearing in mind that pK_a values may be perturbed, one can test whether particular amino acids participate in a reaction by examining the effect of pH on the reaction rate. If the change in rate correlates with the pK_a of a certain ionic amino acid (Section 6.3D), a residue of that amino acid may take part in catalysis.

Only a small number of amino acid residues participate directly in catalyzing reactions. Most residues contribute in an indirect way by helping to maintain the correct three-dimensional structure of a protein. As we saw in Chapter 4, the majority of amino acid residues are not evolutionarily conserved.

In vitro mutagenesis studies of enzymes have confirmed that most amino acid substitutions have little effect on enzyme activity. Nevertheless, every enzyme has a few key residues that are absolutely essential for catalysis. Some of these residues are directly involved in the catalytic mechanism, often by acting as an acid or base catalyst or a nucleophile. Other residues act indirectly to assist or enhance the role of a key residue. Other roles for key catalytic residues include substrate binding, stabilization of the transition state, and interacting with essential cofactors.

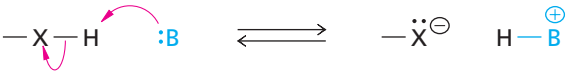
Enzymes usually have between two and six key catalytic residues. The top ten catalytic residues are listed in Table 6.3. The charged residues, His, Asp, Arg, Glu, and Lys account for almost two-thirds of all catalytic residues. This makes sense since charged side chains are more likely to act as acids, bases, and nucleophiles. They are also more likely to play a role in binding substrates or transition states. The number one catalytic residue is histidine. Histidine is 6 times more likely to be involved in catalysis than its abundance in proteins would suggest.

B. Acid–Base Catalysis

In acid–base catalysis, the acceleration of a reaction is achieved by catalytic transfer of a proton. Acid–base catalysis is the most common form of catalysis in organic chemistry and it's also common in enzymatic reactions. Enzymes that employ acid–base catalysis rely on amino acid side chains that can donate and accept protons under the nearly neutral pH conditions of cells. This type of acid–base catalysis, involving proton-transferring agents, is termed *general acid–base catalysis*. (Catalysis by $H^{\oplus}$ or $OH^{\ominus}$ is termed *specific acid* or *specific base catalysis*.) In effect, the active sites of these enzymes provide the biological equivalent of a solution of acid or base.

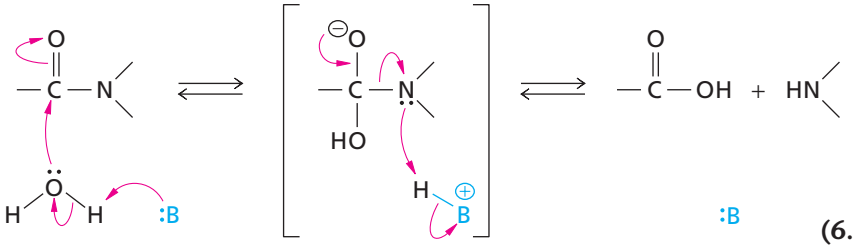
It is convenient to use B: to represent a base, or proton acceptor, and $BH^{\oplus}$ to represent its conjugate acid, a proton donor. (This acid–base pair can also be written as

$HA/A^{\ominus}$.) A proton acceptor can assist reactions in two ways: (1) it can cleave O—H, N—H, or even some C—H bonds by removing a proton



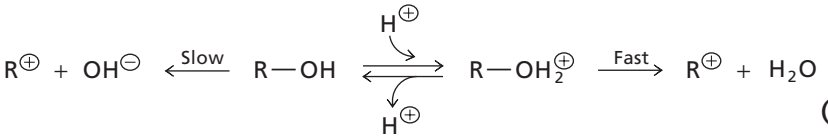
(6.6)

and (2) the general base B: can participate in the cleavage of other bonds involving carbon, such as a C—N bond, by generating the equivalent of $OH^{\ominus}$ in neutral solution through removal of a proton from a molecule of water.



(6.7)

The general acid $BH^{\oplus}$ can also assist in bond cleavage. A covalent bond may break more easily if one of its atoms is protonated. For example,



(6.8)

$BH^{\oplus}$ catalyzes bond cleavage by donating a proton to an atom (such as the oxygen of R—OH in Equation 6.8), thereby making bonds to that atom more labile. In all reactions involving $BH^{\oplus}$ the reverse reaction is catalyzed by B:, and vice versa.

Histidine is an ideal group for proton transfer at neutral pH values because the imidazole/imidazolium of the side chain has a pK_a of about 6 to 7 in most proteins. We have seen that histidine is a common catalytic residue. In the following sections, we will examine some specific roles of histidine side chains.

KEY CONCEPT

In acid–base catalysis, the reaction requires specific amino acid side chains that can donate and accept protons.

C. Covalent Catalysis

In covalent catalysis, a substrate is bound covalently to the enzyme to form a reactive intermediate. The reacting side chain of the enzyme can be either a nucleophile or an electrophile. Nucleophilic catalysis is more common. In the second step of the reaction, a portion of the substrate is transferred from the intermediate to a second substrate. For example, the group X can be transferred from molecule A—X to molecule B in the following two steps via the covalent ES complex X—E:

$$A—X + E \rightleftharpoons X—E + A \qquad (6.9)$$

and

$$X—E + B \rightleftharpoons B—X + E \qquad (6.10)$$

This is a common mechanism for coupling two different reactions in biochemistry. Recall that the ability to couple reactions is one of the important properties of enzymes (Chapter 5; "Introduction"). Transferases, one of the six classes of enzymes (Section 5.1), catalyze group-transfer reactions in this manner and hydrolases catalyze a special kind of group-transfer reaction where water is the acceptor. Transferases and hydrolases together make up more than half of known enzymes.

The reaction catalyzed by bacterial sucrose phosphorylase is an example of group transfer by covalent catalysis. (Sucrose is composed of one glucose residue and one fructose residue.)

$$\text{Sucrose} + P_i \rightleftharpoons \text{Glucose 1-phosphate} + \text{Fructose} \qquad (6.11)$$

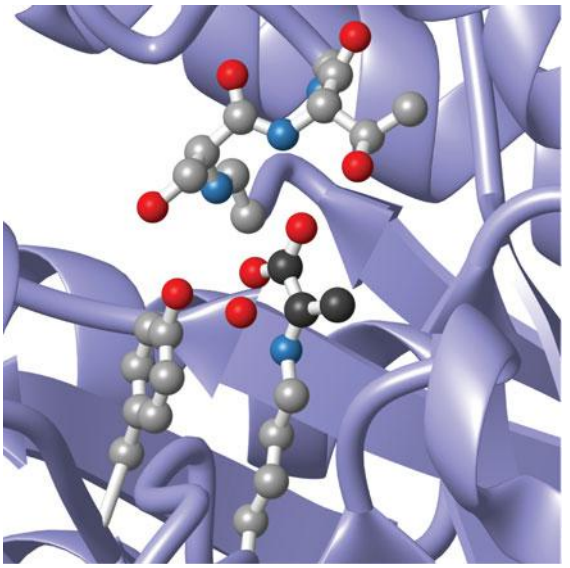


Figure 6.4 ►
Covalent catalysis. The enzyme *N*-acetyl-D-neuraminic acid lyase from *Escherichia coli* catalyzes the condensation of pyruvate and *N*-acetyl-D-mannosamine to form *N*-acetyl-D-neuraminic acid (see Section 8.7C). One of the intermediates in the reaction is a Schiff base (see Fig. 5.15) between pyruvate (black carbon atoms) and a lysine reside. The intermediate is stabilized by hydrogen bonds with other amino acid side chains. [PDB 2WKJ]

KEY CONCEPT

In covalent catalysis mechanisms, the enzyme participates directly in the reaction. It reacts with a substrate and an intermediate containing the enzyme is produced. The reaction is not complete until free enzyme is regenerated.

The first chemical step in the reaction is formation of a covalent glucosyl–enzyme intermediate. In this case, sucrose is equivalent to A—X and glucose is equivalent to X in Reaction 6.9.

$$\text{Sucrose} + \text{Enzyme} \rightleftharpoons \text{Glucosyl}-\text{Enzyme} + \text{Fructose} \qquad \textbf{(6.12)}$$

The covalent ES intermediate can donate the glucose unit either to another molecule of fructose, in the reverse of Reaction 6.12, or to phosphate (which is equivalent to B in Reaction 6.10).

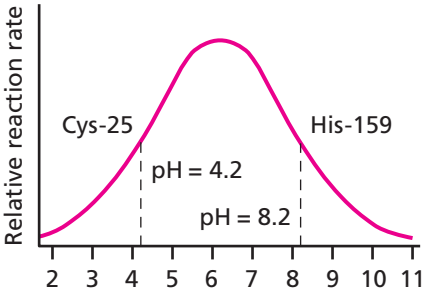
$$\text{Glucosyl}-\text{Enzyme} + P_i \rightleftharpoons \text{Glucose 1}-\text{phosphate} + \text{Enzyme} \qquad \textbf{(6.13)}$$

Proof that an enzyme mechanism relies on covalent catalysis often requires the isolation or detection of an intermediate and demonstration that it is sufficiently reactive. In some cases, the covalently bound intermediate is seen in the crystal structure of an enzyme, and this is direct proof of covalent catalysis (Figure 6.4).

D. pH Affects Enzymatic Rates

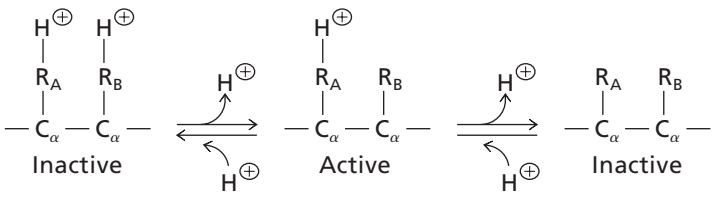
The effect of pH on the reaction rate of an enzyme can suggest which ionizable amino acid residues are in its active site. Sensitivity to pH usually reflects an alteration in the ionization state of one or more residues involved in catalysis, although occasionally substrate binding is affected. A plot of reaction velocity versus pH most often yields a bell-shaped curve provided the enzyme is not denatured when the pH is altered.

A good example is the pH versus rate profile for papain, a protease isolated from papaya fruit (Figure 6.5). The bell-shaped pH profile can be explained by assuming that the ascending portion of the curve represents the deprotonation of an active-site amino acid residue (B) and the descending portion represents the deprotonation of a second active-site amino acid residue (A). The two inflection points approximate the pK_a values of the two ionizable residues. A simple bell-shaped curve is the result of two overlapping



◄ Figure 6.5
pH vs rate profile for papain. The left and right segments of the bell-shaped curve represent the titrations of the side chains of active-site amino acids. The inflection point at pH 4.2 reflects the pK_a of Cys-25, and the inflection point at pH 8.2 reflects the pK_a of His-159. The enzyme is active only when these ionic groups are present as the thiolate–imidazolium ion pair.

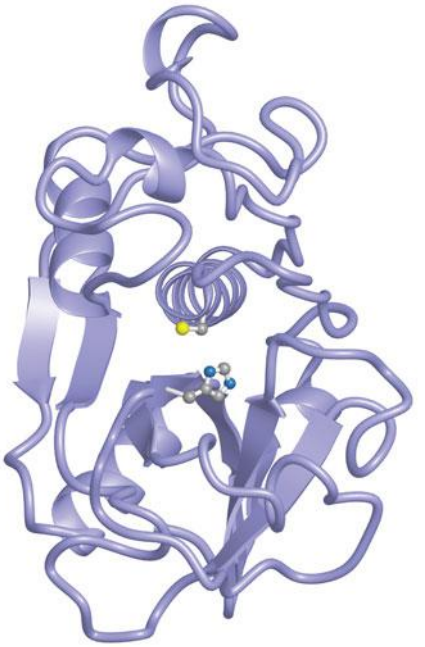
titrations. The side chain of A (R_A) must be protonated for activity and the side chain of B (R_B) must be unprotonated.



(6.14)

At the pH optimum, midway between the two pK_a values, the greatest number of enzyme molecules is in the active form with residue A protonated. Not all pH profiles are bell-shaped. A pH profile is a sigmoidal curve if only one ionizable amino acid residue participates in catalysis and it can have a more complicated shape if more than two ionizable groups participate. Enzymes are routinely assayed near their optimal pH, which is maintained using appropriate buffers.

The pH versus rate graph for papain has inflection points at pH 4.2 and pH 8.2, suggesting that the activity of papain depends on two active-site amino acid residues with pK_a values of about 4 and 8. These ionizable residues are a nucleophilic cysteine (Cys-25) and a proton-donating imidazolium group of histidine (His-159) (Figure 6.6). The side chain of cysteine normally has a pK_a value of 8 to 9.5 but in the active site of papain the pK_a of Cys-25 is greatly perturbed to 3.4. The pK_a of the His-159 residue is perturbed to 8.3. The inflection points on the pH profile do not correspond exactly to the pK_a values of Cys-25 and His-159 because the ionization of additional groups contributes slightly to the overall shape of the curve. Three ionic forms of the catalytic center of papain are shown in Figure 6.7. The enzyme is active only when the thiolate group and the imidazolium group form an ion pair (as in the upper tautomer of the middle pair).



▲ **Figure 6.6 Ionizable residues in papain.** Model of papain, showing ball-and-stick models of the active-site histidine and cysteine side chain. The imidazole nitrogen atoms are blue, and the sulfur atom is yellow.

6.4 Diffusion-Controlled Reactions

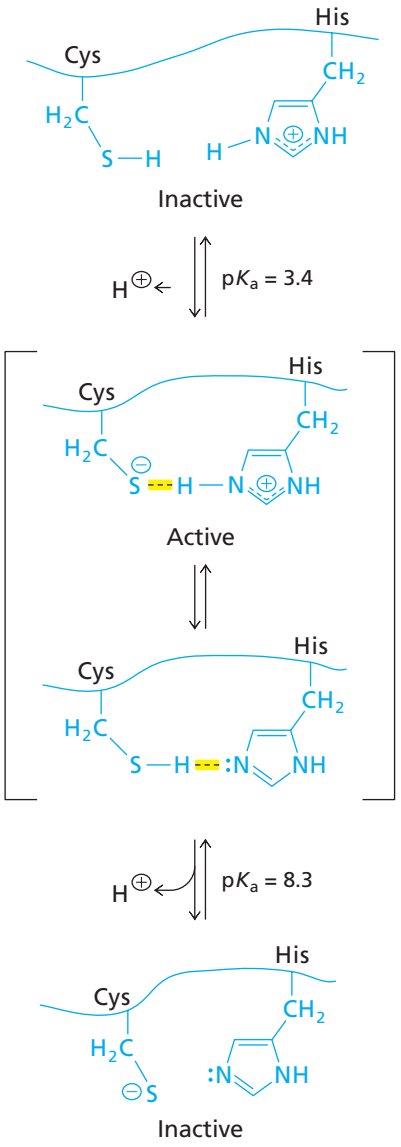
A few enzymes catalyze reactions at rates approaching the upper physical limit of reactions in solution. This theoretical upper limit is the rate of diffusion of reactants into the active site. A reaction that occurs with every collision between reactant molecules is termed a **diffusion controlled reaction** or a diffusion-limited reaction. Under physiological conditions the diffusion-controlled rate is about 10^8 to 10^9 M^{-1} s^{-1}. Compare this theoretical maximum to the apparent second-order rate constants (k_{cat}/K_m) for five very fast enzymes listed in Table 6.4.

The binding of a substrate to an enzyme is a rapid reaction. If the rest of the reaction is simple and fast, the binding step may be the rate-determining step and the overall rate of the reaction may approach the upper limit for catalysis. Only a few types of chemical reactions can proceed this quickly. These include association reactions, some proton transfers, and electron transfers. The reactions catalyzed by all the enzymes listed in Table 6.4 are so simple that the rate-determining steps are roughly as fast as

Table 6.4 Enzymes with second-order rate constants near the upper limit

Enzyme	Substrate	k_{cat}/K_m (M^{-1} s^{-1})*
Catalase	H_2O_2	4×10^7
Acetylcholinesterase	Acetylcholine	2×10^8
Triose phosphate isomerase	D-Glyceraldehyde 3-phosphate	4×10^8
Fumarase	Fumarate	10^9
Superoxide dismutase	$\cdot O_2^{\ominus}$	2×10^9

*The ratio k_{cat}/K_m is the apparent second-order rate constant for the enzyme-catalyzed reaction E + S → E + P. For these enzymes, the formation of the ES complex can be the slowest step.

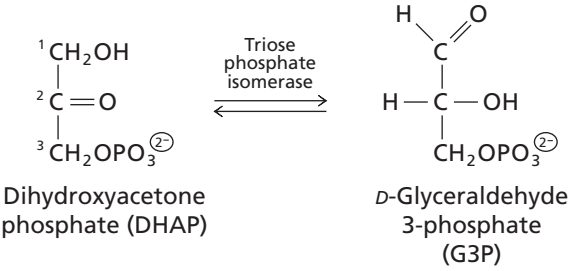


▲ **Figure 6.7 The activity of papain depends on two ionizable residues, histidine (His-159) and cysteine (Cys-25), in the active site.** Three ionic forms of these residues are shown. Only the upper tautomer of the middle pair is active.

binding of substrates to the enzymes. They catalyze diffusion-controlled reactions. We will now look at two of these enzymes in detail: triose phosphate isomerase and superoxide dismutase.

A. Triose Phosphate Isomerase

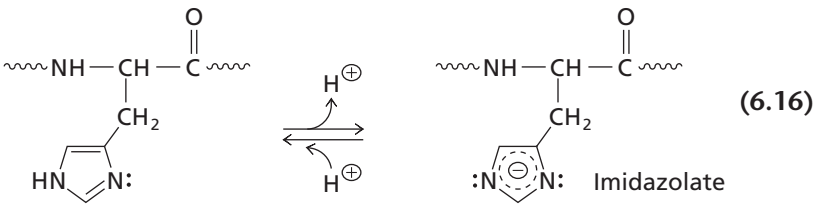
Triose phosphate isomerase catalyzes the rapid interconversion of dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (G3P) in the glycolysis and gluconeogenesis pathways (Chapters 11 and 12).



(6.15)

The reaction proceeds by shifting protons from the carbon atom 1 of DHAP to the carbon atom 2 (Figure 6.8). Triose phosphate isomerase has two ionizable active-site residues: glutamate that acts as a general acid–base catalyst, and histidine that shuttles a proton between oxygen atoms of an enzyme-bound intermediate. When dihydroxyacetone phosphate (DHAP) binds, the carbonyl oxygen forms a hydrogen bond with the imidazole group of His-95. The carboxylate group of Glu-165 removes a proton from C-1 of the substrate to form an enoldiolate transition state (Figure 6.8, top). The transition-state molecule is rapidly converted to a stable enediol intermediate (middle, Figure 6.8). This intermediate is then converted via a second enediolate transition state to D-glyceraldehyde 3-phosphate (G3P).

In this reaction, the proton-donating form of histidine appears to be the neutral species and the proton-accepting species appears to be the imidazolate. The hydrogen bonds formed between histidine and the intermediates in this mechanism appear to be unusually strong.



(6.16)

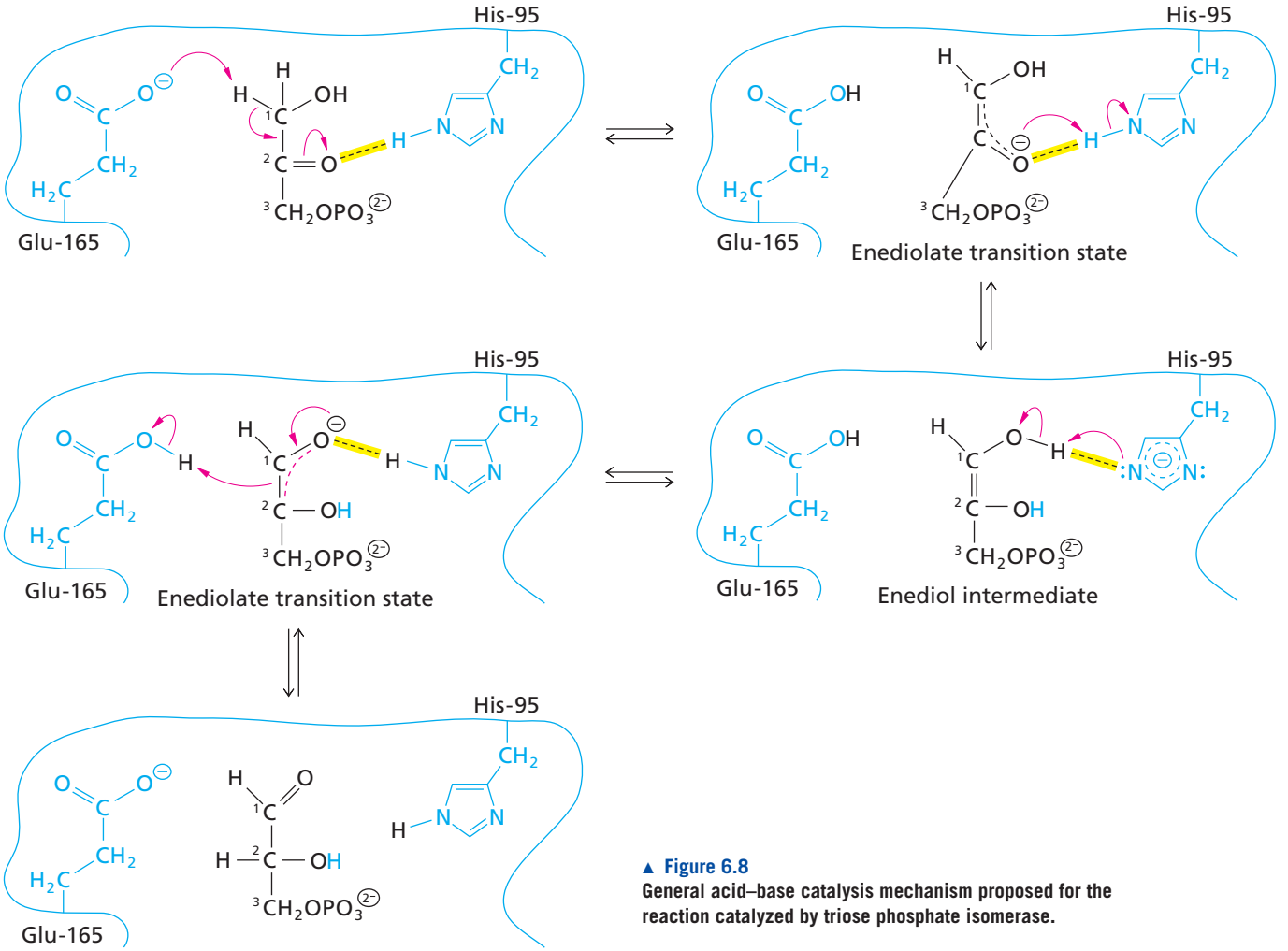
The imidazolate form of a histidine residue is unusual; the triose phosphate isomerase mechanism was the first enzymatic mechanism in which this form was implicated.

The enediol intermediate is stable and in order to prevent it from diffusing out of the active site, triose phosphate isomerase has evolved a "locking" mechanism to seal the active site until the reaction is complete. When substrate binds, a flexible loop of the protein moves to cover the active site and prevent release of the enediol intermediate (Figure 6.9).

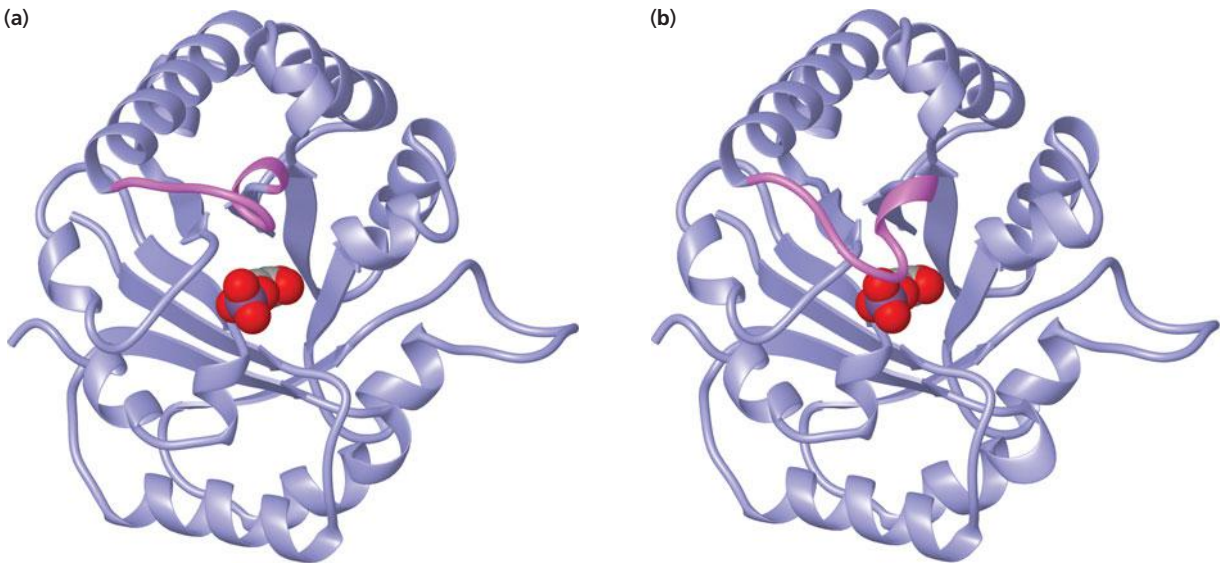
The rate constants of all four kinetically measurable enzymatic steps have been determined.

$$\text{E} + \text{DHAP} \underset{}{\overset{(1)}{\rightleftharpoons}} \text{E-DHAP} \overset{(2)}{\rightleftharpoons} \text{E-Intermediate} \overset{(3)}{\rightleftharpoons}$$

$$\text{E-G3P} \overset{(4)}{\rightleftharpoons} \text{E} + \text{G3P} \qquad \textbf{(6.17)}$$



▲ **Figure 6.8**
General acid–base catalysis mechanism proposed for the reaction catalyzed by triose phosphate isomerase.



▲ **Figure 6.9**
Structure of yeast (*Saccharomyces cerevisiae*) triose phosphate isomerase. The location of the substrate is indicated by the space-filling model of a substrate analog. **(a)** The structure of the "open loop" form of the enzyme when the active site is unoccupied. **(b)** The structure when the loop has closed over the active site to prevent release of the enediol intermediate before the reaction is completed.

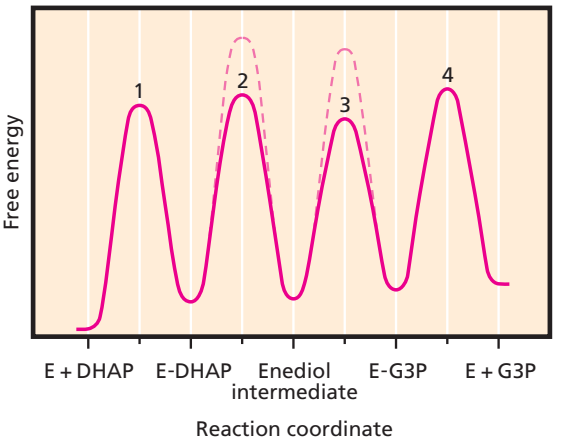


Figure 6.10 ►
Energy diagram for the reaction catalyzed by triose phosphate isomerase. [Adapted from Raines, R. T., Sutton, E. L., Strauss, D. R., Gilbert, W., and Knowles, J. R. (1986). Reaction energetics of a mutant triose phosphate isomerase in which the active-site glutamate has been changed to aspartate. *Biochem.* 25:7142–7154.]

The energy diagram constructed from these rate constants is shown in Figure 6.10. Note that all the barriers for the enzyme are approximately the same height. This means that the steps are balanced, and no single step is rate-limiting. The physical step of S binding to E is rapid but not much faster than the subsequent chemical steps in the reaction sequence. The value of the second-order rate constant k_{cat}/K_m for the conversion of glyceraldehyde 3-phosphate to dihydroxyacetone phosphate is 4×10^8 M^{-1} s^{-1}, which is close to the theoretical rate of a diffusion-controlled reaction. It appears that this isomerase has achieved its maximum possible efficiency as a catalyst.

BOX 6.2 THE "PERFECT ENZYME"?

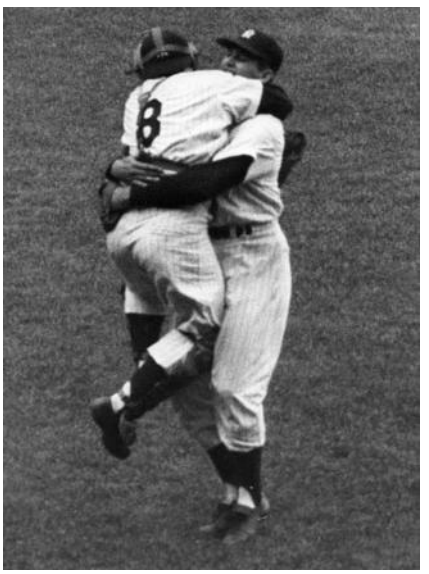
Much of our understanding of the mechanism of triose phosphate isomerase (TPI) comes from the lab of Jeremy Knowles at Harvard University (Cambridge, MA, USA). He points out that the enzyme has achieved catalytic perfection because the overall rate of the reaction is limited only by the rate of diffusion of substrate into the active site. TPI can't work any faster than this!

This has led many people to declare that TPI is the "perfect enzyme" because it has evolved to be so efficient. However, as Knowles and his coworkers have explained, the "perfect enzyme" isn't necessarily one that has evolved the maximum reaction rate. Most enzymes are not under selective pressure to increase their rate of reaction because they are part of a metabolic pathway that meets the cell's needs at less than optimal rates.

Even if it would be beneficial to increase the overall flux in a pathway (i.e., produce more of the end product per second), an individual enzyme need only keep up with the slowest enzyme in the pathway in order to achieve "perfection." The slowest enzyme might be catalyzing a very complicated reaction and might be very efficient. In this case, there will be no selective pressure on the other enzymes to evolve faster mechanisms and they are all "perfect enzymes."

In all species, triose phosphate isomerase is part of the gluconeogenesis pathway leading to the synthesis of glucose. In most species, it also plays a role in the reverse pathway where glucose is degraded (glycolysis). The enzyme is very ancient, and all versions—bacterial and eukaryotic—have achieved catalytic perfection. The two enzymes on either side of the reaction pathway, aldolase and glyceraldehyde 3-phosphate dehydrogenase (Section 11.2), are much slower. Thus, it is by no means obvious why TPI works as fast as it does.

The important point to keep in mind is that the vast majority of enzymes have not evolved catalytic perfection because their *in vivo* rates are "perfectly" adequate for the needs of the cell.



▲ **The Perfect Game.** New York Yankees catcher Yogi Berra congratulates Don Larson for pitching a perfect game in the 1956 World Series against the Brooklyn Dodgers. Perfect games are rare in baseball but there are many "perfect enzymes."

B. Superoxide Dismutase

Superoxide dismutase is an even faster catalyst than triose phosphate isomerase. Superoxide dismutase catalyzes the very rapid removal of the toxic superoxide radical anion, $\cdot O_2^{\ominus}$, a by-product of oxidative metabolism. The enzyme catalyzes the conversion of superoxide to molecular oxygen and hydrogen peroxide, which is rapidly removed by the subsequent action of enzymes such as catalase.

$$4\,\cdot O_2^{\ominus} \xrightarrow[\text{Superoxide dismutase}]{4\,H^{\oplus}\ \ 2\,O_2} 2\,H_2O_2 \xrightarrow[\text{Catalase}]{} 2\,H_2O + O_2 \qquad \textbf{(6.18)}$$

The reaction proceeds in two steps during which an atom of copper bound to the enzyme is reduced and then oxidized.

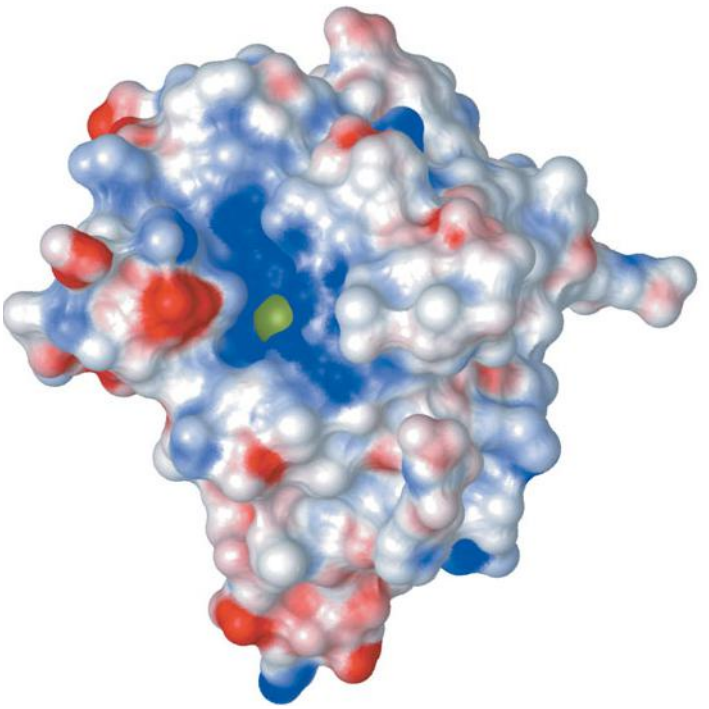
$$\text{E-Cu}^{2+} + \cdot O_2^{\ominus} \longrightarrow \text{E-Cu}^{\oplus} + O_2 \qquad \textbf{(6.19)}$$

$$\text{E-Cu}^{\oplus} + \cdot O_2^{\ominus} + 2H^{\oplus} \longrightarrow \text{E-Cu}^{2+} + H_2O_2 \qquad \textbf{(6.20)}$$

The overall reaction includes binding of the anionic substrate molecules, transfer of electrons and protons, and release of the uncharged products—all very rapid reactions with this enzyme. The k_{cat}/K_m value for superoxide dismutase at 25°C is near $2 \times 10^9\ M^{-1}\ s^{-1}$ (Table 6.4). This rate is even faster than that expected for association of the substrate with the enzyme based on typical diffusion rates.

How can the rate exceed the rate of diffusion? The explanation was revealed when the structure of the enzyme was examined. An electric field around the superoxide dismutase active site enhances the rate of formation of the ES complex about 30-fold. As shown in Figure 6.11, the active-site copper atom lies at the bottom of a deep channel in the protein. Hydrophilic amino acid residues at the rim of the active-site pocket guide negatively charged $\cdot O_2^{\ominus}$ to the positively charged region surrounding the active site. Electrostatic effects allow superoxide dismutase to bind and remove superoxide (radicals) much faster than expected from random collisions of enzyme and substrate.

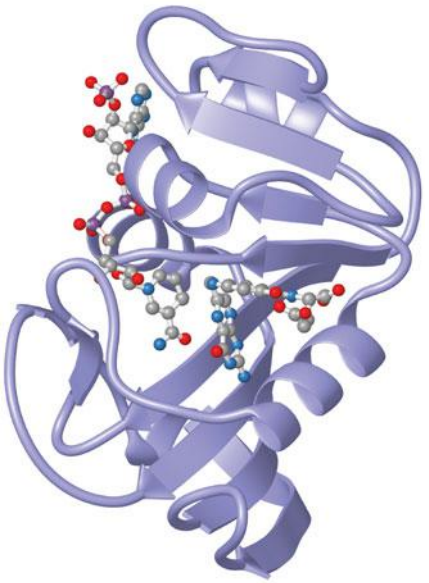
There are probably many enzymes with enhanced rates of binding due to electrostatic effects. In most cases, the rate-limiting step is catalysis so the overall rate (k_{cat}/K_m) is slower than the maximum for a diffusion-controlled reaction. For those enzymes with fast catalytic reactions, natural selection might favor rapid binding to enhance the overall rate. Similarly, an enzyme with rapid binding might evolve a mechanism that favored a faster reaction. However, most biochemical reactions proceed at rates that are more than sufficient to meet the needs of the cell.



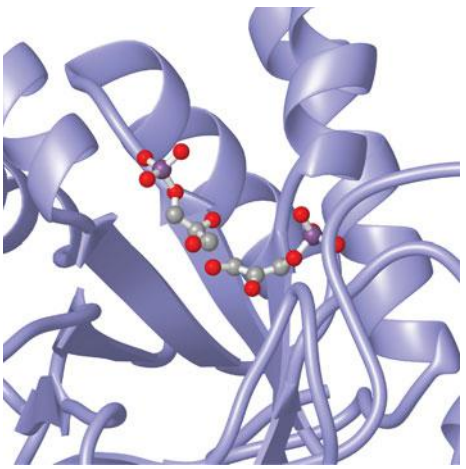
▲ **Figure 6.11**
Surface charge on human superoxide dismutase. The structure of the enzyme is shown as a model that emphasizes the surface of the protein. Positively charged regions are colored blue and negatively charged regions are colored red. The copper atom at the active site is green. Note that the channel leading to the binding site is lined with positively charged residues. [PDB 1HL5]

6.5 Modes of Enzymatic Catalysis

The quantitative effects of various catalytic mechanisms are difficult to assess. We have already seen two chemical mechanisms of enzymatic catalysis, acid–base catalysis and covalent catalysis. From studies of nonenzymatic catalysts it is estimated that acid–base catalysis can accelerate a typical enzymatic reaction by a factor of 10 to 100. Covalent catalysis can provide about the same rate acceleration.



▲ **Figure 6.12**
Substrate binding. Dihydrofolate reductase binds $NADP^+$ (left) and folate (right), positioning them in the active site in preparation for the reductase reaction. Most of the catalytic rate enhancement is due to binding effects. [PDB 7DFR]



▲ **Figure 6.13**
The proximity effect. The enzyme fructose-1,6-*bis*phosphate aldolase catalyzes the biosynthesis of fructose-1,6-*bis*phosphate from DHAP and G3P during gluconeogenesis and the cleavage of fructose-1,6-*bis*phosphate to dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (G3P) during glycolysis (see Section 11.2#4). In the biosynthesis reaction, the two substrates DHAP and G3P must be positioned close together in the active site in an orientation that promotes their joining to form the larger fructose-1,6-*bis*phosphate. This proximity effect is illustrated for the aldolase from *Mycobacterium tuberculosis*. [PDB 2EKZ]

As important as these chemical modes are, they account for only a small portion of the observed rate accelerations achieved by enzymes (typically 10^8 to 10^{12}). The ability of proteins to specifically bind and orient ligands explains the remainder. The proper binding of reactants in the active sites of enzymes provides not only substrate and reaction specificity but also most of the catalytic power of enzymes (Figure 6.12).

There are two catalytic modes based on binding phenomena. First, for multisubstrate reactions the collecting and correct positioning of substrate molecules in the active site raises their effective concentrations over their concentrations in free solution. In the same way, binding of a substrate near a catalytic active-site residue decreases the activation energy by reducing the entropy while increasing the effective concentrations of these two reactants. High effective concentrations favor the more frequent formation of transition states. This phenomenon is called the *proximity effect*. Efficient catalysis requires fairly weak binding of reactants to enzymes since extremely tight binding would inhibit the reaction.

The second major catalytic mode arising from the ligand–enzyme interaction is the increased binding of transition states to enzymes compared to the binding of substrates or products. This catalytic mode is called *transition state stabilization*. There is an equilibrium (not the reaction equilibrium) between ES and the enzymatic transition state, $ES^‡$. Interaction between the enzyme and its ligands in the transition state shifts this equilibrium toward $ES^‡$ and lowers the activation energy.

The effects of proximity and transition-state stabilization were illustrated in Figure 6.3. Experiments suggest that proximity can increase reaction rates more than 10,000-fold, and transition-state stabilization can increase reaction rates at least that much. Enzymes can achieve extraordinary rate accelerations when both of these effects are multiplied by chemical catalytic effects.

The binding forces responsible for formation of ES complexes and for stabilization of $ES^‡$ are familiar from Chapters 2 and 4. These weak forces are charge–charge interactions, hydrogen bonds, hydrophobic interactions, and van der Waals forces. Charge–charge interactions are stronger in nonpolar environments than in water. Because active sites are largely nonpolar, charge–charge interactions in the active sites of enzymes can be quite strong. The side chains of aspartate, glutamate, histidine, lysine, and arginine residues provide negative and positive groups that form ion pairs with substrates in active sites. Next in bond strength are hydrogen bonds that often form between substrates and enzymes. The peptide backbone and the side chains of many amino acids can form hydrogen bonds. Highly hydrophobic amino acids, as well as alanine, proline, tryptophan, and tyrosine, can participate in hydrophobic interactions with the nonpolar groups of ligands. Many weak van der Waals interactions also help bind substrates. Keep in mind that both the chemical properties of the amino acid residues and the shape of the active site of an enzyme determine which substrates will bind.

A. The Proximity Effect

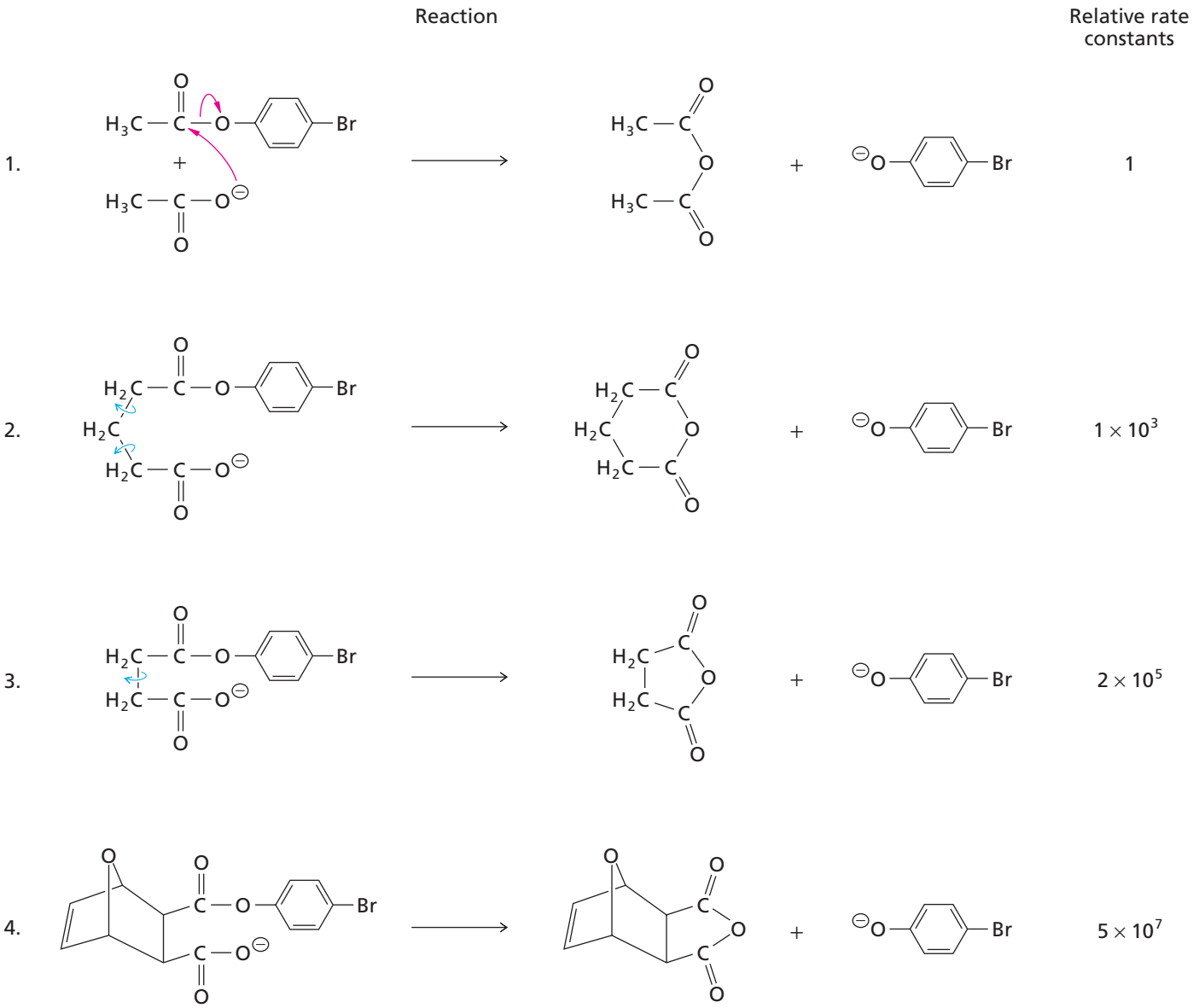
Enzymes are frequently described as entropy traps—agents that collect highly mobile reactants from dilute solution thereby decreasing their entropy and increasing the probability of their interaction. You can think of the reaction of two molecules positioned at the active site as an *intramolecular* (unimolecular) reaction. The correct positioning of two reacting groups in the active site reduces their degrees of freedom and produces a large loss of entropy sufficient to account for a large rate acceleration (Figure 6.13). The acceleration is expressed in terms of the enhanced relative concentration, called the *effective molarity*, of the reacting groups in the unimolecular reaction. The effective molarity can be obtained from the ratio

$$\text{Effective molarity} = \frac{k_1(s^{-1})}{k_2(M^{-1}\,s^{-1})} \qquad \textbf{(6.21)}$$

where k_1 is the rate constant when the reactants are preassembled into a single molecule and k_2 is the rate constant of the corresponding bimolecular reaction. All the units in this equation cancel except M, so the ratio is expressed in molar units. Effective molarities are not real concentrations; in fact, for some reactions the values are impossibly high. Nevertheless, effective molarities indicate how favorably reactive groups are oriented.

The importance of the proximity effect is illustrated by experiments comparing a nonenzymatic bimolecular reaction to a series of chemically similar intramolecular reactions (Figure 6.14). The bimolecular reaction was the two-step hydrolysis of *p*-bromophenyl acetate, catalyzed by acetate and proceeding via the formation of acetic anhydride. (The second step, hydrolysis of acetic anhydride, is not shown in Figure 6.14.) In the unimolecular version, reacting groups were connected by a bridge with progressively greater restriction of rotation. With each restriction placed on the substrate molecules, the relative rate constant (k_1/k_2) increased markedly. The glutarate ester (compound 2) has two bonds that allow rotational freedom whereas the succinate ester (compound 3) has only one. The most restricted compound, the rigid bicyclic compound 4, has no rotational freedom. In this compound, the carboxylate is

▼ Figure 6.14
Reactions of a series of carboxylates with substituted phenyl esters. The proximity effect is illustrated by the increase in rate observed when the reactants are held more rigidly in proximity. Reaction 4 is 50 million times faster than Reaction 1, the bimolecular reaction. [Based on Bruice and Pandit (1960). Intramolecular models depicting the kinetic importance of "fit" in enzymatic catalysis. *Biochem.* 46:402–404.]



KEY CONCEPT

The correct binding and positioning of specific substrates in the active site of an enzyme produces a large acceleration in the rate of a reaction.

close to the ester and the reacting groups are properly aligned. The effective molarity of the carboxylate group is 5×10^7 M. Compound 4 has an extremely high probability of reaction because very little entropy must be lost to reach the transition state. Theoretical considerations suggest that the greatest rate acceleration that can be expected from the proximity effect is about 10^8. This entire rate acceleration can be attributed to the loss of entropy that occurs when two reactants are properly positioned for reaction. These intramolecular reactions can serve as a model of the positioning of two substrates bound in the active site of an enzyme.

B. Weak Binding of Substrates to Enzymes

Reactions of ES complexes are analogous to unimolecular reactions even when two substrates are involved. Although the correct positioning of substrates in an active site produces a large rate acceleration, enzymes do not achieve the maximum 10^8 acceleration theoretically generated by the proximity effect. Typically, the loss in entropy on binding of the substrate allows an acceleration of only 10^4. That's because in ES complexes the reactants are brought toward, but not extremely close to, the transition state. This conclusion is based on both mechanistic reasoning and measurements of the tightness of binding of substrates and inhibitors to enzymes. One major limitation is that binding of substrates to enzymes cannot be extremely tight; that is, K_m values cannot be extremely low.

Figure 6.15 shows energy diagrams for a nonenzymatic unimolecular reaction and the corresponding multistep enzyme-catalyzed reaction. As we will see in the next section, an enzyme increases the rate of a reaction by stabilizing (i.e., tightly binding) the transition state. Therefore, the energy required for ES to reach the transition state ($ES^‡$) in the enzymatic reaction is less than the energy required for S to reach $S^‡$, the transition state in the nonenzymatic reaction.

Recall that the substrate must be bound fairly weakly in the ES complex. If a substrate were bound extremely tightly, it could take just as much energy to reach $ES^‡$ from ES (the arrow labeled 2) as is required to reach $S^‡$ from S in the nonenzymatic reaction (the arrow labeled 1). In other words, extremely tight binding of the substrate would mean little or no catalysis. Excessive ES stability is a thermodynamic pit. The role of enzymes is to bind and position substrates before the transition state is reached but not so tightly that the ES complex is too stable.

The K_m values (representing dissociation constants) of enzymes for their substrates show that enzymes avoid the thermodynamic pit. Most K_m values are on the order of 10^{-4} M, a number that indicates weak binding of the substrate. Enzymes specific for small substrates, such as urea, carbon dioxide, and superoxide anion, exhibit relatively high K_m values for these compounds (10^{-3} to 10^{-4} M) because these molecules can form few noncovalent bonds with enzymes. Enzymes typically have low K_m values

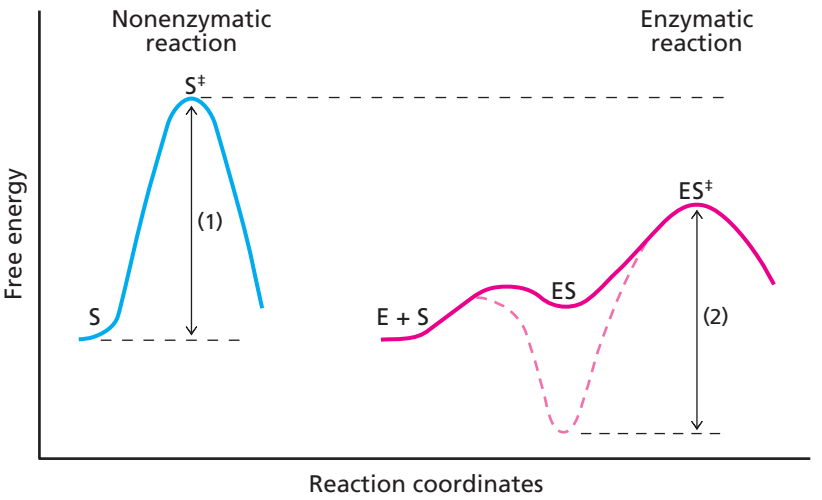


Figure 6.15 ►
Energy of substrate binding. In this hypothetical reaction, the enzyme accelerates the rate of the reaction by stabilizing the transition state. In addition, the activation barrier for formation of the transition state $ES^‡$ from ES must be relatively low. If the enzyme bound the substrate too tightly (dashed profile), the activation barrier (2) would be comparable to the activation barrier of the nonenzymatic reaction (1).

(10^{-6} to 10^{-5} M) for coenzymes, which are bulkier than many substrates. The K_m values for the binding of ATP to most ATP-requiring enzymes are about 10^{-4} M or greater but the muscle-fiber protein myosin (which is not an enzyme) binds ATP a billionfold more avidly. This large difference in binding reflects the fact that in an ES complex not all parts of the substrate are bound.

When the concentration of a substrate inside a cell is below the K_m value of its corresponding enzyme, the equilibrium of the binding reaction E + S $\rightleftharpoons$ ES favors E + S. In other words, the formation of the ES complex is slightly uphill energetically (Figures 6.3 and 6.15), and the ES complex is closer to the energy of the transition state than the ground state is. This weak binding of substrates accelerates reactions. K_m values appear to be optimized by evolution for effective catalysis—low enough that proximity is achieved, but high enough that the ES complex is not too stable. The weak binding of substrates is an important feature of another major force that drives enzymatic catalysis—increased binding of reactants in the $ES^{\ddagger}$ transition state.

The meaning of K_m is discussed in Section 5.3C. In most cases, it represents a good approximation of the *dissociation* constant for the reaction E + S $\rightleftharpoons$ ES. Thus, a K_m of 10^{-4} M means that at equilibrium the concentration of ES will be approximately 10,000-fold higher than the concentration of free substrate.

C. Induced Fit

Enzymes resemble solid catalysts by having limited flexibility but they are not entirely rigid molecules. The atoms of proteins are constantly making small, rapid motions, and small conformational adjustments occur on binding of ligands. An enzyme is most effective if it is in the active form initially so no binding energy is consumed in converting it to an active conformation. In some cases, however, enzymes undergo major shape alterations when substrate molecules bind. The enzyme shifts from an inactive to an active form. Activation of an enzyme by a substrate-initiated conformation change is called **induced fit**. Induced fit is not a catalytic mode but primarily a substrate specificity effect.

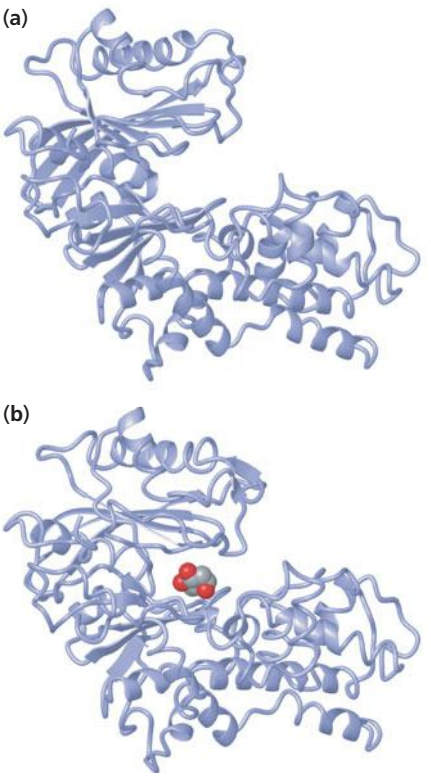
One example of induced fit is seen with hexokinase, an enzyme that catalyzes the phosphorylation of glucose by ATP:

$$\text{Glucose} + \text{ATP} \rightleftharpoons \text{Glucose 6-phosphate} + \text{ADP} \qquad \textbf{(6.22)}$$

Water (HOH), which resembles the alcoholic group at C-6 of glucose (ROH), is small enough and of the proper shape to fit into the active site of hexokinase and therefore it should be a good substrate. If water entered the active site, hexokinase would quickly catalyze the hydrolysis of ATP. However, hexokinase-catalyzed hydrolysis of ATP was shown to be 40,000 times slower than phosphorylation of glucose.

How does the enzyme avoid nonproductive hydrolysis of ATP in the absence of glucose? Structural experiments with hexokinase show that the enzyme exists in two conformations: an open form when glucose is absent, and a closed form when glucose is bound. The angle between the two domains of hexokinase changes considerably when glucose binds, closing the cleft in the enzyme–glucose complex (Figure 6.16). Productive hydrolysis of ATP can only take place in the closed form of the enzyme where the newly formed active site is already occupied by glucose. Water is not a large enough substrate to induce a change in the conformation of hexokinase and this explains why water does not stimulate ATP hydrolysis. Thus, sugar-induced closure of the hexokinase active site prevents wasteful hydrolysis of ATP. A number of other kinases follow induced fit mechanisms.

The substrate specificity that occurs with the induced fit mechanism of hexokinase economizes cellular ATP but exacts a catalytic price. The binding energy consumed in moving the protein molecule into the closed shape—a less-favored conformation—is energy that cannot be used for catalysis. Consequently, an enzyme that uses an induced fit mechanism is less effective as a catalyst than a hypothetical enzyme that is always in an active shape and catalyzes the same reaction. The catalytic cost of induced fit slows kinases so that their k_{cat} values are approximately 10^3 s^{-1} (Table 5.1). We will see another example of induced fit and how it economizes metabolic energy in Section 13.3#1 when we describe citrate synthase. The loop-closing reaction of triose phosphate isomerase is also an example of an induced fit binding mechanism.



▲ **Figure 6.16**
Yeast hexokinase. Yeast hexokinase contains two structural domains connected by a hinge region. On binding of glucose, these domains close, shielding the active site from water. **(a)** Open conformation. **(b)** Closed conformation. [PDB 2YHX and 1HKG].

KEY CONCEPT

Most enzymes exhibit some form of induced fit binding mechanism.

Hexokinase, citrate synthase, and triose phosphate isomerase are extreme examples of induced fit mechanisms. Recent advances in the study of enzyme structures reveal that almost all enzymes undergo some conformational change when substrate binds. The simple concept of a rigid lock and a rigid key is being replaced by a more dynamic interaction where both the "lock" (enzyme) and the "key" (substrate) adjust to each other to form a perfect match.

D. Transition-State Stabilization

Enzymes catalyze reactions by physically or electronically distorting the structures of substrates making them similar to the transition state of the reaction. Transition-state stabilization—the increased interaction of the enzyme with the substrate in the transition state—explains a large part of the rate acceleration of enzymes.

Recall Emil Fischer's lock-and-key theory of enzyme specificity described in Section 5.2B. Fischer proposed that enzymes were rigid templates that accepted only certain substrates as keys. This idea has been replaced by a more dynamic model where both enzyme and substrate change conformations when they interact. Furthermore, the classic lock-and-key model dealt with the interaction between enzyme and substrate but we now think of it in terms of enzyme and transition state—the "key" in the "lock" is the transition state and not the substrate molecule. When a substrate binds to an enzyme the enzyme distorts the structure of the substrate forcing it toward the transition state. Maximal interaction with the substrate molecule occurs only in $ES^{\ddagger}$. A portion of this binding in $ES^{\ddagger}$ can be between the enzyme and nonreacting portions of the substrate.

KEY CONCEPT

The catalytic power of enzymes is explained by binding effects (positioning the substrates together in the correct orientation) and stabilization of the transition state. The result is a lower activation energy and an increased rate of reaction.

An enzyme must be complementary to the transition state in shape and in chemical character. The graph in Figure 6.15 shows that tight binding of the transition state to an enzyme can lower the activation energy. Because the energy difference between E + S and $ES^{\ddagger}$ is significantly less than the energy difference between S and $S^{\ddagger}$, k_{cat} is greater than k_n (the rate constant for the nonenzymatic reaction). The enzyme–substrate transition state ($ES^{\ddagger}$) is lower in absolute energy—and therefore more stable—than the transition state of the reactant in the uncatalyzed reaction. Some transition states may bind to their enzymes 10^{10} to 10^{15} times more tightly than their substrates do. The affinity of other enzymes for their transition states need not be that extreme. A major task for biochemists is to show how transition state stabilization occurs.

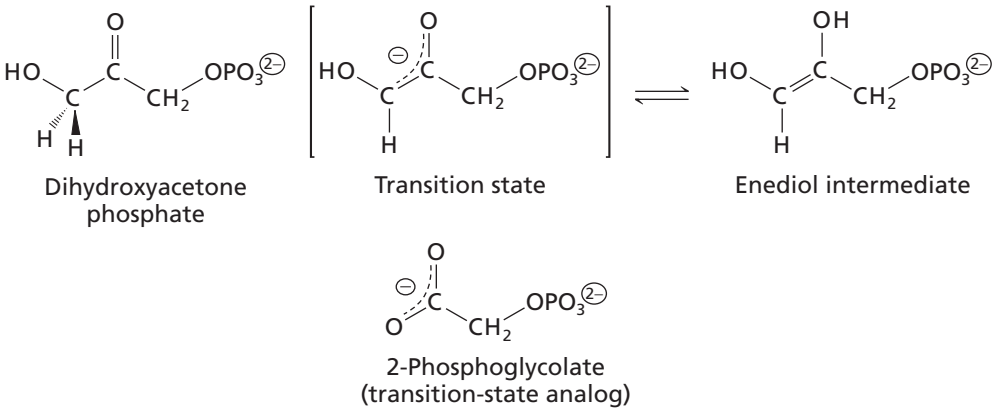
The comparative stabilization of $ES^{\ddagger}$ could occur if an enzyme has an active site with a shape and an electrostatic structure that more closely fits the transition state than the substrate. An undistorted substrate molecule would not be fully bound. For example, an enzyme could have sites that bind the partial charges present only in the unstable transition state.

Transition-state molecules are ephemeral—they have very short half-lives and are difficult to detect. One way in which biochemists can study transition states is to create stable analogs that can bind to the enzyme. These **transition-state analogs** are molecules whose structures resemble presumed transition states. If enzymes prefer to bind to transition states, then a transition-state analog should bind extremely tightly to the appropriate enzyme—much more tightly than substrate—and thus be a potent inhibitor. The dissociation constant for a transition state analog should be about 10^{-13} M or less.

The role of adenosine deaminase is described in Section 18.8.

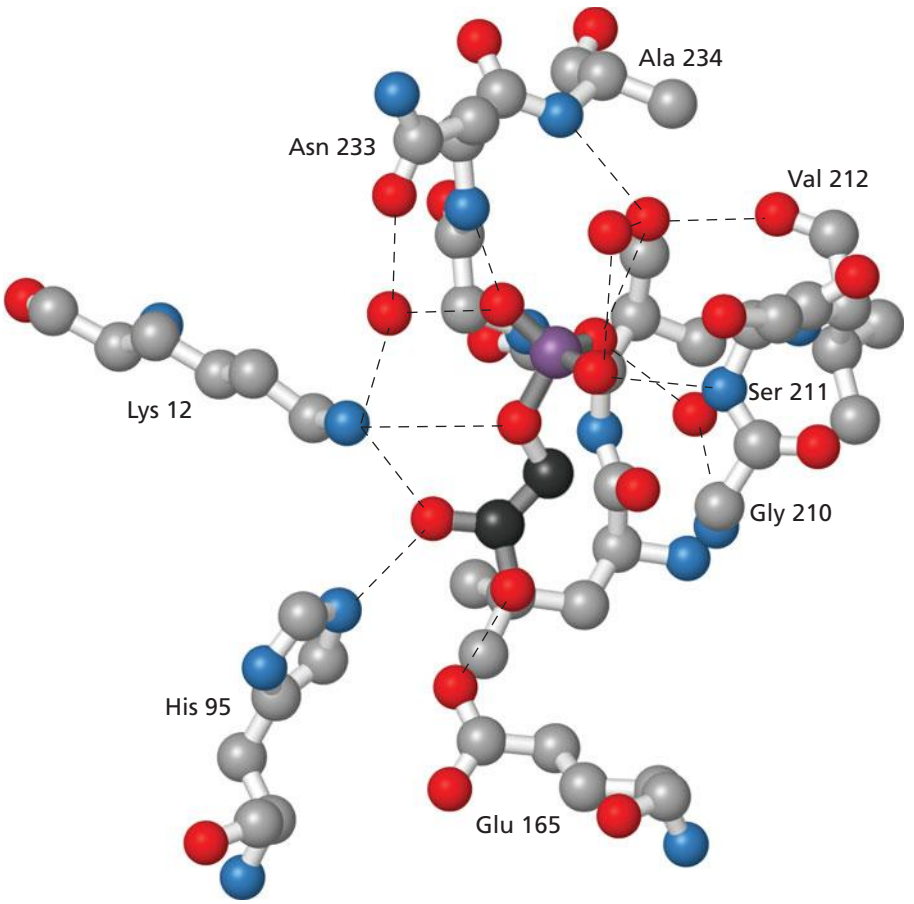
One of the first examples of a transition-state analog was 2-phosphoglycolate (Figure 6.17), whose structure resembles the first enediolate transition state in the reaction catalyzed by triose phosphate isomerase (Section 6.4A). This transition-state analog binds to the isomerase at least 100 times more tightly than either of the substrates of the enzyme (Figure 6.18). Tighter binding results from a partially negative oxygen atom in the carboxylate group of 2-phosphoglycolate, a feature shared with the transition state but not with the substrates.

Experiments with adenosine deaminase have identified a transition-state analog that binds to the enzyme with amazing affinity because it resembles the transition state very closely. Adenosine deaminase catalyzes the hydrolytic conversion of the purine nucleoside adenosine to inosine. The first step of this reaction is the addition of a molecule

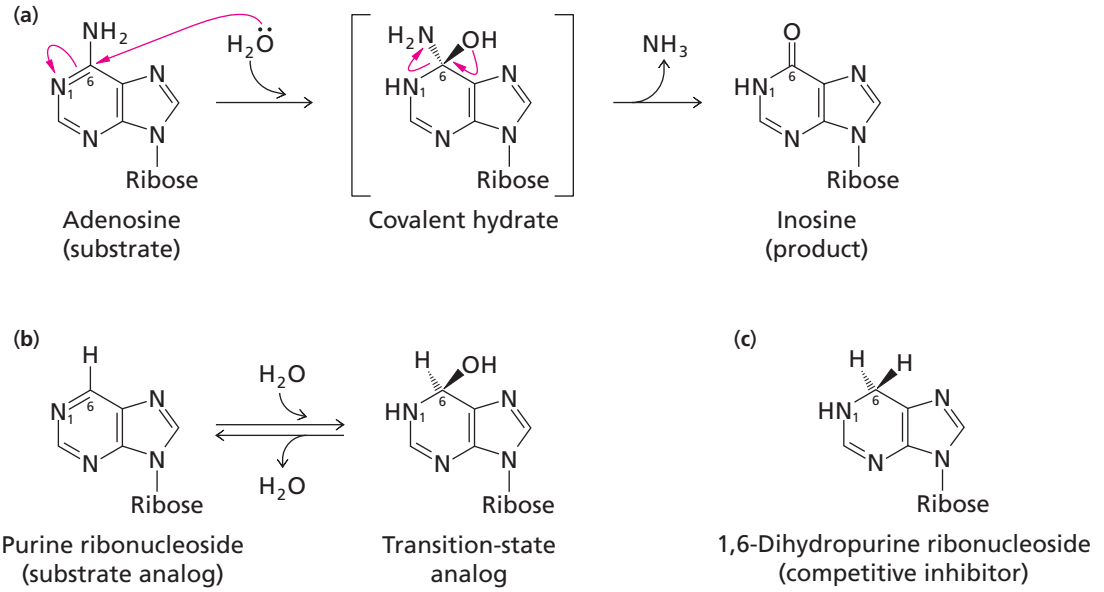


◀ **Figure 6.17**
2-Phosphoglycolate, a transition-state analog for the enzyme triose phosphate isomerase. 2-Phosphoglycolate is presumed to be an analog of C-2 and C-3 of the transition state (center) between dihydroxyacetone phosphate (right) and the initial enediolate intermediate in the reaction.

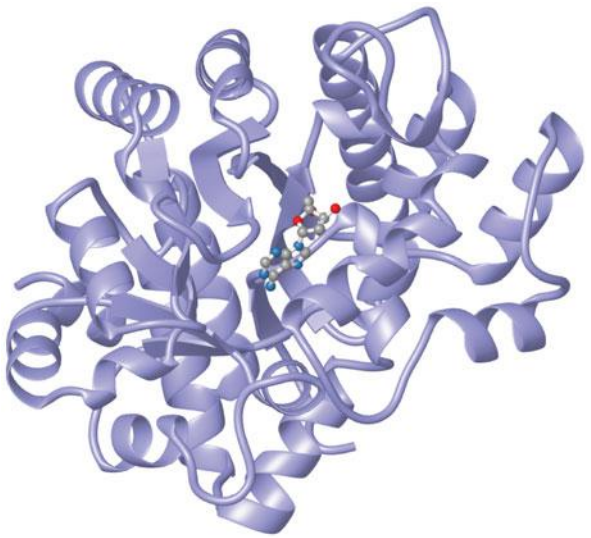
of water (Figure 6.19a). The complex with water, called a covalent hydrate, forms as soon as adenosine is bound to the enzyme and quickly decomposes to products. Adenosine deaminase has broad substrate specificity and catalyzes the hydrolytic removal of various groups from position 6 of purine nucleosides. However, the inhibitor purine ribonucleoside (Figure 6.19b) has just hydrogen at position 6 and undergoes only the first enzymatic step of hydrolysis, addition of the water molecule. The covalent hydrate that's formed is a transition-state analog, a competitive inhibitor having a K_i of 3×10^{-13} M. (For comparison, the affinity constant of adenosine deaminase for its true transition state is expected to be 3×10^{-17} M.). The binding of this analog exceeds the binding of either the substrate or the product by a factor of more than 10^8. A very similar reduced inhibitor, 1,6-dihydropurine ribonucleoside (Figure 6.19c), lacks the hydroxyl group at C-6, and it has a K_i of only 5×10^{-6} M. We can conclude from these studies that adenosine



◀ **Figure 6.18**
Binding of 2-phosphoglycolate to triose phosphate isomerase. The transition state analogue, 2-phosphoglycolate is bound at the active site of *Plasmodium falciparum* triose phosphate isomerase. The molecule is held in position by many hydrogen bonds between the phosphate group and surrounding amino acid side chains. Some of the hydrogen bonds are formed through bridged "frozen" water molecules in the active site. The catalytic residues, Glu-165 and His-95, form hydrogen bonds with the carboxylate group of 2-phosphoglycolate as expected in the transition state. [PDB 1LYZ]



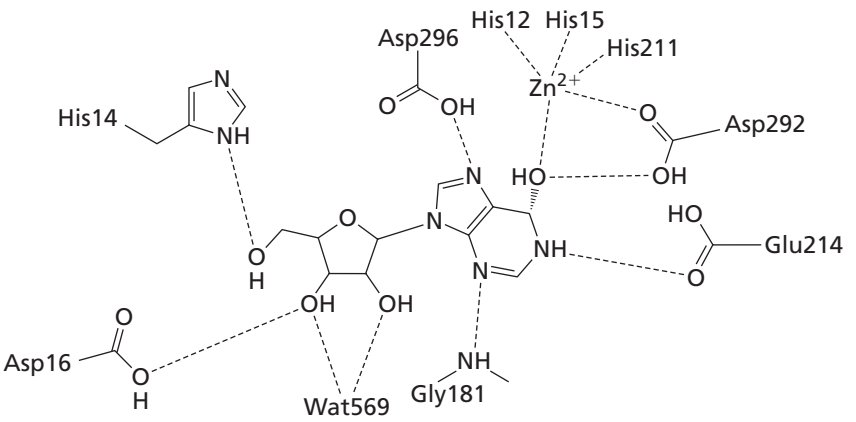
▲ **Figure 6.19**
Inhibition of adenosine deaminase by a transition-state analog. (a) In the deamination of adenosine, a proton is added to N-1 and a hydroxide ion is added to C-6 to form an unstable covalent hydrate, which decomposes to produce inosine and ammonia. **(b)** The inhibitor purine ribonucleoside also rapidly forms a covalent hydrate, 6-hydroxy-1,6-dihydropurine ribonucleoside. This covalent hydrate is a transition-state analog that binds more than a million times more avidly than another competitive inhibitor, 1,6-dihydropurine ribonucleoside **(c)**, which differs from the transition-state analog only by the absence of the 6-hydroxyl group.



▲ **Figure 6.20**
Adenosine deaminase with bound transition-state analog.

deaminase must specifically and avidly bind the transition-state analog—and also the transition state—through interaction with the hydroxyl group at C-6.

The structure of adenosine deaminase with the bound transition-state analog is shown in Figure 6.20 and the interactions between the analog and amino acid side chains in the active site are depicted in Figure 6.21. Notice the hydrogen bonds between Asp-292 and the hydroxyl group on C-6 of 6-hydroxy-1,6-dihydropurine and the interaction between this hydroxyl group and a bound zinc ion in the active site. This confirms the hypothesis that the enzyme specifically binds the transition state in the normal reaction.



▲ **Figure 6.21**
Transition-state analog binding to adenosime deaminase. The interactions between the transition state analog, 6-hydroxy-1,6-dihydropurine, and amino acid side chains in the active site of adenosine deaminase confirms that the enyme recognizes the hydroxyl group at C-6. [PDB 1KRM]

6.6 Serine Proteases

Serine proteases are a class of enzymes that cleave the peptide bond of proteins. As the name implies, they are characterized by the presence of a catalytic serine residue in their active sites. The best-studied serine proteases are the related enzymes trypsin, chymotrypsin, and elastase. These enzymes provide an excellent opportunity to explore the relationship between protein structure and catalytic function. They have been intensively studied for 50 years and form an important part of the history of biochemistry and the elucidation of enzyme mechanisms. In this section, we see how the activity of serine proteases is regulated by zymogen activation and examine a structural basis for the substrate specificity of different serine proteases.

A. Zymogens Are Inactive Enzyme Precursors

Mammals digest food in the stomach and intestines. During this process, food proteins undergo a series of hydrolytic reactions as they pass through the digestive tract. Following mechanical disruption by chewing and moistening with saliva, foods are swallowed and mixed with hydrochloric acid in the stomach. The acid denatures proteins and pepsin (a protease that functions optimally in an acidic environment) catalyzes hydrolysis of these denatured proteins to a mixture of peptides. The mixture passes into the intestine where it is neutralized by sodium bicarbonate and digested by the action of several proteases to amino acids and small peptides that can be absorbed into the bloodstream.

Pepsin is initially secreted as an inactive precursor called pepsinogen. When pepsinogen encounters HCl in the stomach it is activated to cleave itself forming the more active protease, pepsin. The stomach secretions are stimulated by food—or even the anticipation of food—as shown by Ivan Pavlov in his experiments with dogs over 100 years ago. (Pavlov was awarded a Nobel Prize in 1904.) The inactive precursor is called a zymogen. Pavlov was the first to show that zymogens could be converted to active proteases in the stomach and intestines.

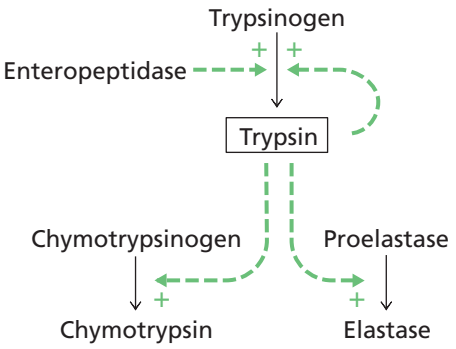
The main serine proteases are trypsin, chymotrypsin, and elastase. Together, they catalyze much of the digestion of proteins in the intestine. Like pepsin, these enzymes are also synthesized and stored as inactive precursors called zymogens. The zymogens, are called trypsinogen, chymotrypsinogen, and proelastase. They are synthesized in the pancreas. It's important to store these hydrolytic enzymes as inactive precursors within the cell since the active proteases would kill the pancreatic cells by cleaving cytoplasmic proteins.

BOX 6.3 KORNBERG'S TEN COMMANDMENTS

1. Rely on enzymology to clarify biologic questions
2. Trust the universality of biochemistry and the power of microbiology
3. Do not believe something because you can explain it
4. Do not waste clean thinking on dirty enzymes
5. Do not waste clean enzymes on dirty substrates
6. Depend on viruses to open windows
7. Correct for extract dilution with molecular crowding
8. Respect the personality of DNA
9. Use reverse genetics and genomics
10. Employ enzymes as unique reagents

Arthur Kornberg, Nobel Laureate in Physiology or Medicine 1959

Kornberg, A. (2000). Ten commandments: lessons from the enzymology of DNA replication. *J. Bacteriol.* 182:3613–3618.
Kornberg, A. (2003). Ten commandments of enzymology, amended. *Trends Biochem. Sci.* 28:515–517.

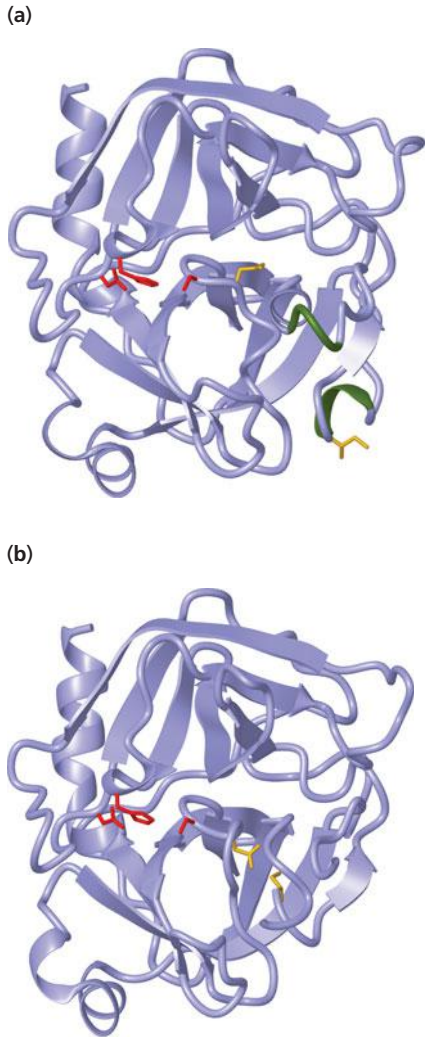


▲ **Figure 6.22**
Activation of some pancreatic zymogens. Initially, enteropeptidase catalyzes the activation of trypsinogen to trypsin. Trypsin then activates chymotrypsinogen, proelastase, and additional trypsinogen molecules.

The enzymes are activated by selective proteolysis—enzymatic cleavage of one or a few specific peptide bonds—when they are secreted from the pancreas into the small intestine. A protease called enteropeptidase specifically activates trypsinogen to trypsin by catalyzing cleavage of the bond between Lys-6 and Ile-7. Once activated by the removal of its N-terminal hexapeptide, trypsin proteolytically cleaves the other pancreatic zymogens, including additional trypsinogen molecules (Figure 6.22).

The activation of chymotrypsinogen to chymotrypsin is catalyzed by trypsin and by chymotrypsin itself. Four peptide bonds (between residues 13 and 14, 15 and 16, 146 and 147, and 148 and 149) are cleaved resulting in the release of two dipeptides. The resulting chymotrypsin retains its three-dimensional shape, despite two breaks in its backbone. This stability is partly due to the presence of five disulfide bonds in the protein.

X-ray crystallography has revealed one major difference between the conformation of chymotrypsinogen and chymotrypsin—the lack of a hydrophobic substrate-binding pocket in the zymogen. The differences are shown in Figure 6.23 where the structures of chymotrypsinogen and chymotrypsin are compared. On zymogen activation, the newly generated α-amino group of Ile-16 turns inward and interacts with the β-carboxyl group of Asp-194 to form an ion pair. This local conformational change generates a relatively hydrophobic substrate-binding pocket near the three catalytic residues with ionizable side chains (Asp-102, His-57, and Ser-195).



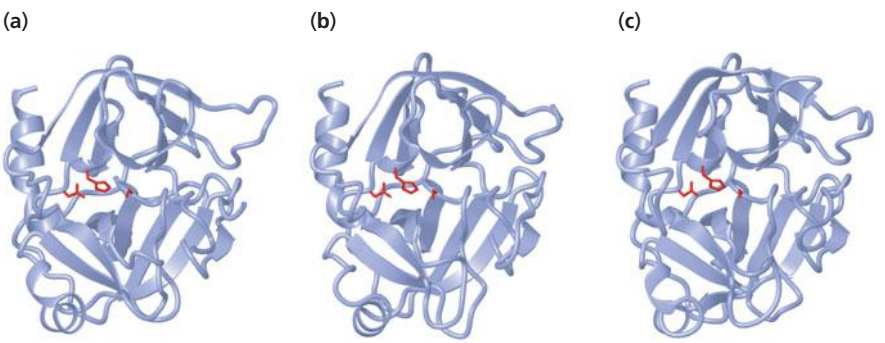
▲ **Figure 6.23**
Polypeptide chains of chymotrypsinogen (left) [PDB 2CGA] and α-chymotrypsin (right) [PDB 5CHA]. Ile-16 and Asp-194 in both zymogen and the active enzyme are shown in yellow. The catalytic-site residues (Asp-102, His-57, and Ser-195) are shown in red. The residues that are removed by processing the zymogen are colored green.

B. Substrate Specificity of Serine Proteases

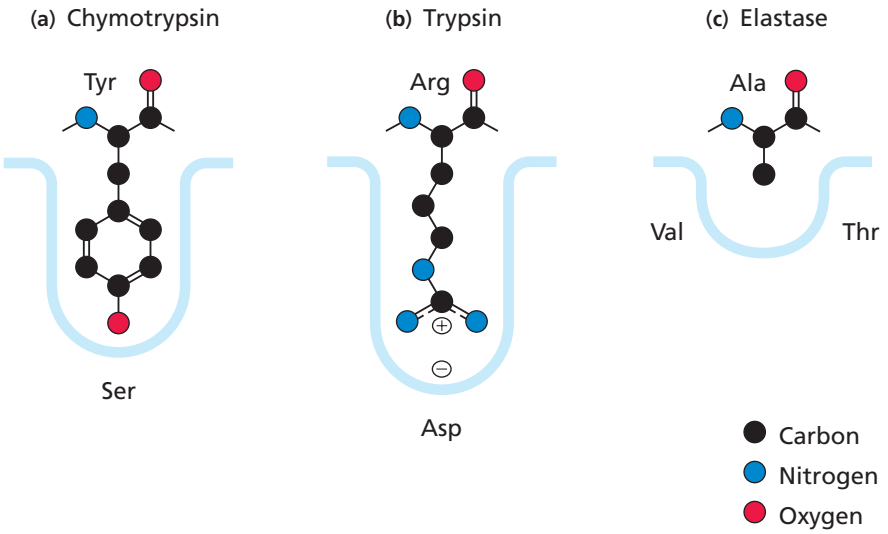
Chymotrypsin, trypsin, and elastase are similar enzymes that share a common ancestor; in other words, they are homologous. Each enzyme has a two-lobed structure with the active site located in a cleft between the two domains. The positions of the catalytically active side chains of the serine, histidine, and aspartate residues in the active sites are almost identical in the three enzymes (Figure 6.24).

The substrate specificities of chymotrypsin, trypsin, and elastase have been explained by relatively small structural differences in the enzymes. Recall that trypsin catalyzes the hydrolysis of peptide bonds whose carbonyl groups are contributed by arginine or lysine (Section 3.10). Both chymotrypsin and trypsin contain a binding pocket that correctly positions the substrates for nucleophilic attack by an active-site serine residue. Each protease has a similar extended region into which polypeptides fit but the so-called specificity pocket near the active-site serine is markedly different for each enzyme. Trypsin differs from chymotrypsin because in chymotrypsin there is an uncharged serine residue at the base of the hydrophobic binding pocket. In trypsin this residue is an aspartate residue (Figure 6.25). This negatively charged aspartate residue forms an ion pair with the positively charged side chain of an arginine or lysine residue of the substrate in the ES complex. Experiments with specifically mutated trypsin indicate that the aspartate residue at the base of its specificity pocket is a major factor in substrate specificity but other parts of the molecule also affect specificity.

Elastase catalyzes the degradation of elastin, a fibrous protein that is rich in glycine and alanine residues. Elastase is similar in tertiary structure to chymotrypsin except that



▲ **Figure 6.24**
Serine proteases. Comparison of the polypeptide backbones of (a) chymotrypsin [PDB 5CHA], (b) trypsin [PDB 1TLD], and (c) elastase [PDB 3EST]. Residues at the catalytic center are shown in red.



◀ **Figure 6.25**
Binding sites of chymotrypsin, trypsin, and elastase. The differing binding sites of these three serine proteases are primary determinants of their substrate specificities. **(a)** Chymotrypsin has a hydrophobic pocket that binds the side chains of aromatic or bulky hydrophobic amino acid residues. **(b)** A negatively charged aspartate residue at the bottom of the binding pocket of trypsin allows trypsin to bind the positively charged side chains of lysine and arginine residues. **(c)** In elastase, the side chains of a valine and a threonine residue at the binding site create a shallow binding pocket. Elastase binds only amino acid residues with small side chains, especially glycine and alanine residues.

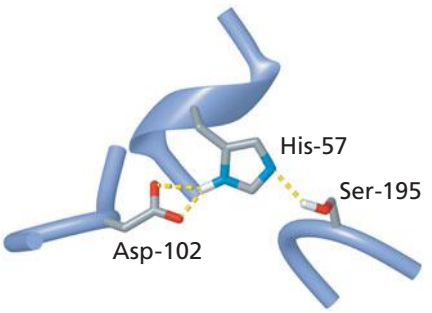
its binding pocket is much shallower. Two glycine residues found at the entrance of the binding site of chymotrypsin and trypsin are replaced in elastase by much larger valine and threonine residues (Figure 6.25c). These residues keep potential substrates with large side chains away from the catalytic center. Thus, elastase specifically cleaves proteins that have small residues such as glycine and alanine.

C. Serine Proteases Use Both the Chemical and the Binding Modes of Catalysis

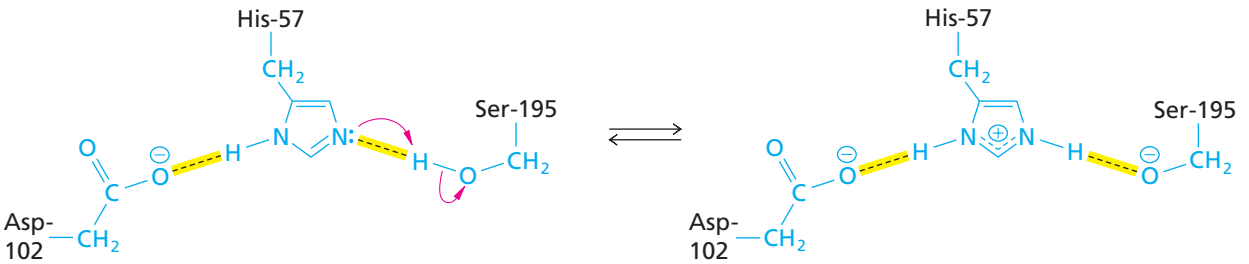
Let's examine the mechanism of chymotrypsin and the roles of three catalytic residues: His-57, Asp-102, and Ser-195. Many enzymes catalyze the cleavage of amide or ester bonds by the same process so study of the chymotrypsin mechanism can be applied to a large family of hydrolases.

Asp-102 is buried in a rather hydrophobic environment. It is hydrogen-bonded to His-57 that in turn is hydrogen-bonded to Ser-195 (Figure 6.26). This group of amino acid residues is called the *catalytic triad.* The reaction cycle begins when His-57 abstracts a proton from Ser-195 (Figure 6.27). This creates a powerful nucleophile (Ser-195) that will eventually attack the peptide bond. Initiation of this part of the reaction is favored because Asp-102 stabilizes the histidine promoting its ability to deprotonate the serine residue.

The discovery that Ser-195 is a catalytic residue of chymotrypsin was surprising because the side chain of serine is usually not sufficiently acidic to undergo deprotonation in order to serve as a strong nucleophile. The hydroxymethyl group of a serine residue generally has a pK_a of about 16 and is similar in reactivity to the hydroxyl group of ethanol. You may recall from organic chemistry that although ethanol can ionize to



▲ **Figure 6.26**
The catalytic site of chymotrypsin. The active-site residues Asp-102, His-57, and Ser-195 are arrayed in a hydrogen-bonded network. The conformation of these three residues is stabilized by a hydrogen bond between the carbonyl oxygen of the carboxylate side chain of Asp-102 and the peptide-bond nitrogen of His-57. Oxygen atoms of the active-site residues are red, and nitrogen atoms are dark blue. [PDB 5CHA].



▲ **Figure 6.27**
Catalytic triad of chymotrypsin. The imidazole ring of His-57 removes the proton from the hydroxymethyl side chain of Ser-195 (to which it is hydrogen-bonded), thereby making Ser-195 a powerful nucleophile. This interaction is facilitated by interaction of the imidazolium ion with its other hydrogen-bonded partner, the buried β-carboxylate group of Asp-102. The residues of the triad are drawn in an arrangement similar to that shown in Figure 6.24.

BOX 6.4 CLEAN CLOTHES

It's a little-known fact that 75% of all laundry detergents contain proteases that are used in helping to remove stubborn protein-based stains from dirty clothes.

All protease additives are based on serine proteases isolated from various *Bacillus* species. These enzymes have been extensively modified in order to be active under the harsh conditions of a detergent solution at high temperature. A successful example of site-directed mutagenesis is the alteration of the serine protease subtilisin from *Bacillus subtilis* (Box 6.4) to make it more resistant to chemical oxidation. It has a methionine residue in the active-site cleft (Met-222) that is readily oxidized leading to inactivation of the enzyme. Resistance to oxidation increases the suitability of subtilisin as a detergent additive. Met-222 was systematically replaced by each of the other common amino acids in a series of mutagenic experiments. All 19 possible mutant subtilisins were isolated and tested and most had greatly diminished peptidase activity. The Cys-222 mutant had high activity but was also subject to oxidation. The Ala-222 and Ser-222 mutants, with nonoxidizable side chains, were not inactivated by oxidation and had relatively high activity. They were the only active, oxygen-stable mutant subtilisin variants.

Site-directed mutagenesis has been performed to alter eight of the 319 amino acid residues of a bacterial protease. The wild-type protease is moderately stable when heated but the suitably mutated enzyme is stable and can function at 100°C. Its denaturation in detergent is prevented by groups, such as a disulfide bridge, that stabilize its conformation.

Recently there has been a trend to lower wash temperatures in order to save energy. The older group of enzymes are not effective at lower wash temperatures so a whole new round of bioengineering has begun creating modified enzymes that can be effective in a modern energy-conscious household.

form an ethoxide this reaction requires the presence of an extremely strong base or treatment with an alkali metal. We see below how the active site of chymotrypsin, achieves this ionization in the presence of a substrate.

A proposed mechanism for chymotrypsin and related serine proteases includes covalent catalysis (by a nucleophilic oxygen) and general acid–base catalysis (donation of a proton to form a leaving group). The steps of the proposed mechanism are illustrated in Figure 6.28.

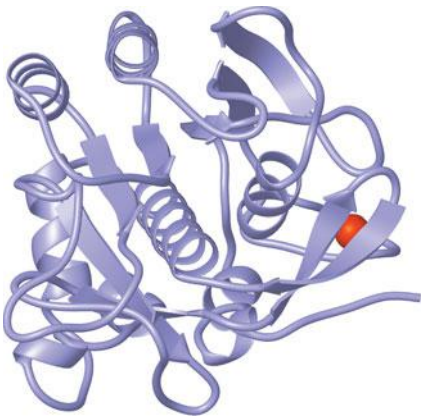
Binding of the peptide substrate causes a slight conformation change in chymotrypsin, sterically compressing Asp-102 and His-57. A low-barrier hydrogen bond is formed between these side chains and the pK_a of His-57 rises from about 7 to about 11. (Formation of this strong, almost covalent, bond drives electrons toward the second N atom of the imidazole ring of His-57 making it more basic.) This increase in basicity makes His-57 an effective general base for abstracting a proton from the $-CH_2OH$ of Ser-195. This mechanism explains how the normally unreactive alcohol group of serine becomes a potent nucleophile.

All the catalytic modes described in this chapter are used in the mechanisms of serine proteases. In the reaction scheme shown in Figure 6.28, steps 1 and 4 in the forward direction use the proximity effect, the gathering of reactants. For example, when a water molecule replaces the amine (P_1) in step 4, it is held by histidine, providing a proximity effect. Acid–base catalysis by histidine lowers the energy barriers for steps 2 and 4. Covalent catalysis using the $-CH_2OH$ of serine occurs in steps 2 through 5. The unstable tetrahedral intermediates at steps 2 and 4 (E-TI_1 and E-TI_2) are believed to be similar to the transition states for these steps. Hydrogen bonds in the oxyanion hole stabilize these intermediates, which are oxyanion forms of the substrate, by binding them more tightly to the enzyme than the substrate was bound. The chemical modes of catalysis (acid–base and covalent catalysis) and the binding modes of catalysis (the proximity effect and transition-state stabilization) all contribute to the enzymatic activity of serine proteases.

BOX 6.5 CONVERGENT EVOLUTION

The protease subtilisin from the bacterium *Bacillus subtilis* is another example of a serine protease. It possesses a catalytic triad consisting of Asp-32, His-64, and Ser-221 at its active site. These are arranged in an alignment similar to the Asp-102, His-57, and Ser-195 residues in chymotrypsin (Figure 6.27). However, as you might deduce from the residue numbers, the structures of subtilisin and chymotrypsin are very different and there is no significant sequence similarity.

This is a remarkable example of convergent evolution. The mammalian intestinal serine proteases and the bacterial subtilisins have independently discovered the catalytic Asp-His-Ser triad.



► **Subtilisin from *Bacillus subtilis*.** The structure of this enzyme is very different from that of serine proteases shown in Figure 6.24. [PDB 1SBC]

6.7 Lysozyme

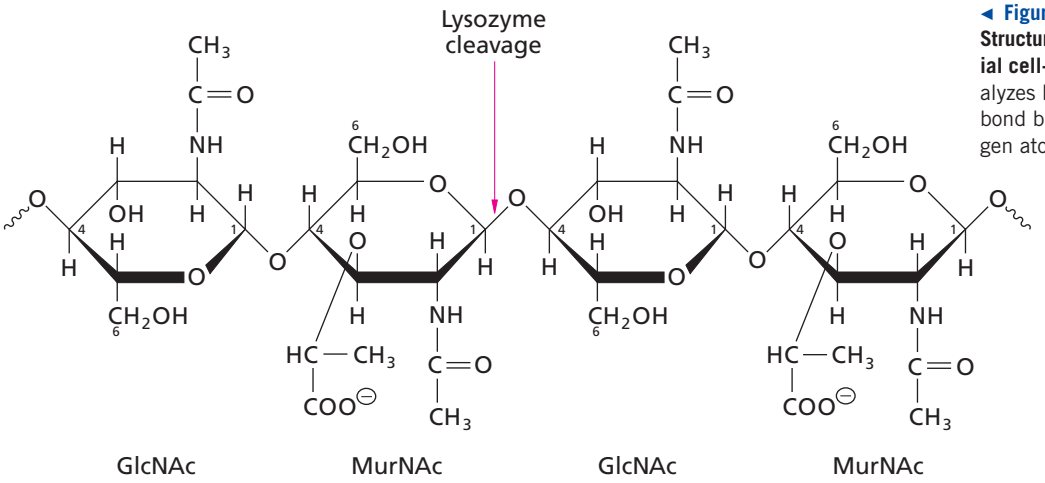
Lysozyme catalyzes the hydrolysis of some polysaccharides, especially those that make up the cell walls of bacteria. It is the first enzyme whose structure was solved and for this reason there has been a long-term interest in working out its precise mechanism of action. Many secretions, such as tears, saliva, and nasal mucus, contain lysozyme activity to help prevent bacterial infection. (Lysozyme causes *lysis*, or disruption, of bacterial cells.) The best-studied lysozyme is from chicken egg white.

The substrate of lysozyme is a polysaccharide composed of alternating residues of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) connected by glycosidic bonds (Figure 6.29). Lysozyme specifically catalyzes hydrolysis of the glycosidic bond between C-1 of a MurNAc residue and the oxygen atom at C-4 of a GlcNAc residue.

The structure of bacterial cell walls is described in Seciton 8.7B.

Models of lysozyme and its complexes with saccharides have been obtained by X-ray crystallographic analysis (Figure 6.30). The substrate-binding cleft of lysozyme accommodates six saccharide residues. Each of the residues binds to a particular part of the active cleft at sites A through E.

Sugar molecules fit easily into all but one site of the structural model. At site D a sugar molecule such as MurNAc does not fit into the model unless it is distorted into a



◄ **Figure 6.29**
Structure of a four-residue portion of a bacterial cell-wall polysaccharide. Lysozyme catalyzes hydrolytic cleavage of the glycosidic bond between C-1 of MurNAc and the oxygen atom involved in the glycosidic bond.

The noncovalent enzyme-substrate complex is formed, orienting the substrate for reaction. Interactions holding the substrate in place include binding of the R_1 group in the specificity pocket (shaded). The binding interactions position the carbonyl carbon of the scissile peptide bond (the bond susceptible to cleavage) next to the oxygen of Ser-195.

E + S

Binding of the substrate compresses Asp-102 and His-57. This strain is relieved by formation of a low-barrier hydrogen bond. The raised pK_a of His-57 enables the imidazole ring to remove a proton from the hydroxyl group of Ser-195. The nucleophilic oxygen of Ser-195 attacks the carbonyl carbon of the peptide bond to form a tetrahedral intermediate (E-TI_1), which is believed to resemble the transition state.

E-S

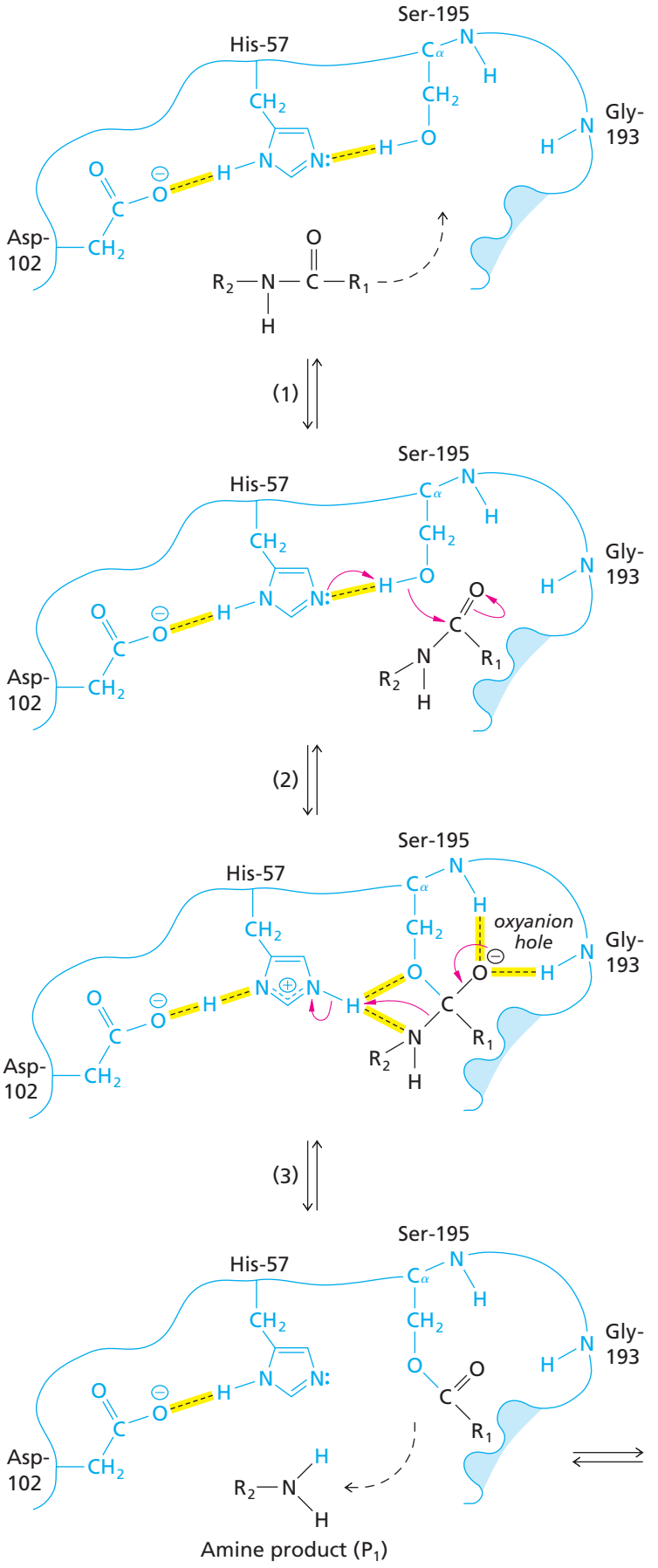
When the tetrahedral intermediate is formed, the substrate C—O bond changes from a double bond to a longer single bond. This allows the negatively charged oxygen (the oxyanion) of the tetrahedral intermediate to move to a previously vacant position, called the oxyanion hole, where it can form hydrogen bonds with the peptide-chain —NH groups of Gly-193 and Ser-195.

The imidazolium ring of His-57 acts as an acid catalyst, donating a proton to the nitrogen of the scissile peptide bond, thus facilitating its cleavage.

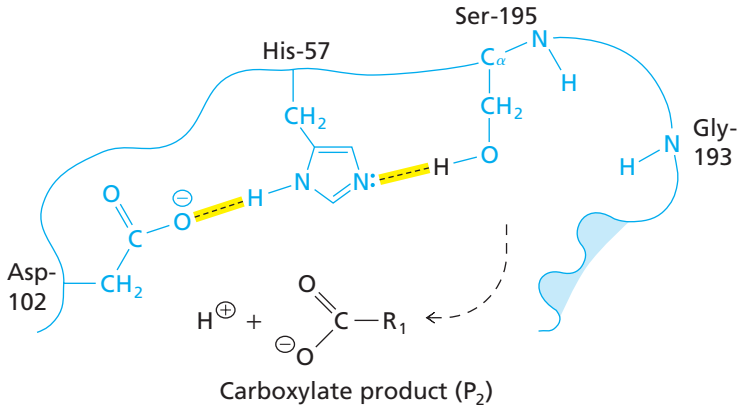
E-TI_1

The carbonyl group from the peptide forms a covalent bond with the enzyme, producing an acyl-enzyme intermediate. After the peptide product (P_1) with the new amino terminus leaves the active site, water enters.

Acyl E + P_1



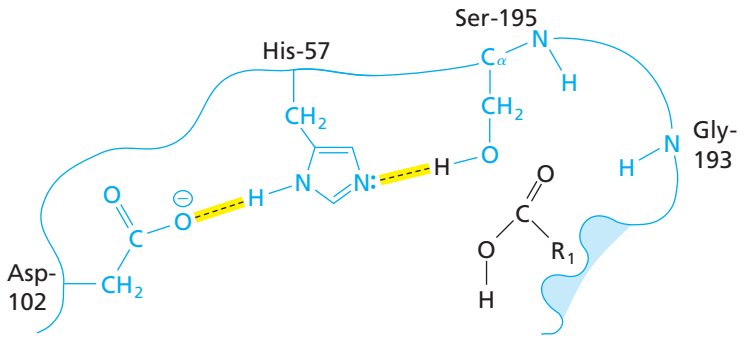
▲ **Figure 6.28**
Mechanism of chymotrypsin-catalyzed cleavage of a peptide bond.



$E + P_2$

The carboxylate product is released from the active site, and free chymotrypsin is regenerated.

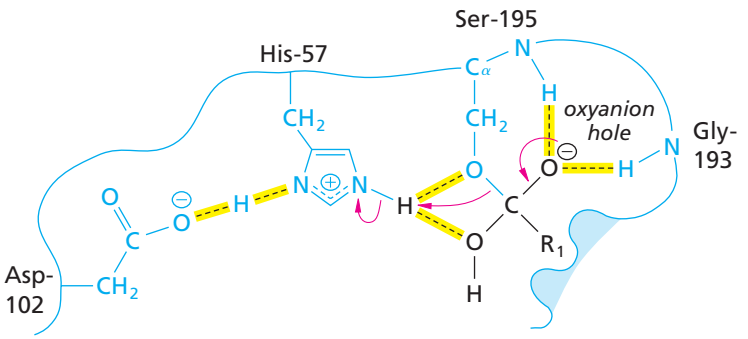
(6)



$E\text{-}P_2$

The second product (P_2)—a polypeptide with a new carboxy terminus—is formed.

(5)

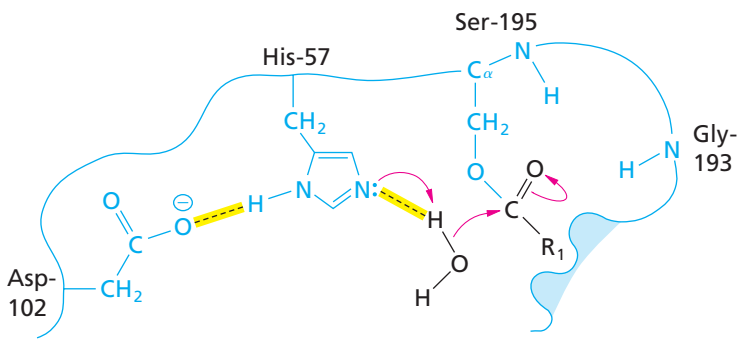


$E\text{-}TI_2$

His-57, once again an imidazolium ion, donates a proton, leading to the collapse of the second tetrahedral intermediate.

A second tetrahedral intermediate ($E\text{-}TI_2$) is formed and stabilized by the oxyanion hole.

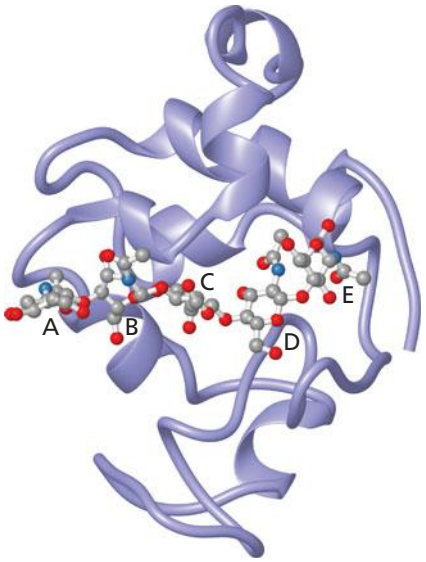
(4)



Acyl E + H_2O

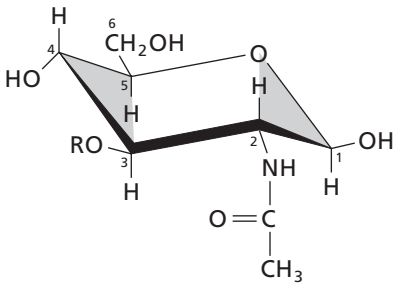
Hydrolysis (deacylation) of the acyl-enzyme intermediate starts when Asp-102 and His-57 again form a low-barrier hydrogen bond and His-57 removes a proton from the water molecule to provide an $OH^{\ominus}$ group to attack the carbonyl group of the ester.

▲ **Figure 6.28** (***continued***)

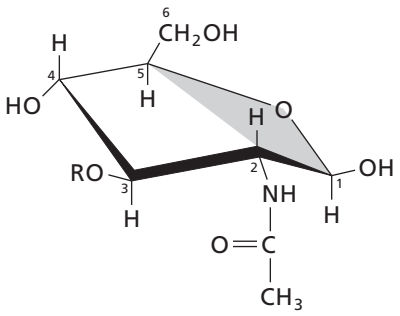


▲ **Figure 6.30**
Lysozyme from chicken with a pentasaccharide molecule (pink). The ligand is bound in sites A, B, C, D and E. Site F is not occupied in this structure. The active site for bond cleavage is between sites D and E. [PDB 1SFB].

(a) Chair conformation



(b) Half-chair conformation



▲ **Figure 6.31**
Conformations of *N*-acetylmuramic acid. **(a)** Chair conformation. **(b)** Half-chair conformation proposed for the sugar bound in site D of lysozyme. R represents the lactyl group of MurNAc.

half-chair conformation (Figure 6.31). Two ionic amino acid residues, Glu-35 and Asp-52, are located close to C-1 of the distorted sugar molecule in the D binding site. Glu-35 is in a nonpolar region of the cleft and has a perturbed pK_a near 6.5. Asp-52, in a more polar environment, has a pK_a near 3.5. The pH optimum of lysozyme is near 5—between these two pK_a values. Recall that the pK_a value of individual amino acid side chains may not be the same as the pK_a value of the free amino acid in solution (Section 3.4).

The proposed mechanism of lysozyme is shown in Figure 6.32. When a molecule of polysaccharide binds to lysozyme, MurNAc residues bind to sites B, D, and F (there is no cavity for the lactyl side chain of MurNAc in site A, C, or E). The extensive binding of the oligosaccharide forces the MurNAc residue in the D site into the half-chair conformation. A near covalent bond forms between Asp-52 and the postulated intermediate (an unstable oxocarbocation). Recent evidence suggests that this interaction might be more like a covalent bond than a strong ion pair but there is much controversy over this point. It's interesting that there are still details of the lysozyme mechanism to be worked out after almost 50 years of effort.

Lysozyme is only one representative of a large group of glycoside hydrolases. Recently, the structures of a bacterial cellulase and its complexes with substrate, intermediate, and product have been determined. This glycosidase has a slightly different mechanism than lysozyme—it forms a covalent glycosyl–enzyme intermediate rather than the strong ion pair postulated for lysozyme. Other aspects of its mechanism, such as distortion of a sugar residue and interaction with active-site —COOH and —$COO^{\ominus}$ side chains, resemble those of the lysozyme mechanism. The structures of the enzyme complexes show that distortion of the substrate forces it toward the transition state.

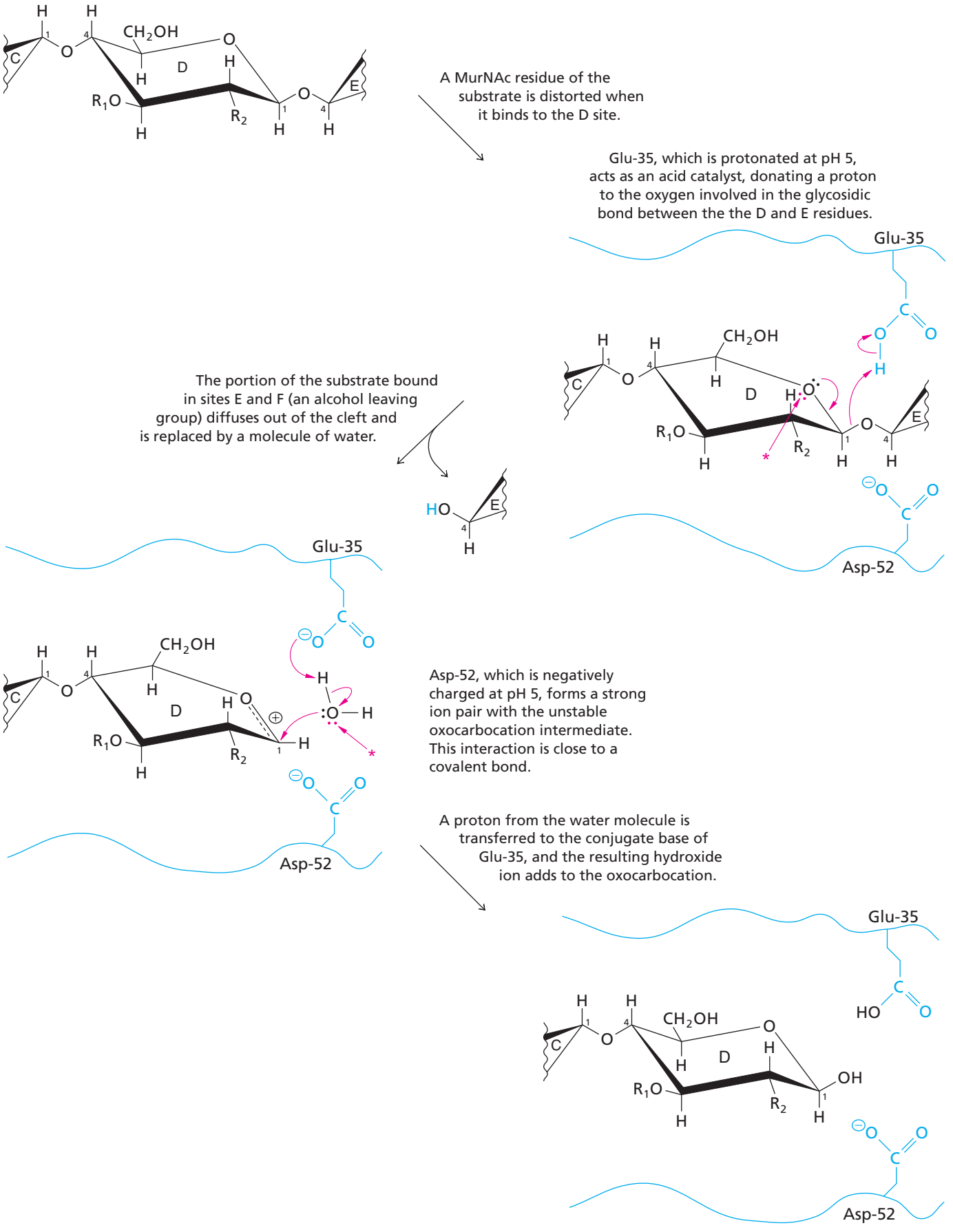
6.8 Arginine Kinase

Most enzymatic reactions for which detailed mechanisms have been elucidated involve fairly simple reactions, such as isomerizations, cleavage reactions, or reactions with water as the second reactant. Therefore, in order to assess proximity effects and the extent of transition state stabilization, it's worthwhile looking at a more complicated reaction, such as that catalyzed by arginine kinase:

$$\text{Arginine} + \text{MgATP} \rightleftharpoons \text{Arginine Phosphate} + \text{MgADP} + H^{\oplus}$$

The structure of a transition-state analog–enzyme complex of arginine kinase has been determined at high resolution (Figure 6.33). However, rather than studying the usual type of transition-state analog in which reactants are fused by covalent bonds, the scientists used three separate components: arginine, nitrate (to model the phosphoryl group transferred between arginine and ADP), and ADP. X-ray crystallographic examination of the active site containing these three compounds led to the proposal of a structure for the transition state and a mechanism for the reaction (see Figure 6.33). The crystallographic results showed that the enzyme has greatly restricted the movement of the bound species (and presumably also of the transition state). For example, the terminal pyrophosphoryl group of ATP is held in place by four arginine side chains and a bound Mg^{2+} ion and the guanidinium group of the arginine substrate molecule is held firmly by two glutamate side chains. The components are precisely and properly aligned by the enzyme.

Arginine kinase, like other kinases, is an induced-fit enzyme (Section 6.5C). It assumes the closed shape when it is crystallized in the presence of arginine, nitrate, and ADP. This enzyme has a k_{cat} of about 2×10^2 s^{-1} and K_m values above 10^{-4} M for both arginine and ATP—values that are quite typical for kinases. The movement that occurs during the induced-fit binding of substrates has precisely aligned the substrates, which had previously been bound fairly weakly, as shown by their moderate K_m values. At least four interrelated catalytic effects participate in this enzymatic reaction: proximity



▲ **Figure 6.32**
Mechanism of lysozyme. R_1 represents the lactyl group, and R_2 represents the *N*-acetyl group of MurNAc.

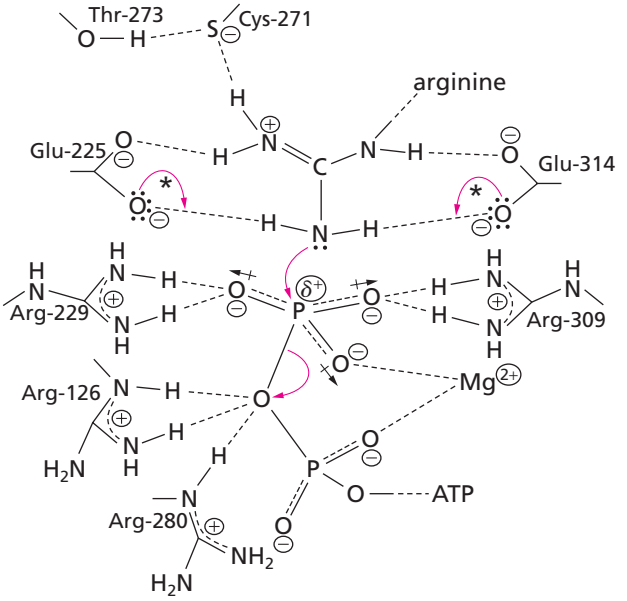


Figure 6.33 ►
Proposed structure of the active site of arginine kinase in the presence of ATP and arginine. The substrate molecules are held firmly and aligned toward the transition state, as shown by the dashed lines. The asterisks (*) show that either Glu-225 or Glu-314 could act as a general acid–base catalyst.

{Adapted from Zhov, G., Somasundaram, T., Blanc, E., Parthasarathy, G., Ellington, W. R., and Chapman, M. S. (1998). Transition state structure of arginine kinase: implications for catalysis of bimolecular reactions. Proc. Natl. Acad. Sci. USA. 95:8453.)

(collection and alignment of substrate molecules), fairly weak initial binding of substrates, acid–base catalysis, and transition-state stabilization (strain of substrates toward the shape of the transition state).

Having gained insight into the general mechanisms of enzymes, we can now go on to examine reactions that include coenzymes. These reactions require groups not supplied by the side chains of amino acids.

Summary

1. The four major modes of enzymatic catalysis are acid–base catalysis and covalent catalysis (chemical modes) and proximity and transition-state stabilization (binding modes). The atomic details of reactions are described by reaction mechanisms, which are based on the analysis of kinetic experiments and protein structures.
2. For each step in a reaction, the reactants pass through a transition state. The energy difference between stable reactants and the transition state is the activation energy. Catalysts allow faster reactions by lowering the activation energy.
3. Ionizable amino acid residues in active sites form catalytic centers. These residues may participate in acid–base catalysis (proton addition or removal) or covalent catalysis (covalent attachment of a portion of the substrate to the enzyme). The effects of pH on the rate of an enzymatic reaction can suggest which residues participate in catalysis.
4. The catalytic rates for a few enzymes are so high that they approach the upper physical limit of reactions in solution, the rate at which reactants approach each other by diffusion.
5. Most of the rate acceleration achieved by an enzyme arises from the binding of substrates to the enzyme.
6. The proximity effect is acceleration of the reaction rate due to the formation of a noncovalent ES complex that collects and orients reactants resulting in a decrease in entropy.
7. An enzyme binds its substrates fairly weakly. Excessively strong binding would stabilize the ES complex and slow the reaction.
8. An enzyme binds a transition state with greater affinity than it binds substrates. Evidence for transition state stabilization is provided by transition-state analogs that are enzyme inhibitors.
9. Some enzymes use induced fit (substrate-induced activation that involves a conformation change) to prevent wasteful hydrolysis of a reactive substrate.
10. Many serine proteases are synthesized as inactive zymogens that are activated extracellularly under appropriate conditions by selective proteolysis. The examination of serine proteases by X-ray crystallography shows how the three-dimensional structures of proteins can reveal information about the active sites, including the binding of specific substrates.
11. The active sites of serine proteases contain a hydrogen-bonded Ser–His–Asp catalytic triad. The serine residue serves as a covalent catalyst, and the histidine residue serves as an acid–base catalyst. Anionic tetrahedral intermediates are stabilized by hydrogen bonds with the enzyme.
12. The proposed mechanism for lysozyme, an enzyme that catalyzes the hydrolysis of bacterial cell walls, includes substrate distortion and stabilization of an unstable oxocarbocation intermediate.

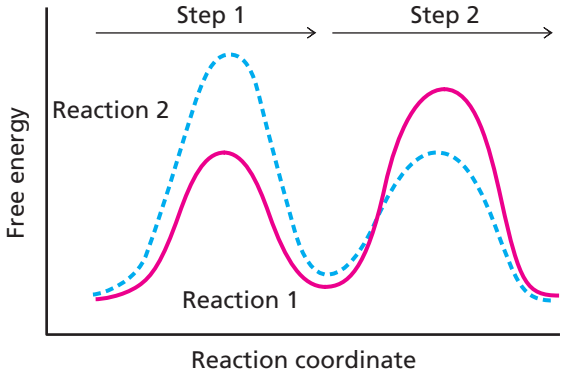
Problems

1. (a) What forces are involved in binding substrates and intermediates to the active sites of enzymes?

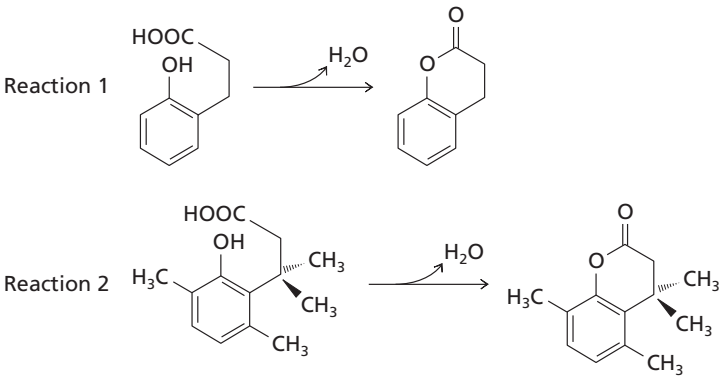
(b) Explain why very tight binding of a substrate to an enzyme is not desirable for enzyme catalysis, whereas tight binding of the transition state is desirable.

2. The enzyme orotodine 5-phosphate decarboxylase is one of the most proficient enzymes known, accelerating the rate of decarboxylation of orotidine 5′ monophosphate by a factor of 10^{23} (Section 5.4). Nitrogen-15 isotope effect studies have shown that two major participating mechanisms are (1) destabilization of the ground state ES complex by electrostatic repulsion between the enzyme and substrate, and (2) stabilization of the transition state by favorable electrostatic interactions between the enzyme and ES‡. Draw an energy diagram that shows how these two effects promote catalysis.

3. The energy diagrams for two multistep reactions are shown below. What is the rate-determining step in each of these reactions?



4. Reaction 2 below occurs 2.5×10^{11} times faster than Reaction 1. What is likely to be a major reason for this enormous rate increase in Reaction 2? How is this model relevant for interpreting possible mechanisms for enzyme rate increases?



5. List three major catalytic effects for lysozyme and explain how each is used during the enzyme-catalyzed hydrolysis of a glycosidic bond.

6. There are multiple serine residues in α-chymotrypsin but only serine 195 reacts rapidly when the enzyme is treated with active phosphate inhibitors such as diisopropyl fluorophosphate (DFP). Explain.

7. (a) Identify the residues in the catalytic triad of α-chymotrypsin and indicate the type of catalysis mediated by each residue.

(b) What additional amino acid groups are found in the oxyanion hole and what role do they play in catalysis?

(c) Explain why site-directed mutagenesis of aspartate to asparagine in the active site of trypsin decreases the catalytic activity 10,000-fold.

8. Catalytic triad groupings of amino acid residues increase the nucleophilic character of active-site serine, threonine, or cysteine residues present in many enzymes involved in catalyzing the cleavage of substrate amide or ester bonds. Using α-chymotrypsin as a model system, diagram the expected arrangements of the catalytic triads in the enzymes below.

(a) Human cytomegalovirus protease: His, His, Ser

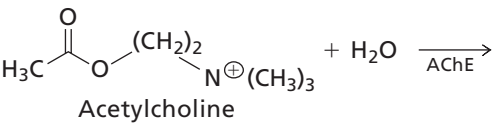
(b) β-lactamase: Glu, Lys, Ser

(c) Asparaginase: Asp, Lys, Thr

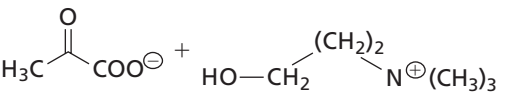
(d) Hepatitis A protease: Asp, (H_2O), His, Cys (a water molecule is situated between the Asp and His residues)

9. Human dipeptidyl peptidase IV (DDP-IV) is a serine protease that catalyzes hydrolysis of prolyl peptide bonds at the next-to-last position at the N terminus of a protein. Many physiological peptides have been identified as substrates, including proteins involved in the regulation of glucose metabolism. DDP-IV contains a catalytic triad at the active site (Glu-His-Ser) and a tyrosine residue in the oxyanion hole. Site-directed mutagenesis of this tyrosine residue in DPP-IV was performed, and the ability of the enzyme to cleave a peptide substrate was compared to that of the wild-type enzyme. The tyrosine residue found in the oxyanion hole was changed to a phenylalanine. The phenylalanine mutant had less than 1% of the activity of the wild-type enzyme (Bjelke, J. R., Christensen, J., Branner, S., Wagtmann, N., Olsen, C. Kanstrup, A. B., and Rasmussen, H. B. (2004). Tyrosine 547 constitutes an essential part of the catalytic mechanism of dipeptidyl peptidase IV. *J. Biol. Chem.* 279:34691–34697). Is this tyrosine required for activity of DDP-IV? Why does the replacement of a tyrosine with a phenylalanine abolish the enzyme activity?

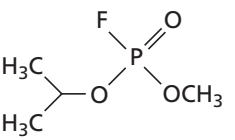
10. Acetylcholinesterase (AChE) catalyzes the breakdown of the neurotransmitter acetylcholine to acetate and choline. This enzyme contains a catalytic triad with the residues His, Glu, and Ser. The catalytic triad enhances the nucleophilicity of the serine residue. The nucleophilic oxygen of serine attacks the carbonyl carbon of acetylcholine to form a tetrahedral intermediate.



Acetylcholine



The nerve agent sarin is an extremely potent inactivator of AChE. Sarin is an irreversible inhibitor that covalently modifies the serine residue in the active site of AChE.

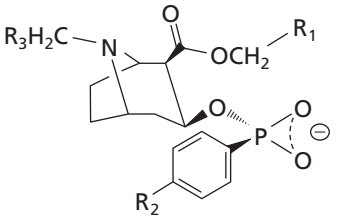


Sarin

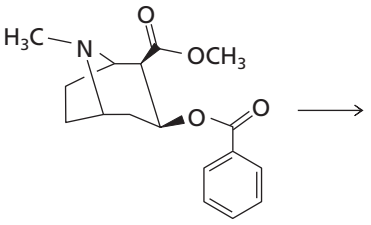
(a) Diagram the expected arrangement of the amino acids in the catalytic triad.

(b) Propose a mechanism for the covalent modification of AChE by sarin.

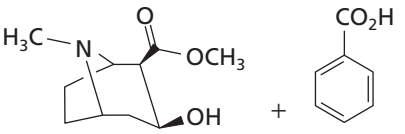
11. Catalytic antibodies are potential therapeutic agents for drug overdose and addiction. For example, a catalytic antibody that catalyzes the breakdown of cocaine before it reached the brain would be an effective detoxification treatment for drug abuse and addiction. The phosphonate analog below was used to raise an anticocaine antibody that catalyzes the rapid hydrolysis of cocaine. Explain why this phosphonate ester was chosen to produce a catalytic antibody.



Phosphonate analog



(−) - Cocaine



Ecgonine methyl ester

Benzoic acid

12. In the chronic lung disease emphysema, the lung's air sacs (alveoli), where oxygen from the air is exchanged for carbon dioxide in the blood, degenerate. α1-Proteinase inhibitor deficiency is a genetic condition that runs in certain families and results from mutations in critical amino acids in the sequence of α1-proteinase inhibitor. The individuals with mutations are more likely to develop emphysema. α1-Proteinase inhibitor is produced by the liver and then circulates in the blood. α1-Proteinase inhibitor is a protein that serves as the major inhibitor of neutrophil elastase, a serine protease present in the lung. Neutrophil elastase cleaves the protein elastin, which is an important component for lung function. The increased rate of elastin breakdown in lung tissue is believed to cause emphysema. One treatment for α1-proteinase inhibitor deficiency is to give the patient human wild-type α1-proteinase inhibitor (derived from large pools of human plasma) intravenously by injecting the protein directly into the bloodstream.

(a) Explain the rational for the treatment with wild-type α1-proteinase inhibitor.

(b) This treatment involves the intravenous administration of the wild-type α1-proteinase inhibitor. Explain why α1-proteinase inhibitor cannot be taken orally.

Selected Readings

General

Fersht, A. (1985). *Enzyme Structure and Mechanism*, 2nd ed. (New York: W. H. Freeman).

Binding and Catalysis

Bartlett, G. J., Porter, C. T., Borkakoti, N. and Thornton, J. M. (2002). Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* 324:105–121.

Bruice, T. C. and Pandrit, U. K. (1960). Intramolecular models depicting the kinetic importance of "fit" in enzymatic catalysis. *Proc. Natl. Acad. Sci. USA.* 46:402–404.

Hackney, D. D. (1990). Binding energy and catalysis. In *The Enzymes*, Vol. 19, 3rd ed., D. S. Sigman and P. D. Boyer, eds. (San Diego: Academic Press), pp. 1–36.

Jencks, W. P. (1987). Economics of enzyme catalysis. *Cold Spring Harbor Symp. Quant. Biol.* 52:65–73.

Kraut, J. (1988). How do enzymes work? *Science* 242:533–540.

Neet, K. E. (1998). Enzyme catalytic power mini-review series. *J. Biol. Chem.* 273:25527–25528, and related papers on pages 25529–25532, 26257–26260, and 27035–27038.

Pauling, L. (1948) Nature of forces between large molecules of biological interest. *Nature* 161:707–709.

Schiøtt, B., Iversen, B. B., Madsen, G. K. H., Larsen, F. K., and Bruice, T. C. (1998). On the electronic nature of low-barrier hydrogen bonds in enzymatic reactions. *Proc. Natl. Acad. Sci. USA* 95:12799–12802.

Shan, S.-U., and Herschlag, D. (1996). The change in hydrogen bond strength accompanying charge rearrangement: implications for enzymatic catalysis. *Proc. Natl. Acad. Sci. USA* 93:14474–14479.

Transition-State Analogs

Schramm, V. L. (1998). Enzymatic transition states and transition state analog design. *Annu. Rev. Biochem.* 67:693–720.

Wolfenden, R., and Radzicka, A. (1991). Transition-state analogues. *Curr. Opin. Struct. Biol.* 1:780–787.

Specific Enzymes

Cassidy, C. S., Lin, J., and Frey, P. A. (1997). A new concept for the mechanism of action of chymotypsin: the role of the low-barrier hydrogen bond. *Biochem.* 36:4576–4584.

Blacklow, S. C., Raines, R. T., Lim, W. A., Zamore, P. D., and Lnowles, J. R. (1988). Triosephosphate isomerase catalysis is diffusion controlled. *Biochem.* 27:1158–1167.

Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schülein, M., and Withers, S. G. (1998). Snapshots along an enzymatic reaction coordinate: analysis of a retaining β-glycoside hydrolase. *Biochem.* 37:11707–11713.

Dodson, G., and Wlodawer, A. (1998). Catalytic triads and their relatives. *Trends Biochem. Sci.* 23:347–352.

Frey, P. A., Whitt, S. A., and Tobin, J. B. (1994). A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science.* 264:1927–1930.

Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L., and Hallewell, R. A. (1992). Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature.* 358:347–351.

Harris, T. K., Abeygunawardana, C., and Mildvan, A. S. (1997). NMR studies of the role of hydrogen bonding in the mechanism of triosephosphate isomerase. *Biochem.* 36:14661–14675.

Huber, R., and Bode, W. (1978). Structural basis of the activation and action of trypsin. *Acc. Chem. Res.* 11:114–122.

Kinoshita, T., Nishio, N., Nakanishi, I., Sato, A., and Fujii, T. (2003). Structure of bovine adenosine deaminase complexed with 6-hydroxy-1,6-dihydropurine riboside. *Acta Cryst.* D59:299–303.

Kirby, A. J. (2001). The lysozyme mechanism sorted—after 50 years. *Nature Struct. Biol.* 8:737–739.

Knolwes, J. R. (1991) Enzyme catalysis: not different, just better. *Nature.* 350:121–124.

Knowles, J. R., and Albery, W. J. (1977). Perfection in enzyme catalysis: the energetics of triosephosphate isomerase. *Acc. Chem. Res.* 10:105–111.

Kuser, P., Cupri, F., Bleicher, L., and Polikarpov, I. (2008). Crystal structure of yeast hexokinase P1 in complex with glucose: a classical "induced fit" example revisited. *Proteins.* 72:731–740.

Lin, J., Cassidy, C. S., and Frey, P. A. (1998). Correlations of the basicity of His-57 with transition state analogue binding, substrate reactivity, and the strength of the low-barrier hydrogen bond in chymotrypsin. *Biochem.* 37:11940–11948.

Lodi, P. J., and Knowles, J. R. (1991). Neutral imidazole is the electrophile in the reaction catalyzed by triosephosphate isomerase: structural origins and catalytic implications. *Biochem.* 30:6948–6956.

Parthasarathy, S., Ravinda, G., Balaram, H., Balaram, P., and Murthy, M. R. N. (2002). Structure of the plasmodium falciparum triosephosphate isomerase—phosphoglycolate complex in two crystal forms: characterization of catalytic open and closed conformations in the ligand-bound state. *Biochem.* 41:13178–13188.

Paetzel, M., and Dalbey, R. E. (1997). Catalytic hydroxyl/amine dyads within serine proteases. *Trends Biochem. Sci.* 22:28–31.

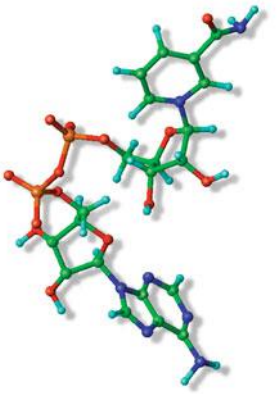
Perona, J. J., and Craik, C. S. (1997). Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J. Biol. Chem.* 272:29987–29990.

Schäfer T., Borchert T. W., Nielsen V. S., Skagerlind P., Gibson K., Wenger K., Hatzack F., Nilsson L. D., Salmon S., Pedersen S., Heldt-Hansen H. P., Poulsen P. B., Lund H., Oxenbøll K. M., Wu, G. F., Pedersen H. H., Xu, H. (2007). Industrial enzymes. *Adv. Biochem. Eng. Biotechnol. 2007* 105:59–131.

Steitz, T. A., and Shulman, R. G. (1982). Crystallographic and NMR studies of the serine proteases. *Annu. Rev. Biophys. Bioeng.* 11:419–444.

Von Dreele, R. B. (2005). Binding of *N*-acetylglucosamine oligosaccharides to hen egg-white lysozyme: a powder diffraction study. *Acta Crystallographic.* D61:22–32.

Zhou, G., Somasundaram, T., Blanc, E., Parthasarathy, G., Ellington, W. R., and Chapman, M. S. (1998). Transition state structure of arginine kinase: implications for catalysis of bimolecular reactions. *Proc. Natl. Acad. Sci. USA* 95:8449–8454.



Coenzymes and Vitamins

Evolution has produced a spectacular array of protein catalysts but the catalytic repertoire of an organism is not limited by the reactivity of amino acid side chains. Other chemical species, called **cofactors**, often participate in catalysis. Cofactors are required by inactive apoenzymes (proteins only) to convert them to active holoenzymes. There are two types of cofactors: **essential ions** (mostly metal ions) and organic compounds known as **coenzymes** (Figure 7.1). Both inorganic and organic cofactors become essential portions of the active sites of certain enzymes.

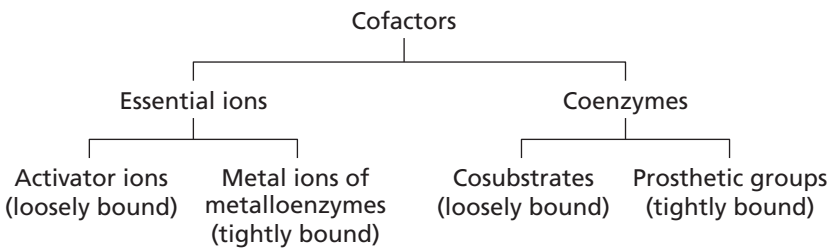
Many of the minerals required by all organisms are essential because they are cofactors. Some essential ions, called *activator ions*, are reversibly bound and often participate in the binding of substrates. In contrast, some cations are tightly bound and frequently participate directly in catalytic reactions.

Coenzymes act as group-transfer reagents. They accept and donate specific chemical groups. For some coenzymes, the group is simply hydrogen or an electron but other coenzymes carry larger, covalently attached chemical groups. These mobile metabolic groups are attached at the **reactive center** of the coenzyme. (Either the mobile metabolic group or the reactive center is shown in red in the structures presented in this chapter.) We can simplify our study of coenzymes by focusing on the chemical properties of their reactive centers. The two classes of coenzymes are described in Section 7.2.

We begin this chapter with a discussion of essentialion cofactors. Much of the rest of the chapter is devoted to the more complex organic cofactors. In mammals, many of these coenzymes are derived from dietary precursors called vitamins. We therefore discuss vitamins in this chapter. We conclude with a look at a few proteins that are coenzymes. Most of the structures and reactions presented here will be encountered in later chapters when we discuss particular metabolic pathways.

Finally, we come to a group of compounds which have only been known for a relatively short time, but which during this short time have attracted very considerable attention, both from chemists and from the public at large. Who today is unacquainted with vitamins, these mysterious substances which are of such immense significance for life, vita, itself and which have thus justifiably taken their name from it?

—H.G. Söderbaum Presentation speech for the Nobel Prize in chemistry to Adolf Windaus, 1928



◀ **Figure 7.1**
Types of cofactors. Essential ions and coenzymes can be further distinguished by the strength of interaction with their apoenzymes.

Top: Nicotinamide adenine dinucleotide ($NAD^{\oplus}$), a coenzyme derived from the vitamin nicotinic acid (niacin). $NAD^{\oplus}$ is an oxidizing agent.

7.1 Many Enzymes Require Inorganic Cations

Over a quarter of all known enzymes require metallic cations to achieve full catalytic activity. These enzymes can be divided into two groups: metal-activated enzymes and metalloenzymes. **Metal-activated enzymes** either have an absolute requirement for added metal ions or are stimulated by the addition of metal ions. Some of these enzymes require monovalent cations such as $K^{\oplus}$ and others require divalent cations such as $Ca^{2\oplus}$ or $Mg^{2\oplus}$. Kinases, for example, require magnesium ions for the magnesium-ATP complex they use as a phosphoryl group donating substrate. Magnesium shields the negatively charged phosphate groups of ATP making them more susceptible to nucleophilic attack (Section 10.6).

Metalloenzymes contain firmly bound metal ions at their active sites. The ions most commonly found in metalloenzymes are the transition metals, iron and zinc, and less often, copper and cobalt. Metal ions that bind tightly to enzymes are usually required for catalysis. The cations of some metalloenzymes can act as electrophilic catalysts by polarizing bonds. For example, the cofactor for the enzyme carbonic anhydrase is an electrophilic zinc atom bound to the side chains of three histidine residues and to a molecule of water. Binding to $Zn^{2\oplus}$ causes the water to ionize more readily. A basic carboxylate group of the enzyme removes a proton from the bound water molecule, producing a nucleophilic hydroxide ion that attacks the substrate (Figure 7.2). This enzyme has a very high catalytic rate partly because of the simplicity of its mechanism (Section 6.4). Many other zinc metalloenzymes activate bound water molecules in this fashion.

The ions of other metalloenzymes can undergo reversible oxidation and reduction by transferring electrons from a reduced substrate to an oxidized substrate. For example, iron is part of the heme group of catalase, an enzyme that catalyzes the degradation of H_2O_2. Similar heme groups also occur in cytochromes, electron-transferring proteins found associated with specific metalloenzymes in mitochondria and chloroplasts. Nonheme iron is often found in metalloenzymes in the form of iron-sulfur clusters (Figure 7.3). The most common iron-sulfur clusters are the [2 Fe–2 S] and [4 Fe–4 S] clusters in which the iron atoms are complexed with an equal number of sulfide ions from H_2S and $—S^{\ominus}$ groups from cysteine residues. Iron-sulfur clusters mediate some oxidation-reduction reactions. Each cluster, whether it contains two or four iron atoms, can accept only one electron in an oxidation reaction.

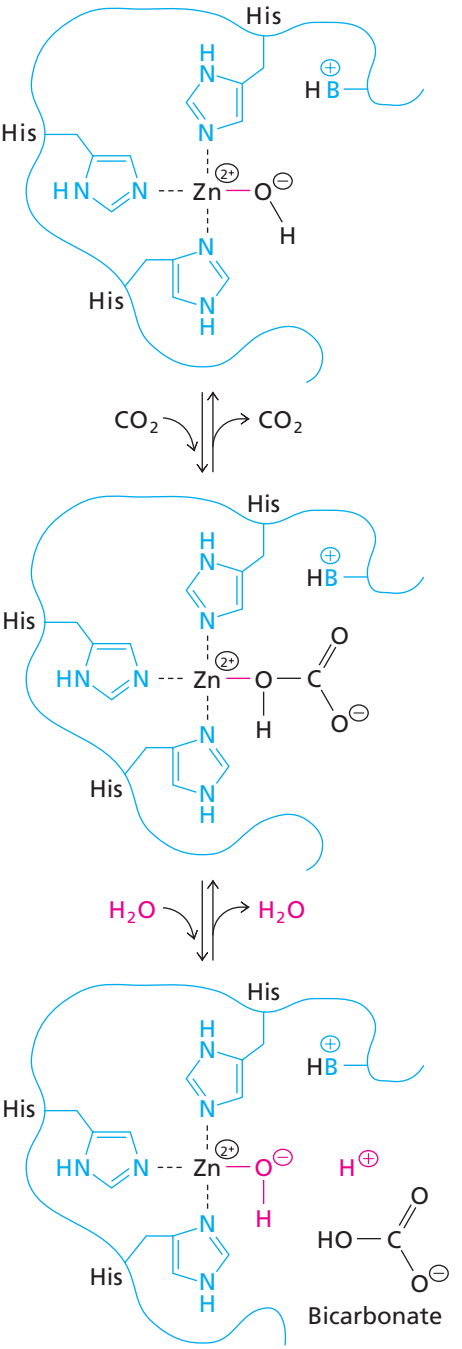
7.2 Coenzyme Classification

Coenzymes can be classified into two types based on how they interact with the apoenzyme (Figure 7.1). Coenzymes of one type—often called **cosubstrates**—are actually substrates in enzyme-catalyzed reactions. A cosubstrate is altered in the course of the reaction and dissociates from the active site. The original structure of the cosubstrate is regenerated in a subsequent reaction catalyzed by another enzyme. The cosubstrate is recycled repeatedly within the cell, unlike an ordinary substrate whose product typically undergoes further transformation. Cosubstrates shuttle mobile metabolic groups among different enzyme-catalyzed reactions.

The second type of coenzyme is called a **prosthetic group**. A prosthetic group remains bound to the enzyme during the course of the reaction. In some cases the prosthetic group is covalently attached to its apoenzyme, while in other cases it is tightly bound to the active site by many weak interactions. Like the ionic amino acid residues of the active site, a prosthetic group must return to its original form during each full catalytic event or the holoenzyme will not remain catalytically active. Cosubstrates and prosthetic groups are part of the active site of enzymes. They supply reactive groups that are not available on the side chains of amino acid residues.

Every living species uses coenzymes in a diverse number of important enzyme-catalyzed reactions. Most of these species are capable of synthesizing their coenzymes from simple precursors. This is especially true in four of the five kingdoms—prokaryotes, protists, fungi, and plants—but animals have lost the ability to synthesize some

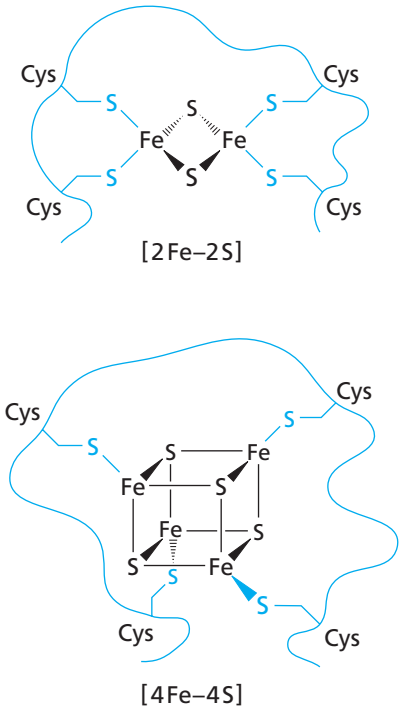
Refer to Figure 1.1 for a table of the essential elements.



▲ **Figure 7.2**
Mechanism of carbonic anhydrase. The zinc ion in the active site promotes the ionization of a bound water molecule. The resulting hydroxide ion attacks the carbon atom of carbon dioxide, producing bicarbonate, which is released from the enzyme.

Review Section 4.12 for the structure of heme.

Cytochromes will be discussed in Section 7.16.



▲ **Figure 7.3**
Iron-sulfur clusters. In each type of iron-sulfur cluster, the iron atoms are complexed with an equal number of sulfide ions (S^{2-}) and with the thiolate groups of the side chains of cysteine residues.

Table 7.1 Some vitamins and their associated deficiency diseases

Vitamin	Disease
Ascorbate (C)	Scurvy
Thiamine (B_1)	Beriberi
Riboflavin (B_2)	Growth retardation
Nicotinic acid (B_3)	Pellagra
Pantothenate (B_5)	Dermatitis in chickens
Pyridoxal (B_6)	Dermatitis in rats
Biotin (B_7)	Dermatitis in humans
Folate (B_9)	Anemia
Cobalamin (B_{12})	Pernicious anemia

The structure and chemistry of nucleotides is discussed in more detail in Chapter 19.

◀ **Brown rice and white rice.** Brown rice (top left) has been processed to remove the outer husks but it retains part of the outer skin or "bran." This skin contains thiamine (vitamin B_1). Further processing of the grain yields white rice (middle left), which lacks thiamine.

coenzymes. Mammals (including humans) need a source of coenzymes in order to survive. The ones they can't synthesize are supplied by nutrients, usually in small amounts (micrograms or milligrams per day). These essential compounds are called **vitamins** and animals rely on other organisms to supply these micronutrients. The ultimate sources of vitamins are usually plants and microorganisms. Most vitamins are coenzyme precursors—they must be enzymatically transformed to their corresponding coenzymes.

A vitamin-deficiency disease can result when a vitamin is deficient or absent in the diet of an animal. Such diseases can be overcome or prevented by consuming the appropriate vitamin. Table 7.1 lists nine vitamins and the diseases associated with their deficiencies. Each of these vitamins and their metabolic roles are discussed below. Most of them are converted to coenzymes, sometimes after a reaction with ATP.

The word *vitamin* (originally spelled "vitamine") was coined by Casimir Funk in 1912 to describe a "vital amine" from brown rice that cured beriberi, a nutritional-deficiency disease that results in neural degeneration. The term *vitamin* has been retained even though many vitamins proved not to be amines. Beriberi was first described in birds and then in humans whose diets consisted largely of polished rice. Christiaan Eijkman, a Dutch physician working in what was then the Dutch East Indies (now Indonesia), was the first to notice that chickens fed polished rice leftover from the local hospital developed beriberi but they recovered when they were fed brown rice. This discovery led eventually to isolation of an antiberiberi substance from the skin that covers brown rice. This substance became known as vitamin B_1 (thiamine).

Two broad classes of vitamins have since been identified: water-soluble (such as B vitamins) and fat-soluble (also called lipid vitamins). Water-soluble vitamins are required daily in small amounts because they are readily excreted in the urine and the cellular stores of their coenzymes are not stable. Conversely, lipid vitamins such as vitamins A, D, E, and K, are stored by animals and excessive intakes can result in toxic conditions known as hypervitaminoses. It's important to note that not all vitamins are coenzymes or their precursors (see Box 7.4 and Section 7.14).

The most common coenzymes are listed in Table 7.2 along with their metabolic role and their vitamin source. The following sections describe the structures and functions of these common coenzymes.

7.3 ATP and Other Nucleotide Cosubstrates

A number of nucleosides and nucleotides are coenzymes. Adenosine triphosphate (ATP) is by far the most abundant. Other common examples are GTP, *S*-adenosylmethionine, and nucleotide sugars such as uridine diphosphate glucose (UDP-glucose). ATP (Figure 7.4) is a versatile reactant that can donate its phosphoryl, pyrophosphoryl, adenylyl (AMP), or adenosyl groups in group-transfer reactions.

The most common reaction involving ATP is phosphoryl group transfer. In reactions catalyzed by kinases, for example, the γ-phosphoryl group of ATP is transferred to a nucleophile leaving ADP. The second most common reaction is nucleotidyl group transfer (transfer of the AMP moiety) leaving pyrophosphate (PP_i). ATP plays a central role in metabolism. Its role as a "high energy" cofactor is described in more detail in Chapter 10, "Introduction to Metabolism."

ATP is also the source of several other metabolite coenzymes. One, *S*-adenosylmethionine (Figure 7.5), is synthesized by the reaction of methionine with ATP.

$$\text{Methionine} + \text{ATP} \longrightarrow \text{S-Adenosylmethionine} + P_i + PP_i \qquad (7.1)$$

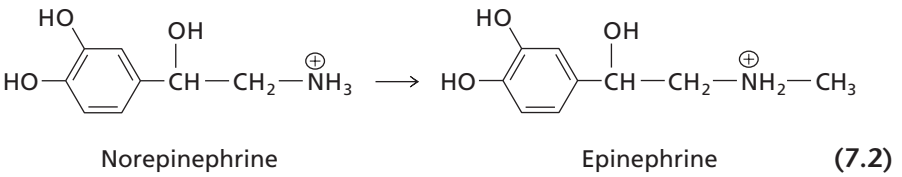
The normal thiomethyl group of methionine (—S—CH_3) is not very reactive but the positively charged sulfonium of *S*-adenosylmethionine is highly reactive. *S*-adenosylmethionine

Table 7.2 Major coenzymes

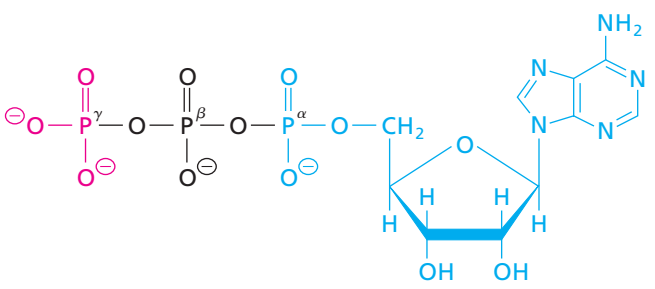
Coenzyme	Vitamin source	Major metabolic roles	Mechanistic role
Adenosine triphosphate (ATP)	—	Transfer of phosphoryl or nucleotidyl groups	Cosubstrate
S-Adenosylmethionine	—	Transfer of methyl groups	Cosubstrate
Uridine diphosphate glucose	—	Transfer of glycosyl groups	Cosubstrate
Nicotinamide adenine dinucleotide ($NAD^{\oplus}$) and nicotinamide adenine dinucleotide phosphate ($NADP^{\oplus}$)	Niacin (B_3)	Oxidation-reduction reactions involving two-electron transfer	Cosubstrate
Flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD)	Riboflavin (B_2)	Oxidation-reduction reactions involving one- and two-electron transfers	Prosthetic group
Coenzyme A (CoA)	Pantothenate (B_5)	Transfer of acyl groups	Cosubstrate
Thiamine pyrophosphate (TPP)	Thiamine (B_1)	Transfer of multi-carbon fragments containing a carbonyl group	Prosthetic group
Pyridoxal phosphate (PLP)	Pyridoxine (B_6)	Transfer of groups to and from amino acids	Prosthetic group
Biotin	Biotin (B_7)	ATP-dependent carboxylation of substrates or carboxyl-group transfer between substrates	Prosthetic group
Tetrahydrofolate	Folate	Transfer of one-carbon substituents, especially formyl and hydroxymethyl groups; provides the methyl group for thymine in DNA	Cosubstrate
Cobalamin	Cobalamin (B_{12})	Intramolecular rearrangements, transfer of methyl groups.	Prosthetic group
Lipoamide	—	Oxidation of a hydroxyalkyl group from TPP and subsequent transfer as an acyl group	Prosthetic group
Retinal	Vitamin A	Vision	Prosthetic group
Vitamin K	Vitamin K	Carboxylation of some glutamate residues	Prosthetic group
Ubiquinone (Q)	—	Lipid-soluble electron carrier	Cosubstrate
Heme Group	—	Electron transfer	Prosthetic group

reacts readily with nucleophilic acceptors and is the donor of almost all the methyl groups used in biosynthetic reactions. For example, it is required for conversion of the hormone norepinephrine to epinephrine.

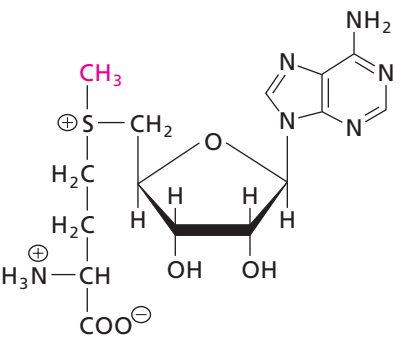
The thermodynamics of reactions involving ATP is explained in Section 10.6.



Norepinephrine Epinephrine **(7.2)**



▲ **Figure 7.4**
ATP. The nitrogenous base adenine is linked to a ribose bearing three phosphoryl groups. Transfer of a phosphoryl group (red) generates ADP, and transfer of a nucleotidyl group (AMP, blue) generates pyrophosphate.



▲ **Figure 7.5**
***S*-Adenosylmethionine.** The activated methyl group of this coenzyme is shown in red.

BOX 7.1 MISSING VITAMINS

Whatever happened to vitamin B_4 and vitamin B_8? They are never listed in the textbooks but you'll often find them sold in stores that cater to the demand for supplements that might make you feel better and live longer.

Vitamin B_4 was adenine, the base found in DNA and RNA. We now know that it's not a vitamin. All species, including humans, can make copious quantities of adenine whenever it's needed (Sections 18.1 and 18.2). Vitamin B_8 was inositol, a precursor of several important lipids (Figure 8.16 and Section 9.12C). It's no longer considered a vitamin.

If you know anyone who is paying money for vitamin B_4 and B_8 supplements then here's your chance to be helpful. Tell them why they're wasting their money.

▲ **P.T. Barnum.** P.T. Barnum was a famous American showman. He's credited with saying, "There's a sucker born every minute." It's likely that the memorable phrase was coined by one of his rivals and later attributed to Barnum in order to discredit him.

Methylation reactions that require *S*-adenosylmethionine include methylation of phospholipids, proteins, DNA, and RNA. In plants, *S*-adenosylmethionine—as a precursor of the plant hormone ethylene—is involved in regulating the ripening of fruit.

Nucleotide-sugar coenzymes are involved in carbohydrate metabolism. The most common nucleotide sugar, uridine diphosphate glucose (UDP-glucose), is formed by the reaction of glucose 1-phosphate with uridine triphosphate (UTP) (Figure 7.6). UDP-glucose can donate its glycosyl group (shown in red) to a suitable acceptor, releasing UDP. UDP-glucose is regenerated when UDP accepts a phosphoryl group from ATP and the resulting UTP reacts with another molecule of glucose 1-phosphate.

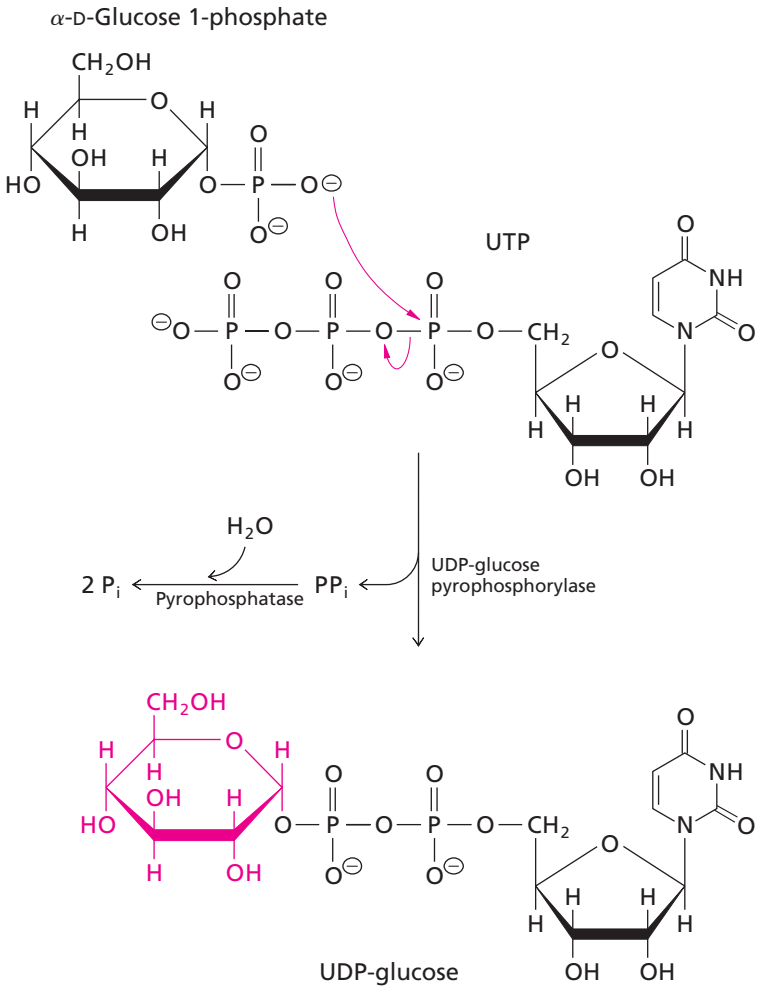
Both the sugar and the nucleoside of nucleotide-sugar coenzymes may vary. Later on, we will encounter CDP, GDP, and ADP variants of this coenzyme.

7.4 $NAD^{\oplus}$ and $NADP^{\oplus}$

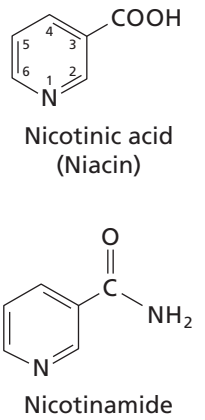
The nicotinamide coenzymes are nicotinamide adenine dinucleotide ($NAD^{\oplus}$) and the closely related nicotinamide adenine dinucleotide phosphate ($NADP^{\oplus}$). These were the first coenzymes to be recognized. Both contain nicotinamide, the amide of nicotinic acid (Figure 7.7). Nicotinic acid (also called niacin) is the factor missing in the disease pellagra. Nicotinic acid or nicotinamide is essential as a precursor of $NAD^{\oplus}$ and $NADP^{\oplus}$. (In many species, tryptophan is degraded to nicotinic acid. Dietary tryptophan can therefore spare some of the requirement for niacin or nicotinamide.)

The nicotinamide coenzymes play a role in many oxidation–reduction reactions. They assist in the transfer of electrons to and from metabolites (Section 10.9). The oxidized forms, $NAD^{\oplus}$ and $NADP^{\oplus}$, are electron deficient and the reduced forms, NADH and NADPH, carry an extra pair of electrons in the form of a covalently bound hydride ion. The structures of these coenzymes are shown in Figure 7.8 . Both coenzymes contain a phosphoanhydride linkage that joins two 5′-nucleotides: AMP and the ribonucleotide of nicotinamide, called nicotinamide mononucleotide (NMN) (formed from nicotinic acid). In the case of $NADP^{\oplus}$, a phosphoryl group is present on the 2′-oxygen atom of the adenylate moiety.

Note that the $^{\oplus}$ sign in $NAD^{\oplus}$ simply indicates that the nitrogen atom carries a positive charge. This does not mean that the entire molecule is a positively charged ion; in fact, it is negatively charged due to the phosphates. A nitrogen atom normally has



◀ **Figure 7.6**
Formation of UDP-glucose catalyzed by UDP-glucose pyrophosphorylase. An oxygen of the phosphate group of α-D-glucose 1-phosphate attacks the α-phosphorus of UTP. The PP_i released is rapidly hydrolyzed to $2P_i$ by the action of pyrophosphatase. This hydrolysis helps drive the pyrophosphorylase-catalyzed reaction toward completion. The mobile glycosyl group of UDP-glucose is shown in red.



▲ **Figure 7.7**
Nicotinic acid (niacin) and nicotinamide.

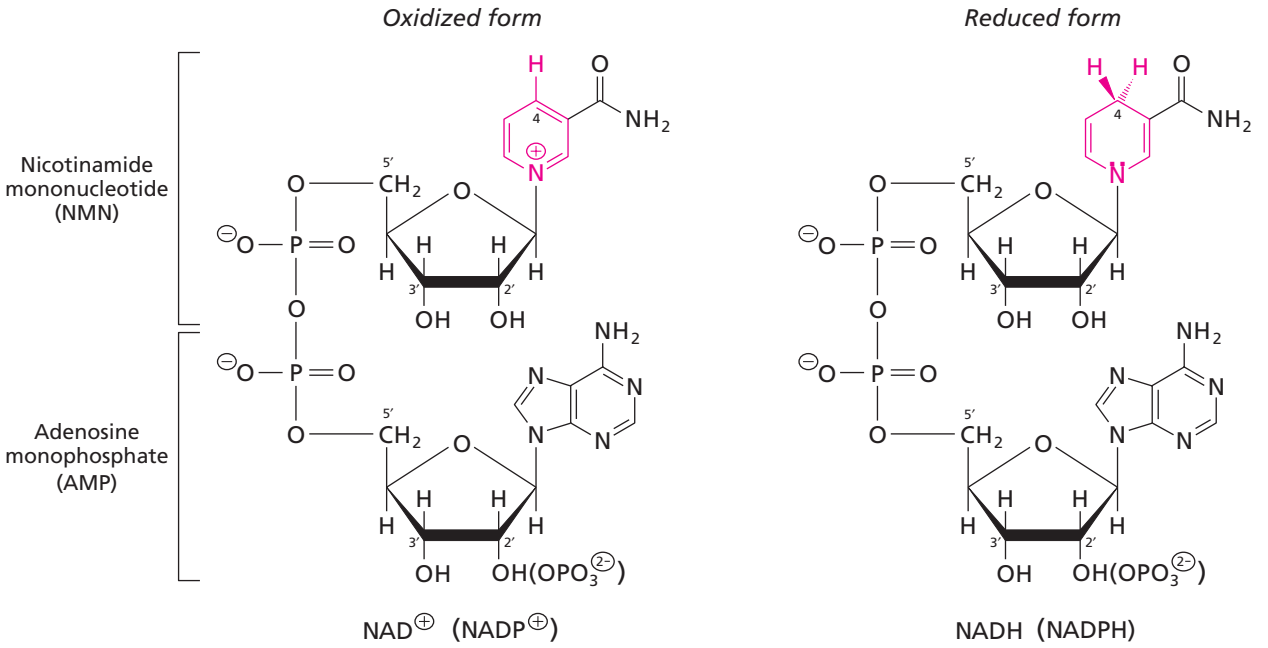
seven protons and seven electrons. The outer shell has five electrons that can participate in bond formation. In the oxidized form of the coenzyme (NAD⊕ and NADP⊕) the nicotinamide nitrogen is missing one of its electrons. It has only four electrons in the outer shell and those are shared with adjacent carbon atoms to form a total of four covalent bonds. (Each bond has a pair of electrons so the outer shell of the nitrogen atom is filled with eight shared electrons.) This is why we normally associate the positive charge with the ring nitrogen atom as shown in Figure 7.8. In fact, the charge is distributed over the entire aromatic ring.

The reduced form of the nitrogen atom has its normal, full complement of electrons. In particular, the nitrogen atom has five electrons in its outer shell. Two of these electrons (represented by dots in Figure 7.8) are a free pair of electrons. The other three electrons participate in three covalent bonds.

NAD⊕ and NADP⊕ almost always act as cosubstrates for dehydrogenases. Pyridine nucleotide-dependent dehydrogenases catalyze the oxidation of their substrates by transferring two electrons and a proton in the form of a hydride ion ($H^{\ominus}$) to C-4 of the nicotinamide group of NAD⊕ or NADP⊕. This generates the reduced form, NADH or NADPH, where a new C—H bond has formed at C-4 (one pair of electrons) and the electron previously associated with the ring double bond has delocalized to the ring nitrogen atom. Thus, oxidation by pyridine nucleotides (or reduction, the reverse reaction) always occurs two electrons at a time.

NADH and NADPH are said to possess reducing power (i.e., they are biological reducing agents). The stability of reduced pyridine nucleotides allows them to carry their reducing power from one enzyme to another, a property not shared by flavin

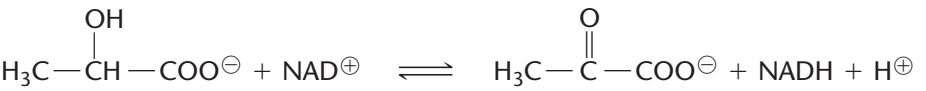
NADH and NADPH exhibit a peak of ultraviolet absorbance at 340 nm due to the dihydropyridine ring, whereas NAD⊕ and NADP⊕ do not absorb light at this wavelength. The appearance and disappearance of absorbance at 340 nm are useful for measuring the rates of oxidation and reduction reactions if they involve NAD⊕ or NADP⊕. (see Box 10.1).



▲ **Figure 7.8**
Oxidized and reduced forms of NAD (and NADP). The pyridine ring of $NAD^{\oplus}$ is reduced by the addition of a hydride ion to C-4 when $NAD^{\oplus}$ is converted to NADH (and when $NADP^{\oplus}$ is converted to NADPH). In $NADP^{\oplus}$, the 2′-hydroxyl group of the sugar ring of adenosine is phosphorylated. The reactive center of these coenzymes is shown in red.

coenzymes (Section 7.5). Most reactions forming NADH and NADPH are catabolic reactions and the subsequent oxidation of NADH by the membrane-associated electron transport system is coupled to the synthesis of ATP. Most NADPH is used as a reducing agent in biosynthetic reactions. The concentration of NADH is about ten times higher than that of NADPH.

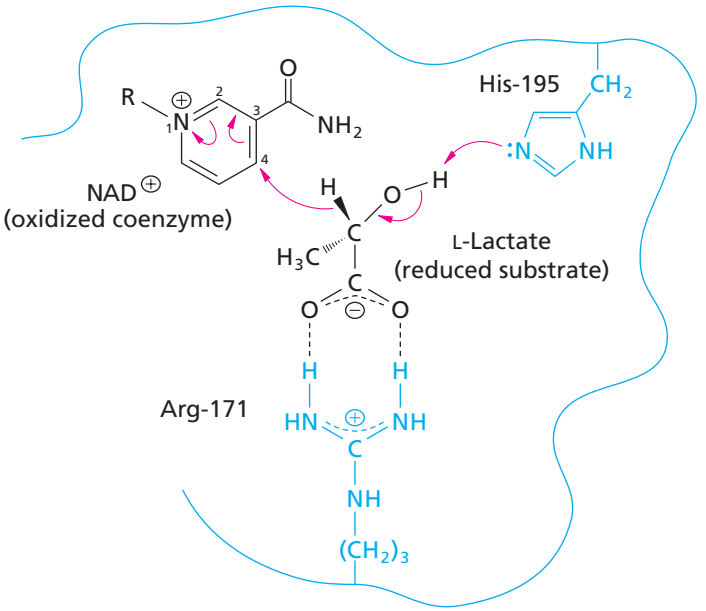
Lactate dehydrogenase is an oxidoreductase that catalyzes the reversible oxidation of lactate. The enzyme is a typical NAD-dependent dehydrogenase. A proton is released from lactate when $NAD^{\oplus}$ is reduced.



(7.3)

NADH is a cosubtrate, like ATP. When the reaction is complete, the structure of the cosubstrate is altered and the original form must be regenerated in a separate reaction. In this example, $NAD^{\oplus}$ is reduced to NADH and the reaction will soon reach equilibrium unless NADH is used up in a separate reaction where $NAD^{\oplus}$ is regenerated. We describe one example of how this is accomplished in Section 11.3B.

Figure 7.9 shows how both the enzyme and the coenzyme participate in the oxidation of lactate to pyruvate catalyzed by lactate dehydrogenase. In this mechanism, the coenzyme accepts a hydride ion at C-4 in the nicotinamide group. This leads to a rearrangement of bonds in the ring as electrons are shuffled to the positively charged nitrogen atom. The enzyme provides an acid–base catalyst and suitable binding sites for both the coenzyme and the substrate. Note that two hydrogens are removed from lactate to produce pyruvate (Equation 7.3). One of these hydrogens is transferred to $NAD^{\oplus}$ as a hydride ion carrying two electrons and the other is transferred to His-195 as a proton. The second hydrogen is subsequently released as $H^{\oplus}$ in order to regenerate the base catalyst (His-195). There are many examples of NAD-dependent reactions where the reduction of $NAD^{\oplus}$ is accompanied by release of a proton so it's quite common to see NADH + $H^{\oplus}$ on one side of the equation.



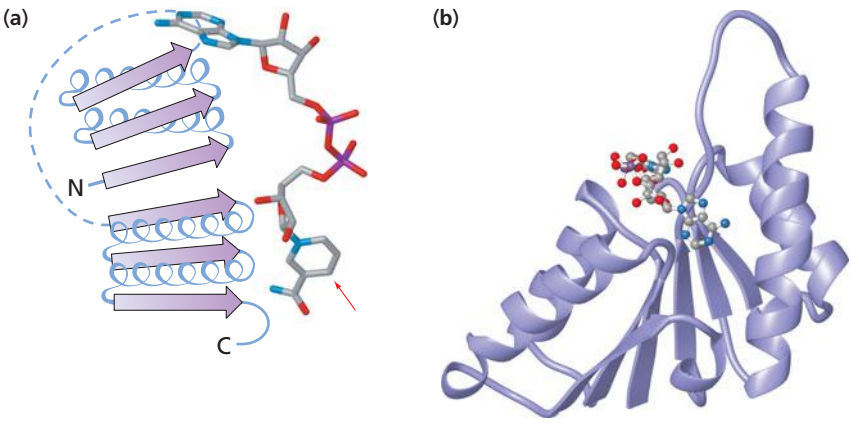
◀ **Figure 7.9**
Mechanism of lactate dehydrogenase. His-195, a base catalyst in the active site, abstracts a proton from the C-2 hydroxyl group of lactate, facilitating transfer of the hydride ion ($H^{\ominus}$) from C-2 of the substrate to C-4 of the bound $NAD^{\oplus}$. Arg-171 forms an ion pair with the carboxylate group of the substrate. In the reverse reaction, $H^{\ominus}$ is transferred from the reduced coenzyme, NADH, to C-2 of the oxidized substrate, pyruvate.

BOX 7.2 NAD BINDING TO DEHYDROGENASES

In the 1970s, structures were determined for four NAD-dependent dehydrogenases: lactate dehydrogenase, malate dehydrogenase, alcohol dehydrogenase, and glyceraldehyde 3-phosphate dehydrogenase. Each of these enzymes is oligomeric, with a chain length of about 350 amino acid residues. These chains all fold into two distinct domains—one to bind the coenzyme and one to bind the specific substrate. For each enzyme, the active site is in the cleft between the two domains.

As structures of more dehydrogenases were determined, several conformations of the coenzyme-binding motif were observed. Many of them possess one or more similar NAD- or NADP-binding structures consisting of a pair of $\beta\alpha\beta\alpha\beta$ units known as the Rossman fold after Michael Rossman, who first observed them in nucleotide-binding proteins (see figure). Each of the Rossman fold motifs binds to one half of the $NAD^{\oplus}$ dinucleotide. All of these enzymes bind the coenzyme in the same orientation and in a similar extended conformation.

Although many different dehydrogenases contain the Rossman fold motif, the rest of the structures may be very different and the dehydrogenases may not share significant sequence similarity. It's possible that all Rossman fold–containing enzymes descend from a common ancestor, but it's also possible that the motifs evolved independently in different dehydrogenases. That would be another example of convergent evolution.



◀ **NAD-binding region of some dehydrogenases.** **(a)** The coenzyme is bound in an extended conformation through interaction with two side-by-side motifs known as Rossman folds. The extended protein motifs form a β sheet of six parallel β strands. The arrow indicates the site where the hydride ion is added to C-4 of the nicotinamide group. **(b)** NADH bound to a Rossmann fold motif in rat lactate dehydrogenase [PDB 3H3F].

[Adapted from Rossman et al. (1975). *The Enzymes*, Vol. 11, Part A, 3rd ed., P. D., Boyer, ed. (New York: Academic Press), pp. 61–102.]

▲ These yellow FADs are not flavins but Fish Aggregating Devices. They are buoys tethered to the sea floor in order to attract fish. This one has been deployed by the government of New South Wales off the east coast of Australia. The strong ocean current is threatening to carry it off.

Crystals of Old Yellow Enzyme, a typical flavoprotein, are shown in the introduction to Chapter 5.

7.5 FAD and FMN

The coenzymes flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) are derived from riboflavin, or vitamin B_2. Riboflavin is synthesized by bacteria, protists, fungi, plants, and some animals. Mammals obtain riboflavin from food. Riboflavin consists of the five-carbon alcohol ribitol linked to the N-10 atom of a heterocyclic ring system called isoalloxazine (Figure 7.10a). The riboflavin-derived coenzymes are shown in Figure 7.11b. Like $NAD^{\oplus}$ and $NADP^{\oplus}$, FAD contains AMP and a diphosphate linkage.

Many oxidoreductases require FAD or FMN as a prosthetic group. Such enzymes are called flavoenzymes or flavoproteins. The prosthetic group is very tightly bound, usually noncovalently. By binding the prosthetic groups tightly, the apoenzymes protect the reduced forms from wasteful reoxidation.

FAD and FMN are reduced to $FADH_2$ and $FMNH_2$ by taking up a proton and two electrons in the form of a hydride ion (Figure 7.11). The oxidized enzymes are bright yellow as a result of the conjugated double-bond system of the isoalloxazine ring system. The color is lost when the coenzymes are reduced to $FMNH_2$ and $FADH_2$.

$FMNH_2$ and $FADH_2$ donate electrons either one or two at a time, unlike NADH and NADPH that participate exclusively in two-electron transfers. A partially oxidized compound, FADH• or FMNH•, is formed when one electron is donated. These intermediates are relatively stable free radicals called semiquinones. The oxidation of $FADH_2$ and $FMNH_2$ is often coupled to reduction of a metalloprotein containing Fe^{3+} (in an [Fe–S] cluster). Because an iron–sulfur cluster can accept only one electron, the reduced flavin must be oxidized in two one-electron steps via the semiquinone intermediate. The ability of FMN to couple two-electron transfers with one-electron transfers is important in many electron transfer systems.

7.6 Coenzyme A and Acyl Carrier Protein

Many metabolic processes depend on coenzyme A (CoA, or HS-CoA) including the oxidation of fuel molecules and the biosynthesis of some carbohydrates and lipids. This coenzyme is involved in acyl-group–transfer reactions in which simple carboxylic acids and fatty acids are the mobile metabolic groups. Coenzyme A has three major components: a 2-mercaptoethylamine unit that bears a free —SH group, the vitamin pantothenate (vitamin B_5, an amide of β-alanine and pantoate), and an ADP moiety

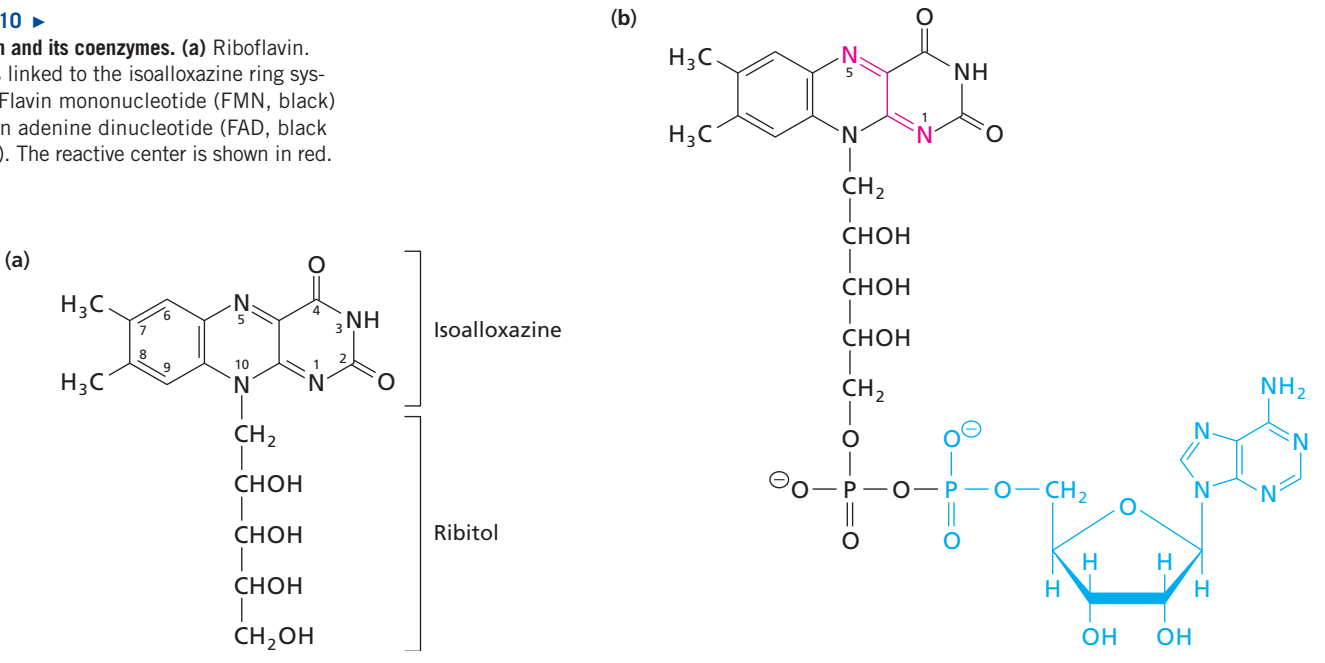
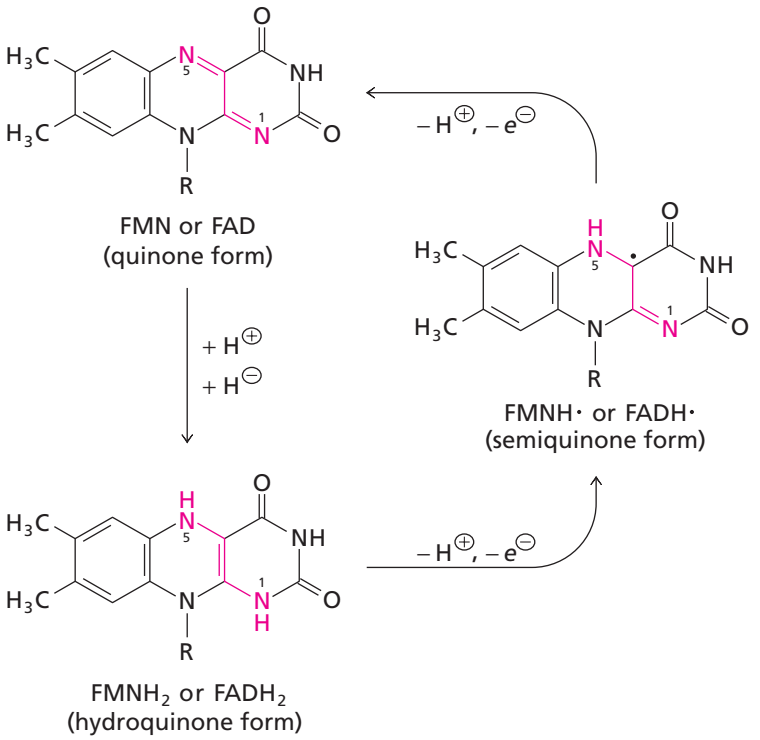


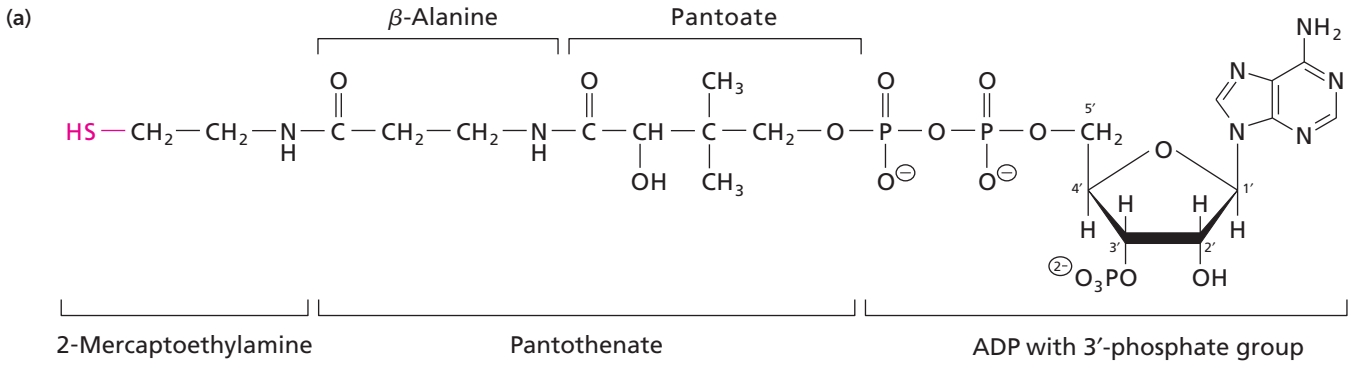
Figure 7.10 ►
Riboflavin and its coenzymes. (a) Riboflavin. Ribitol is linked to the isoalloxazine ring system. **(b)** Flavin mononucleotide (FMN, black) and flavin adenine dinucleotide (FAD, black and blue). The reactive center is shown in red.

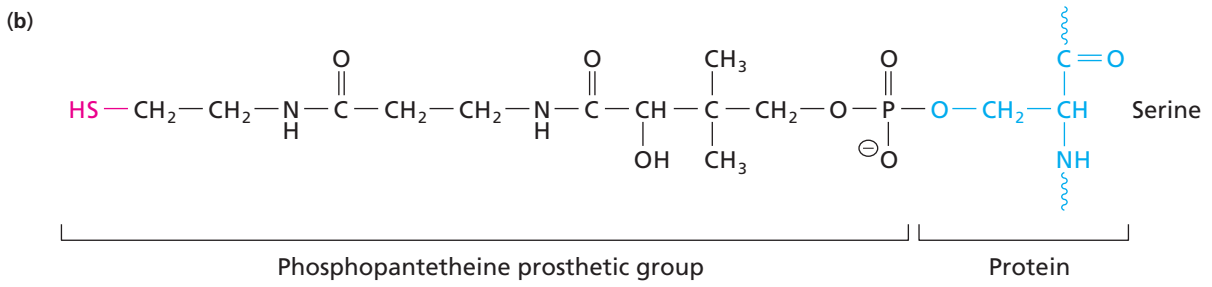


◀ **Figure 7.11**
Reduction and reoxidation of FMN or FAD. The conjugated double bonds between N-1 and N-5 are reduced by addition of a hydride ion and a proton to form $FMNH_2$ or $FADH_2$, respectively, the hydroquinone form of each coenzyme. Oxidation occurs in two steps. A single electron is removed by a one-electron oxidizing agent, with loss of a proton, to form a relatively stable free-radical intermediate. This semiquinone is then oxidized by removal of a proton and an electron to form fully oxidized FMN or FAD. These reactions are reversible.

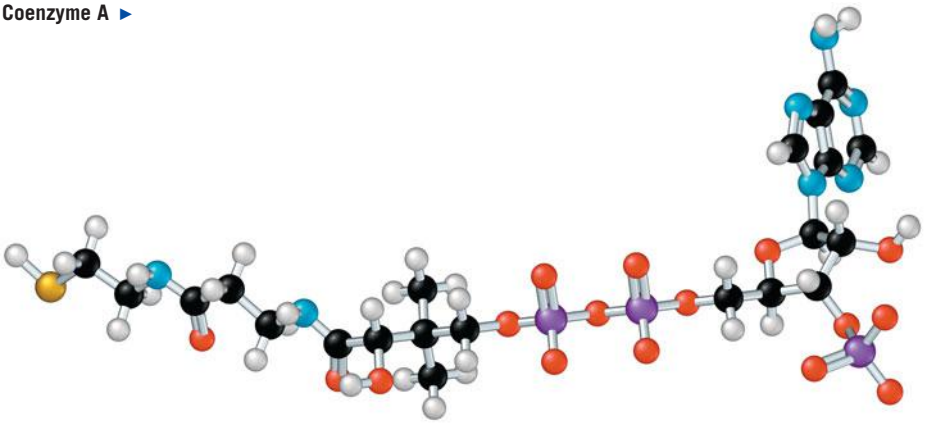
whose 3′-hydroxyl group is esterified with a third phosphate group (Figure 7.12a). The reactive center of CoA is the —SH group. Acyl groups covalently attach to the —SH group to form thioesters. A common example is acetyl CoA (Figure 7.13), where the acyl group is an acetyl moiety. Acetyl CoA is a "high energy" compound due to the thioester linkage (Section 19.8). Coenzyme A was originally named for its role as the

▼ **Figure 7.12**
Coenzyme A and acyl carrier protein (ACP). **(a)** In coenzyme A, 2-mercaptoethylamine is bound to the vitamin pantothenate, which in turn is bound via a phosphoester linkage to an ADP group that has an additional 3′-phosphate group. The reactive center is the thiol group (red). **(b)** In acyl carrier protein, the phosphopantetheine prosthetic group, which consists of the 2-mercaptoethylamine and pantothenate moieties of coenzyme A, is esterified to a serine residue of the protein.



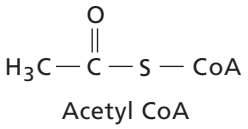


Coenzyme A ▶



acetylation coenzyme. We will see acetyl CoA frequently when we discuss the metabolism of carbohydrates, fatty acids, and amino acids.

Phosphopantetheine, a phosphate ester containing the 2-mercaptoethylamine and pantothenate moieties of coenzyme A, is the prosthetic group of a small protein (77 amino acid residues) known as the acyl carrier protein (ACP). The prosthetic group is esterified to ACP via the side-chain oxygen of a serine residue (Figure 7.12b). The —SH of the prosthetic group of ACP is acylated by intermediates in the biosynthesis of fatty acids (Chapter 16).



▲ Figure 7.13
Acetyl CoA

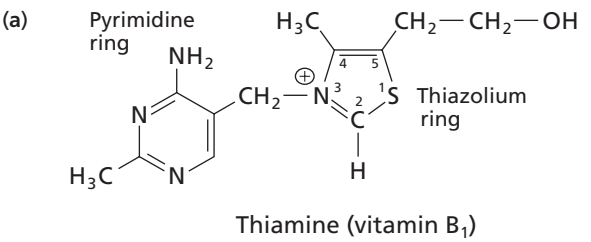
7.7 Thiamine Diphosphate

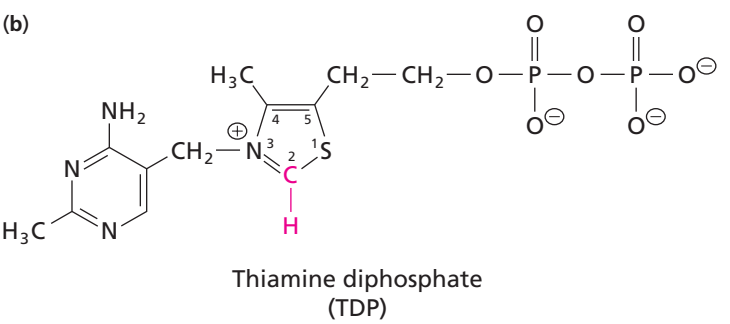
Thiamine (or vitamin B_1) contains a pyrimidine ring and a positively charged thiazolium ring (Figure 7.14a). The coenzyme is thiamine diphosphate (TDP), also called thiamine pyrophosphate (TPP) in the older literature (Figure 7.14b). TDP is synthesized from thiamine by enzymatic transfer of a pyrophosphoryl group from ATP.

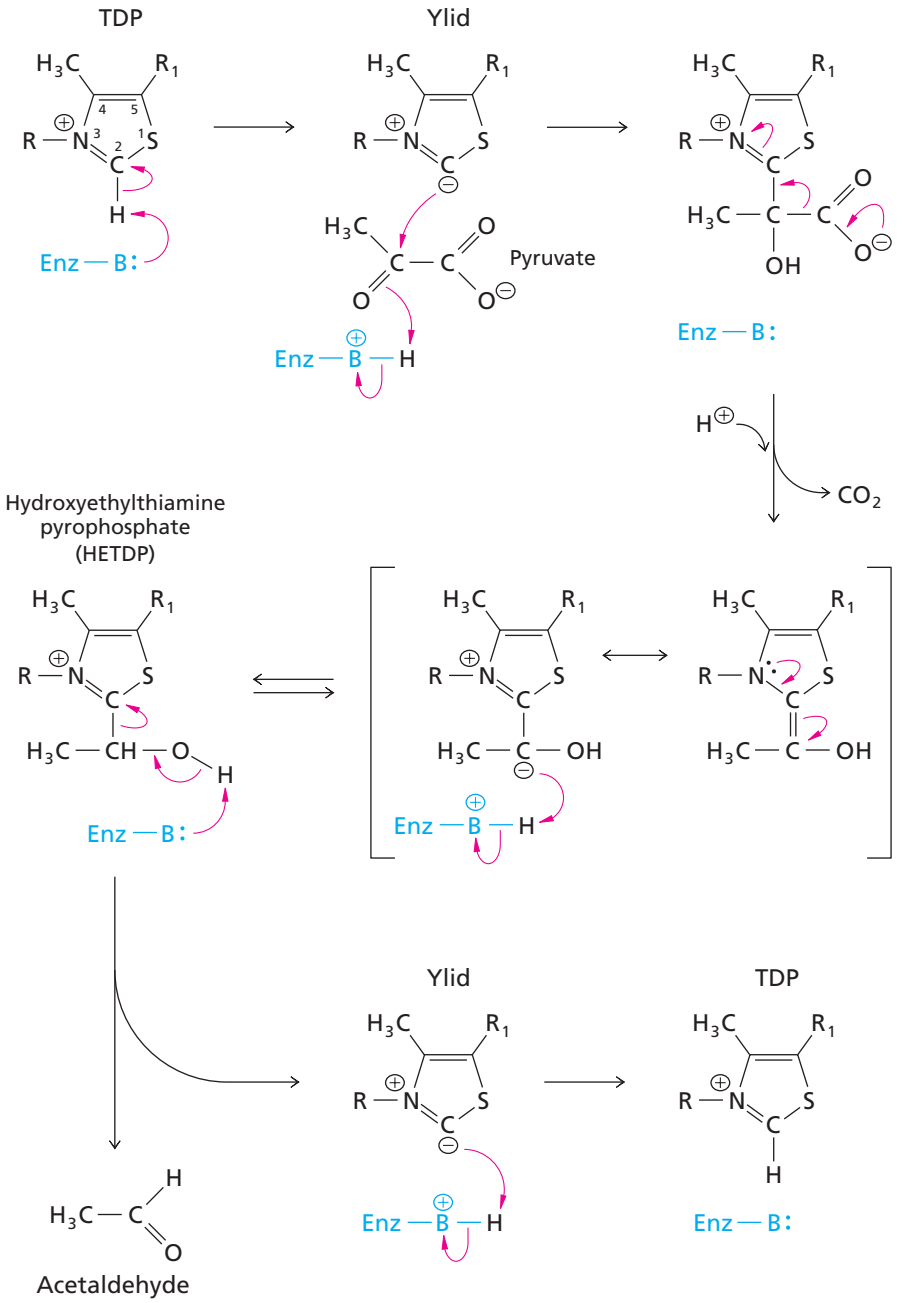
The metabolic role of pyruvate decarboxylase will be encountered in Section 11.3. Transketolases are discussed in Section 12.9. The role of TDP as a coenzyme in pyruvate dehydrogenase is described in Section 13.2.

About half a dozen decarboxylases (carboxylases) are known to require TDP as a coenzyme. For example, TDP is the prosthetic group of yeast pyruvate decarboxylase whose mechanism is shown in Figure 7.15. TDP is also a coenzyme involved in the oxidative decarboxylation of α-keto acids other than pyruvate. The first steps in those reactions proceed by the mechanism shown in Figure 7.15. In addition, TDP is a prosthetic group for enzymes known as transketolases that catalyze transfer between sugar molecules of two-carbon groups that contain a keto group.

Figure 7.14 ▶
Thiamine diphosphate (TDP). (a) Thiamine (vitamin B_1). **(b)** Thiamine diphosphate (TDP). The thiazolium ring of the coenzyme contains the reactive center (red).

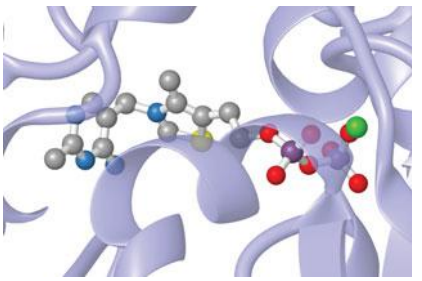






◀ **Figure 7.15**
Mechanism of yeast pyruvate decarboxylase. The positive charge of the thiazolium ring of TDP attracts electrons, weakening the bond between C-2 and hydrogen. This proton is presumably removed by a basic residue of the enzyme. Ionization generates a dipolar carbanion known as an ylid (a molecule with opposite charges on adjacent atoms). The negatively charged C-2 attacks the electron-deficient carbonyl carbon of the substrate pyruvate and the first product (CO_2) is released. Two carbons of pyruvate are now attached to the thiazole ring as part of a resonance-stabilized carbanion. In the following step, protonation of the carbanion produces hydroxyethylthiamine diphosphate (HETDP). HETDP is cleaved, releasing acetaldehyde (the second product) and regenerating the ylid form of the enzyme-TDP complex. TDP re-forms when the ylid is protonated by the enzyme.

The thiazolium ring of the coenzyme contains the reactive center. C-2 of TDP has unusual reactivity; it is acidic despite its extremely high pK_a in aqueous solution. Similarly, recent experiments indicate that the pK_a value for the ionization of hydroxyethylthiamine diphosphate (HETDP) (i.e., formation of the dipolar carbanion) is changed from 15 in water to 6 at the active site of pyruvate decarboxylase. This increased acidity is attributed to low polarity of the active site, which also accounts for the reactivity of TDP.

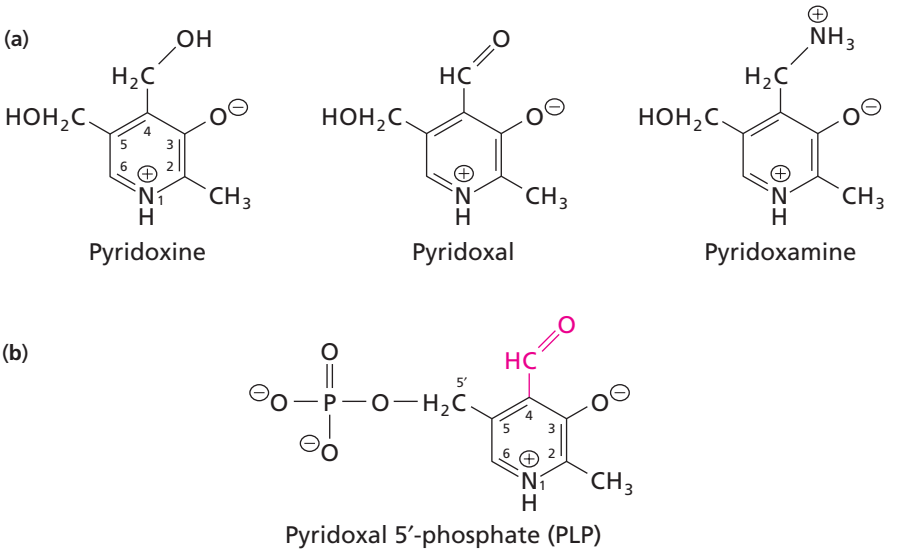


▲ **Thiamine diphosphate bound to pyruvate dehydrogenase.** The coenzyme is bound in an extended conformation and the diphosphate group is chelated to a magnesium ion (green). [PDB 1PYD]

7.8 Pyridoxal Phosphate

The B_6 family of water-soluble vitamins consists of three closely related molecules that differ only in the state of oxidation or amination of the carbon bound to position 4 of the pyridine ring (Figure 7.16a). Vitamin B_6—most often pyridoxal or pyridoxamine—is widely available from plant and animal sources. Induced B_6 deficiencies in rats result in dermatitis and various disorders related to protein metabolism but actual vitamin

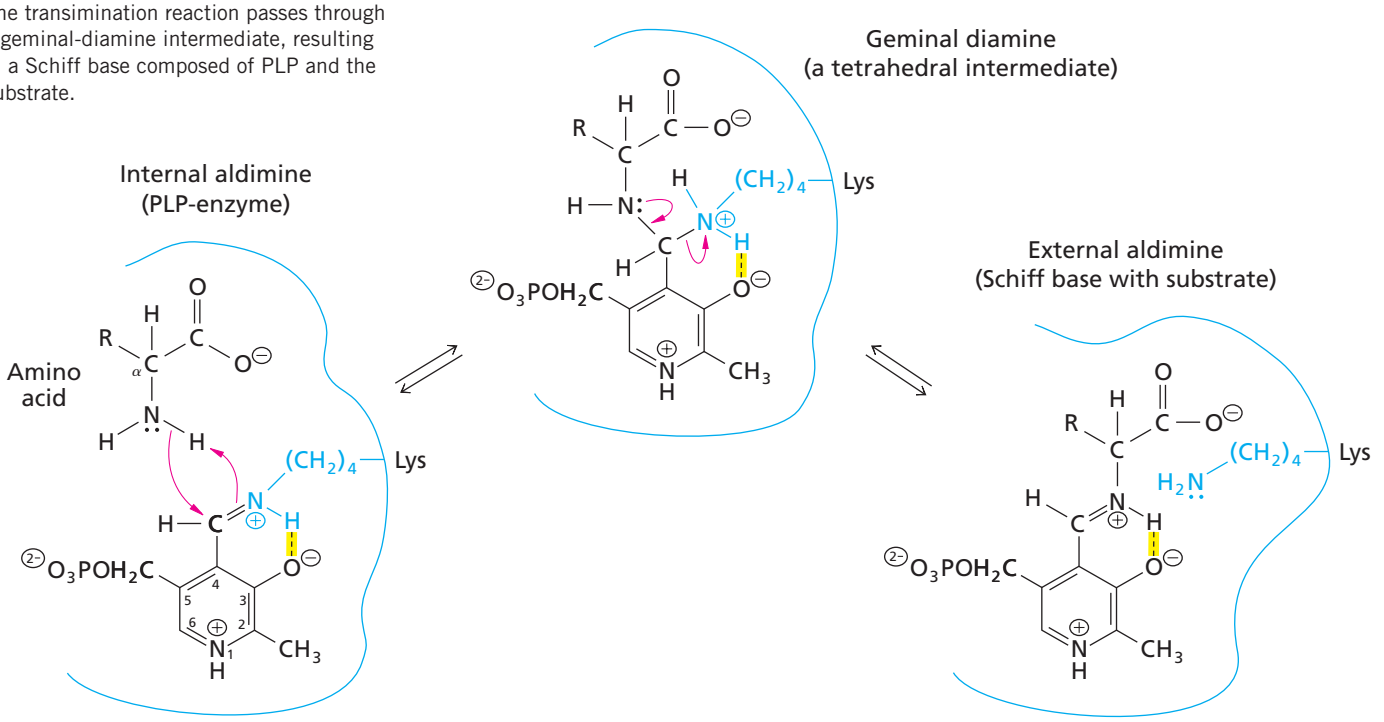
Figure 7.16 ►
B_6 vitamins and pyridoxal phosphate. (a) Vitamins of the B_6 family: pyridoxine, pyridoxal, and pyridoxamine. **(b)** Pyridoxal 5′-phosphate (PLP). The reactive center of PLP is the aldehyde group (red).

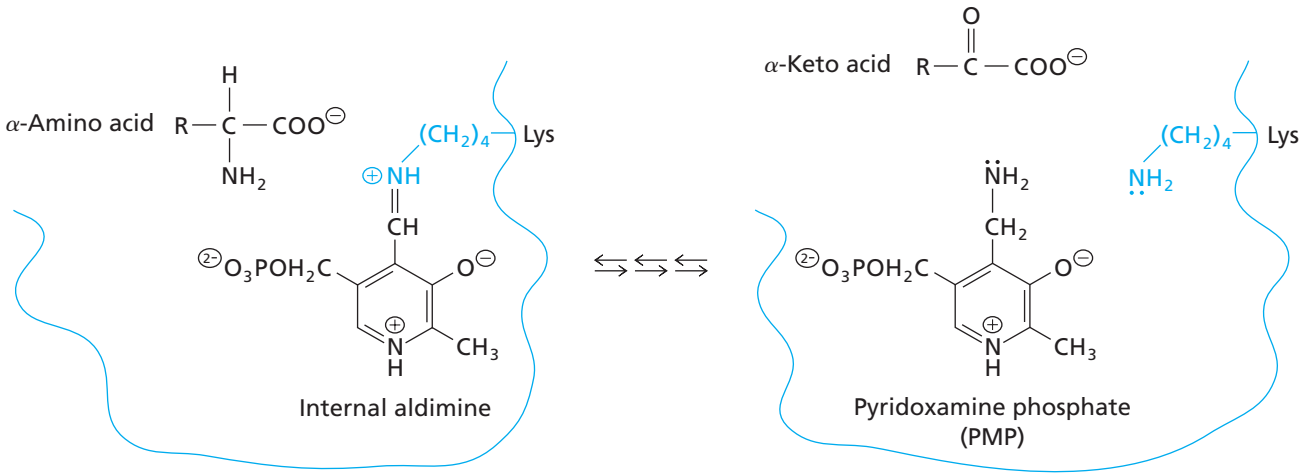


B_6 deficiencies in humans are rare. Enzymatic transfer of the γ-phosphoryl group from ATP forms the coenzyme pyridoxal 5′-phosphate (PLP) once vitamin B_6 enters a cell (Figure 7.16b).

Pyridoxal phosphate is the prosthetic group for many enzymes that catalyze a variety of reactions involving amino acids such as isomerizations, decarboxylations, and side-chain eliminations or replacements. In PLP-dependent enzymes, the carbonyl group of the prosthetic group is bound as a Schiff base (imine) to the ε-amino group of a lysine residue at the active site. (A Schiff base results from condensation of a primary amine with an aldehyde or ketone.) The enzyme-coenzyme Schiff base, shown on the left in Figure 7.17, is sometimes referred to as an internal aldimine. PLP is tightly bound to the enzyme by many weak noncovalent interactions; the additional covalent linkage of the internal aldimine helps prevent loss of the scarce coenzyme when the enzyme is not functioning.

Figure 7.17 ►
Binding of substrate to a PLP-dependent enzyme. The Schiff base linking PLP to a lysine residue of the enzyme is replaced by reaction of the substrate molecule with PLP. The transimination reaction passes through a geminal-diamine intermediate, resulting in a Schiff base composed of PLP and the substrate.





▲ **Figure 7.18**
Mechanism of transaminases. An amino acid displaces lysine from the internal aldimine that links PLP to the enzyme, generating an external aldimine. Subsequent steps lead to the transfer of the amino group to PLP yielding an α-keto acid, which dissociates, and PMP, which remains bound to the enzyme. If another α-keto acid enters, each step proceeds in reverse. The amino group is transferred to the α-keto acid producing a new amino acid and regenerating the original PLP form of the enzyme.

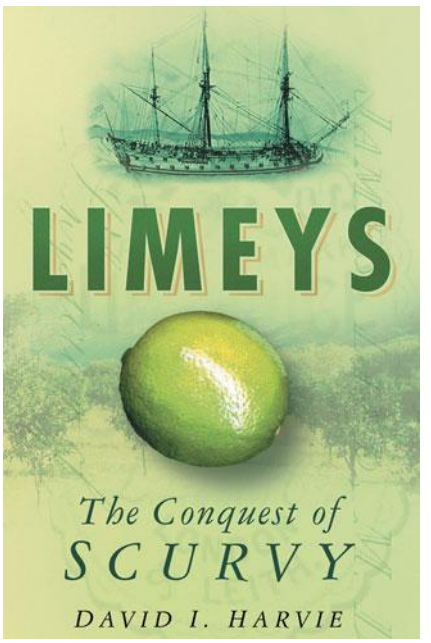
The initial step in all PLP-dependent enzymatic reactions with amino acids is the linkage of PLP to the α-amino group of the amino acid (formation of an external aldimine). When an amino acid binds to a PLP-enzyme, a transimination reaction takes place (Figure 7.17). This transfer reaction proceeds via a geminal-diamine intermediate rather than via formation of the free-aldehyde form of PLP. Note that the Schiff bases contain a system of conjugated double bonds in the pyridine ring ending with the positive charge on N-1. Similar ring structures with positively charged nitrogen atoms are present in $NAD^{\oplus}$. The prosthetic group serves as an electron sink during subsequent steps in the reactions catalyzed by PLP-enzymes. Once an α-amino acid forms a Schiff base with PLP, electron withdrawal toward N-1 weakens the three bonds to the α-carbon. In other words, the Schiff base with PLP stabilizes a carbanion formed when one of the three groups attached to the α-carbon of the amino acid is removed. Which group is lost depends on the chemical environment of the enzyme active site.

Removal of the α-amino group from amino acids is catalyzed by transaminases that participate in both the biosynthesis and degradation of amino acids (Chapter 17). Transamination is the most frequently encountered PLP-dependent reaction. The mechanism involves formation of an external aldimine (Figure 17.17) followed by release of the α-keto acid. The amino group remains bound to PLP forming pyridoxamine phosphate (PMP) (Figure 7.18). The next step in transaminase reactions is the reverse of the reaction shown in Figure 7.18 using a different α-keto acid as a substrate.

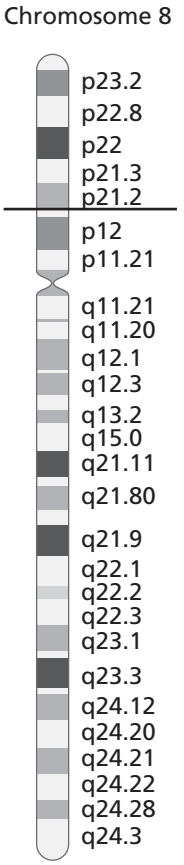
A specific transaminase is described in Section 17.2B.

7.9 Vitamin C

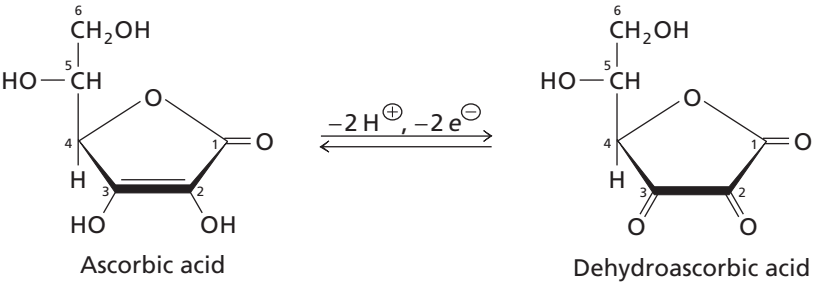
The simplest vitamin is the antiscurvy agent ascorbic acid (vitamin C). Scurvy is a disease whose symptoms include skin lesions, fragile blood vessels, loose teeth, and bleeding gums.The link between scurvy and nutrition was recognized four centuries ago when British navy physicians discovered that citrus juice in limes and lemons were a remedy for scurvy in sailors whose diet lacked fresh fruits and vegetables. It was not until 1919, however, that ascorbic acid was isolated and shown to be the essential dietary component supplied by citrus juices.



► *Limeys* is the story of Dr. James Lind and his attempt to promote citrus fruit as a cure for scurvy in the 1700s.



▲ The human GULO pseudogene is located on the short arm of chromosome 8.



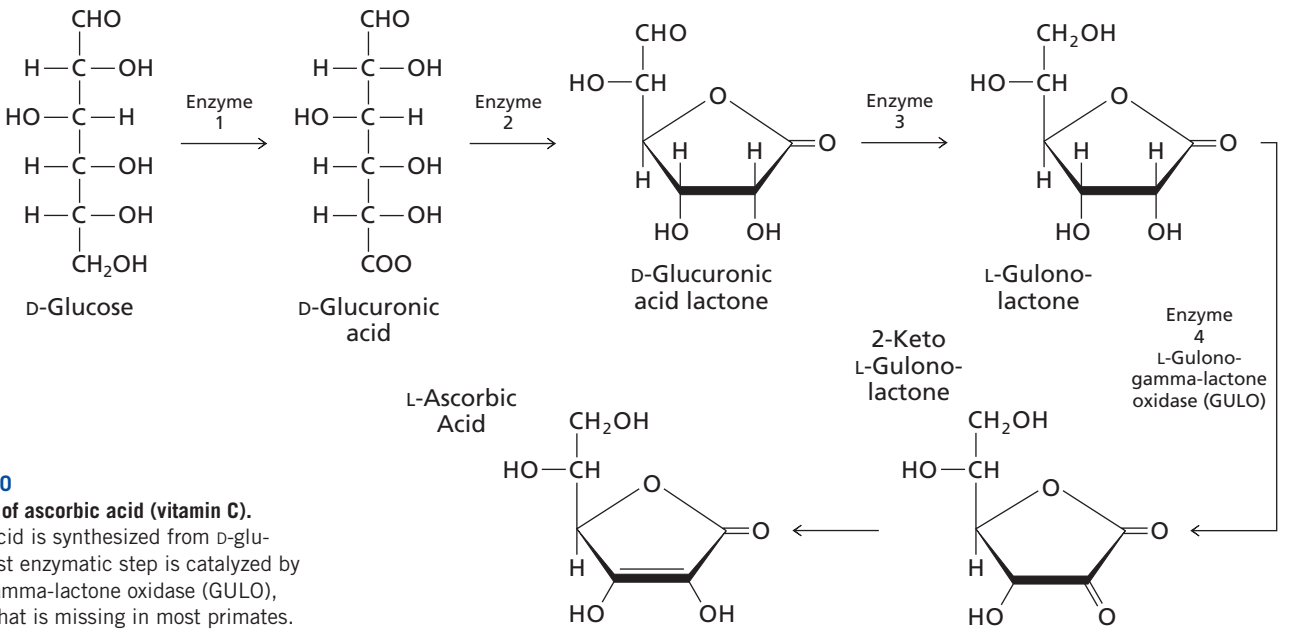
▲ **Figure 7.19**
Ascorbic acid (vitamin C) and its dehydro, oxidized form.

Back in the 18th century it was not easy to convince authorities that a simple solution like citrus fruit would solve the problem of scurvy because there were many competing theories. The story of Dr. James Lind and his efforts to convince the British navy is just one of many stories associated with vitamin C. It shows us that scientific evidence is not all that's required in order to make changes in the way we do things. Eventually, British sailors began to eat lemons and limes on a regular basis when they were at sea. Not only did this reduce the incidences of scurvy but it also gave rise to a famous nickname for British sailors. They were called "limeys" although lemons were much more effective than limes.

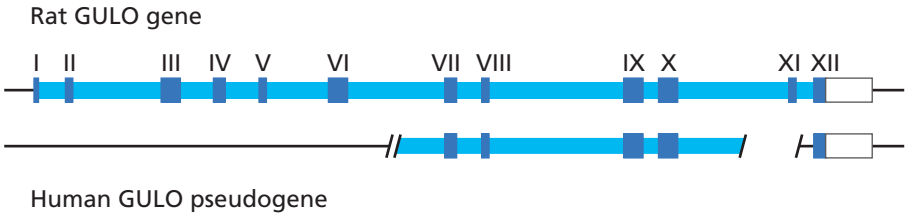
Ascorbic acid is a lactone, an internal ester in which the C-1 carboxylate group is condensed with the C-4 hydroxyl group, forming a ring structure. We now know that ascorbic acid is not a coenzyme but acts as a reducing agent in several different enzymatic reactions (Figure 7.19). The most important of these reactions is the hydroxylation of collagen (Section 4.12). Most mammals can synthesize ascorbic acid but guinea pigs, bats, and some primates (including humans) lack this ability and must therefore rely on dietary sources.

In most cases, we don't know very much about how certain enzymes disappeared from some species leading to a reliance on external sources for some essential metabolites. Most of the presumed gene disruption events happened so far in the distant past that few traces remain in modern genomes. The loss of ability to make vitamin C is an exception to that rule and serves as an instructive example of evolution.

Ascorbic acid is synthesized from D-glucose in a five-step pathway involving four enzymes (the last step is spontaneous). The last enzyme in the pathway is L-glucono-



▲ **Figure 7.20**
Biosynthesis of ascorbic acid (vitamin C). L-ascorbic acid is synthesized from D-glucose. The last enzymatic step is catalyzed by L-glucono-gamma-lactone oxidase (GULO), an enzyme that is missing in most primates.



◀ **Figure 7.21**
Comparison of the intact rat GULO gene and the human pseudogene. The human pseudogene is missing the first six exons and exon 11. In addition, there are many mutations in the remaining exons that prevent them from producing protein product.

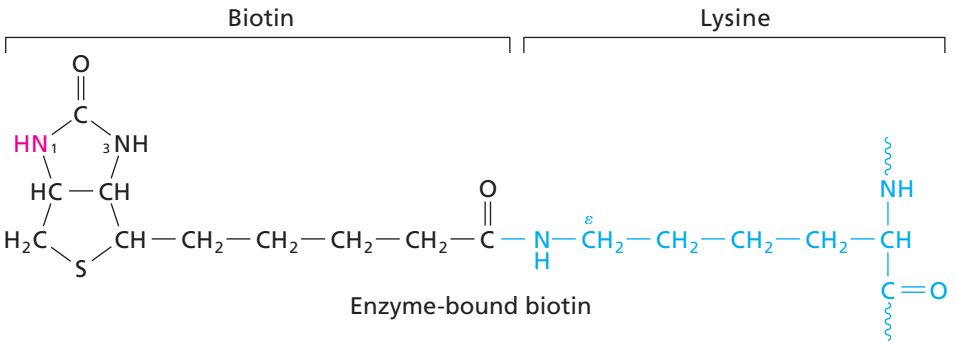
gamma-lactone oxidase (GULO) (Figure 7.20). GULO (the enzyme) is not present in primates of the haplorrhini family (monkeys and apes), but it is present in the strepsirrhini (lemurs, lorises etc.). These groups diverged about 80 million years ago. This led to the prediction that the GULO *gene* would be absent or defective in the monkeys and apes but intact in the other primates.

The prediction was confirmed with the discovery of a human GULO pseudogene on chromosome 8 in a block of genes that contains an active GULO gene in other animals. A comparison of the human pseudogene and a functional rat gene reveals many differences (Figure 7.21). The human pseudogene is missing the first six exons of the normal gene plus exon 11. The pseduogene in other apes is also missing these exons indicating that the ancestor of all apes had a similar defective GULO gene.

The original mutation that made the GULO gene inactive isn't known. Once inactivated, the pseudogene accumulated additional mutations that became fixed by random genetic drift. We can assume that lack of ability to synthesize vitamin C was not detrimental in these species because they obtained sufficient quantities in their normal diet.

7.10 Biotin

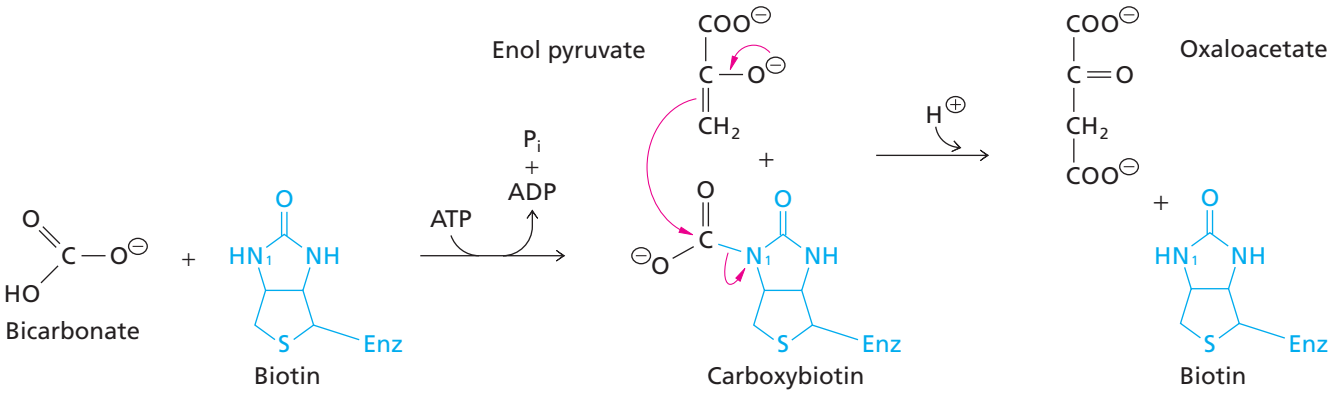
Biotin is a prosthetic group for enzymes that catalyze carboxyl group transfer reactions and ATP-dependent carboxylation reactions. Biotin is covalently linked to the active site of its host enzyme by an amide bond to the ε-amino group of a lysine residue (Figure 7.22).



◀ **Figure 7.22**
Enzyme-bound biotin. The carboxylate group of biotin is covalently bound via amide linkage to the ε-amino group of a lysine residue (blue). The reactive center of the biotin moiety is N-1 (red).

The pyruvate carboxylase reaction demonstrates the role of biotin as a carrier of carbon dioxide (Figure 7.23). In this ATP-dependent reaction, pyruvate, a three-carbon acid, reacts with bicarbonate forming the four-carbon acid oxaloacetate. Enzyme-bound biotin is the intermediate carrier of the mobile carboxyl metabolic group. The pyruvate carboxylase reaction is an important CO_2 fixation reaction. It is required in the gluconeogenesis pathway (Chapter 11).

Biotin was first identified as an essential factor for the growth of yeast. Biotin deficiency is rare in humans or animals on normal diets because biotin is synthesized by intestinal bacteria and is required only in very small amounts (micrograms per day). A biotin deficiency can be induced, however, by ingesting raw egg whites that contain a protein called avidin. Avidin binds tightly to biotin making it unavailable for absorption



▲ **Figure 7.23**
Reaction catalyzed by pyruvate carboxylase. First, biotin, bicarbonate, and ATP react to form carboxybiotin. The carboxybiotinyl-enzyme complex provides a stable, activated form of CO_2 that can be transferred to pyruvate. Next, the enolate form of pyruvate attacks the carboxyl group of carboxybiotin, forming oxaloacetate and regenerating biotin.

from the intestinal tract. Avidin is denatured when eggs are cooked and it loses its affinity for biotin.

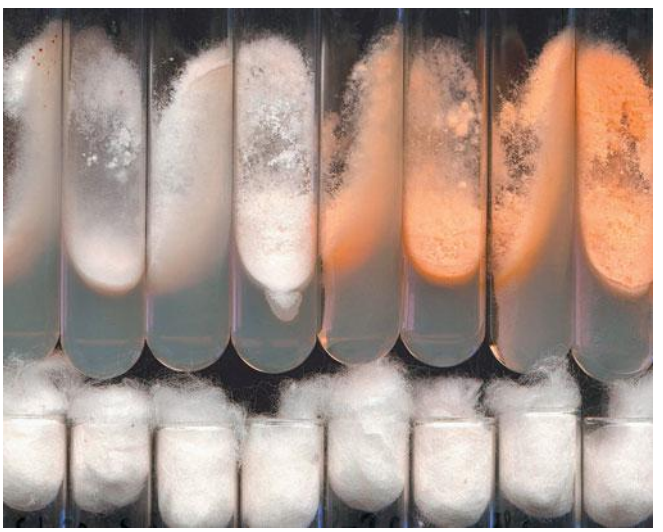
A variety of laboratory techniques take advantage of the high affinity of avidin for biotin. For example, a substance to which biotin is covalently attached can be extracted from a complex mixture by affinity chromatography (Section 3.6) on a column of immobilized avidin. The association constant for biotin and avidin is about 10^{15} M^{-1}—one of the tightest binding interactions known in biochemistry (see Section 4.9).

BOX 7.3 ONE GENE: ONE ENZYME

George Beadle and Edward Tatum wanted to test the idea that each gene encoded a single enzyme in a metabolic pathway. It was back in the late 1930s and this correspondence, which we now take for granted, was still a hypothesis. Remember, this was a time when it wasn't even clear whether genes were proteins or some other kind of chemical.

Beadle and Tatum chose the fungus *Neurospora crassa* for their experiments. *Neurospora* grows on a well-defined medium needing only sugar and biotin (vitamin B_7) as supplements. They reasoned that by irradiating *Neurospora* spores with X rays they could find mutants that would grow on rich supplemented medium but not on the simple defined medium. All they had to do next was identify the one supplement that needed to be added to the minimal medium to correct the defect. This would identify a gene for an enzyme that synthesized the now-essential supplement.

The 299th mutant required vitamin B_6 and the 1085th mutant required vitamin B_1. The B_6 and B_1 biosynthesis pathways were the first two pathways to be identified in this set of experiments. Later on, they worked out the genes/enzymes used in the tryptophan pathway. The results were published in 1941 and Beadle and Tatum received the Nobel Prize in Physiology or Medicine in 1958.



▲ ***Neurospora crassa* growing on defined medium in a test tube.** The strains on the right are producing orange carotenoid and the ones on the left are nonproducing strains.

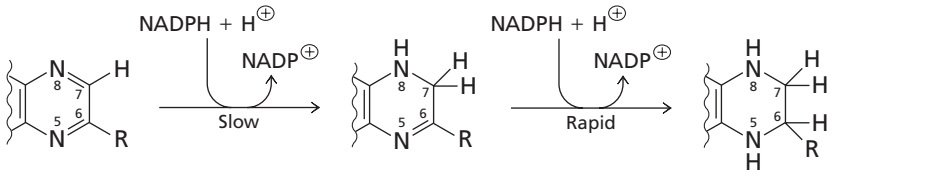
(Source: Courtesy of Manchester University, United Kingdom).

7.11 Tetrahydrofolate

The vitamin folate was first isolated in the early 1940s from green leaves, liver, and yeast. Folate has three main components: pterin (2-amino-4-oxopteridine), a *p*-aminobenzoic acid moiety, and a glutamate residue. The structures of pterin and folate are shown in Figures 7.24a and 7.24b. Humans require folate in their diets because we cannot synthesize the pterin-*p*-aminobenzoic acid intermediate (PABA) and we cannot add glutamate to exogenous PABA.

The coenzyme forms of folate, known collectively as tetrahydrofolate, differ from the vitamin in two respects: they are reduced compounds (5,6,7,8-tetrahydropterins), and they are modified by the addition of glutamate residues bound to one another through γ-glutamyl amide linkages (Figure 7.24c). The anionic polyglutamyl moiety, usually five to six residues long, participates in the binding of the coenzymes to enzymes. When using the term *tetrahydrofolate*, keep in mind that it refers to compounds that have polyglutamate tails of varying lengths.

Tetrahydrofolate is formed from folate by adding hydrogen to positions 5, 6, 7, and 8 of the pterin ring system. Folate is reduced in two NADPH-dependent steps in a reaction catalyzed by dihydrofolate reductase (DHFR).

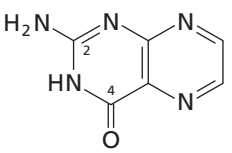


(7.4)

The primary metabolic function of dihydrofolate reductase is the reduction of dihydrofolate produced during the formation of the methyl group of thymidylate (dTMP) (Chapter 18). This reaction, which uses a derivative of tetrahydrofolate, is an essential step in the biosynthesis of DNA. Because cell division cannot occur when DNA synthesis is interrupted, dihydrofolate reductase has been extensively studied as a target for chemotherapy in the treatment of cancer (Box 18.4). In most species, dihydrofolate reductase is a relatively small monomeric enzyme that has evolved efficient binding sites for the two large substrates (folate and NADPH) (Figure 6.12).

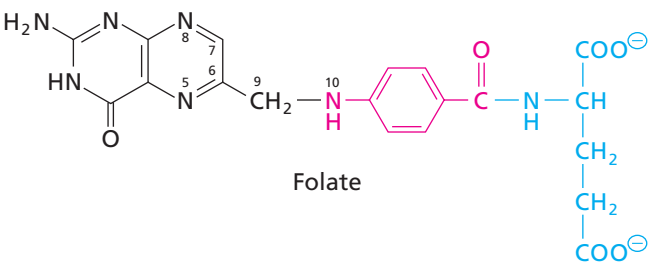
▼ **Figure 7.24**
Pterin, folate, and tetrahydrofolate. Pterin **(a)** is part of folate **(b)**, a molecule containing *p*-aminobenzoate (red) and glutamate (blue). **(c)** The polyglutamate forms of tetrahydrofolate usually contain five or six glutamate residues. The reactive centers of the coenzyme, N-5 and N-10, are shown in red.

(a)



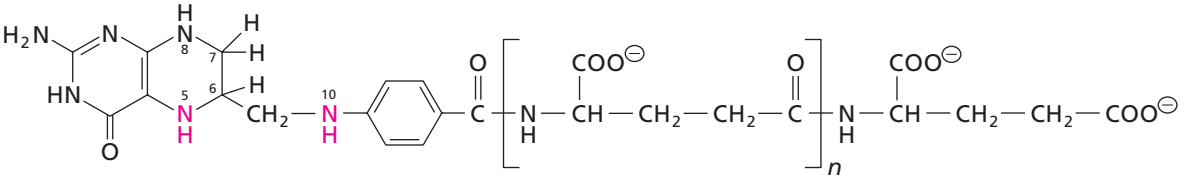
Pterin
(2-Amino-4-oxopteridine)

(b)



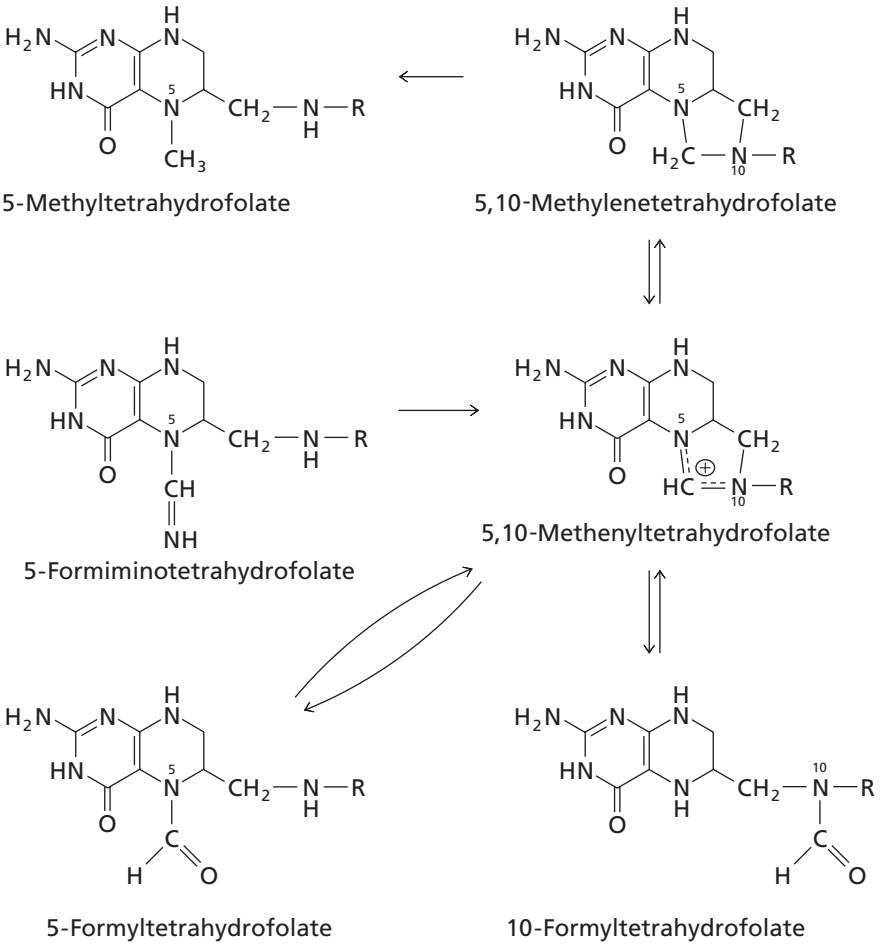
Folate

(c)

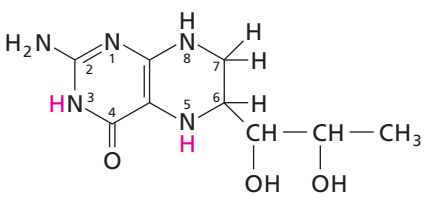


Tetrahydrofolate (Tetrahydrofolyl polyglutamate)

Figure 7.25 ►
One-carbon derivatives of tetrahydrofolate. The derivatives can be interconverted enzymatically by the routes shown. (R represents the benzoyl polyglutamate portion of tetrahydrofolate.)



▲ **Many fruits and vegetables contain adequate supplies of folate.** Yeast and liver products are also excellent sources of folate.



▲ **Figure 7.26**
5,6,7,8-Tetrahydrobiopterin. The hydrogen atoms lost on oxidation are shown in red.

5,6,7,8-Tetrahydrofolate is required by enzymes that catalyze biochemical transfers of several one-carbon units. The groups bound to tetrahydrofolate are methyl, methylene, or formyl groups. Figure 7.25 shows the structures of several one-carbon derivatives of tetrahydrofolate and the enzymatic interconversions that occur among them. The one-carbon metabolic groups are covalently bound to the secondary amine N-5 or N-10 of tetrahydrofolate, or to both in a ring form. 10-Formyltetrahydrofolate is the donor of formyl groups and 5,10-methylenetetrahydrofolate is the donor of hydroxymethyl groups.

Another pterin coenzyme, 5,6,7,8-tetrahydrobiopterin, has a three-carbon side chain at C-6 of the pterin moiety in place of the large side chain found in tetrahydrofolate (Figure 7.26). This coenzyme is not derived from a vitamin but is synthesized by animals and other organisms. Tetrahydrobiopterin is the cofactor for several hydroxylases and will be encountered as a reducing agent in the conversion of phenylalanine to tyrosine (Chapter 17). It also is required by the enzyme that catalyzes the synthesis of nitric oxide from arginine (Section 17.12).

The sale of vitamins and supplements is big business in developed nations. It's often difficult to decide whether an extra supply of vitamins is necessary for good health because the scientific evidence is often missing or contradictory. Folate (vitamin B_9) deficiency is uncommon in normal, healthy adults and children in developed nations but there are documented cases of folate deficiency in pregnant women. A lack of tetrahydrofolate can lead to anemia and to severe defects in the developing fetus. While there are many fruits and vegetables that contain folate, it's a good idea for pregnant women to supplement their diet with folate in order to ensure their own health and that of the baby.

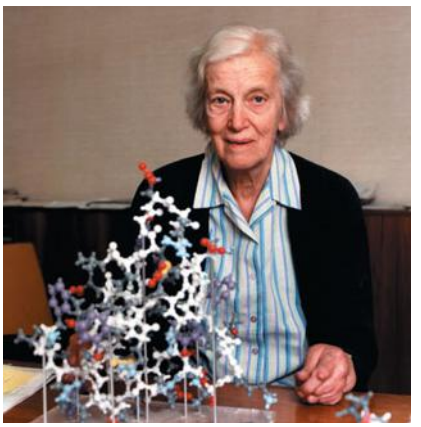
7.12 Cobalamin

Cobalamin (vitamin B_{12}) is the largest B vitamin and was the last to be isolated. The structure of cobalamin (Figure 7.27a) includes a corrin ring system that resembles the porphyrin ring system of heme (Figure 4.37). Note that cobalamin contains cobalt rather than the iron found in heme. The abbreviated structure shown in Figure 7.27b emphasizes the positions of two axial ligands bound to the cobalt, a benzimidazole ribonucleotide below the corrin ring and an R group above it. In the coenzyme forms of cobalamin, the R group is either a methyl group (in methylcobalamin) or a 5′-deoxyadenosyl group (in adenosylcobalamin).

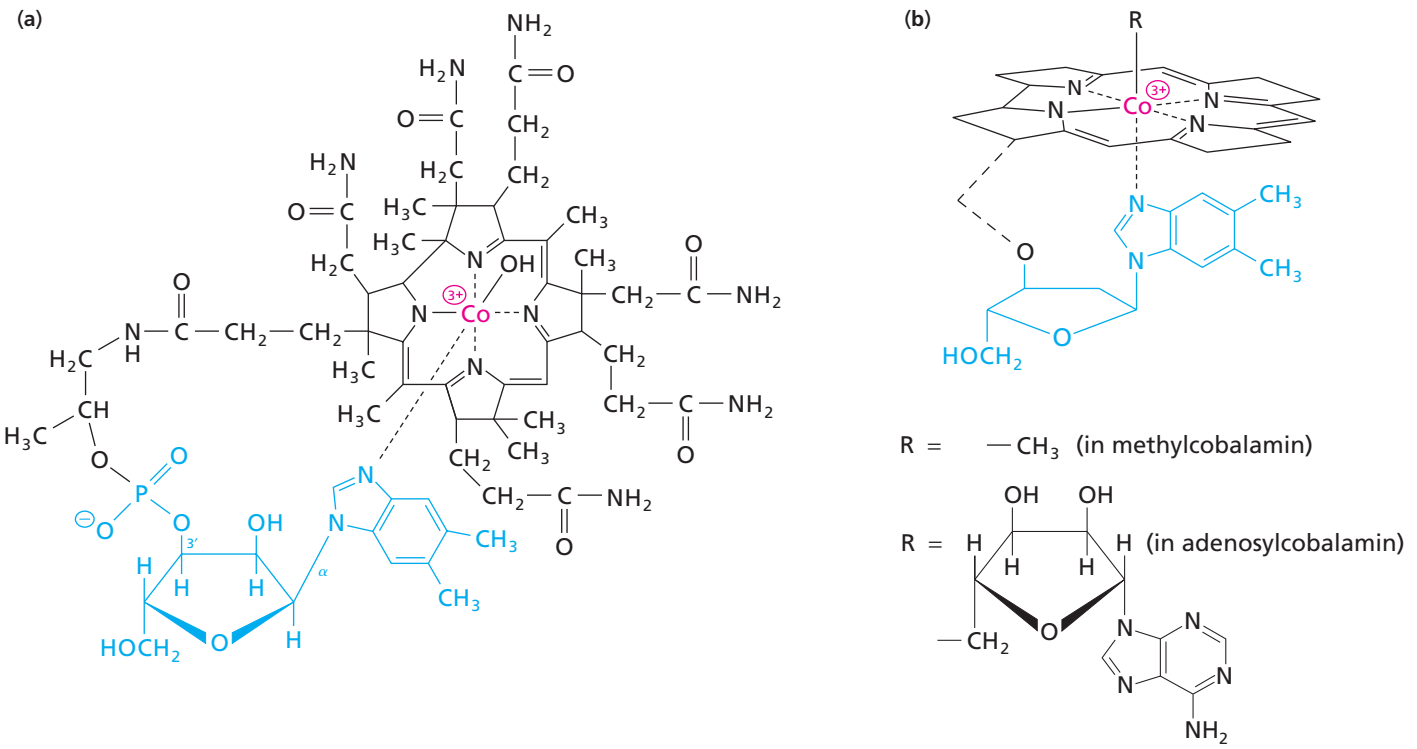
Cobalamin is synthesized by only a few microorganisms. It is required as a micronutrient by all animals and by some bacteria and algae. Humans obtain cobalamin from foods of animal origin. A deficiency of cobalamin can lead to pernicious anemia, a potentially fatal disease in which there is a decrease in the production of blood cells by bone marrow. Pernicious anemia can also cause neurological disorders. Most victims of pernicious anemia do not secrete a necessary glycoprotein (called intrinsic factor) from the stomach mucosa. This protein specifically binds cobalamin and the complex is absorbed by cells of the small intestine. Impaired absorption of cobalamin is now treated by regular injections of the vitamin.

The role of adenosylcobalamin reflects the reactivity of its C—Co bond. The coenzyme participates in several enzyme-catalyzed intramolecular rearrangements in which a hydrogen atom and a second group, bound to adjacent carbon atoms within a substrate, exchange places (Figure 7.28a). An example is the methylmalonyl–CoA mutase reaction (Figure 7.28b) that is important in the metabolism of odd-chain fatty acids (Chapter 16) and leads to the formation of succinyl CoA, an intermediate of the citric acid cycle.

Methylcobalamin participates in the transfer of methyl groups, as in the regeneration of methionine from homocysteine in mammals.



▲ **Dorothy Crowfoot Hodgkin (1910–1994).** Hodgkin received the Nobel Prize in 1964 for determining the structure of vitamin B_{12} (cobalamin). The structure of insulin, shown in the photograph, was published in 1969.



▲ **Figure 7.27**
Cobalamin (vitamin B_{12}) and its coenzymes. (a) Detailed structure of cobalamin showing the corrin ring system (black) and 5,6-dimethylbenzimidazole ribonucleotide (blue). The metal coordinated by corrin is cobalt (red). The benzimidazole ribonucleotide is coordinated with the cobalt of the corrin ring and is also bound via a phosphoester linkage to a side chain of the corrin ring system. **(b)** Abbreviated structure of cobalamin coenzymes. A benzimidazole ribonucleotide lies below the corrin ring, and an R group lies above the ring.

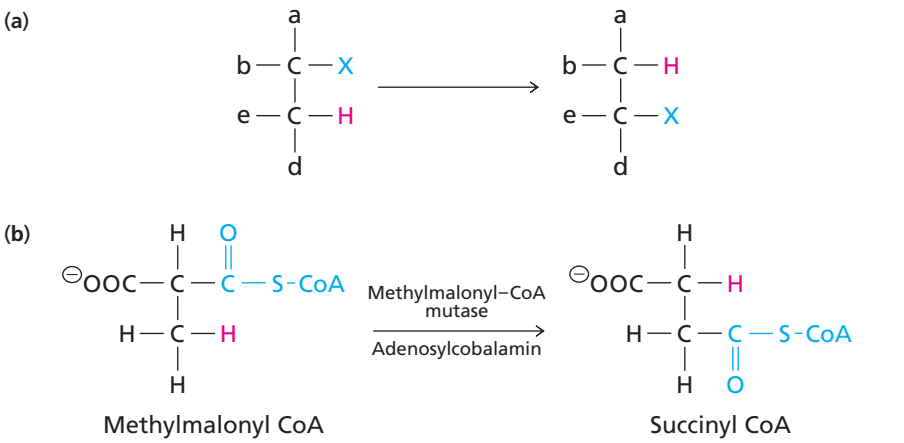
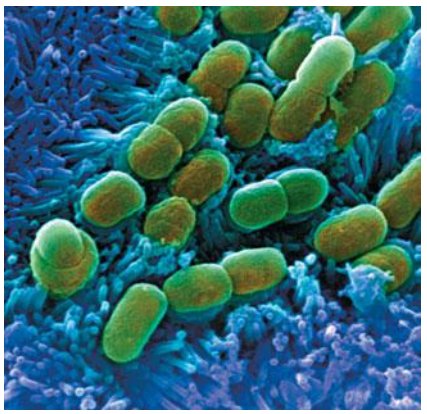
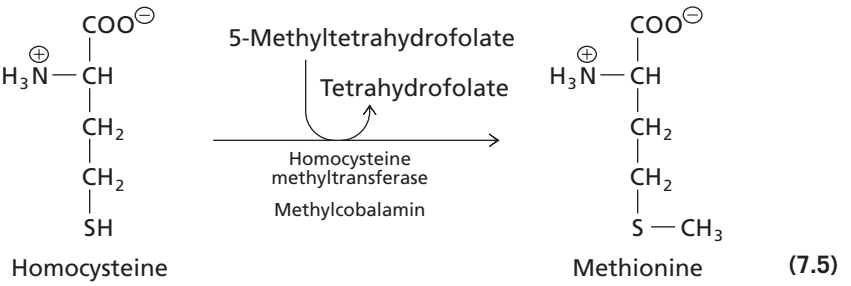


Figure 7.28 ►
Intramolecular rearrangements catalyzed by adenosylcobalamin-dependent enzymes. **(a)** Rearrangement in which a hydrogen atom and a substituent on an adjacent carbon atom exchange places. **(b)** Rearrangement of methylmalonyl CoA to succinyl CoA, catalyzed by methylmalonyl–CoA mutase.



▲ **Intestinal bacteria.** Normal, healthy humans harbor billions of bacteria in their intestines. There are at least several dozen different species. The one shown here is *Helicobacter pylori*, which causes stomach ulcers when it invades the stomach. The bacteria are sitting on the surface of the intestine that has many projections for absorbing nutrients. Other common species are *Escherichia coli* and various species of *Actinomyces* and *Streptococcus*. These bacteria help break down ingested food and they supply many of the essential vitamins and amino acids that humans need, especially cobalamin.



(7.5)

In this reaction, the methyl group of 5-methyltetrahydrofolate is passed to a reactive, reduced form of cobalamin to form methylcobalamin that can transfer the methyl group to the thiol side chain of homocysteine.

7.13 Lipoamide

The lipoamide coenzyme is the protein-bound form of lipoic acid. Lipoic acid is sometimes described as a vitamin but animals appear to be able to synthesize it. It is required by certain bacteria and protozoa for growth. Lipoic acid is an eight-carbon carboxylic acid (octanoic acid) in which two hydrogen atoms, on C-6 and C-8, have been replaced by sulfhydryl groups in disulfide linkage. Lipoic acid does not occur free—it is covalently attached via an amide linkage through its carboxyl group to the ε-amino group of a lysine residue of a protein (Figure 7.29). This structure is found in dihydrolipoamide acyltransferases that are components of the pyruvate dehydrogenase complex and related enzymes.

Lipoamide carries acyl groups between active sites in multienzyme complexes. For example, in the pyruvate dehydrogenase complex (Section 12.2), the disulfide ring of

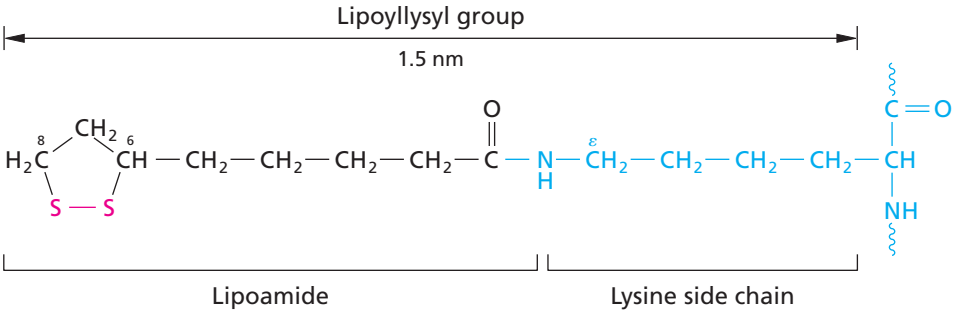
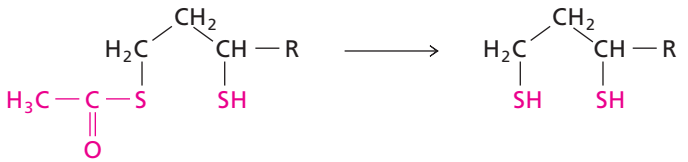


Figure 7.29 ►
Lipoamide. Lipoic acid is bound in amide linkage to the ε-amino group of a lysine residue (blue) of dihydrolipoamide acyltransferases. The dithiolane ring of the lipoyllysyl groups is extended 1.5 nm from the polypeptide backbone. The reactive center of the coenzyme is shown in red.

the lipoamide prosthetic group reacts with HETDP (Figure 7.15) binding its acetyl group to the sulfur atom attached to C-8 of lipoamide and forming a thioester. The acyl group is then transferred to the sulfur atom of a coenzyme A molecule generating the reduced (dihydrolipoamide) form of the prosthetic group.



(7.6)

The final step catalyzed by the pyruvate dehydrogenase complex is the oxidation of dihydrolipoamide. In this reaction, NADH is formed by the action of a flavoprotein component of the complex. The actions of the multiple coenzymes of the pyruvate dehydrogenase complex show how coenzymes, by supplying reactive groups that augment the catalytic versatility of proteins, are used to conserve both energy and carbon building blocks.

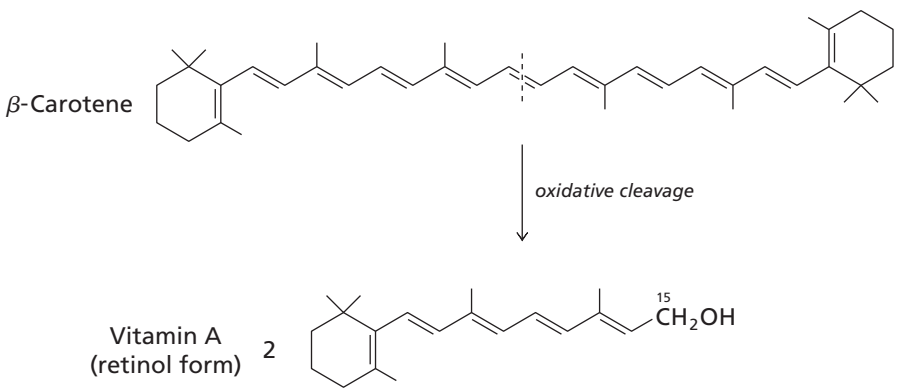
7.14 Lipid Vitamins

The structures of the four lipid vitamins (A, D, E, and K) contain rings and long aliphatic side chains. The lipid vitamins are highly hydrophobic although each possesses at least one polar group. In humans and other mammals, ingested lipid vitamins are absorbed in the intestine by a process similar to the absorption of other lipid nutrients (Section 16.1a). After digestion of any proteins that may bind them, they are carried to the cellular interface of the intestine as micelles formed with bile salts. The study of these hydrophobic molecules has presented several technical difficulties so research on their mechanisms has progressed more slowly than that on their water-soluble counterparts. Lipid vitamins differ widely in their functions, as we will see below.

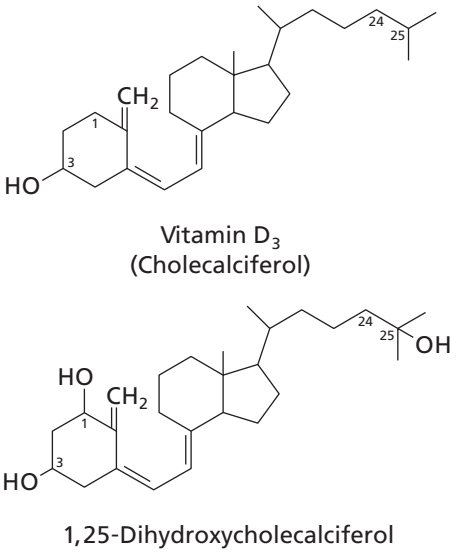
A. Vitamin A

Vitamin A, or retinol, is a 20-carbon lipid molecule obtained in the diet either directly or indirectly from β-carotene. Carrots and other yellow vegetables are rich in β-carotene, a 40-carbon plant lipid whose enzymatic oxidative cleavage yields vitamin A (Figure 7.30). Vitamin A exists in three forms that differ in the oxidation state of the terminal functional group: the stable alcohol retinol, the aldehyde retinal, and retinoic acid. Their hydrophobic side chain is formed from repeated isoprene units (Section 9.6).

All three vitamin A derivatives have important biological functions. Retinoic acid is a signal compound that binds to receptor proteins inside cells; the ligand–receptor



◀ **Figure 7.30**
Formation of vitamin A from β-carotene.



▲ **Figure 7.31**
Vitamin D_3 (cholecalciferol) and 1,25-dihydroxycholecalciferol. (Vitamin D_2 has an additional methyl group at C-24 and a *trans* double bond between C-22 and C-23.) 1,25-Dihydroxycholecalciferol is produced from vitamin D_3 by two separate hydroxylations.

complexes then bind to chromosomes and can regulate gene expression during cell differentiation. The aldehyde retinal is a light-sensitive compound with an important role in vision. Retinal is the prosthetic group of the protein rhodopsin; absorption of a photon of light by retinal triggers a neural impulse.

B. Vitamin D

Vitamin D is the collective name for a group of related lipids. Vitamin D_3 (cholecalciferol) is formed nonenzymatically in the skin from the steroid 7-dehydrocholesterol when humans are exposed to sufficient sunlight. Vitamin D_2, a compound related to vitamin D_3 (D_2 has an additional methyl group), is the additive in fortified milk. The active form of vitamin D_3, 1,25-dihydroxycholecalciferol, is formed from vitamin D_3 by two hydroxylation reactions (Figure 7.31); vitamin D_2 is similarly activated. The active compounds are hormones that help control Ca^{2+} utilization in humans—vitamin D regulates both intestinal absorption of calcium and its deposition in bones. In vitamin D–deficiency diseases, such as rickets in children and osteomalacia in adults, bones are weak because calcium phosphate does not properly crystallize on the collagen matrix of the bones.

C. Vitamin E

Vitamin E, or α-tocopherol (Figure 7.32), is one of several closely related tocopherols, compounds having a bicyclic oxygen-containing ring system with a hydrophobic side chain. The phenol group of vitamin E can undergo oxidation to a stable free radical. Vitamin E is believed to function as a reducing agent that scavenges oxygen and free radicals. This antioxidant action may prevent damage to fatty acids in biological membranes. A deficiency of vitamin E is rare but may lead to fragile red blood cells and neurological damage. The deficiency is almost always caused by genetic defects in absorption of fat molecules. There is currently no scientific evidence to support claims that vitamin E supplements in the diet of normal, healthy individuals will improve health.

D. Vitamin K

Phylloquinone (vitamin K) are important components of photosynthesis reaction centers in bacteria, algae, and plants.

Vitamin K (phylloquinone) (Figure 7.32) is a lipid vitamin from plants that is required for the synthesis of some of the proteins involved in blood coagulation. It is a coenzyme for a mammalian carboxylase that catalyzes the conversion of specific glutamate residues to γ-carboxyglutamate residues (Equation 7.7). The reduced (hydroquinone) form of vitamin K participates in the carboxylation as a reducing agent. Oxidized vitamin K has to be regenerated in order to support further modifications of clotting factors. This is accomplished by vitamin K reductase.

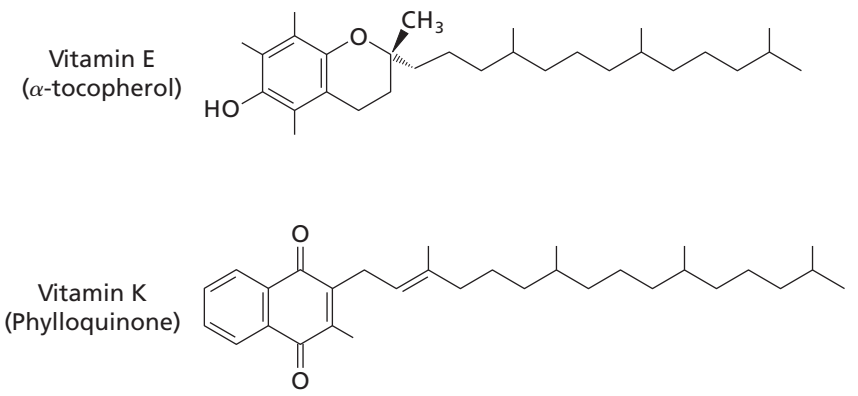
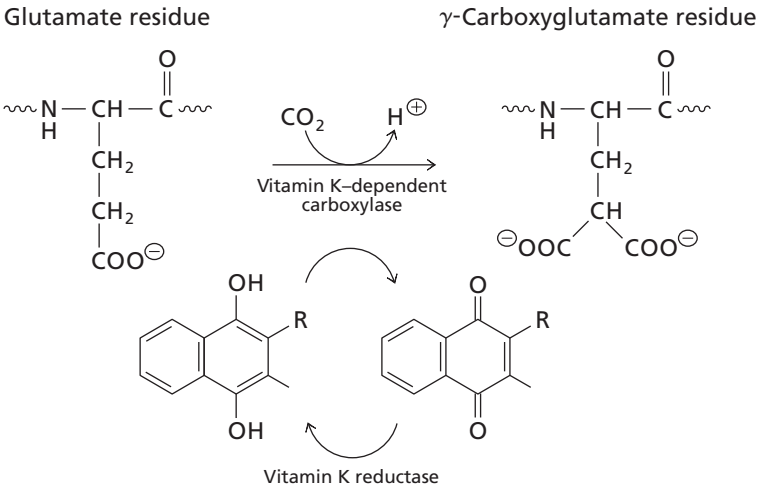


Figure 7.32 ►
Structures of vitamin E and vitamin K.

▲ **Vitamin D and the evolution of skin color.** Black skin protects cells from damage by sunlight but it may inhibit formation of vitamin D. This isn't a problem in Nairobi, Kenya (left) but it might be in Stockholm, Sweden (right). One hypothesis for the evolution of skin color suggests that light-colored skin evolved in northern climates in order to increase vitamin D production.



(7.7)

When calcium binds to the γ-carboxyglutamate residues of the coagulation proteins, the proteins adhere to platelet surfaces where many steps of the coagulation process take place.

7.15 Ubiquinone

Ubiquinone—also called coenzyme Q and therefore abbreviated "Q"—is a lipid-soluble coenzyme synthesized by almost all species. Ubiquinone is a benzoquinone with four substituents, one of which is a long hydrophobic chain. This chain of 6 to 10 isoprenoid units allows ubiquinone to dissolve in lipid membranes. In the membrane, ubiquinone transports electrons between enzyme complexes. Some bacteria use menaquinone instead of ubiquinone (Figure 7.33 a). An analog of ubiquinone, plastoquinone (Figure 7.33b), serves a similar function in photosynthetic electron transport in chloroplasts (Chapter 15).

Ubiquinone is a stronger oxidizing agent than either $NAD^{\oplus}$ or the flavin coenzymes. Consequently, it can be reduced by NADH or $FADH_2$. Like FMN and FAD, ubiquinone can accept or donate two electrons one at a time because it has three oxidation states: oxidized Q, a partially reduced semiquinone free radical, and fully reduced QH_2, called ubiquinol (Figure 7.34). Coenzyme Q plays a major role in membrane-associated electron transport. It is responsible for moving protons from one side of the membrane to the other by a process known as the Q cycle. (Chapter 14). The resulting proton gradient contributes to ATP synthesis.

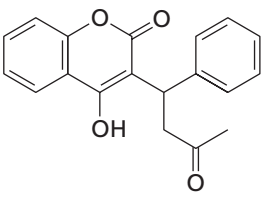
BOX 7.4 RAT POISON

Warfarin is an effective rat poison that has been used for many decades. It's a competitive inhibitor of vitamin K reductase, the enzyme that regenerates the reduced form of vitamin K (Equation 7.7). Blocking the formation of blood clotting factors leads to death in the rodents by internal bleeding. Rodents are very sensitive to inhibition of vitamin K reductase.

Later on it was discovered that low concentrations of warfarin were effective in individuals who suffer from excessive blood clotting. The drug was renamed (e.g., Coumadin®) for use in humans since its association with rat poison had a somewhat negative connotation.

Vitamin K analogs are widely used as anticoagulants in patients who are prone to thrombosis where they can prevent strokes and other embolisms. Like all medications, the dosage must be carefully regulated and controlled in order to prevent adverse effects, but in this case the dosage is even more critical.

Since the drugs only affect the synthesis of new clotting factors, they often take several days to have an effect.This is why patients will often be started at low dosages of these analogs and the amount of drug will be increased slowly over the course of many months.



▲ Warfarin.

▲ A rat (*Rattus norvegicus*).

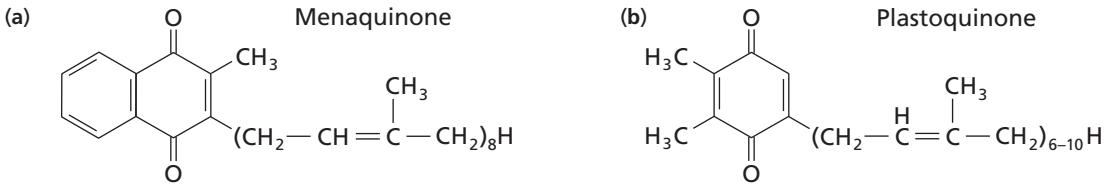


Figure 7.33 ►
Structures of (a) menaquinone and (b) plastoquinone. The hydrophobic tail of each molecule is composed of 6 to 10 five-carbon isoprenoid units.

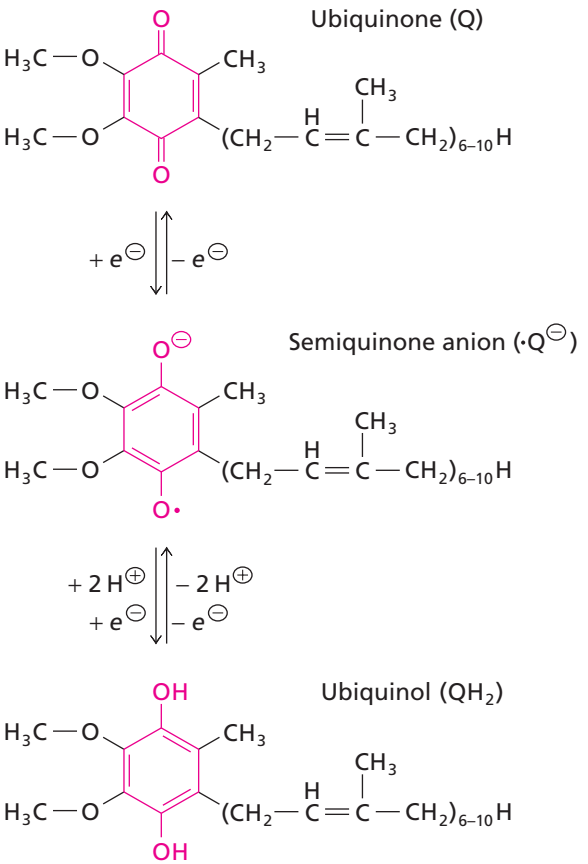


Figure 7.34 ►
Three oxidation states of ubiquinone. Ubiquinone is reduced in two one-electron steps via a semiquinone free-radical intermediate. The reactive center of ubiquinone is shown in red.

Unlike FAD or FMN, ubiquinone and its derivatives cannot accept or donate a pair of electrons in a single step.

The strength of coenzyme oxidizing agents (standard reduction potential) is described in Section 10.9.

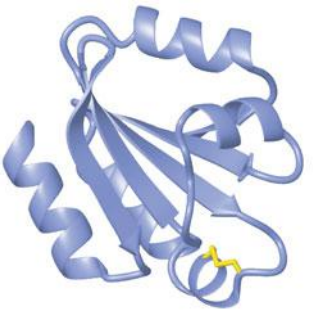
7.16 Protein Coenzymes

Some proteins act as coenzymes. They do not catalyze reactions by themselves but are required by certain other enzymes. These coenzymes are called either group transfer proteins or **protein coenzymes**. They contain a functional group either as part of their protein backbone or as a prosthetic group. Protein coenzymes are generally smaller and more heat-stable than most enzymes. They are called *coenzymes* because they participate in many different reactions and associate with a variety of different enzymes.

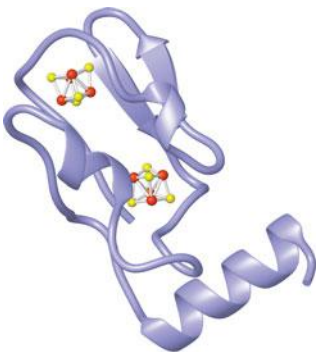
Some protein coenzymes participate in group transfer reactions or in oxidation–reduction reactions in which the transferred group is hydrogen or an electron. Metal ions, iron-sulfur clusters, and heme groups are reactive centers commonly found in these protein coenzymes. (Cytochromes are an important class of protein coenzymes that contain heme prosthetic groups. See Section 7.17.) Several protein coenzymes have two reactive thiol side chains that cycle between their dithiol and disulfide forms. For example, thioredoxins have cysteines three residues apart (—Cys—X—X—Cys—). The thiol side chains of these cysteine residues undergo reversible oxidation to form the disulfide bond of a cystine unit. We will encounter thioredoxins as reducing agents when we examine the citric acid cycle (Chapter 13), photosynthesis (Chapter 15), and deoxyribonucleotide synthesis (Chapter 18). The disulfide reactive center of thioredoxin is on the surface of the protein where it is accessible to the active sites of appropriate enzymes (Figure 7.35).

Ferredoxin is another common oxidation-reduction coenzyme. It contains two iron-sulfur clusters that can accept or donate electrons (Figure 7.36).

Some other protein coenzymes contain firmly bound coenzymes or portions of coenzymes. In *Escherichia coli*, a carboxyl carrier protein containing covalently bound biotin is one of three protein components of acetyl CoA carboxylase that catalyzes the first committed step of fatty acid synthesis. (In animal acetyl CoA carboxylases, the three protein components are fused into one protein chain.) ACP, introduced in Section 7.6, contains a phosphopantetheine moiety as its reactive center. The reactions of ACP therefore resemble those of coenzyme A. ACP is a component of all fatty acid synthases that have been tested. A protein coenzyme necessary for the degradation of glycine in mammals, plants, and bacteria (Chapter 17) contains a molecule of covalently bound lipoamide as a prosthetic group.



▲ **Figure 7.35**
Oxidized thioredoxin. Note that the cystine group is on the exposed surface of the protein. The sulfur atoms are shown in yellow. See Figure 4.24m for another view of thioredoxin. [PDB 1ERU].



▲ **Figure 7.36**
Ferredoxin. This ferredoxin from *Pseudomonas aeruginosa* contains two [4 Fe–4 S] iron-sulfur clusters that can be oxidized and reduced. Ferredoxin is a common cosubstrate in many oxidation–reduction reactions. [PDB 2FGO]

7.17 Cytochromes

Cytochromes are heme-containing protein coenzymes whose Fe(III) atoms undergo reversible one-electron reduction. Some structures of cytochromes were shown in Figures 4.21 and 4.24b. Cytochromes are classified as *a*, *b*, and *c* on the basis of their visible absorption spectra. The absorption spectra of reduced and oxidized cytochrome *c* are shown in Figure 7.37. Although the most strongly absorbing band is the Soret (or γ) band, the band labeled α is used to characterize cytochromes as either *a*, *b*, or *c*. Cytochromes in the same class may have slightly different spectra; therefore, a subscript number denoting the peak wavelength of the α absorption band of the reduced cytochrome often differentiates the cytochromes of a given class (e.g., cytochrome b_{560}). Wavelengths of maximum absorption for reduced cytochromes are given in Table 7.3.

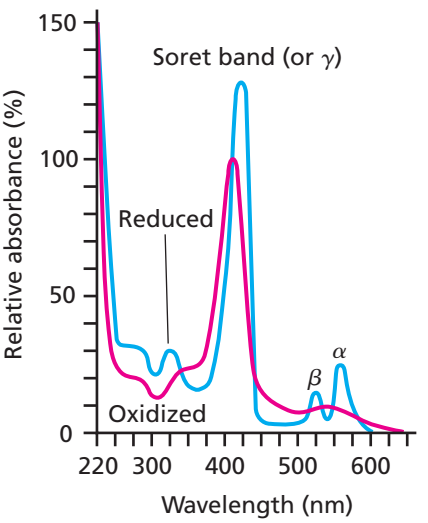


Figure 7.37 ►
Comparison of the absorption spectra of oxidized (red) and reduced (blue) horse cytochrome *c*. The reduced cytochrome has three absorbance peaks, designated α, β, and γ. On oxidation, the Soret (or γ) band decreases in intensity and shifts to a slightly shorter wavelength, whereas the α and β peaks disappear, leaving a single broad band of absorbance.

Table 7.3 Absorption maxima (in nm) of major spectral bands in the visible absorption spectra of the reduced cytochromes

	Absorption band		
Heme protein	α	β	γ
Cytochrome *c*	550–558	521–527	415–423
Cytochrome *b*	555–567	526–546	408–449
Cytochrome *a*	592–604	Absent	439–443

The classes have slightly different heme prosthetic groups (Figure 7.38). The heme of *b*-type cytochromes is the same as that of hemoglobin and myoglobin (Figure 4.44). The heme of cytochrome *a* has a 17-carbon hydrophobic chain at C-2 of the porphyrin ring and a formyl group at C-8, whereas the *b*-type heme has a vinyl group attached to C-2 and a methyl group at C-8. In *c*-type cytochromes, the heme is covalently attached to the apoprotein by two thioether linkages formed by addition of the thiol groups of two cysteine residues to vinyl groups of the heme.

The tendency to transfer an electron to another substance, measured as a reduction potential, varies among individual cytochromes. The differences arise from the different environment each apoprotein provides for its heme prosthetic group. The reduction potentials of iron-sulfur clusters also vary widely depending on the chemical and physical environment provided by the apoprotein. The range of reduction potentials among prosthetic groups is an important feature of membrane-associated electron transport pathways (Chapter 14) and photosynthesis (Chapter 15).

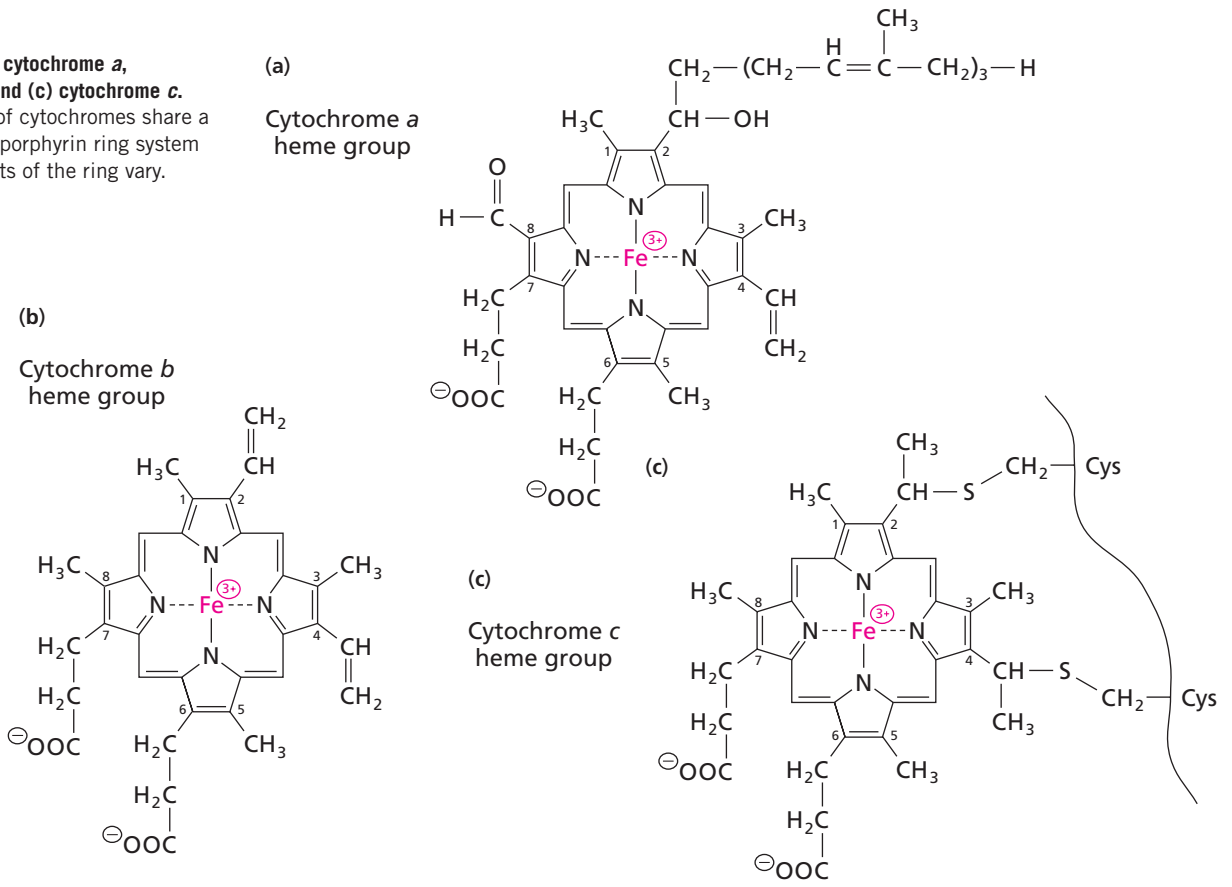


Figure 7.38 ▶
Heme groups of (a) cytochrome *a*, (b) cytochrome *b*, and (c) cytochrome *c*. The heme groups of cytochromes share a highly conjugated porphyrin ring system but the substituents of the ring vary.

BOX 7.5 NOBEL PRIZES FOR VITAMINS AND COENZYMES

The discovery of vitamins in the first part of the 20th century stimulated an enormous amount of biochemistry research. What were these mysterious chemicals that seemed essential for life? Why were they essential?

We now take vitamins and coenzymes for granted but that doesn't do justice to the workers who discovered their role in metabolism. Here's a list of the scientists who received Nobel Prizes for their work on vitamins and coenzymes.

Chemistry 1928: **Adolf Otto Reinhold Windaus** "for the services rendered through his research into the constitution of the sterols and their connection with the vitamins."

Physiology or Medicine 1929: **Christiaan Eijkman** "for his discovery of the antineuritic vitamin." **Sir Frederick Gowland Hopkins** "for his discovery of the growth-stimulating vitamins."

Chemistry 1937: **Paul Karrer** "for his investigations on carotenoids, flavins and vitamins A and B_2." **Walter Norman Haworth** "for his investigations on carbohydrates and vitamin C."

Physiology or Medicine 1937: **Albert von Szent-Györgyi Nagyrapolt** "for his discoveries in connection with the biological combustion processes, with special reference to vitamin C and the catalysis of fumaric acid."

Chemistry 1938: **Richard Kuhn** "for his work on carotenoids and vitamins."

Physiology or Medicine 1943: **Henrik Carl Peter Dam** "for his discovery of vitamin K." **Edward Adelbert Doisy** "for his discovery of the chemical nature of vitamin K."

Physiology or Medicine 1953: **Fritz Albert Lipmann** "for his discovery of co-enzyme A and its importance for intermediary metabolism."

Chemistry 1964: **Dorothy Crowfoot Hodgkin** "for her determinations by X-ray techniques of the structures of important biochemical substances."

Chemistry 1970: **Luis F. Leloir** "for his discovery of sugar nucleotides and their role in the biosynthesis of carbohydrates."

Chemistry 1997: **Paul D. Boyer and John E. Walker** "for their elucidation of the enzymatic mechanism underlying the synthesis of adenosine triphosphate (ATP)."

▲ **Nobel Medals.** Chemistry (left), Physiology or Medicine (right).

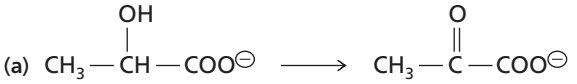
Summary

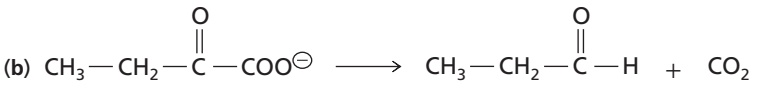
1. Many enzyme-catalyzed reactions require cofactors. Cofactors include essential inorganic ions and group-transfer reagents called coenzymes. Coenzymes can either function as cosubstrates or remain bound to enzymes as prosthetic groups.
2. Inorganic ions, such as $K^{\oplus}$, $Mg^{2\oplus}$, $Ca^{2\oplus}$, $Zn^{2\oplus}$, and $Fe^{3\oplus}$, may participate in substrate binding or in catalysis.
3. Some coenzymes are synthesized from common metabolites; others are derived from vitamins. Vitamins are organic compounds that must be supplied in small amounts in the diets of humans and other animals.
4. The pyridine nucleotides, $NAD^{\oplus}$ and $NADP^{\oplus}$, are coenzymes for dehydrogenases. Transfer of a hydride ion ($H^{\ominus}$) from a specific substrate reduces $NAD^{\oplus}$ or $NADP^{\oplus}$ to NADH or NADPH, respectively, and releases a proton.
5. The coenzyme forms of riboflavin—FAD and FMN—are tightly bound as prosthetic groups. FAD and FMN are reduced by hydride (two-electron) transfers to form $FADH_2$ and $FMNH_2$, respectively. The reduced flavin coenzymes donate electrons one or two at a time.
6. Coenzyme A, a derivative of pantothenate, participates in acyl-group–transfer reactions. Acyl carrier protein is required in the synthesis of fatty acids.
7. The coenzyme form of thiamine is thiamine diphosphate (TDP), whose thiazolium ring binds the aldehyde generated on decarboxylation of an α-keto acid substrate.
8. Pyridoxal 5′-phosphate is a prosthetic group for many enzymes in amino acid metabolism. The aldehyde group at C-4 of PLP forms a Schiff base with an amino acid substrate, through which it stabilizes a carbanion intermediate.
9. Vitamin C is a vitamin but not a coenzyme. It's a substrate in several reactions including those required in the synthesis of collagen. Vitamin C deficiency causes scurvy. Primates need an external source of vitamin C because they have lost one of the key enzymes required for its synthesis. The gene for this enzyme is a pseudogene in certain primate genomes.
10. Biotin, a prosthetic group for several carboxylases and carboxyltransferases, is covalently linked to a lysine residue at the enzyme active site.
11. Tetrahydrofolate is a reduced derivative of folate and participates in the transfer of one-carbon units at the oxidation levels of methanol, formaldehyde, and formic acid. Tetrahydrobiopterin is a reducing agent in some hydroxylation reactions.
12. The coenzyme forms of cobalamin—adenosylcobalamin and methylcobalamin—contain cobalt and a corrin ring system. These coenzymes participate in a few intramolecular rearrangements and methylation reactions.
13. Lipoamide, a prosthetic group for α-keto acid dehydrogenase multienzyme complexes, accepts an acyl group, forming a thioester.
14. The four fat-soluble, or lipid, vitamins are A, D, E, and K. These vitamins have diverse functions.

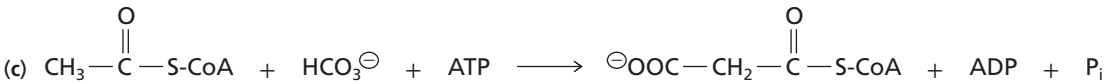
15. Ubiquinone is a lipid-soluble electron carrier that transfers electrons one or two at a time.
16. Some proteins, such as acyl carrier protein and thioredoxin, act as coenzymes in group-transfer reactions or in oxidation–reduction reactions in which the transferred group is hydrogen or an electron.
17. Cytochromes are small, heme-containing protein coenzymes that participate in electron transport. They are differentiated by their absorption spectra.

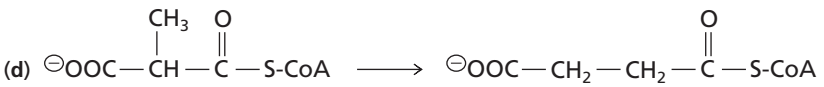
Problems

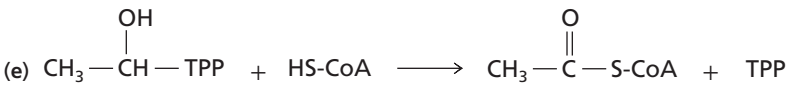
1. For each of the following enzyme-catalyzed reactions, determine the type of reaction and the coenzyme that is likely to participate.



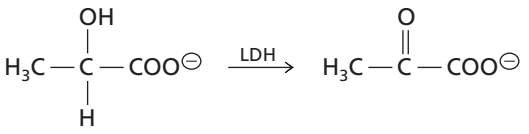






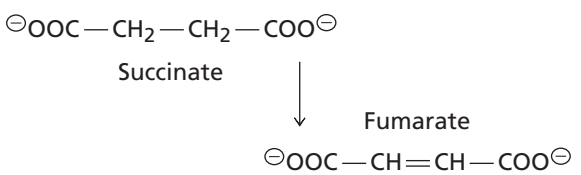


2. List the coenzymes that
   (a) participate as oxidation–reduction reagents.
   (b) act as acyl carriers.
   (c) transfer methyl groups.
   (d) transfer groups to and from amino acids.
   (e) are involved in carboxylation or decarboxylation reactions.
3. In the oxidation of lactate to pyruvate by lactate dehydrogenase (LDH), $NAD^{\oplus}$ is reduced in a two-electron transfer process from lactate. Since two protons are removed from lactate as well, is it correct to write the reduced form of the coenzyme as $NADH_2$? Explain.

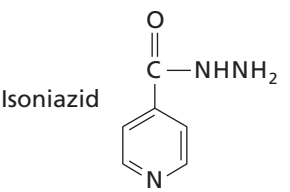


L-Lactate Pyruvate

4. Succinate dehydrogenase requires FAD to catalyze the oxidation of succinate to fumarate in the citric acid cycle. Draw the isoalloxazine ring system of the cofactor resulting from the oxidation of succinate to fumarate and indicate which hydrogens in $FADH_2$ are lacking in FAD.



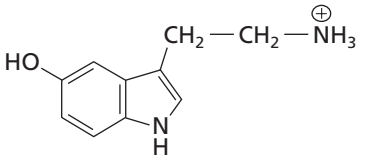
5. What is the common structural feature of $NAD^{\oplus}$, FAD, and coenzyme A?
6. Certain nucleophiles can *add* to C-4 of the nicotinamide ring of $NAD^{\oplus}$, in a manner similar to the addition of a hydride in the reduction of $NAD^{\oplus}$ to NADH. Isoniazid is the most widely used drug for the treatment of tuberculosis. X-ray studies have shown that isoniazid inhibits a crucial enzyme in the tuberculosis bacterium where a covalent adduct is formed between the carbonyl of isoniazid and the 4′ position of the nicotinamide ring of a bound $NAD^{\oplus}$ molecule. Draw the structure of this NAD-isoniazid inhibitory adduct.



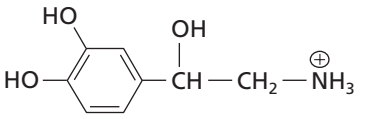
Isoniazid

7. A vitamin B_6 deficiency in humans can result in irritability, nervousness, depression, and sometimes convulsions. These symptoms may result from decreased levels of the neurotransmitters serotonin and norepinephrine, which are metabolic derivatives of tryptophan and tyrosine, respectively. How could a deficiency of vitamin B_6 result in decreased levels of serotonin and norepinephrine?

Serotonin



Norepinephrine



8. Macrocytic anemia is a disease in which red blood cells mature slowly due to a decreased rate of DNA synthesis. The red blood cells are abnormally large (macrocytic) and are more easily ruptured. How could the anemia be caused by a deficiency of folic acid?

9. A patient suffering from methylmalonic aciduria (high levels of methylmalonic acid) has high levels of homocysteine and low levels of methionine in the blood and tissues. Folic acid levels are normal.

(a) What vitamin is likely to be deficient?

(b) How could the deficiency produce the symptoms listed above?

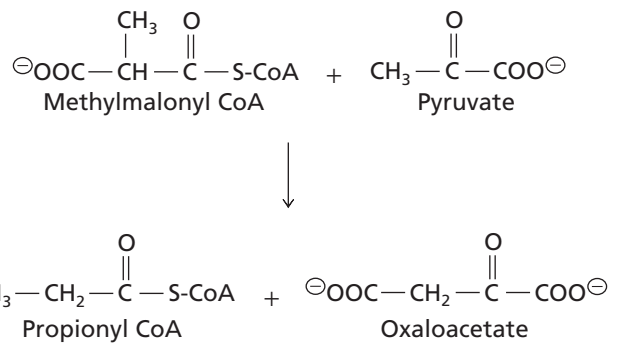
(c) Why is this vitamin deficiency more likely to occur in a person who follows a strict vegetarian diet?

10. Alcohol dehydrogenase (ADH) from yeast is a metalloenzyme that catalyzes the $NAD^{\oplus}$-dependent oxidation of ethanol to acetaldehyde. The mechanism of yeast ADH is similar to that of lactate dehydrogenase (LDH) (Figure 7.9) except that the zinc ion of ADH occupies the place of His-195 in LDH.

(a) Draw a mechanism for the oxidation of ethanol to acetaldehyde by yeast ADH.

(b) Does ADH require a residue analogous to Arg-171 in LDH?

11. In biotin-dependent transcarboxylase reactions, an enzyme transfers a carboxyl group between substrates in a two-step process without the need for ATP or bicarbonate. The reaction catalyzed by the enzyme methylmalonyl CoA-pyruvate transcarboxylase is shown below. Draw the structures of the products expected from the first step of the reaction.



12. (a) Histamine is produced from histidine by the action of a decarboxylase. Draw the external aldimine produced by the reaction of histidine and pyridoxal phosphate at the active site of histidine decarboxylase.

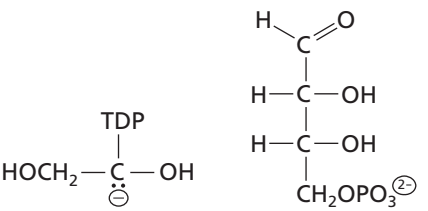
(b) Since racemization of amino acids by PLP-dependent enzymes proceeds via Schiff base formation, would racemization of L-histidine to D-histidine occur during the histidine decarboxylase reaction?

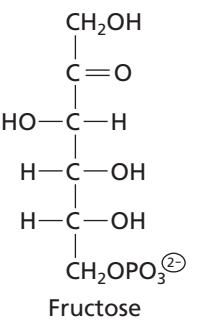
13. (a) Thiamine pyrophosphate is a coenzyme for oxidative decarboxylation reactions in which the keto carbonyl carbon is oxidized to an acid or an acid derivative. Oxidation occurs by removal of two electrons from a resonance-stabilized carbanion intermediate. What is the mechanism for the reaction pyruvate + HS-CoA → acetyl CoA + CO_2, beginning from the resonance-stabilized carbanion intermediate formed after decarboxylation (Figure 7.15) (such as a thioester in the case below)?

(b) Pyruvate dehydrogenase (PDH) is an enzyme complex that catalyzes the oxidative decarboxylation of pyruvate to acetyl CoA and CO_2 in a multistep reaction. The oxidation and acetyl-group transfer steps require TDP and lipoic acid in addition to other coenzymes. Draw the chemical structures for the molecules in the following two steps in the PDH reaction.

HETDP + lipoamide ⟶ acetyl-TDP + dihydrolipoamide ⟶
TDP + acetyl-dihydrolipoamide

(c) In a transketolase enzyme TDP-dependent reaction, the resonance-stabilized carbanion intermediate shown adjacent is generated as an intermediate. This intermediate is then involved in a condensation reaction (resulting in C—C bond formation) with the aldehyde group of erythrose 4-phosphate (E4P) to form fructose 6-phosphate (F6P). Starting from the carbanion intermediate, show a mechanism for this transketolase reaction. (Fischer projections of carbohydrate structures are sometimes drawn as shown here.)





Selected Readings

Metal Ions

Berg, J. M. (1987). Metal ions in proteins: structural and functional roles. *Cold Spring Harbor Symp. Quant. Biol.* 52:579–585.

Rees, D. C. (2002). Great metalloclusters in enzymology. *Annu. Rev. Biochem.* 71: 221–246.

Specific Cofactors

Banerjee, R., and Ragsdale, S.W. (2003). The many faces of vitamin B_{12}: catalysis by cobalmin-dependent enzymes. *Annu. Rev. Biochem.* 72:209–247.

Bellamacina, C. R. (1996). The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. *FASEB J.* 10:1257–1268.

Blakley, R. L., and Benkovic, S. J., eds. (1985). *Folates and Pterins*, Vol. 1 and Vol. 2. (New York: John Wiley & Sons).

Chiang, P. K., Gordon, R. K., Tal, J., Zeng, G. C., Doctor, B. P., Pardhasaradhi, K., and McCann, P. P. (1996). *S*-Adenosylmethionine and methylation. *FASEB J.* 10:471–480.

Coleman, J. E. (1992). Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins. *Annu. Rev. Biochem.* 61:897–946.

Ghisla, S., and Massey, V. (1989). Mechanisms of flavoprotein-catalyzed reactions. *Eur. J. Biochem.* 181:1–17.

Hayashi, H., Wada, H., Yoshimura, T., Esaki, N., and Soda, K. (1990). Recent topics in pyridoxal 5′-phosphate enzyme studies. *Annu. Rev. Biochem.* 59:87–110.

Jordan, F. (1999). Interplay of organic and biological chemistry in understanding coenzyme mechanisms: example of thiamin diphosphate-dependent decarboxylations of 2-oxo acids. *FEBS Lett.* 457:298–301.

Jordan, F., Li, H., and Brown, A. (1999). Remarkable stabilization of zwitterionic intermediates may account for a billion-fold rate acceleration by thiamin diphosphate-dependent decarboxylases. *Biochem.* 38:6369–6373.

Jurgenson, C. T., Begley, T. P. and Ealick, S. E. (2009). The structural and biochemical foundations of thiamin biosynthesis. *Ann. Rev. Biochem.* 78:569–603.

Knowles, J. R. (1989). The mechanism of biotin-dependent enzymes. *Annu. Rev. Biochem.* 58:195–221.

Ludwig, M. L., and Matthews, R. G. (1997). Structure-based perspectives on B_{12}-dependent enzymes. *Annu. Rev. Biochem.* 66:269–313.

Palfey, B. A., Moran, G. R., Entsch, B., Ballou, D. P., and Massey, V. (1999). Substrate recognition by "password" in *p*-hydroxybenzoate hydroxylase. *Biochem.* 38:1153–1158.

NAD-Binding Motifs

Bellamacina, C. R. (1996). The nictotinamide d inucleotide binding motif: a comparison of nucleotide binding proteins. *FASEB J.* 10:1257–1269.

Rossman, M. G., Liljas, A., Brändén, C.-I., and Banaszak, L. J. (1975). Evolutionary and structural relationships among dehydrogenases. In *The Enzymes.* Vol. 11, Part A, 3rd ed., P. D., Boyer, ed. (New York: Academic Press), pp. 61–102.

Wilks, H. M., Hart, K. W., Feeney, R., Dunn, C. R., Muirhead, H., Chia, W. N., Barstow, D. A., Atkinson, T., Clarke, A. R., and Holbrook, J. J. (1988). A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science* 242:1541–1544.

Carbohydrates

Carbohydrates (also called saccharides) are—on the basis of mass—the most abundant class of biological molecules on Earth. Although all organisms can synthesize carbohydrate, much of it is produced by photosynthetic organisms, including bacteria, algae, and plants. These organisms convert solar energy to chemical energy that is then used to make carbohydrate from carbon dioxide. Carbohydrates play several crucial roles in living organisms. In animals and plants, carbohydrate polymers act as energy storage molecules. Animals can ingest carbohydrates that can then be oxidized to yield energy for metabolic processes. Polymeric carbohydrates are also found in cell walls and in the protective coatings of many organisms. Other carbohydrate polymers are marker molecules that allow one type of cell to recognize and interact with another type. Carbohydrate derivatives are found in a number of biological molecules, including some coenzymes (Chapter 7) and the nucleic acids (Chapter 19).

The name *carbohydrate*, "hydrate of carbon," refers to their empirical formula $(CH_2O)_n$, where n is 3 or greater (n is usually 5 or 6 but can be up to 9). Carbohydrates can be described by the number of monomeric units they contain. **Monosaccharides** are the smallest units of carbohydrate structure. **Oligosaccharides** are polymers of two to about 20 monosaccharide residues. The most common oligosaccharides are disaccharides, which consist of two linked monosaccharide residues. **Polysaccharides** are polymers that contain many (usually more than 20) monosaccharide residues. Oligosaccharides and polysaccharides do not have the empirical formula $(CH_2O)_n$ because water is eliminated during polymer formation. The term *glycan* is a more general term for carbohydrate polymers. It can refer to a polymer of identical sugars (homoglycan) or of different sugars (heteroglycan).

Glycoconjugates are carbohydrate derivatives in which one or more carbohydrate chains are linked covalently to a peptide, protein, or lipid. These derivatives include proteoglycans, peptidoglycans, glycoproteins, and glycolipids.

In this chapter, we discuss nomenclature, structure, and function of monosaccharides, disaccharides, and the major homoglycans—starch, glycogen, cellulose, and

Molecular biology has dealt largely on the triad of DNA, RNA and protein. Biochemistry is concerned with all the molecules of the cell. Excluded from the province of molecular biology have been most of the structures and functions essential for growth and maintenance: carbohydrates, coenzymes, lipids, and membranes.

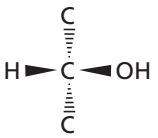
—Arthur Kornberg
"For the love of enzymes: the odyssey of a biochemist" (1989)

Photosynthesis is described in detail in Chapter 15.

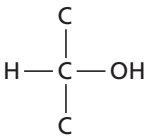
Top: Darkling beetle. The exoskeletons of insects contain chitin, a homoglycan.

KEY CONCEPT

A Fischer projection is a convention designed to convey information about the stereochemistry of a molecule. It does not resemble the actual conformation of the molecule in solution.



Stereo view



Fischer projection

For *each* chiral carbon atom in a Fischer projection the vertical bonds project into the plane of the page and the horizontal bonds project upward toward the viewer.

chitin. We then consider proteoglycans, peptidoglycans, and glycoproteins, all of which contain heteroglycan chains.

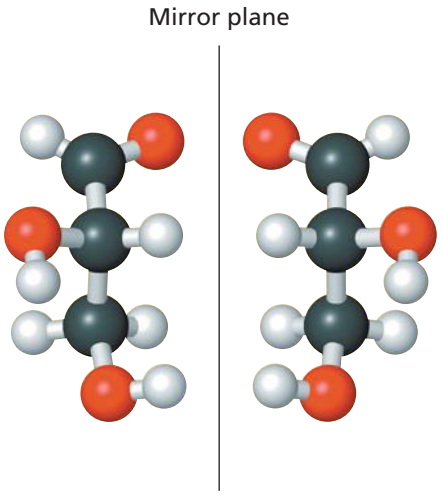
8.1 Most Monosaccharides Are Chiral Compounds

Monosaccharides are water-soluble, white, crystalline solids that have a sweet taste. Examples include glucose and fructose. Chemically, monosaccharides are polyhydroxy aldehydes, or **aldoses**, or polyhydroxy ketones, or **ketoses**. They are classified by their type of carbonyl group and their number of carbon atoms. As a rule, the suffix *-ose* is used in naming carbohydrates, although there are a number of exceptions. All monosaccharides contain at least three carbon atoms. One of these is the carbonyl carbon, and each of the remaining carbon atoms bears a hydroxyl group. In aldoses, the most oxidized carbon atom is designated C-1 and is drawn at the top of a Fischer projection. In ketoses, the most oxidized carbon atom is usually C-2.

We've encountered Fischer projections before but now it's time to present the convention in more detail. A Fischer projection is a two-dimensional representation of a three-dimensional molecule. It is designed to preserve information about the stereochemistry of a molecule. In a Fischer projection of sugars, the C-1 atom is always at the top of the figure. For each separate chiral carbon atom, the two horizontal bonds project upward from the page toward you. The two vertical bonds project downward into the page. Remember, this applies to each chiral carbon atom, so in a carbohydrate with multiple carbon atoms the Fischer projection represents a molecule that curls back into the page. For longer molecules, the top and bottom groups may even come in virtual contact, forming a loop. The Fischer projection is a convention for preserving stereochemical information; it does not represent a realistic model of how a molecule might look in solution.

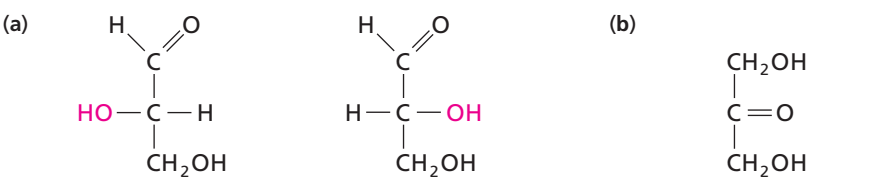
The smallest monosaccharides are **trioses**, or three-carbon sugars. One- or two-carbon compounds having the general formula $(CH_2O)_n$ do not have properties typical of carbohydrates (such as sweet taste and the ability to crystallize). The aldehydic triose, or aldotriose, is glyceraldehyde (Figure 8.1a). Glyceraldehyde is chiral because its central carbon, C-2, has four different groups attached to it, (Section 3.1). The ketonic triose, or ketotriose, is dihydroxyacetone (Figure 8.1b). It is achiral because it has no asymmetric carbon atom. All other monosaccharides, longer-chain versions of these two sugars, are chiral.

The stereoisomers D- and L-glyceraldehyde are shown as ball-and-stick models in Figure 8.2. Chiral molecules are optically active; that is, they rotate the plane of polarized light. The convention for designating D and L isomers was originally based on the optical properties of glyceraldehyde. The form of glyceraldehyde that caused rotation to the right (dextrorotatory) was designated D and the form that caused rotation to the left (levorotatory) was designated L. Structural knowledge was limited when this convention was established in the late 19th century so the configurations for the enantiomers of glyceraldehyde were assigned arbitrarily, with a 50% probability of error. X-ray crystallographic experiments later proved that the original structural assignments were correct.



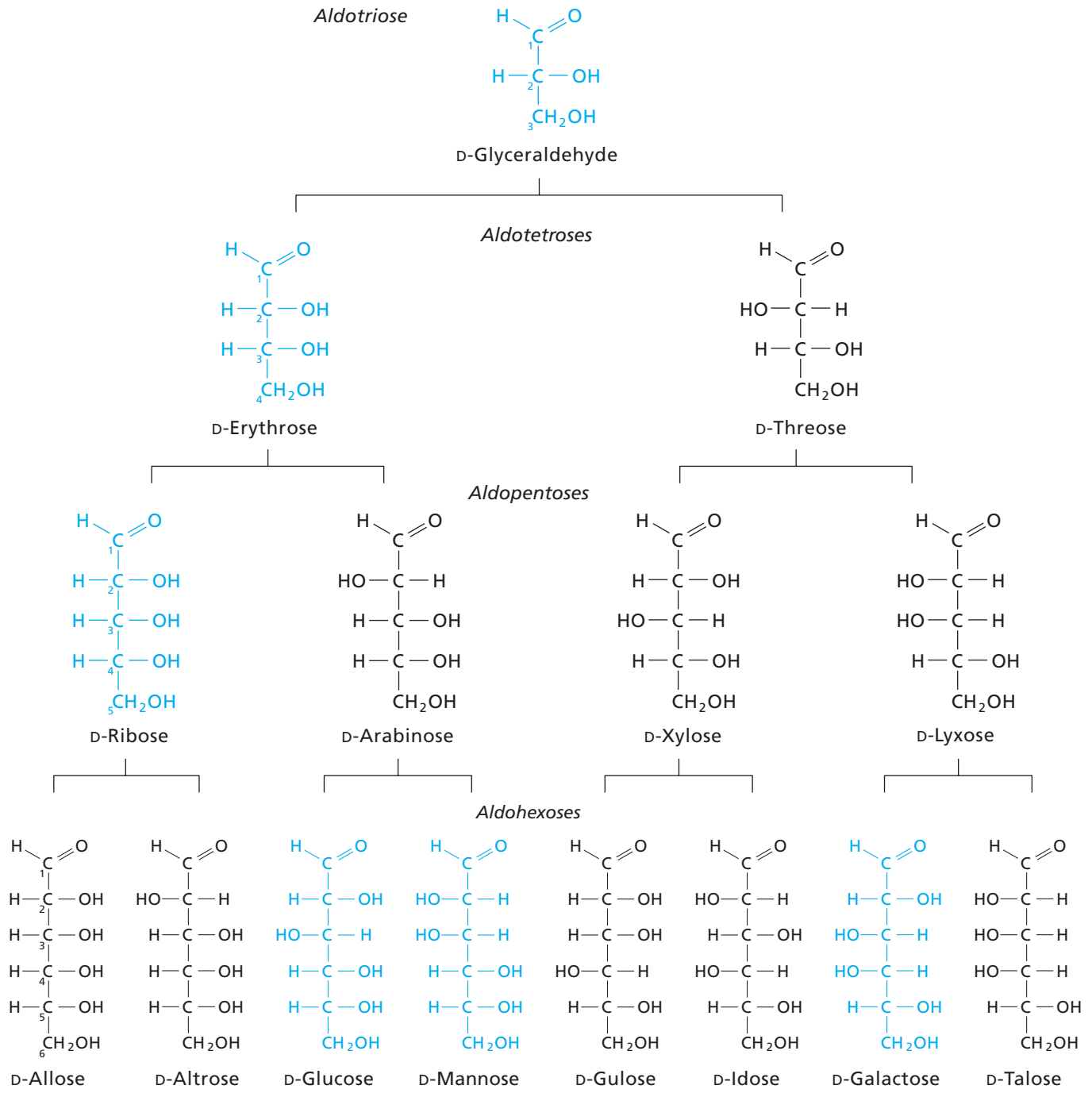
L-Glyceraldehyde D-Glyceraldehyde

▲ **Figure 8.2**
View of L-glyceraldehyde (left) and D-glyceraldehyde (right). These molecules are drawn in a conformation that corresponds to the Fischer projections in Figure 8.1.



L-Glyceraldehyde D-Glyceraldehyde Dihydroxyacetone

▲ **Figure 8.1**
Fischer projections of (a) glyceraldehyde and (b) dihydroxyacetone. The designations L (for left) and D (for right) for glyceraldehyde refer to the configuration of the hydroxyl group of the chiral carbon (C-2). Dihydroxyacetone is achiral.



▲ **Figure 8.3**
Fischer projections of the three- to six-carbon D-aldoses. The aldoses shown in blue are the most important in our study of biochemistry.

Longer aldoses and ketoses can be regarded as extensions of glyceraldehyde and dihydroxyacetone, respectively, with chiral H—C—OH groups inserted between the carbonyl carbon and the primary alcohol group. Figure 8.3 shows the complete list of the names and structures of the tetroses (four-carbon aldoses), pentoses (five-carbon aldoses), and hexoses (six-carbon aldoses) related to D-glyceraldehyde. Many of these monosaccharides are not synthesized by most organisms and we will not encounter them again in this book.

Note that the carbon atoms are numbered from the carbon of the aldehyde group that is assigned the number 1. By convention, sugars are said to have the D configuration when the configuration of the chiral carbon with the highest number—the chiral carbon most distant from the carbonyl carbon—is the same as that of C-2 of D-glyceraldehyde

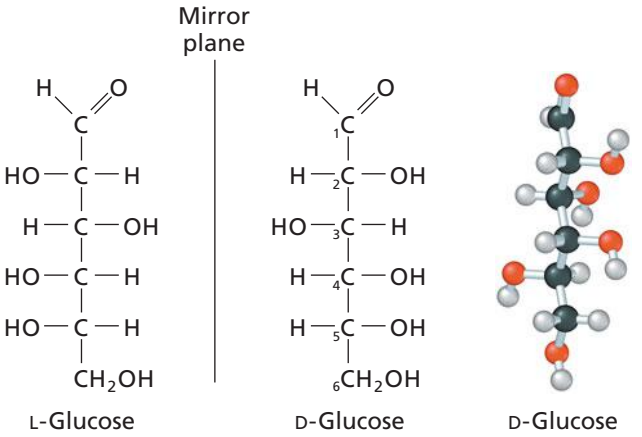


Figure 8.4 ►
L- and D-glucose. Fischer projections (left) showing that L- and D-glucose are mirror images. Conformation of the extended form of D-glucose in solution.

(i.e., the —OH group attached to this carbon atom is on the right side in a Fischer projection). The arrangement of asymmetric carbon atoms is unique for each monosaccharide, giving each its distinctive properties. Except for glyceraldehyde (which was used as the standard), there is no predictable association between the absolute configuration of a sugar and whether it is dextrorotatory or levorotatory.

It is mostly the D enantiomers that are synthesized in living cells—just as the L enantiomers of amino acids are more common. The L enantiomers of the 15 aldoses in Figure 8.3 are not shown. Recall that pairs of enantiomers are mirror images; in other words, the configuration at each chiral carbon is opposite. For example, the hydroxyl groups bound to carbon atoms 2, 3, 4, and 5 of D-glucose point right, left, right, and right, respectively, in the Fischer projection; those of L-glucose point left, right, left, and left (Figure 8.4).

The three-carbon aldose, glyceraldehyde, has only a single chiral atom (C-2) and therefore only two stereoisomers. There are four stereoisomers for aldotetroses (D- and L-erythrose and D- and L-threose) because erythrose and threose each possess two chiral carbon atoms. In general, there are 2^n possible stereoisomers for a compound with n chiral carbons. Aldohexoses, which possess four chiral carbons, have a total of 2^4, or 16, stereoisomers (the eight D aldohexoses in Figure 8.3 and their L enantiomers).

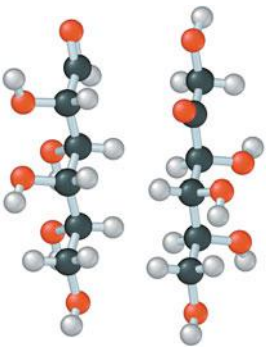
Sugar molecules that differ in configuration at only one of several chiral centers are called **epimers**. For example, D-mannose and D-galactose are epimers of D-glucose (at C-2 and C-4, respectively), although they are not epimers of each other (Figure 8.3).

Longer-chain ketoses (Figure 8.5) are related to dihydroxyacetone in the same way that longer-chain aldoses are related to glyceraldehyde. Note that a ketose has one fewer chiral carbon atom than the aldose of the same empirical formula. For example, there are only two stereoisomers for the one ketotetrose (D- and L-erythrulose), and four stereoisomers for ketopentoses (D- and L-xylulose and D- and L-ribulose). Ketotetrose and ketopentoses are named by inserting *-ul-* in the name of the corresponding aldose. For example, the ketose xylulose corresponds to the aldose xylose. This nomenclature does not apply to the ketohexoses (tagatose, sorbose, psicose, and fructose) because they have traditional (trivial) names.

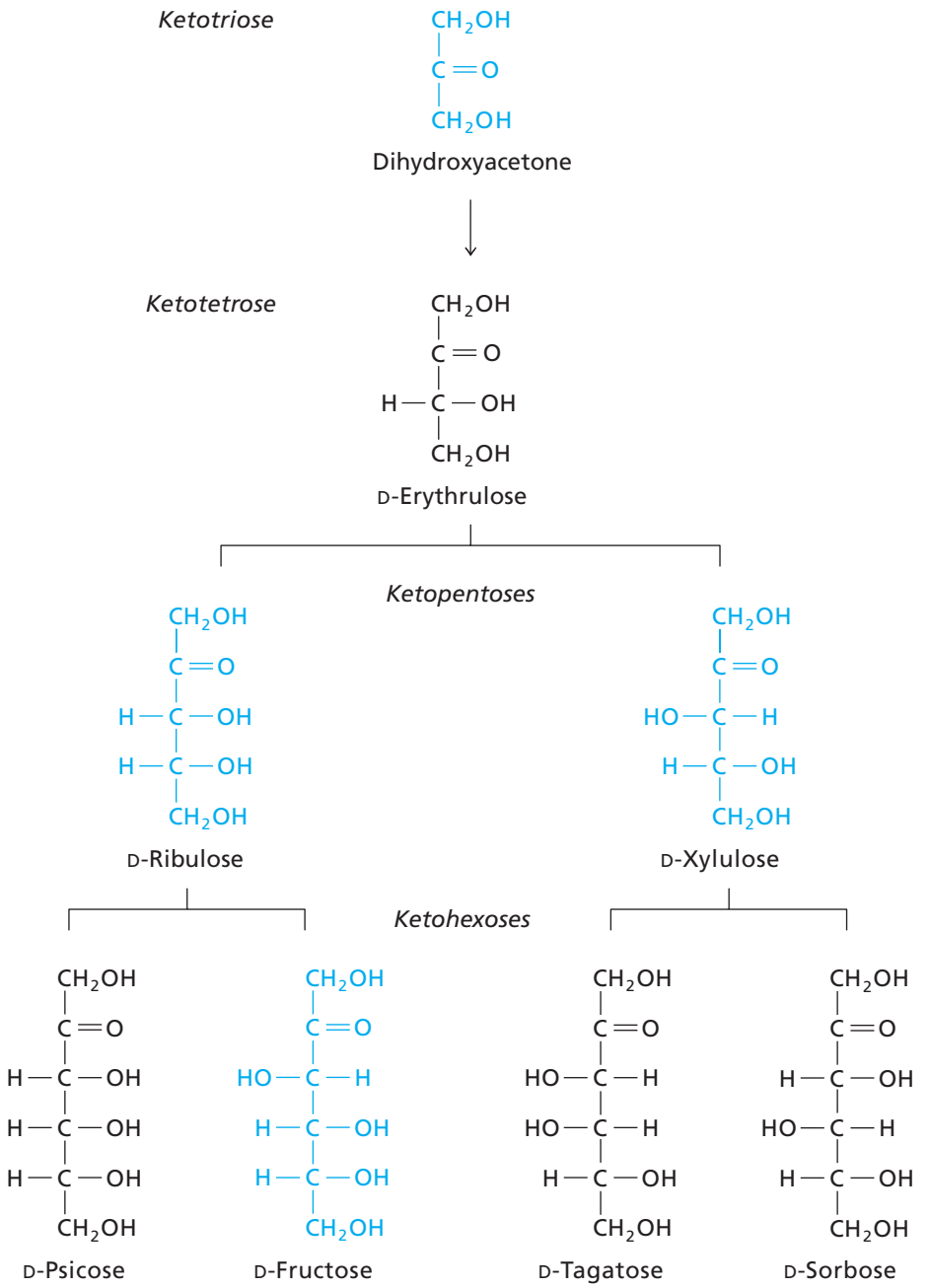
8.2 Cyclization of Aldoses and Ketoses

The optical behavior of some monosaccharides suggests they have one more chiral carbon atom than is evident from the structures shown in Figures 8.3 and 8.5. D-Glucose, for example, exists in two forms that contain five (not four) asymmetric carbons. The source of this additional asymmetry is an intramolecular cyclization reaction that produces a new chiral center at the carbon atom of the carbonyl group. This cyclization resembles the reaction of an alcohol with an aldehyde to form a hemiacetal or with a ketone to form a hemiketal (Figure 8.7).

The carbonyl carbon of an aldose containing at least five carbon atoms or of a ketose containing at least six carbon atoms can react with an intramolecular hydroxyl



▲ **Who am I?** The structures of the D sugars are shown in Figures 8.3 and 8.5. You can deduce the structures of the L configurations. Knowing the convention for Fischer projections, you should have no trouble identifying these molecules.



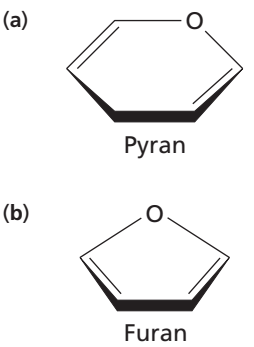
◄ **Figure 8.5**
Fischer projections of the three- to six-carbon D-ketoses. The ketoses shown in blue are the most important in our study of biochemistry.

group to form a cyclic hemiacetal or cyclic hemiketal, respectively. The oxygen atom from the reacting hydroxyl group becomes a member of the five- or six-membered ring structures (Figure 8.8).

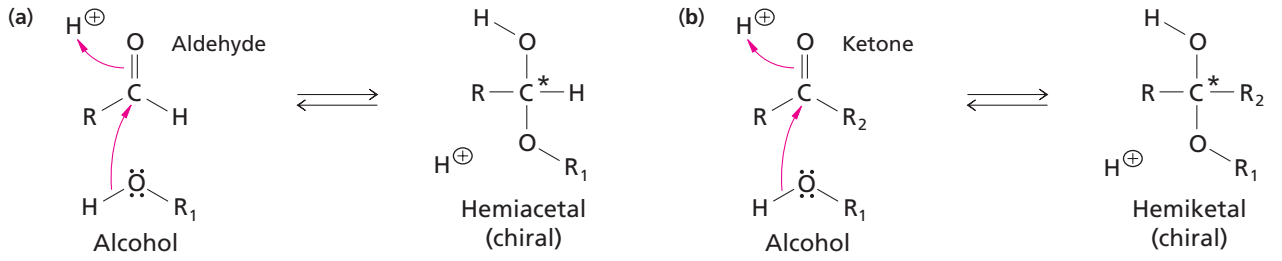
Because it resembles the six-membered heterocyclic compound pyran (Figure 8.6a), the six-membered ring of a monosaccharide is called a **pyranose**. Similarly, because the five-membered ring of a monosaccharide resembles furan (Figure 8.6b), it is called a **furanose**. Note that, unlike pyran and furan, the rings of carbohydrates do not contain double bonds.

The most oxidized carbon of a cyclized monosaccharide, the one attached to two oxygen atoms, is referred to as the **anomeric carbon**. In ring structures, the anomeric carbon is chiral. Thus, the cyclized aldose or ketose can adopt either of two configurations (designated α or β), as illustrated for D-glucose in Figure 8.8. The α and β isomers are called **anomers**.

In solution, aldoses and ketoses that form ring structures equilibrate among their various cyclic and open-chain forms. At 31°C, for example, D-glucose exists in an equilibrium



▲ **Figure 8.6**
(a) Pyran and (b) furan.

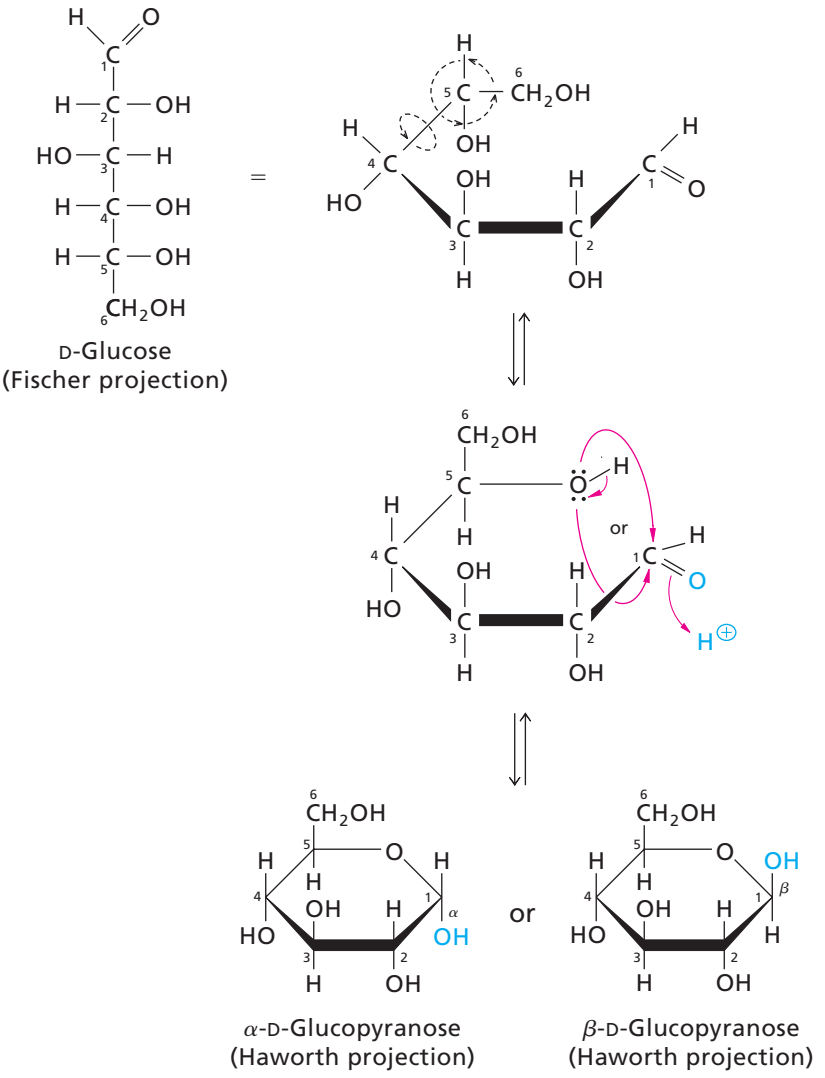


▲ **Figure 8.7**
Hemiacetal and hemiketal. (a) Reaction of an alcohol with an aldehyde to form a hemiacetal. **(b)** Reaction of an alcohol with a ketone to form a hemiketal. The asterisks indicate the newly formed chiral centers.

mixture of approximately 64% β-D-glucopyranose and 36% α-D-glucopyranose, with very small amounts of the furanose (Figure 8.9) and open-chain (Figure 8.4) forms. Similarly, D-ribose exists as a mixture of approximately 58.5% β-D-ribopyranose, 21.5% α-D-ribopyranose, 13.5% β-D-ribofuranose, and 6.5% α-D-ribofuranose, with a tiny fraction in the open-chain form (Figure 8.10). The relative abundance of the various forms of monosaccharides at equilibrium reflects the relative stabilities of each form. Although unsubstituted D-ribose is most stable as the β-pyranose, its structure in nucleotides (Section 8.5c) is the β-furanose form.

The ring drawings shown in these figures are called Haworth projections, after Norman Haworth who worked on the cyclization reactions of carbohydrates and first

Figure 8.8 ►
Cyclization of D-glucose to form glucopyranose. The Fischer projection (top left) is rearranged into a three-dimensional representation (top right). Rotation of the bond between C-4 and C-5 brings the C-5 hydroxyl group close to the C-1 aldehyde group. Reaction of the C-5 hydroxyl group with one side of C-1 gives α-D-glucopyranose; reaction of the hydroxyl group with the other side gives β-D-glucopyranose. The glucopyranose products are shown as Haworth projections in which the lower edges of the ring (thick lines) project in front of the plane of the paper and the upper edges project behind the plane of the paper. In the α-D-anomer of glucose, the hydroxyl group at C-1 points down; in the β-D-anomer, it points up.

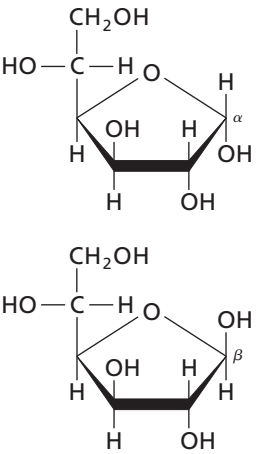


proposed these representations. He received the Nobel Prize in Chemistry in 1937 for his work on carbohydrate structure and the synthesis of vitamin C.

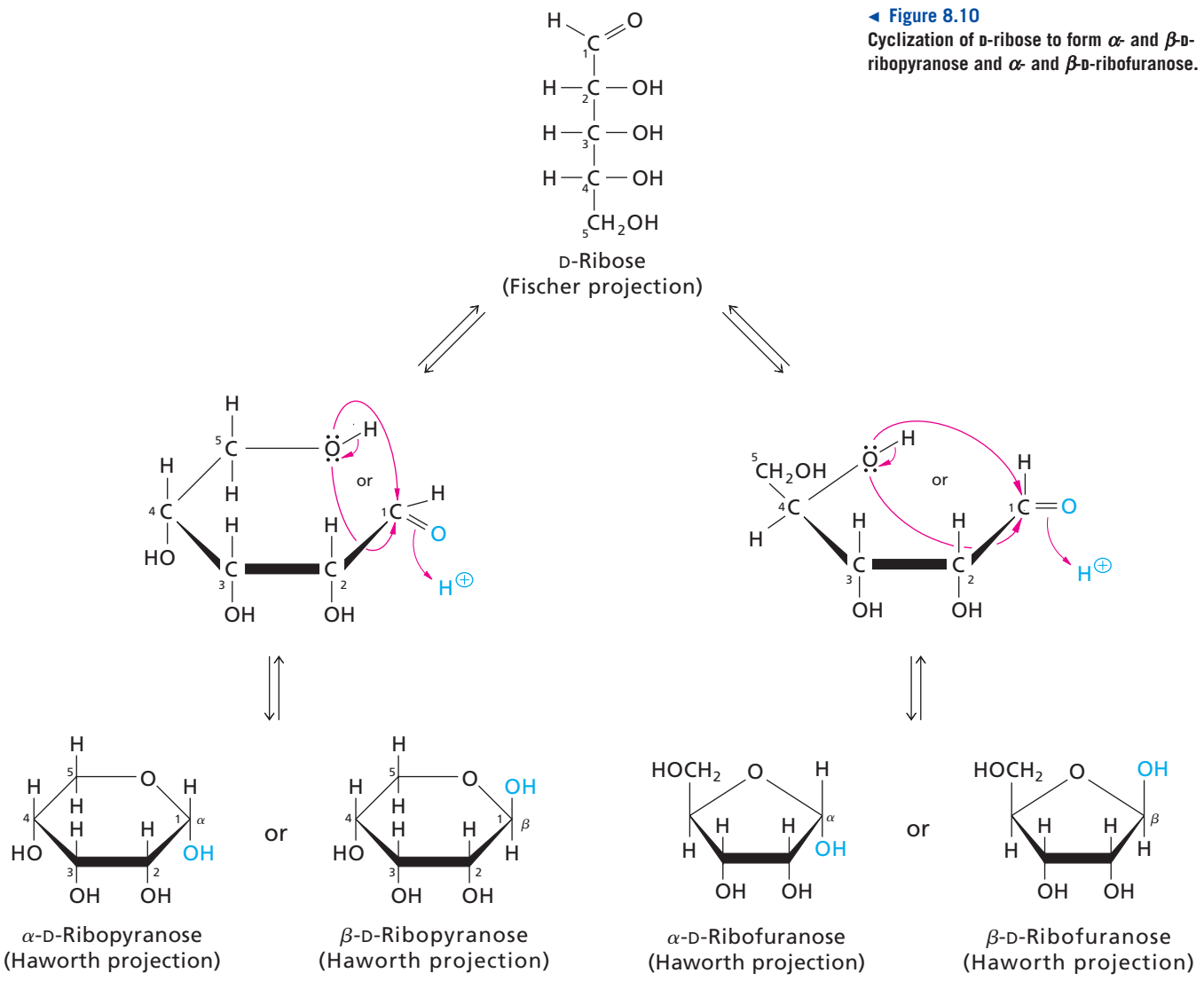
A Haworth projection adequately indicates stereochemistry and can be easily related to a Fischer projection: groups on the right in a Fischer projection point downwards in a Haworth projection. Because rotation around carbon–carbon bonds is constrained in the ring structure, the Haworth projection is a much more faithful representation of the actual conformation of sugars.

By convention, a cyclic monosaccharide is drawn so the anomeric carbon is on the right and the other carbons are numbered in a clockwise direction. In a Haworth projection, the configuration of the anomeric carbon atom is designated α if its hydroxyl group is *cis* to (on the same side of the ring as) the oxygen atom of the highest-numbered chiral carbon atom. It is β if its hydroxyl group is *trans* to (on the opposite side of the ring from) the oxygen attached to the highest-numbered chiral carbon. With α-D-glucopyranose, the hydroxyl group at the anomeric carbon points down; with β-D-glucopyranose, it points up.

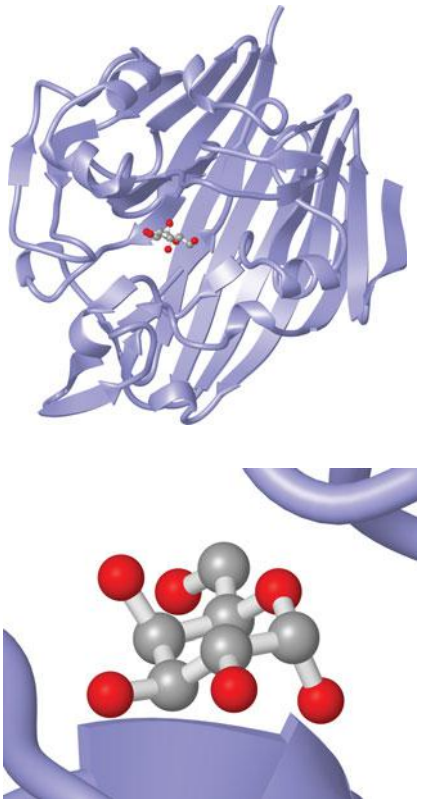
Monosaccharides are often drawn in either the α- or β-D-furanose or the α- or β-D-pyranose form. However, you should remember that the anomeric forms of five- and six-carbon sugars are in rapid equilibrium. Throughout this chapter and the rest of the book, we draw sugars in the correct anomeric form if it is known. We refer to sugars in a nonspecific way (e.g., glucose) when we are discussing an equilibrium



▲ **Figure 8.9**
α-D-glucofuranose (top) and β-D-glucofuranose (bottom).



◀ **Figure 8.10**
Cyclization of D-ribose to form α- and β-D-ribopyranose and α- and β-D-ribofuranose.



▲ **Galactose mutarotase.** Mutarotases are enzymes that catalyze the interconversion of α and β configurations. This interconversion involves the breaking and remaking of covalent bonds, which is why they are different configurations. The enzyme shown here is galactose mutarotase from *Lactococcus lactis* with a molecule of α-D-galactose in the acitve site. The bottom figure shows the *conformation* of this molecule. Can you identify this conformation? [PDB 1L7K]

mixture of the various anomeric forms as well as the open-chain forms. When we are discussing a specific form of a sugar, however, we will refer to it precisely (e.g., β-D-glucopyranose). Also, since the D enantiomers of carbohydrates predominate in nature, we always assume that a carbohydrate has the D configuration unless specified otherwise.

8.3 Conformations of Monosaccharides

Haworth projections are commonly used in biochemistry because they accurately depict the *configuration* of the atoms and groups at each carbon atom of the sugar's backbone. However, the geometry of the carbon atoms of a monosaccharide ring is tetrahedral (bond angles near 110°), so monosaccharide rings are not actually planar. Cyclic monosaccharides can exist in a variety of *conformations* (three-dimensional shapes having the same configuration). Furanose rings adopt envelope conformations in which one of the five ring atoms (either C-2 or C-3) is out-of-plane and the remaining four are approximately coplanar (Figure 8.11). Furanoses can also form twist conformations where two of the five ring atoms are out-of-plane—one on either side of the plane formed by the other three atoms. The relative stability of each conformer depends on the degree of steric interference between the hydroxyl groups. The various conformers of unsubstituted monosaccharides can rapidly interconvert.

Pyranose rings tend to assume one of two conformations, the chair conformation or the boat conformation (Figure 8.12). There are two distinct chair conformers and six distinct boat conformers for each pyranose. The chair conformations minimize steric repulsion among the ring substituents and are generally more stable than boat conformations. The —H, —OH, and —CH_2OH substituents of a pyranose ring in the chair conformation may occupy two different positions. In the axial position the substituent is above or below the plane of the ring, while in the equatorial position the substituent lies in the plane of the ring. In pyranoses, five substituents are axial and five are equatorial. Whether a group is axial or equatorial depends on which carbon atom (C-1 or C-4) extends above the plane of the ring when the ring is in the chair conformation. Figure 8.13 shows the two different chair conformers of β-D-glucopyranose. The more stable conformation is the one in which the bulkiest ring substituents are equatorial (top structure). In fact, this conformation of β-D-glucose has the least steric strain of any aldohexose. Pyranose rings are occasionally forced to adopt slightly different conformations, such as the unstable half-chair adopted by a polysaccharide residue in the active site of lysozyme (Section 6.6).

KEY CONCEPT

Different *configurations* can only be formed by breaking and reforming covalent bonds. Molecules can adopt different *conformations* without breaking covalent bonds.

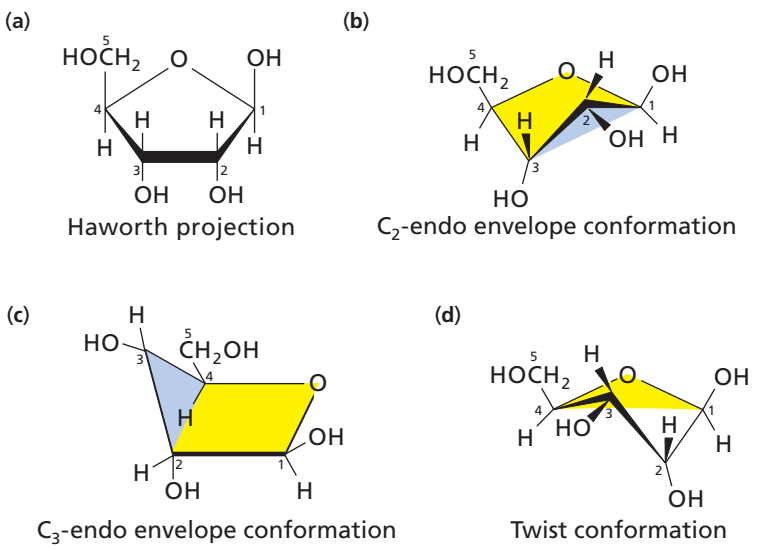
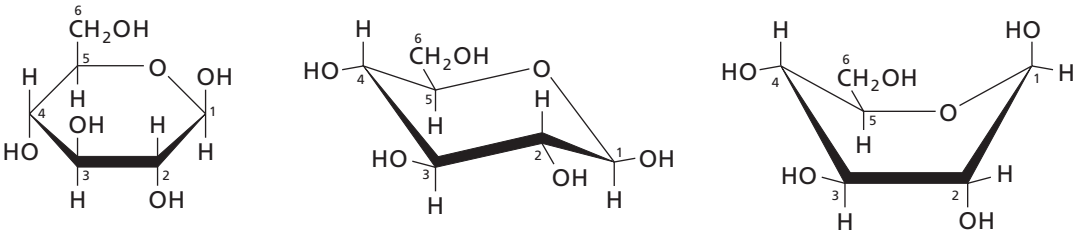
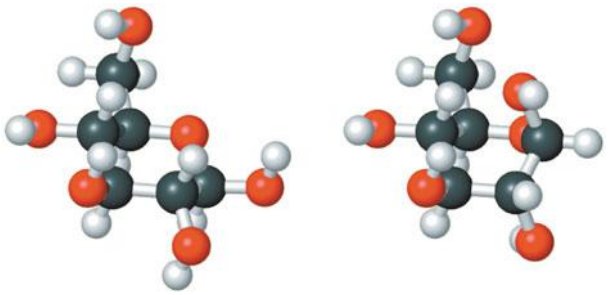


Figure 8.11 ►
Conformations of β-D-ribofuranose. (a) Haworth projection. **(b)** C_2-endo envelope conformation. **(c)** C_3-endo envelope conformation. **(d)** Twist conformation. In the C_2-endo conformation, C-2 lies above the plane defined by C-1, C-3, C-4, and the ring oxygen. In the C_3-endo conformation, C-3 lies above the plane defined by C-1, C-2, C-4, and the ring oxygen. In the twist conformation shown, C-3 lies above and C-2 lies below the plane defined by C-1, C-4, and the ring oxygen. The planes are shown in yellow.

(a)



(b)

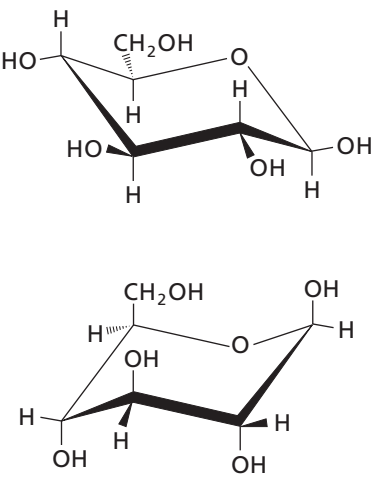


◀ **Figure 8.12**
Conformations of β-D-glucopyranose. **(a)** Haworth projection, a chair conformation, and a boat conformation. **(b)** Ball-and-stick model of a chair (left) and a boat (right) conformation.

8.4 Derivatives of Monosaccharides

There are many known derivatives of the basic monosaccharides. They include polymerized monosaccharides, such as oligosaccharides and polysaccharides, as well as several classes of nonpolymerized compounds. In this section, we introduce a few monosaccharide derivatives, including sugar phosphates, deoxy and amino sugars, sugar alcohols, and sugar acids.

Like other polymer-forming biomolecules, monosaccharides and their derivatives have abbreviations used in describing more complex polysaccharides. The accepted abbreviations contain three letters, with suffixes added in some cases. The abbreviations for some pentoses and hexoses and their major derivatives are listed in Table 8.1. We use these abbreviations later in this chapter.



▲ **Figure 8.13**
The two chair conformers of β-D-glucopyranose. The top conformer is more stable.

A. Sugar Phosphates

Monosaccharides are often converted to phosphate esters. Figure 8.14 shows the structures of several of the sugar phosphates we will encounter in our study of carbohydrate metabolism. The triose phosphates, ribose 5-phosphate, and glucose 6-phosphate are simple alcohol-phosphate esters. Glucose 1-phosphate is a hemiacetal phosphate, which is more reactive than an alcohol phosphate. The ability of UDP-glucose to act as a glucosyl donor (Section 7.3) is evidence of this reactivity.

B. Deoxy Sugars

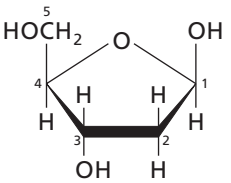
The structures of two deoxy sugars are shown in Figure 8.15. In these derivatives, a hydrogen atom replaces one of the hydroxyl groups in the parent monosaccharide. 2-Deoxy-D-ribose is an important building block for DNA. L-Fucose (6-deoxy-L-galactose) is widely distributed in plants, animals, and microorganisms. Despite its unusual L configuration, fucose is derived metabolically from D-mannose.

C. Amino Sugars

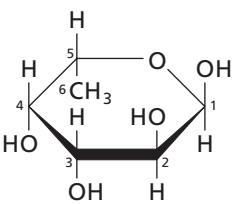
In a number of sugars, an amino group replaces one of the hydroxyl groups in the parent monosaccharide. Sometimes the amino group is acetylated. Three examples of amino

Table 8.1 Abbreviations for some monosaccharides and their derivatives

Monosaccharide or derivative	Abbreviation
Pentoses	
Ribose	Rib
Xylose	Xyl
Hexoses	
Fructose	Fru
Galactose	Gal
Glucose	Glc
Mannose	Man
Deoxy sugars	
Abequose	Abe
Fucose	Fuc
Amino sugars	
Glucosamine	GlcN
Galactosamine	GalN
N-Acetylglucosamine	GlcNAc
N-Acetylgalactosamine	GalNAc
N-Acetylneuraminic acid	NeuNAc
N-Acetylmuramic acid	MurNAc
Sugar acids	
Glucuronic acid	GlcUA
Iduronic acid	IdoA

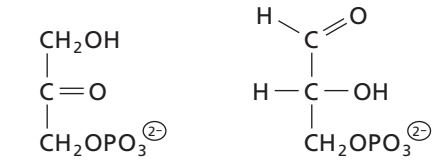


β-2-Deoxy-D-ribose



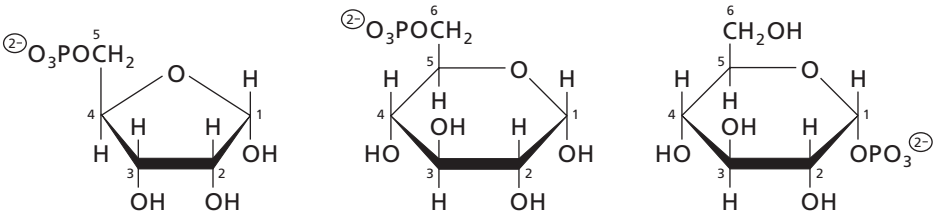
α-L-Fucose
(6-Deoxy-L-galactose)

▲ **Figure 8.15**
Structures of the deoxy sugars 2-deoxy-D-ribose and L-fucose.



Dihydroxyacetone phosphate

D-Glyceraldehyde 3-phosphate



α-D-Ribose 5-phosphate

α-D-Glucose 6-phosphate

α-D-Glucose 1-phosphate

▲ **Figure 8.14**
Structures of several metabolically important sugar phosphates.

sugars are shown in Figure 8.16. Amino sugars formed from glucose and galactose commonly occur in glycoconjugates. *N*-Acetylneuraminic acid (NeuNAc) is an acid formed from *N*-acetylmannosamine and pyruvate. When this compound cyclizes to form a pyranose, the carbonyl group at C-2 (from the pyruvate moiety) reacts with the hydroxyl group of C-6. NeuNAc is an important constituent of many glycoproteins and of a family of lipids called gangliosides (Section 9.5). Neuraminic acid and its derivatives, including NeuNAc, are collectively known as sialic acids.

D. Sugar Alcohols

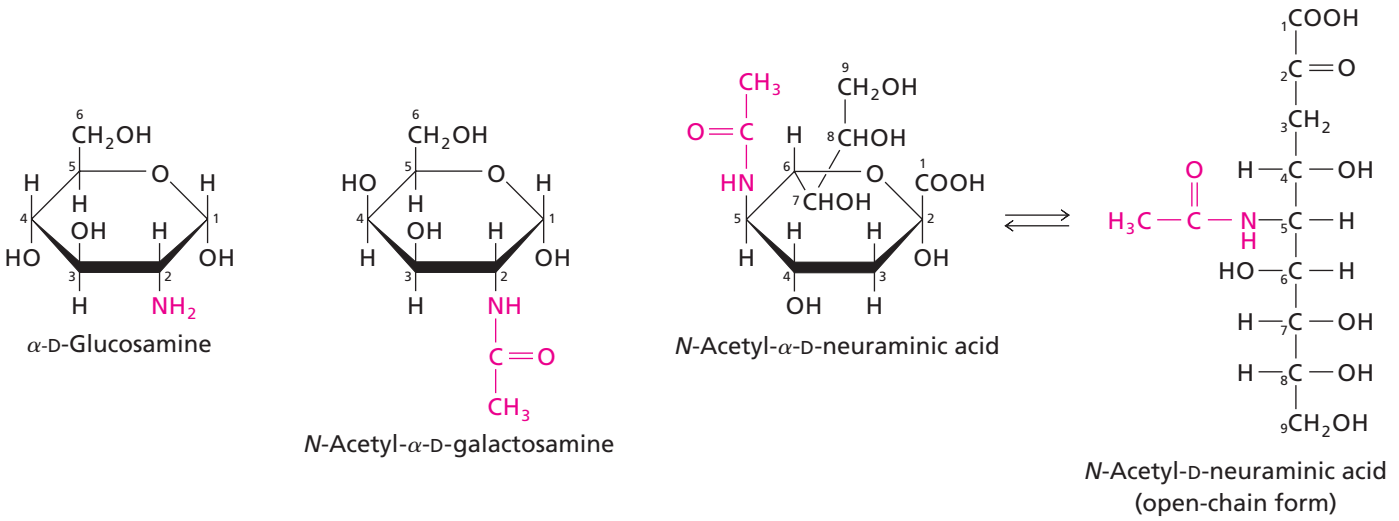
In a sugar alcohol, the carbonyl oxygen of the parent monosaccharide has been reduced, producing a polyhydroxy alcohol. Figure 8.17 shows three examples of sugar alcohols. Glycerol and *myo*-inositol are important components of lipids (Section 10.4). Ribitol is a component of flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) (Section 7.4). In general, sugar alcohols are named by replacing the suffix *-ose* of the parent monosaccharides with *-itol*.

E. Sugar Acids

Sugar acids are carboxylic acids derived from aldoses, either by oxidation of C-1 (the aldehydic carbon) to yield an aldonic acid or by oxidation of the highest-numbered carbon (the carbon bearing the primary alcohol) to yield an alduronic acid. The structures of the aldonic and alduronic derivatives of glucose—gluconate and glucuronate—are shown in Figure 8.18. Aldonic acids exist in the open-chain form in alkaline solution and form lactones (intramolecular esters) on acidification. Alduronic acids can exist as pyranoses and therefore possess an anomeric carbon. Note that *N*-acetylneuraminic acid (Figure 8.16) is a sugar acid as well as an amino sugar. Sugar acids are important components of many polysaccharides. L-Ascorbic acid or vitamin C, is an enediol of a lactone derived from D-glucuronate (Section 7.9).

8.5 Disaccharides and Other Glycosides

The **glycosidic bond** is the primary structural linkage in all polymers of monosaccharides. A glycosidic bond is an acetal linkage in which the anomeric carbon of a sugar is condensed with an alcohol, an amine, or a thiol. As a simple example, glucopyranose



▲ **Figure 8.16**
Structures of several amino sugars. The amino and acetylamino groups are shown in red.

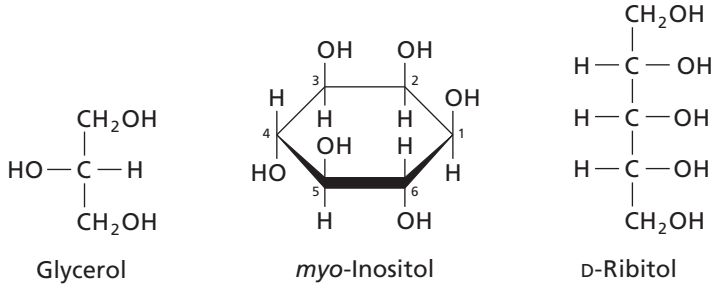
can react with methanol in an acidic solution to form an acetal (Figure 8.19). Compounds containing glycosidic bonds are called **glycosides**; if glucose supplies the anomeric carbon, they are specifically termed **glucosides**. The glycosides include disaccharides, polysaccharides, and some carbohydrate derivatives.

A. Structures of Disaccharides

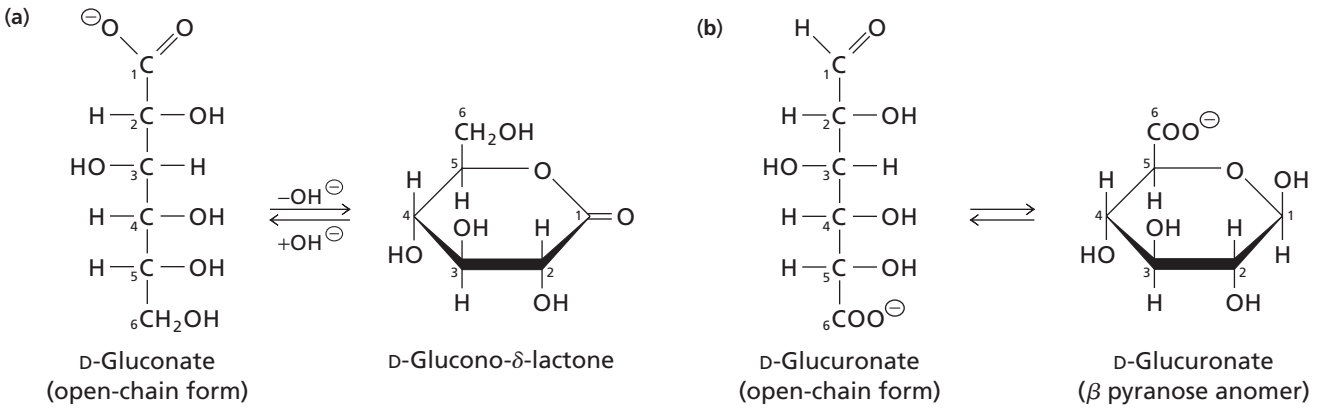
Disaccharides are formed when the anomeric carbon of one sugar molecule interacts with one of several hydroxyl groups in the other sugar molecule. For disaccharides and other carbohydrate polymers, we must note both the types of monosaccharide residues that are present and the atoms that form the glycosidic bonds. In the systematic description of a disaccharide we must specify the linking atoms, the configuration of the glycosidic bond, and the name of each monosaccharide residue (including its designation as a pyranose or furanose). Figure 8.20 presents the structures and nomenclature for four common disaccharides.

Maltose (Figure 8.20a) is a disaccharide released during the hydrolysis of starch, which is a polymer of glucose residues. It is present in malt, a mixture obtained from corn or grain that is used in malted milk and in brewing. Maltose is composed of two D-glucose residues joined by an α-glycosidic bond. The glycosidic bond links C-1 of one residue (on the left in Figure 8.20a) to the oxygen atom attached to C-4 of the second residue (on the right). Maltose is therefore α-D-glucopyranosyl-$(1 \rightarrow 4)$-D-glucose. Note that the glucose residue on the left, whose anomeric carbon is involved in the glycosidic bond, is fixed in the α configuration, whereas the glucose residue on the right (the reducing end, as explained in Section 8.5B) freely equilibrates among the α, β, and open-chain structures. (The open-chain form is present in very small amounts). The structure shown in Figure 8.20a is the β-pyranose anomer of maltose (the anomer whose reducing end is in the β configuration, the predominant anomeric form).

Cellobiose [β-D-glucopyranosyl-$(1 \rightarrow 4)$-D-glucose] is another glucose dimer (Figure 8.20b). Cellobiose is the repeating disaccharide in the structure of cellulose, a



◀ **Figure 8.17**
Structures of several sugar alcohols. Glycerol (a reduced form of glyceraldehyde) and myo-inositol (metabolically derived from glucose) are important constituents of many lipids. Ribitol (a reduced form of ribose) is a constituent of the vitamin riboflavin and its coenzymes.



▲ **Figure 8.18**
Structures of sugar acids derived from D-glucose. (a) Gluconate and its δ-lactone. **(b)** The open-chain and pyranose forms of glucuronate.

plant polysaccharide, and is released during cellulose degradation. The only difference between cellobiose and maltose is that the glycosidic linkage in cellobiose is β (it is α in maltose). The glucose residue on the right in Figure 8.20b, like the residue on the right in Figure 8.20a, equilibrates among the α, β, and open-chain structures.

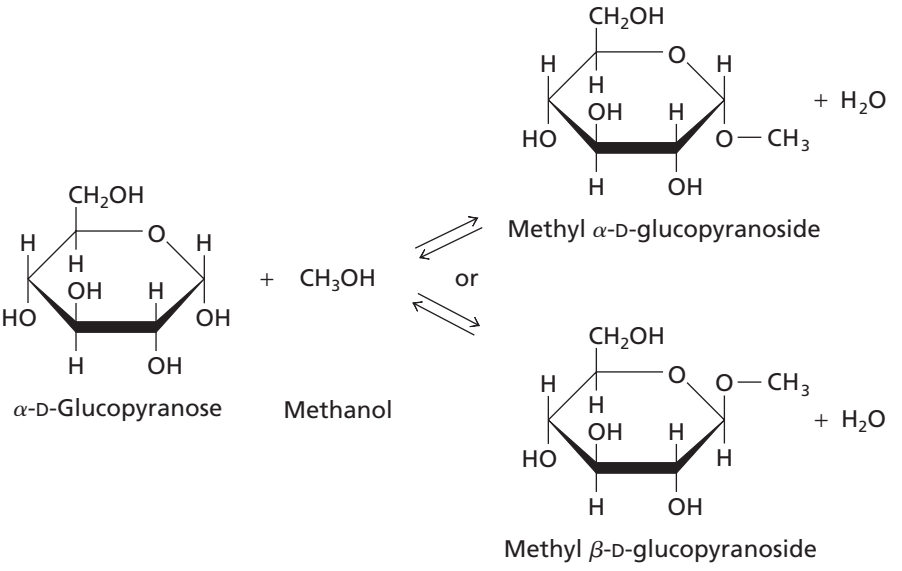
Lactose [β-D-galactopyranosyl-(1 → 4)-D-glucose], a major carbohydrate in milk, is a disaccharide synthesized only in lactating mammary glands (Figure 8.20c). Note that lactose is an epimer of cellobiose. The naturally occurring α anomer of lactose is sweeter and more soluble than the β anomer. The β anomer can be found in stale ice cream, where it has crystallized during storage and given a gritty texture to the ice cream.

Sucrose [α-D-glucopyranosyl-(1 → 2)-β-D-fructofuranoside], or table sugar, is the most abundant disaccharide found in nature (Figure 8.20d). Sucrose is synthesized only in plants. Sucrose is distinguished from the other three disaccharides in Figure 8.20 because its glycosidic bond links the anomeric carbon atoms of two monosaccharide residues. Therefore, the configurations of both the glucopyranose and fructofuranose residues in sucrose are fixed, and neither residue is free to equilibrate between α and β anomers.

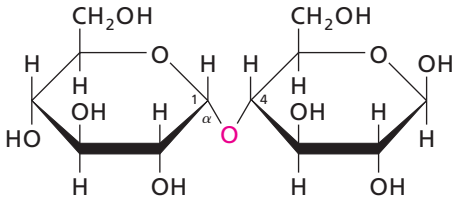
B. Reducing and Nonreducing Sugars

Monosaccharides, and most disaccharides, are hemiacetals with a reactive carbonyl group. They are readily oxidized to diverse products, a property often used in their analysis. Such carbohydrates, including glucose, maltose, cellobiose, and lactose, are sometimes called reducing sugars. Historically, reducing sugars were detected by their ability

Figure 8.19 ►
Reaction of glucopyranose with methanol produces a glycoside. In this acid-catalyzed condensation reaction, the anomeric —OH group of the hemiacetal is replaced by an —OCH_3 group, forming methyl glucoside, an acetal. The product is a mixture of the α and β anomers of methyl glucopyranoside.

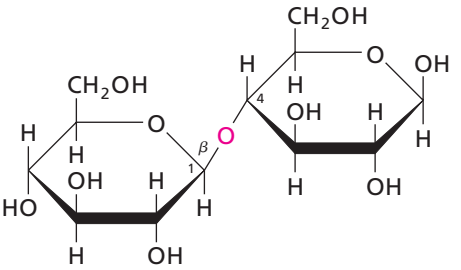


(a)



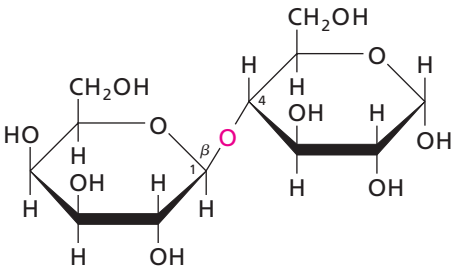
β anomer of maltose
(α-D-Glucopyranosyl-(1→4)-β-D-glucopyranose)

(b)



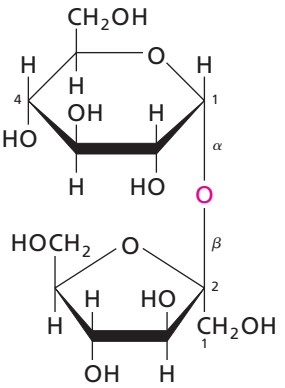
β anomer of cellobiose
(β-D-Glucopyranosyl-(1→4)-β-D-glucopyranose)

(c)



α anomer of lactose
(β-D-Galactopyranosyl-(1→4)-α-D-glucopyranose)

(d)



Sucrose
(α-D-Glucopyranosyl-(1→2)-β-D-fructofuranoside)

▲ **Figure 8.20**
Structures of (a) maltose, (b) cellobiose, (c) lactose, and (d) sucrose. The oxygen atom of each glycosidic bond is shown in red.

to reduce metal ions such as Cu^{2+} or Ag^{+} to insoluble products. Carbohydrates that are not hemiacetals, such as sucrose, are not readily oxidized because both anomeric carbon atoms are fixed in a glycosidic linkage. These are classified as nonreducing sugars.

The reducing ability of a sugar polymer is of more than analytical interest. The polymeric chains of oligosaccharides and polysaccharides show directionality based on their reducing and nonreducing ends. There is usually one reducing end (the residue containing the free anomeric carbon) and one nonreducing end in a linear polymer. All the internal glycosidic bonds of a polysaccharide involve acetals. The internal residues are not in equilibrium with open-chain forms and thus cannot reduce metal ions. A branched polysaccharide has a number of nonreducing ends but only one reducing end.

C. Nucleosides and Other Glycosides

The anomeric carbons of sugars form glycosidic linkages not only with other sugars but also with a variety of alcohols, amines, and thiols. The most commonly encountered glycosides, other than oligosaccharides and polysaccharides, are the nucleosides, in which a purine or pyrimidine is attached by its secondary amino group to a β-D-ribofuranose or β-D-deoxyribofuranose moiety. Nucleosides are called *N*-glycosides because a nitrogen atom participates in the glycosidic linkage. Guanosine (β-D-ribofuranosylguanine) is a typical nucleoside (Figure 8.21). We have already discussed ATP and other nucleotides that are metabolite coenzymes (Section 7.3). NAD and FAD also are nucleotides.

Two other examples of naturally occurring glycosides are shown in Figure 8.21. Vanillin glucoside (Figure 8.21b) is the flavored compound in natural vanilla extract. β-Galactosides constitute an abundant class of glycosides. In these compounds, a variety of nonsugar molecules are joined in β linkage to galactose. For example, galactocerebrosides (see Section 9.5) are glycolipids common in eukaryotic cell membranes and can be hydrolyzed readily by the action of enzymes called β-galactosidases.

▲ **Sugar cane is a major source of commercial sucrose.**

There is a more complete discussion of nucleosides and nucleotides in Chapter 19.

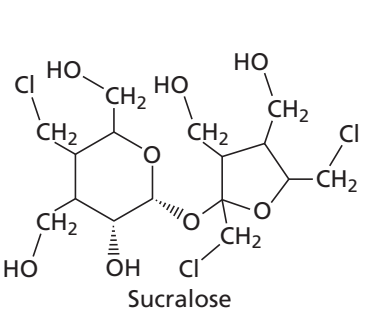
BOX 8.1 THE PROBLEM WITH CATS

One of the characteristics of sugars is that they taste sweet. You certainly know the taste of sucrose and you probably know that fructose and lactose also taste sweet. So do many of the other sugars and their derivatives, although we don't recommend that you go into a biochemistry lab and start tasting all the carbohydrates in those white plastic bottles on the shelves.

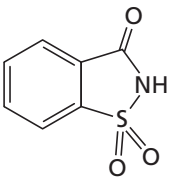
Sweetness is not a physical property of molecules. It's a subjective interaction between a chemical and taste receptors in your mouth. There are five different kinds of taste receptors: sweet, sour, salty, bitter, and umami (umami is like the taste of glutamate in monosodium glutamate). In order to trigger the sweet taste, a molecule like sucrose has to bind to the receptor and initiate a response that eventually makes it to your brain. Sucrose elicits a moderately strong response that serves as the standard for sweetness. The response to fructose is almost twice as strong and the response to lactose is only about one-fifth as strong as that of sucrose. Artificial sweeteners such as saccharin (Sweet'N Low®), sucralose (Splenda®), and aspartame (NutraSweet®) bind to the sweetness receptor and cause the sensation of sweetness. They are hundreds of times more sweet than sucrose.

The sweetness receptor is encoded by two genes called *Tas1r2* and *Tas1r3*. We don't know how sucrose and the other ligands bind to this receptor even though this is a very active area of research. In the case of sucrose and the artifical sweeteners, how can such different molecules elicit the taste of sweet?

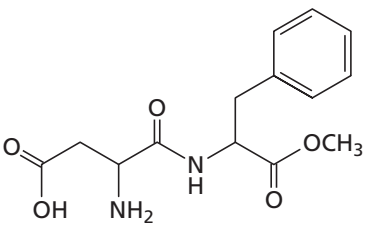
Cats, including lions, tigers and cheetahs, do not have a functional *Tas1r2* gene. It has been converted to a pseudogene because of a 247 bp deletion in exon 3. It's very likely that your pet cat has never experienced the taste of sweetness. That explains a lot about cats.



Sucralose



Saccharin



Aspartame

▲ **Cats are carnivores.** They probably can't taste sweetness.

8.6 Polysaccharides

Polysaccharides are frequently divided into two broad classes. **Homoglycans,** or homopolysaccharides, are polymers containing residues of only one type of monosaccharide. **Heteroglycans**, or heteropolysaccharides, are polymers containing residues of more than one type of monosaccharide. Polysaccharides are created without a template by the addition of particular monosaccharide and oligosaccharide residues. As a result, the lengths and compositions of polysaccharide molecules may vary within a population of these molecules. Some common polysaccharides and their structures are listed in Table 8.2.

Most polysaccharides can also be classified according to their biological roles. For example, starch and glycogen are storage polysaccharides while cellulose and chitin are structural polysaccharides. We will see additional examples of the variety and versatility of carbohydrates when we discuss the heteroglycans in the next section."

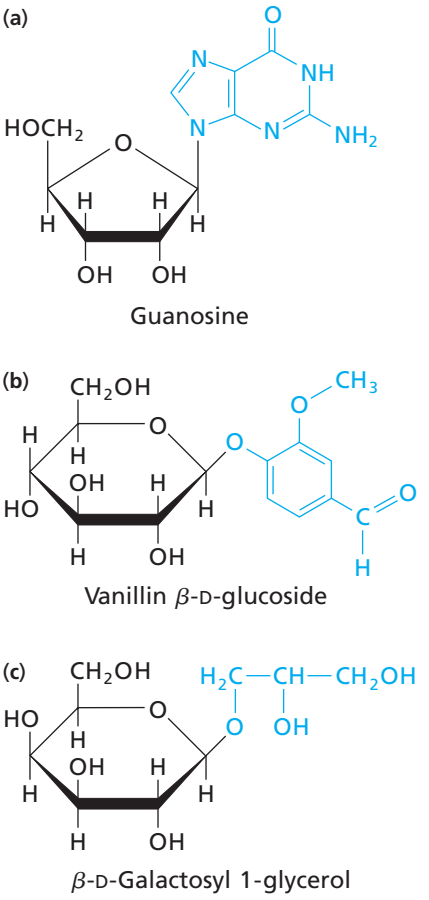
A. Starch and Glycogen

D-Glucose is synthesized in all species. Excess glucose can be broken down to produce metabolic energy. Glucose residues are stored as polysaccharides until they are needed for energy production. The most common storage homoglycan of glucose in plants and fungi is starch and in animals it is glycogen. Both types of polysaccharides occur in bacteria.

Table 8.2 Structures of some common polysaccharides

Polysaccharide[a]	Component(s)[b]	Linkage(s)
Storage homoglycans		
Starch		
Amylose	Glc	α-(1 → 4)
Amylopectin	Glc	α-(1 → 4), α-(1 → 6) (branches)
Glycogen	Glc	α-(1 → 4), α-(1 → 6) (branches)
Structural homoglycans		
Cellulose	Glc	β(1 → 4)
Chitin	GlcNAc	β(1 → 4)
Heteroglycans		
Glycosaminoglycans	Disaccharides (amino sugars, sugar acids)	Various
Hyaluronic acid	GlcUA and GlcNAc	β(1 → 3), β(1 → 4)

[a]Polysaccharides are unbranched unless otherwise indicated.
[b]Glc, Glucose; GlcNAc, *N*-acetylglucosamine; GlcUA, D-glucuronate.



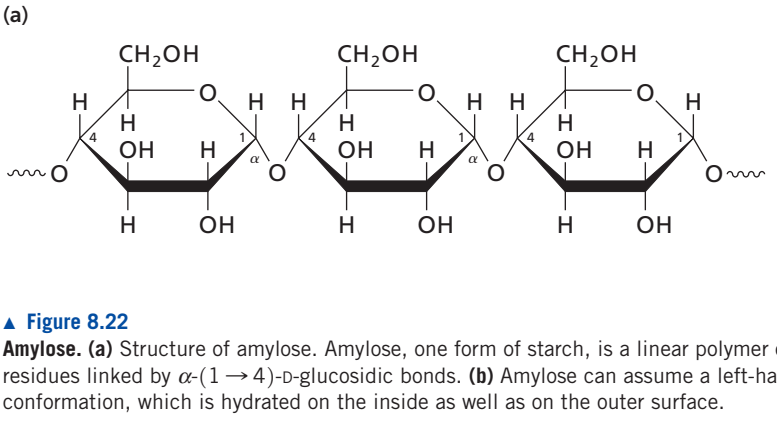
▲ **Figure 8.21**
Structures of three glycosides. The nonsugar components are shown in blue. **(a)** Guanosine. **(b)** Vanillin glucoside, the flavored compound in vanilla extract. **(c)** β-D-Galactosyl 1-glycerol, derivatives of which are common in eukaryotic cell membranes.

Starch is present in plant cells as a mixture of amylose and amylopectin and is stored in granules whose diameters range from 3 to 100 μm. Amylose is an unbranched polymer of about 100 to 1000 D-glucose residues connected by α-(1 → 4) glycosidic linkages, specifically termed α-(1 → 4) *glucosidic bonds* because the anomeric carbons belong to glucose residues (Figure 8.22a). The same type of linkage connects glucose monomers in the disaccharide maltose (Figure 8.20a). Although it is not truly soluble in water, amylose forms hydrated micelles in water and can assume a helical structure under some conditions (Figure 8.22b).

Amylopectin is a branched version of amylose (Figure 8.23). Branches, or polymeric side chains, are attached via α-(1 → 6) glucosidic bonds to linear chains of residues linked by α-(1 → 4) glucosidic bonds. Branching occurs, on average, once every 25 residues and the side chains contain about 15 to 25 glucose residues. Some side chains themselves are branched. Amylopectin molecules isolated from living cells may contain 300 to 6000 glucose residues.

An adult human consumes about 300 g of carbohydrate daily, much of which is in the form of starch. Raw starch granules resist enzymatic hydrolysis but cooking causes them to absorb water and swell. The swollen starch is a substrate for two different glycosidases. Dietary starch is degraded in the gastrointestinal tract by the actions of α-amylase and a debranching enzyme. α-Amylase, which is present in both animals and

Starch metabolism is described in Chapter 15.



▲ **Figure 8.22**
Amylose. (a) Structure of amylose. Amylose, one form of starch, is a linear polymer of glucose residues linked by α-(1 → 4)-D-glucosidic bonds. **(b)** Amylose can assume a left-handed helical conformation, which is hydrated on the inside as well as on the outer surface.

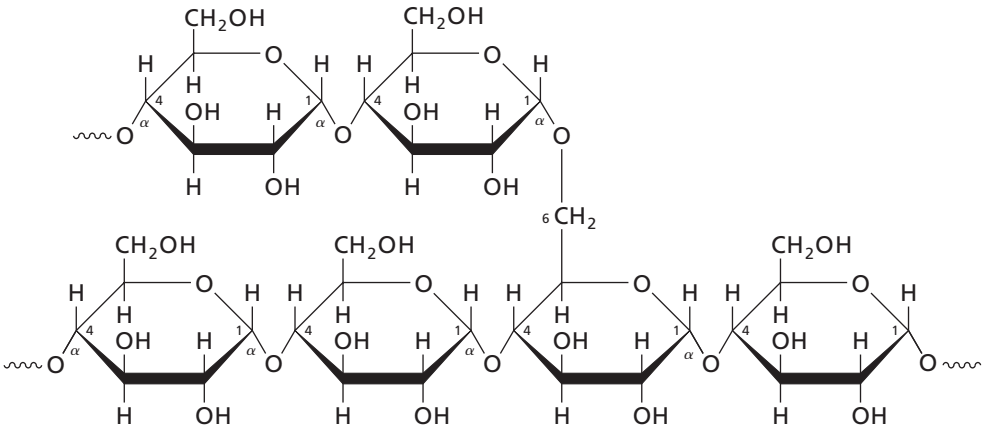


Figure 8.23 ►
Structure of amylopectin. Amylopectin, a second form of starch, is a branched polymer. The linear glucose residues of the main chain and the side chains of amylopectin are linked by α-(1 → 4)-D-glucosidic bonds, and the side chains are linked to the main chain by α-(1 → 6)-D-glucosidic bonds.

plants, is an endoglycosidase (it acts on internal glycosidic bonds). The enzyme catalyzes random hydrolysis of the α-(1 → 4) glucosidic bonds of amylose and amylopectin.

Another hydrolase, β-amylase, is found in the seeds and tubers of some plants. β-Amylase is an exoglycosidase (it acts on terminal glycosidic bonds). It catalyzes sequential hydrolytic release of maltose from the free, nonreducing ends of amylopectin.

Despite their α and β designations, both types of amylases act only on α-(1 → 4)-D-glycosidic bonds. Figure 8.24 shows the action of α-amylase and β-amylase on amylopectin. The α-(1 → 6) linkages at branch points are not substrates for either α- or β-amylase. After amylase-catalyzed hydrolysis of amylopectin, highly branched cores resistant to further hydrolysis, called **limit dextrins**, remain. Limit dextrins can be further degraded only after debranching enzymes have catalyzed hydrolysis of the α-(1 → 6) linkages at branch points.

Glycogen is also a branched polymer of glucose residues. Glycogen contains the same types of linkages found in amylopectin but the branches in glycogen are smaller and more frequent, occurring every 8–12 residues. In general, glycogen molecules are larger than starch molecules, Glycogen up to contains 50,000 glucose residues. In mammals,

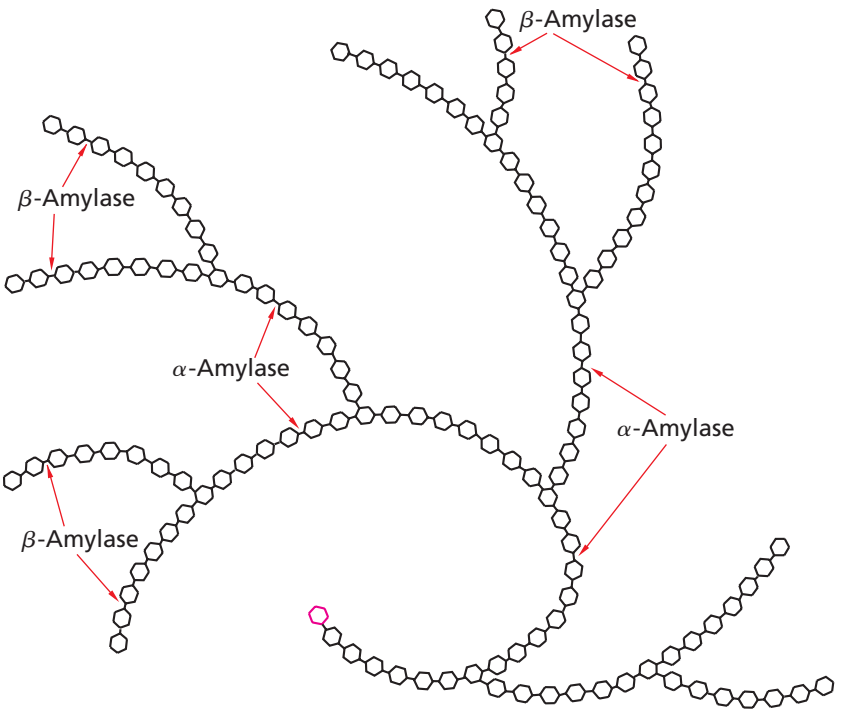


Figure 8.24 ►
Action of α-amylase and β-amylase on amylopectin. α-Amylase catalyzes random hydrolysis of internal α-(1 → 4) glucosidic bonds; β-amylase acts on the nonreducing ends. Each hexagon represents a glucose residue; the single reducing end of the branched polymer is red. (An actual amylopectin molecule contains many more glucose residues than shown here.)

depending on the nutritional state, glycogen can account for up to 10% of the mass of the liver and 2% of the mass of muscle.

The branched structures of amylopectin and glycogen possess only one reducing end but many nonreducing ends. The reducing end of glycogen is covalently attached to a protein called glycogenin (Section 12.5A). Enzymatic lengthening and degradation of polysaccharide chains occurs at the nonreducing ends.

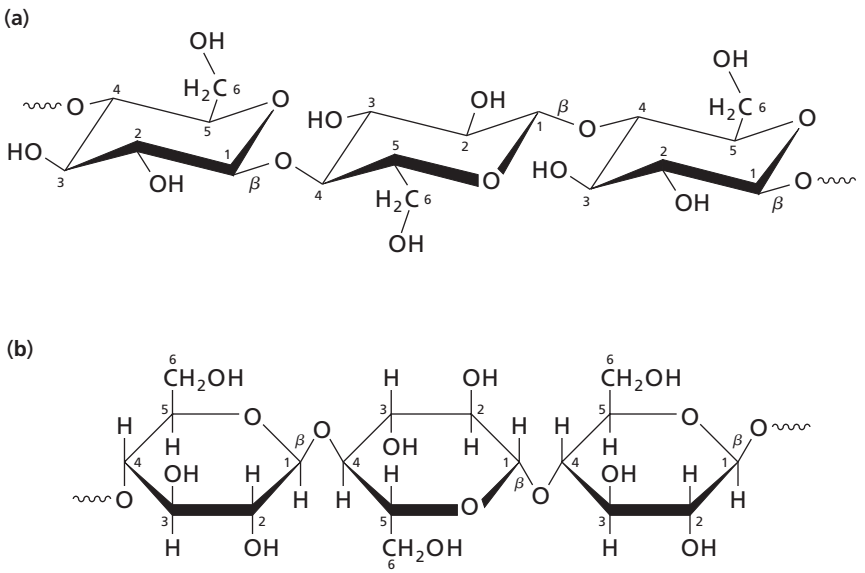
Enzymes that catalyze the intracellular synthesis and breakdown of glycogen are described in Chapter 12.

B. Cellulose

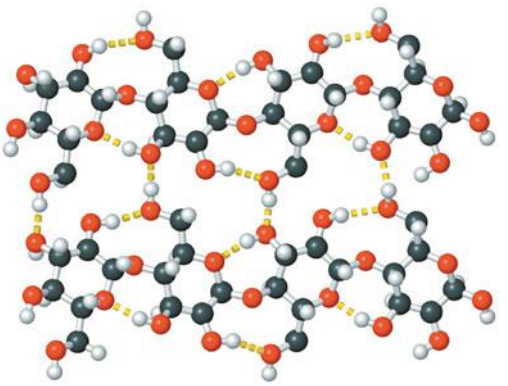
Cellulose is a structural polysaccharide. It is a major component of the rigid cell walls that surround many plant cells. The stems and branches of many plants consist largely of cellulose. This single polysaccharide accounts for a significant percentage of all organic matter on Earth. Like amylose, cellulose is a linear polymer of glucose residues, but in cellulose the glucose residues are joined by β-(1 → 4) linkages rather than α-(1 → 4) linkages. The two glucose residues of the disaccharide cellobiose also are connected by a β-(1 → 4) linkage (Figure 8.20b). Cellulose molecules vary greatly in size, ranging from about 300 to more than 15,000 glucose residues.

The β linkages of cellulose result in a rigid extended conformation in which each glucose residue is rotated 180° relative to its neighbors (Figure 8.25). Extensive hydrogen bonding within and between cellulose chains leads to the formation of bundles, or fibrils (Figure 8.26). Cellulose fibrils are insoluble in water and are quite strong and rigid. Cotton fibers are almost entirely cellulose and wood is about half cellulose. Because of its strength, cellulose is used for a variety of purposes and is a component of a number of synthetic materials including cellophane and the fabric rayon. We are most familiar with cellulose as the main component of paper.

Enzymes that catalyze the hydrolysis of α-D-glucosidic bonds (α-glucosidases, such as α- and β-amylase) do not catalyze the hydrolysis of β-D-glucosidic bonds. Similarly, β-glucosidases (such as cellulase) do not catalyze the hydrolysis of α-D-glucosidic bonds. Humans and other mammals can metabolize starch, glycogen, lactose, and sucrose and use the monosaccharide products in a variety of metabolic pathways. Mammals cannot metabolize cellulose because they lack enzymes capable of catalyzing the hydrolysis of β-glucosidic linkages. Ruminants such as cows and sheep have microorganisms in their rumen (a compartment in their multichambered stomachs) that produce β-glucosidases. Thus, ruminants can obtain glucose from grass and other plants that are rich in cellulose. Because they have cellulase-producing bacteria in their digestive tracts, termites also can obtain glucose from dietary cellulose.

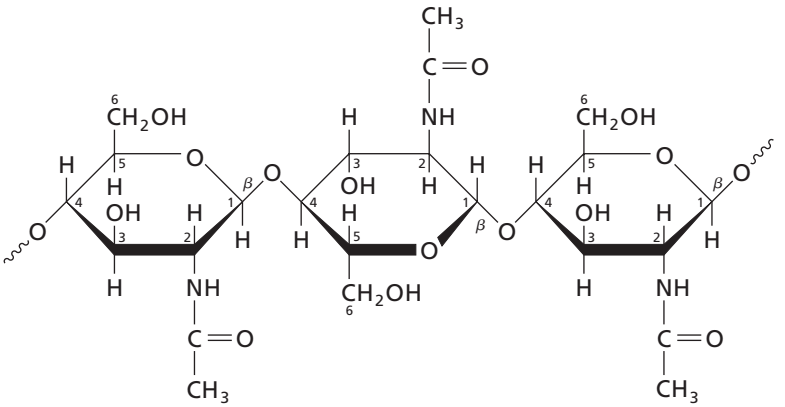


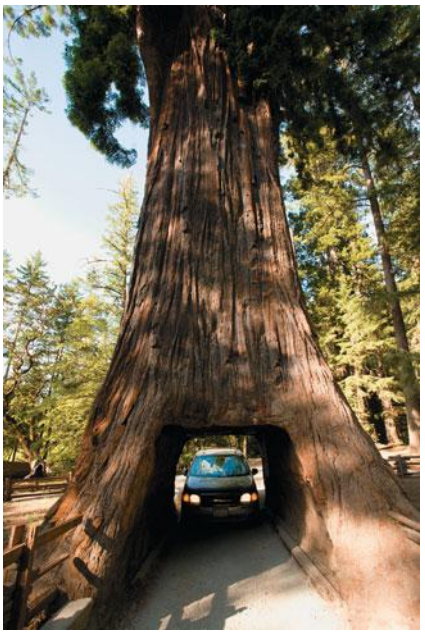
▲ **Figure 8.25**
Structure of cellulose. Note the alternating orientation of successive glucose residues in the cellulose chain. **(a)** Chair conformation. **(b)** Modified Haworth projection.



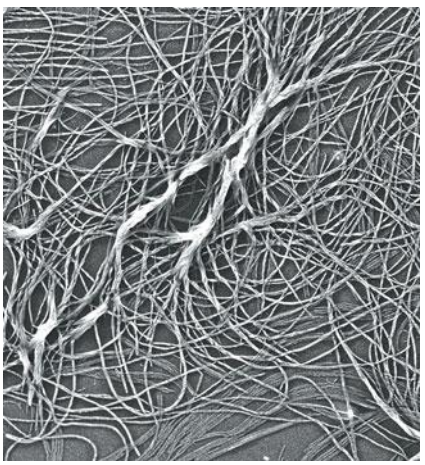
▲ **Figure 8.26**
Cellulose fibrils. Intra- and interchain hydrogen bonding gives cellulose its strength and rigidity.

Figure 8.27 ►
Structure of chitin. The linear homoglycan chitin consists of repeating units of β-(1 → 4)-linked GlcNAc residues. Each residue is rotated 180° relative to its neighbors.





▲ The giant redwood trees of California contains tons of cellulose.



▲ **Cellulose fibers.** Plants make large cellulose fibers that serve as structural support. A scanning electron micrograph of these fibers shows how they overlap to form a large net-like sheet. These cellulose fibers are about 253 million years old. They were recovered from deep within a salt mine in New Mexico.

C. Chitin

Chitin, probably the second most abundant organic compound on Earth, is a structural homoglycan found in the exoskeletons of insects and crustaceans and also in the cell walls of most fungi and red algae. Chitin is a linear polymer similar to cellulose. It is made up of β-(1 → 4)-linked GlcNAc residues rather than glucose residues (Figure 8.27). Each GlcNAc residue is rotated 180° relative to its neighbors. The GlcNAc residues in adjacent strands of chitin form hydrogen bonds with each other resulting in linear fibrils of great strength. Chitin is often closely associated with nonpolysaccharide compounds, such as proteins and inorganic material.

8.7 Glycoconjugates

Glycoconjugates consist of polysaccharides linked to (conjugated with) proteins or peptides. In most cases, the polysaccharides are composed of several different monosaccharide units. Thus, they are heteroglycans. (Starch, glycogen, cellulose, and chitin are homoglycans.) Heteroglycans appear in three types of glycoconjugates—proteoglycans, peptidoglycans, and glycoproteins. In this section, we see how the chemical and physical properties of the heteroglycans in glycoconjugates are suited to various biological functions.

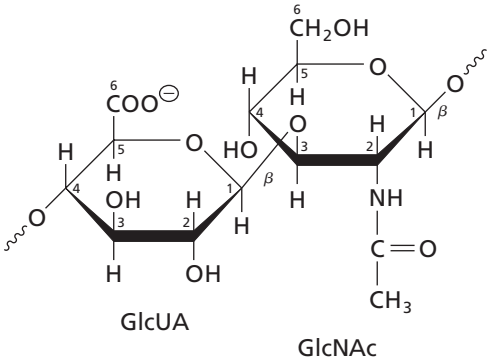
A. Proteoglycans

Proteoglycans are complexes of proteins and a class of polysaccharides called *glycosaminoglycans*. These glycoconjugates occur predominately in the extracellular matrix (connective tissue) of multicellular animals.

Glycosaminoglycans are unbranched heteroglycans of repeating disaccharide units. As the name *glycosaminoglycan* indicates, one component of the disaccharide is an amino sugar, either D-galactosamine (GalN) or D-glucosamine (GlcN). The amino group of the amino-sugar component can be acetylated forming *N*-acetylgalactosamine (GalNAc) or GlcNAc. The other component of the repeating disaccharide is usually an alduronic acid. Specific hydroxyl and amino groups of many glycosaminoglycans are sulfated. These sulfate groups and the carboxylate groups of alduronic acids make glycosaminoglycans polyanionic.

Several types of glycosaminoglycans have been isolated and characterized. Each type has its own sugar composition, linkages, tissue distribution, and function and each is attached to a characteristic protein. Hyaluronic acid is an example of a glycosaminoglycan composed of the repeating disaccharide unit shown in Figure 8.28. It is found in the fluid of joints where it forms a viscous solution that is an excellent lubricant. Hyaluronic acid is also a major component of cartilage.

Up to100 glycosaminoglycan chains can be attached to the protein of a proteoglycan. These heteroglycan chains are usually covalently bound by a glycosidic linkage to



◀ **Figure 8.28**
Structure of the repeating disaccharide of hyaluronic acid. The repeating disaccharide of this glycosaminoglycan contains D-glucuronate (GlcUA) and GlcNAc. Each GlcUA residue is linked to a GlcNAc residue through β-(1 → 3) linkage; each GlcNAc residue is in turn linked to the next GlcUA residue through a β-(1 → 4) linkage.

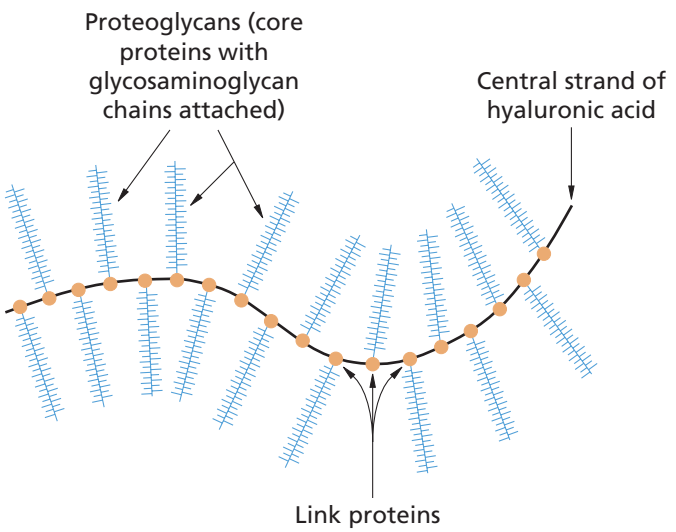
the hydroxyl oxygens of serine residues. (Not all glycosaminoglycans are covalently linked to proteins.) Glycosaminoglycans can account for up to 95% of the mass of a proteoglycan.

Proteoglycans are highly hydrated and occupy a large volume because their glycosaminoglycan component contains polar and ionic groups. These features confer elasticity and resistance to compression—important properties of connective tissue. For example, the flexibility of cartilage allows it to absorb shocks. Some of the water can be pressed out when cartilage is compressed but relief from pressure allows cartilage to rehydrate. In addition to maintaining the shapes of tissues, proteoglycans can also act as extracellular sieves and help direct cell growth and migration.

Examination of the structure of cartilage shows how proteoglycans are organized in this tissue. Cartilage is a mesh of collagen fibers (Section 4.11) interspersed with large proteoglycan aggregates ($M_r \sim 2 \times 10^8$). Each aggregate assumes a characteristic shape that resembles a bottle brush (Figure 8.29). These aggregates contain hyaluronic acid and several other glycosaminoglycans, as well as two types of proteins—core proteins and link proteins. A central strand of hyaluronic acid runs through the aggregate and many proteoglycans—core proteins with glycosaminoglycan chains attached—branch from its sides. The core proteins interact noncovalently with the hyaluronic acid strand, mostly by electrostatic interactions. Link proteins stabilize the core protein–hyaluronic acid interactions.

The major proteoglycan of cartilage is called *aggrecan*. The protein core of aggrecan ($M_r \sim 220{,}000$) carries approximately 30 molecules of keratan sulfate (a glycosaminoglycan composed chiefly of alternating *N*-acetylglucosamine 6-sulfate and galactose residues) and approximately 100 molecules of chondroitin sulfate (a glycosaminoglycan

▲ **Lobsters have an exoskeleton made of chitin.** The color of the exoskeleton is determined by the foods that the lobster eats. When it ingests β-carotene derivatives they are converted to a complex mixture of protein-bound carotenes called crustacayanin that has a greenish-brown color. When lobsters are cooked, the crustacyanin breaks down, releasing free β-carotene derivatives that are red in color, like the red color of maple leaves in autumn (see Section 15.1).



◀ **Figure 8.29**
Proteoglycan aggregate of cartilage. Core proteins carrying glycosaminoglycan chains are associated with a central strand of a single hyaluronic acid molecule. These proteins have many covalently attached glycosaminoglycan chains (keratan sulfate and chondroitin sulfate molecules). The interactions of the core proteins with hyaluronic acid are stabilized by link proteins, which interact noncovalently with both types of molecules. The aggregate has the appearance of a bottle brush.

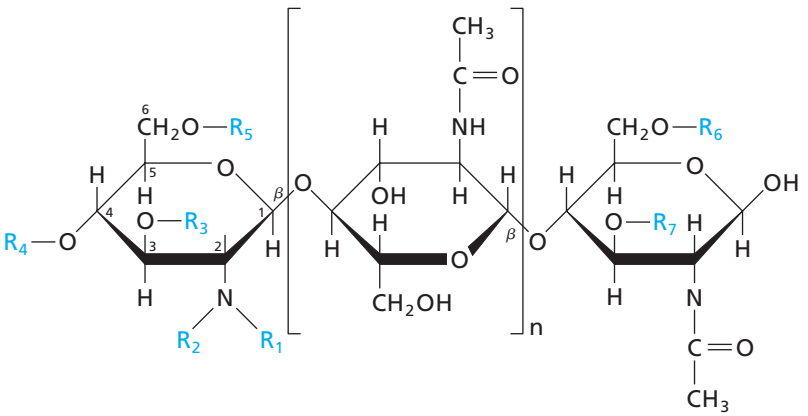
BOX 8.2 NODULATION FACTORS ARE LIPO-OLIGOSACCHARIDES

Legumes such as alfalfa, peas, and soybeans develop organs called *nodules* on their roots. Certain soil bacteria (rhizobia) infect the nodules and, in a symbiosis with the plants, carry out nitrogen fixation (reduction of atmospheric nitrogen to ammonia). The symbiosis is highly species-specific: only certain combinations of legumes and bacteria can cooperate and therefore these organisms must recognize each other. Rhizobia produce extracellular signal molecules that are oligosaccharides called nodulation factors. Extremely low concentrations of these compounds can induce their plant hosts to develop the nodules that the rhizobia can infect. A host plant responds only to a nodulation factor of a characteristic composition.

Infection begins when the plant root hair recognizes the nodulation factor via surface Nod-factor receptors. This results in a response that allows the bacteria to enter the root hair and migrate down to the cells in the root where the nodule forms.

All the nodulation factors studied to date are oligosaccharides that have a linear chain of β-(1 → 4) *N*-acetylglucosamine (GlcNAc)—the same repeating structure as in chitin (Section 8.6b). Most nodulation factors are sugar pentamers although the number of residues can vary between three and six (see figure below). Species specificity is provided by variation in polymer length and potential substitution on five sites at the nonreducing end (R1 to R5) and two sites at the reducing end (R6 and R7). R1, an acyl group substituting the nitrogen atom at C-2 of the nonreducing end, is a fatty acid, usually 18 carbons long. Thus, the nodulation factors are lipo-oligosaccharides. R6, bound to the alcohol at C-6 of the reducing end, can have a wide variety of structures, including sulfate or methyl fucose. Research on these growth regulators for legumes has stimulated the search for biological activities of other oligosaccharides.

See Section 17.1 for details about nitrogen fixation.



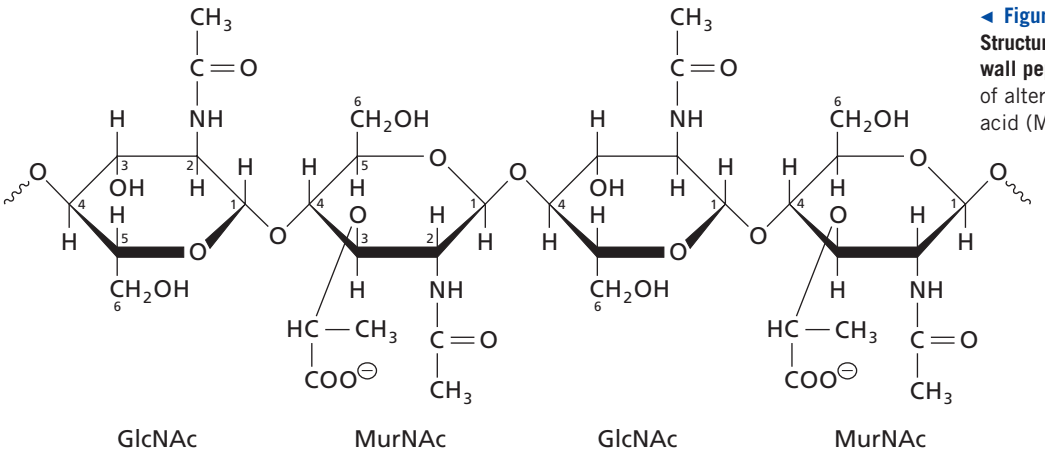
▲ General structure of nodulation factors, lipo-oligosaccharides with an *N*-acetylglucosamine (GlcNAc) backbone. The number of internal residues of *N*-acetylglucosamine is shown by *n*, which is usually 3 but can sometimes be 1, 2, or 4. R1 is a fatty acyl substituent, usually 18 carbons long.

▲ **Formation of nodules in the legume *Lotus japonicus*.** Rhizobia (blue) have secreted nodulation factor leading to endocytosis by root hair cells and formation of an infection thread connecting the point of uptake (top) to the root nodule cells (below).

composed of alternating *N*-acetylgalactosamine sulfate and glucuronate residues). Aggrecan is a member of a small family of hyalectans, proteoglycans that bind to hyaluronic acid. Other hyalectans provide elasticity to blood vessel walls and modulate cell-cell interactions in the brain.

B. Peptidoglycans

Peptidoglycans are polysaccharides linked to small peptides. The cell walls of many bacteria contain a special class of peptidoglycan with a heteroglycan component attached to a four or five residue peptide. The heteroglycan component is composed of alternating residues of GlcNAc and *N*-acetylmuramic acid (MurNAc) joined by β-(1 → 4) linkages (Figure 8.30). MurNAc is a nine-carbon sugar found only in bacteria. MurNAc consists of the three-carbon acid D-lactate joined by an ether linkage to C-3 of GlcNAc.



◀ **Figure 8.30**
Structure of the polysaccharide in bacterial cell wall peptidoglycan. The glycan is a polymer of alternating GlcNAc and *N*-acetylmuramic acid (MurNAc) residues.

The polysaccharide moiety of peptidoglycans resembles chitin except that every second GlcNAc residue is modified by addition of lactate to form MurNAc. The antibacterial action of lysozyme (Section 6.6) results from its ability to catalyze hydrolysis of the polysaccharide chains of peptidoglycans.

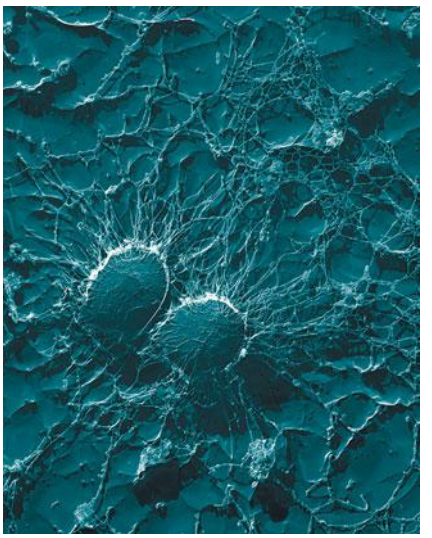
The peptide component of peptidoglycans varies among bacteria. The peptide component in *Staphylococcus aureus* is a tetrapeptide with alternating L and D amino acids: L-Ala–D-Isoglu–L-Lys–D-Ala. (Isoglu represents isoglutamate, a form of glutamate in which the γ-carboxyl group—not the α-carboxyl group—is linked to the next residue.) Other species have a different amino acid at the third position. An amide bond links the amino group of the L-alanine residue to the lactyl carboxylate group of a MurNAc residue of the glycan polymer (Figure 8.31). The tetrapeptide is cross-linked to another tetrapeptide on a neighboring peptidoglycan molecule by a chain of five glycine residues (pentaglycine). Pentaglycine joins the L-lysine residue of one tetrapeptide to the carboxyl group of the D-alanine residue of the other tetrapeptide. Extensive cross-linking essentially converts the peptidoglycan to one huge, rigid, macromolecule that defines the shape of the bacterium by covering its plasma membrane and protecting the cell from fluctuations in osmotic pressure.

Most bacteria have an additional exterior layer of dense polysaccharide called the *capsule*. The capsule is made up of chains of polysaccharide composed mainly of *N*-acetylglucosamine (GlcNAc) residues but various other amino sugars are present. The capsule protects the bacterial cell from injury. The capsule in pathogenic bacteria help cells avoid destruction by the immune system.

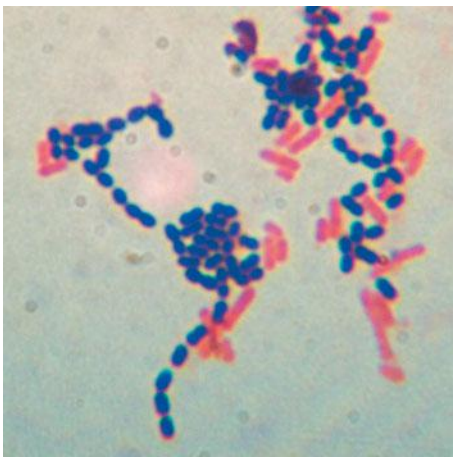
In gram-negative bacteria, the peptidoglycan cell wall lies between the inner plasma membrane and the outer membrane. In gram-positive bacteria, there is no outer membrane and the cell wall is much thicker. This is one of the reasons why the Gram stain (named after Christian Gram) will color the surfaces of some bacteria (gram positive) and not others (gram negative).

During peptidoglycan biosynthesis, a five-residue peptide—L-Ala–D-Isoglu–L-Lys–D-Ala–D-Ala—is attached to a MurNAc residue. In subsequent steps, five glycine residues are added sequentially to the ε-amino group of the lysine residue forming the pentaglycine bridge. In the final step of synthesis, a transpeptidase catalyzes formation of a peptide linkage between the penultimate alanine residue and a terminal glycine residue of a pentaglycine bridge of a neighboring peptidoglycan strand. This reaction is driven by release of the terminal D-alanine residue.

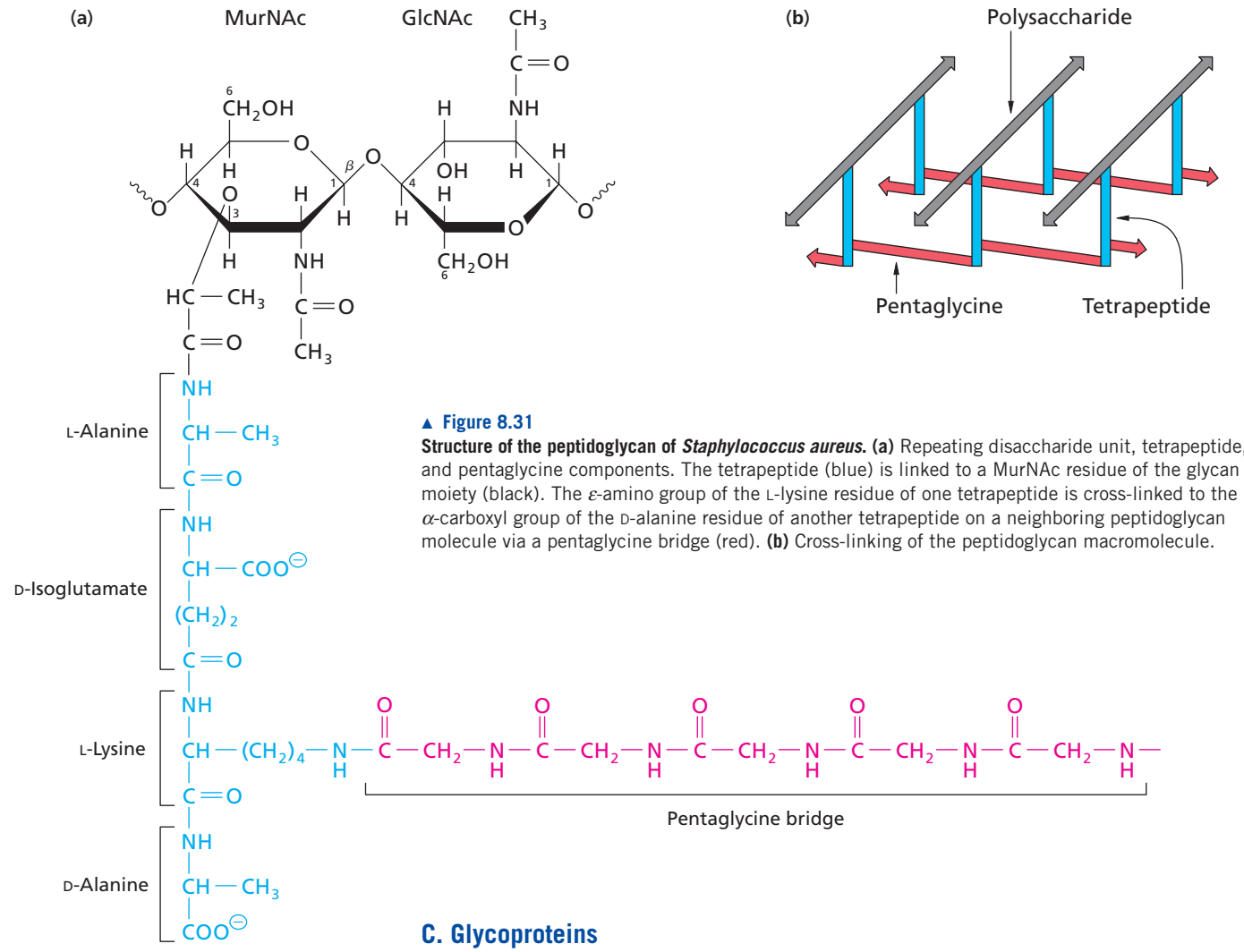
The structure of the antibiotic penicillin (Figure 8.32) resembles the terminal D-Ala–D-Ala residues of the immature peptidoglycan. Penicillin binds, probably irreversibly, to the transpeptidase active site inhibiting the activity of the enzyme and thereby blocking further peptidoglycan synthesis. The antibiotic prevents growth and proliferation of bacteria. Penicillin is selectively toxic to bacteria because the reaction it affects occurs only in certain bacteria, not in eukaryotic cells.



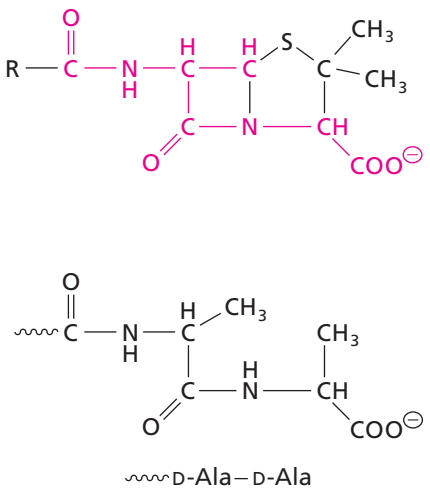
▲ ***Staphylococcus aureus* cells.** These bacterial cells have extensive polysaccharide capsules that protect them from their host's immune system.



▲ **The Gram stain.** The Gram staining procedure distinguishes between gram-positive bacteria (left, purple) and gram-negative bacteria (right, pink).



▲ **Figure 8.31**
Structure of the peptidoglycan of *Staphylococcus aureus*. (a) Repeating disaccharide unit, tetrapeptide, and pentaglycine components. The tetrapeptide (blue) is linked to a MurNAc residue of the glycan moiety (black). The ε-amino group of the L-lysine residue of one tetrapeptide is cross-linked to the α-carboxyl group of the D-alanine residue of another tetrapeptide on a neighboring peptidoglycan molecule via a pentaglycine bridge (red). **(b)** Cross-linking of the peptidoglycan macromolecule.



▲ **Figure 8.32**
Structures of penicillin and –D-Ala–D-Ala. The portion of penicillin that resembles the dipeptide is shown in red. R can be a variety of substituents.

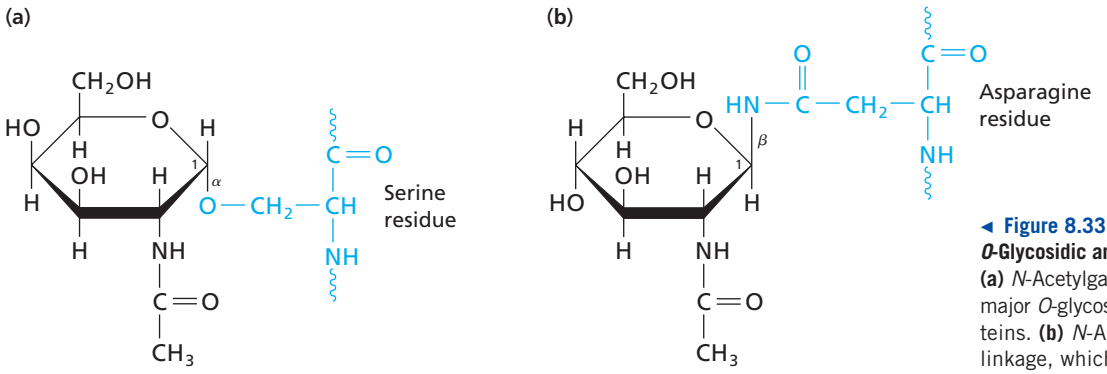
C. Glycoproteins

Glycoproteins, like proteoglycans, are proteins that contain covalently bound oligosaccharides (i.e., proteins that are glycosylated). In fact, proteoglycans are a type of glycoprotein. The carbohydrate chains of a glycoprotein vary in length from one to more than 30 residues and can account for as much as 80% of the total mass of the molecule. Glycoproteins are an extraordinarily diverse group of proteins that includes enzymes, hormones, structural proteins, and transport proteins.

The oligosaccharide chains of different glycoproteins exhibit great variability in composition. The composition of oligosaccharide chains can vary even among molecules of the same protein, a phenomenon called *microheterogeneity*.

Several factors contribute to the structural diversity of the oligosaccharide chains of glycoproteins.

1. An oligosaccharide chain can contain several different sugars. Eight sugars predominate in eukaryotic glycoproteins: the hexoses L-fucose, D-galactose, D-glucose, and D-mannose; the hexosamines *N*-acetyl-D-galactosamine and *N*-acetyl-D-glucosamine; the nine-carbon sialic acids (usually *N*-acetylneuraminic acid); and the pentose D-xylose. Many different combinations of these sugars are possible.
2. The sugars can be joined by either α- or β-glycosidic linkages.
3. The linkages can also join various carbon atoms in the sugars. In hexoses and hexosamines, the glycosidic linkages always involve C-1 of one sugar but can involve C-2, C-3, C-4, or C-6 of another hexose or C-3, C-4, or C-6 of an amino sugar (C-2 is usually *N*-acetylated in this class of sugar). C-2 of sialic acid, not C-1, is linked to other sugars.
4. Oligosaccharide chains of glycoproteins can contain up to four branches.



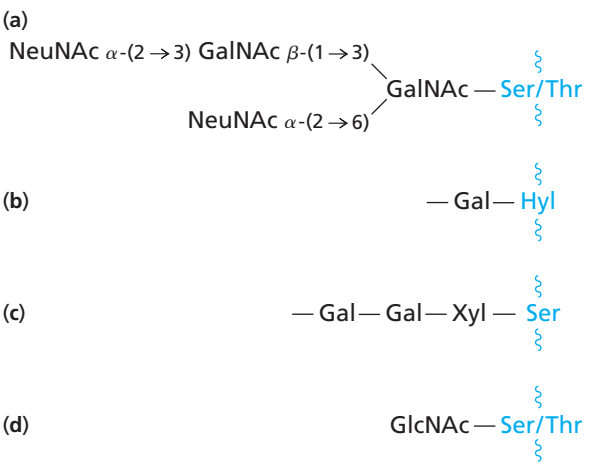
◀ **Figure 8.33**
***O*-Glycosidic and *N*-glycosidic linkages.** **(a)** *N*-Acetylgalactosamine–serine linkage, the major *O*-glycosidic linkage found in glycoproteins. **(b)** *N*-Acetylglucosamine–asparagine linkage, which characterizes N-linked glycoproteins. The *O*-glycosidic linkage is α, whereas the *N*-glycosidic linkage is β.

The astronomical number of possible oligosaccharide structures afforded by these four factors is not realized in cells because cells do not possess specific glycosyltransferases to catalyze the formation of all possible glycosidic linkages. In addition, individual glycoproteins—through their unique conformations—modulate their own interactions with the glycosylating enzymes so that most glycoproteins possess a heterogeneous but reproducible oligosaccharide structure.

The oligosaccharide chains of most glycoproteins are either *O*- or *N*-linked. In **O-linked oligosaccharides**, a GalNAc residue is typically linked to the side chain of a serine or threonine residue. In **N-linked oligosaccharides**, a GlcNAc residue is linked to the amide nitrogen of an asparagine residue. The structures of an *O*-glycosidic and an *N*-glycosidic linkage are compared in Figure 8.33. Additional sugar residues can be attached to the GalNAc or the GlcNAc residue. An individual glycoprotein can contain both *O*- and *N*-linked oligosaccharides and some glycoproteins contain a third type of linkage. In these glycoproteins, the protein is attached to ethanolamine that is linked to a branched oligosaccharide to which lipid is also attached (Section 9.10).

There are four important subclasses of *O*-glycosidic linkages in glycoproteins.

1. The most common *O*-glycosidic linkage is the GalNAc-Ser/Thr linkage mentioned above. Other sugars—for example, galactose and sialic acid—are frequently linked to the GalNAc residue (Figure 8.34a).
2. Some of the 5-hydroxylysine (Hyl) residues of collagen (Figure 4.35) are joined to D-galactose via an *O*-glycosidic linkage (Figure 8.34b). This structure is unique to collagen.
3. The glycosaminoglycans of certain proteoglycans are joined to the core protein via a Gal–Gal–Xyl–Ser structure (Figure 8.34c).
4. In some proteins, a single residue of GlcNAc is linked to serine or threonine (Figure 8.34d).



◀ **Figure 8.34**
Four subclasses of *O*-glycosidic linkages. **(a)** Example of a typical linkage in which *N*-acetylgalactosamine (GalNAc) with attached residues is linked to a serine or threonine residue. **(b)** Linkage found in collagen, where a galactose residue, usually attached to a glucose residue, is linked to hydroxylysine (Hyl). **(c)** Trisaccharide linkage found in certain proteoglycans. **(d)** GlcNAc linkage found in some proteins.

BOX 8.3 ABO BLOOD GROUP

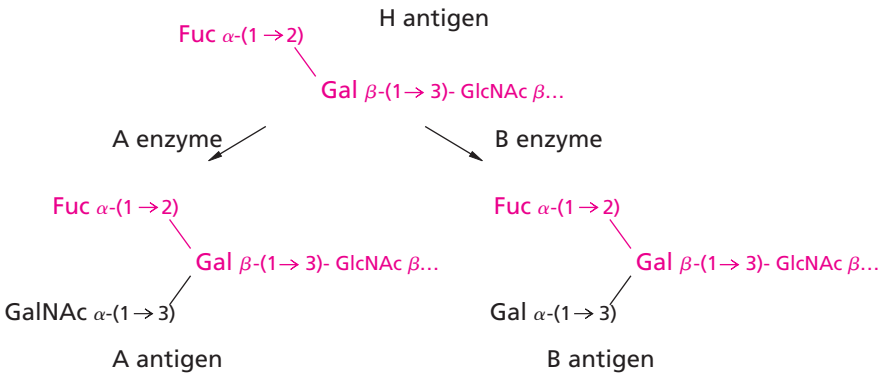
The ABO blood group was first discovered in 1901 by Karl Landsteiner, who received the Nobel Prize in Physiology or Medicine in 1930. Most primates display three different kinds of *O*- or *N*-linked oligosaccharides on their cell surfaces. The core structure of these oligosaccharides is called H antigen. It consists of various combinations of galactose (Gal), fucose (Fuc), *N*-acetylglucosamine (GlcNac), and *N*-acetylneuraminic acid (sialic acid, NeuNAc). These monosaccharides are linked in various ways to form a short branched structure that exhibits considerable microheterogeneity. One of the most common H antigen structures is shown in the figure.

The core structure (H antigen) can be modified in various ways. The addition of a GalNAc residue in α-(1 ⟶ 3) linkage forms A antigen. This reaction is catalyzed by A enzyme. The addition of Gal in α-(1 ⟶ 3) linkage is catalyzed by B enzyme.

If only A antigen is present, a person will have A blood type. If only B antigen is present, the blood type will be B. The AB blood type indicates that both A antigen and B antigen are present on cell surfaces. If neither GalNAc or Gal have been added to the H antigen structure, then neither A antigen nor B antigen will be present and the blood type is O.

The ABO blood group is determined by a single gene on chromosome 9. Human (and other primate) populations contain many alleles of this gene. The original gene encoded A enzyme, which transfers GalNAc. Variants of this gene have altered the specificity of the enzyme so that it no longer recognizes GalNAc but, instead, transfers Gal. These B enzymes differ by several amino acid residues from the allele that encodes the A enzyme. The structures of both types of glycosyltransferase enzymes have been solved and they reveal that only a single amino acid substitution is required to change the specificity from *N*-acetylaminogalactosyltransferase to galactosyltransferase.

The chromosome 9 locus can also contain several alleles that encode nonfunctional proteins. One of the most common mutations is a single base pair deletion near the *N*-terminus



O-Linked oligosaccharides may account for 80% of the mass of mucins. These large glycoproteins are found in mucus, the viscous fluid that protects and lubricates the epithelium of the gastrointestinal, genitourinary, and respiratory tracts. The oligosaccharide chains of mucins contain an abundance of NeuNAc residues and sulfated sugars. The negative charges of these residues are responsible in part for the extended shape of mucins, which contributes to the viscosity of solutions containing mucins.

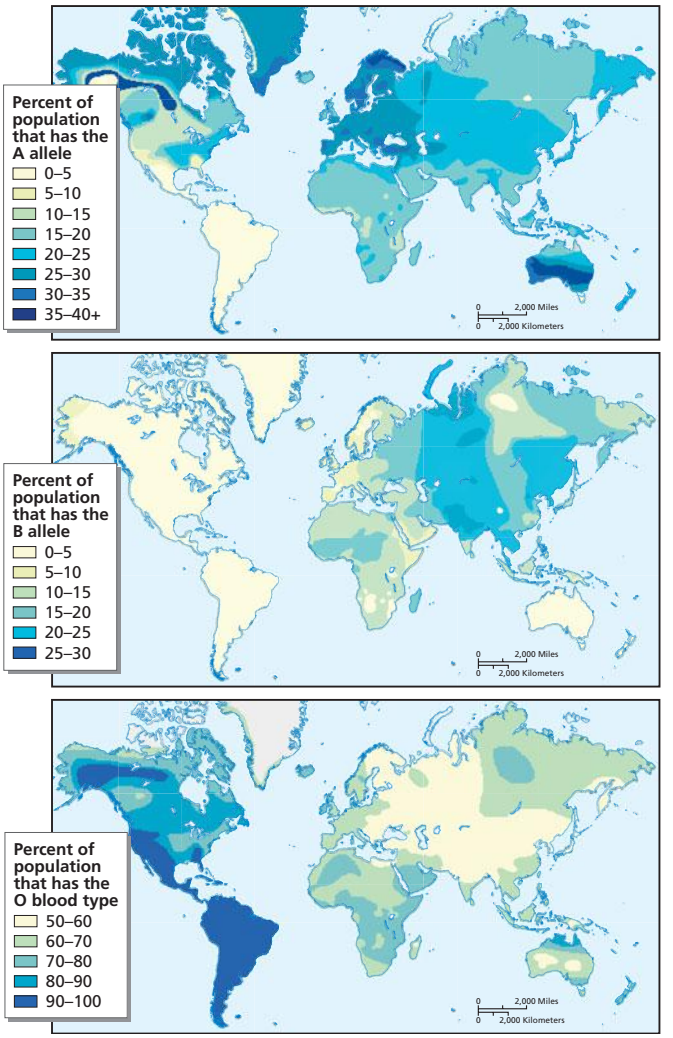
The biosynthesis of the oligosaccharide chains of glycoproteins requires a battery of specific enzymes in distinct compartments of the cell. In the stepwise synthesis of *O*-linked oligosaccharides, glycosyltransferases catalyze the addition of glycosyl groups donated by nucleotide–sugar coenzymes. The oligosaccharide chains are assembled by addition of the first sugar molecule to the protein, followed by subsequent single-sugar additions to the nonreducing end.

N-Linked oligosaccharides, like *O*-linked oligosaccharides, exhibit great variety in sugar sequence and composition. Most *N*-linked oligosaccharides can be divided into

of the coding region. This deletion shifts the reading frame for translation (Section 22.1) making it impossible to synthesize a functional enzyme of either type. This is another example of a human pseudogene. People who are homozygous for these nonfunctional O alleles will not synthesize either A antigen or B antigen and their blood type will be O. (See the Online Medelian Inheritance in Man (OMIM: **ncbi.nlm.nih.gov/omim**) database entry 110300 for an excellent and complete summary of all ABO variants.)

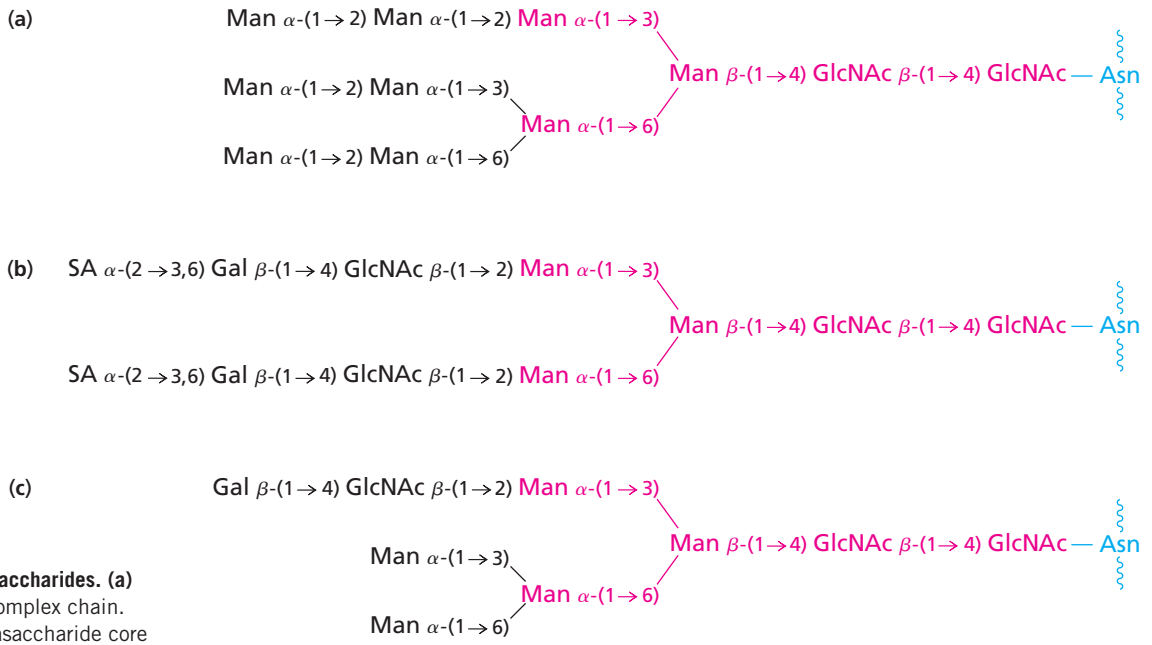
All of your blood cells display some of the unmodified core oligosaccharide (H antigen) even if your blood type is A, B, or AB. This is because not all of the H antigen structures are modified. Under normal circumstances, human plasma will not contain antibodies against H antigen. However, O-type individuals will have antibodies against A antigen and B antigen because these structures are recognized as nonself. If O-type individuals receive a blood transfusion from someone with A, B, or AB blood, they will mount an immune response and reject it. Similarly, if you have A-type blood you will have anti-B antibodies and cannot receive a transfusion from someone with B or AB blood type.

The O allele (pseudogene) is the most common allele in most human populations and the B allele is the most rare. Some Native American populations are homogeneous for the O allele and everyone has type O blood. Type O individuals are perfectly normal, indicating that the absence of the A and B oligosaccharide structures has no effect on normal growth and development (i.e., the allele is neutral in most environments). However, there are some correlations between blood type and disease. People with type O blood, for example, are more susceptible to cholera caused by infections of the bacterium *Vibrio cholerae*. Such selective pressures may be responsible for maintaining the frequencies of A and B alleles in some populations.

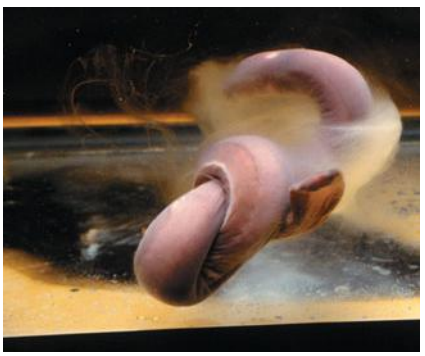


▲ **ABO blood group: distribution of alleles in humans.**

three subclasses: high mannose, complex, and hybrid (Figure 8.35). The appearance of a common core pentasaccharide ($GlcNAc_2Man_3$) in each class reflects a common initial pathway for biosynthesis. The synthesis of *N*-linked oligosaccharides begins with the assembly of a compound consisting of a branched oligosaccharide with 14 residues (nine of which are mannose residues) linked to the lipid dolichol. The entire oligosaccharide chain is transferred to an asparagine residue of a newly synthesized protein, after which the chain is trimmed by the action of glycosidases. High-mannose chains represent an early stage in the biosynthesis of *N*-linked oligosaccharides. Complex oligosaccharide chains result from further removal of sugar residues from high-mannose chains and the addition of other sugar residues, such as fucose, galactose, GlcNAc, and sialic acid (a phenomenon called oligosaccharide processing). These additional sugar residues are donated by nucleotide sugars in reactions catalyzed by glycosyltransferases as in the synthesis of *O*-linked oligosaccharides. In certain cases, a glycoprotein can contain a hybrid oligosaccharide chain, a branched oligosaccharide in which one branch is of the high-mannose type and the other is of the complex type.



▲ **Figure 8.35**
Structures of *N*-linked oligosaccharides. (a) High-mannose chain. **(b)** Complex chain. **(c)** Hybrid chain. The pentasaccharide core common to all *N*-linked structures is shown in red. SA represents sialic acid, usually NeuNAc.



▲ **Mucins.** Mucins are heavily glycosylated proteins secreted by the epithelial cells of animals. You are probably familiar with the mucins secreted by cells lining your mouth (saliva), nasal cavity ("snot"), and intestine. The mucin shown here is being secreted by a hagfish.

The synthesis of glycoproteins is discussed in Section 22.10.

Most glycoproteins are secreted from the cell or are bound to the outer surface of the plasma membrane. There are very few glycoproteins in the cytoplasm. With rare exceptions, none of the basic metabolic enzymes are glycosylated. The addition of oligosaccharide chains is tightly coupled to sorting and secretion in eukaryotic cells. The oligosaccharides are attached to specific proteins in the lumen of the endoplasmic reticulum and the groups are modified by various glycosyltransferase enzymes as the proteins move from the ER through the Golgi to the cell surface. The structure of the linked oligosaccharide serves as a marker for sorting proteins into various compartments. For example, some proteins are targeted to the lysosomes, depending on the structure of the oligosaccharide, while others are marked for secretion.

In addition to their roles as markers in sorting and secretion, the presence of one or more oligosaccharide chains on a protein can alter its physical properties, including its size, shape, solubility, electric charge, and stability. Biological properties that can be altered include rate of secretion, rate of folding, and immunogenicity. In a few cases, specific roles for the oligosaccharide chains of glycoproteins have been identified. For example, a number of mammalian hormones are dimeric glycoproteins whose oligosaccharide chains facilitate assembly of the dimer and confer resistance to proteolysis. Also, the recognition of one cell by another that occurs during cell migration or oocyte fertilization can depend in part on the binding of proteins on the surface of one cell to the carbohydrate portions of certain glycoproteins on the surface of the other cell.

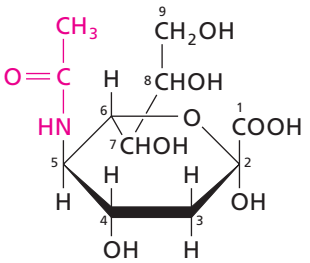
Summary

1. Carbohydrates include monosaccharides, oligosaccharides, and polysaccharides. Monosaccharides are classified as aldoses or ketoses or their derivatives.
2. A monosaccharide is designated D or L, according to the configuration of the chiral carbon farthest from the carbonyl carbon atom. Each monosaccharide has 2^n possible stereoisomers, where n is the number of chiral carbon atoms. Enantiomers are nonsuperimposable mirror images of each other. Epimers differ in configuration at only one of several chiral centers.
3. Aldoses with at least five carbon atoms and ketoses with at least six carbon atoms exist principally as cyclic hemiacetals or hemiketals known as furanoses and pyranoses. In these ring structures, the configuration of the anomeric (carbonyl) carbon is designated either α or β. Furanoses and pyranoses can adopt several conformations.
4. Derivatives of monosaccharides include sugar phosphates, deoxy sugars, amino sugars, sugar alcohols, and sugar acids.

5. Glycosides are formed when the anomeric carbon of a sugar forms a glycosidic linkage with another molecule. Glycosides include disaccharides, polysaccharides, and some carbohydrate derivatives.
6. Homoglycans are polymers containing a single type of sugar residue. Examples of homoglycans include the storage polysaccharides starch and glycogen and the structural polysaccharides cellulose and chitin.
7. Hetero glycans contain more than one type of sugar residue. They are found in glycoconjugates such as proteoglycans, peptidoglycans, and glycoproteins.
8. Proteoglycans are proteins linked to chains of repeating disaccharides. Proteoglycans are prominent in the extracellular matrix and in connective tissues such as cartilage.
9. The cell walls of many bacteria are made of peptidoglycans that are heteroglycan chains linked to peptides. Peptidoglycan molecules are extensively cross-linked, essentially converting peptidoglycan into a rigid macromolecule that defines the shape of a bacterium and protects the plasma membrane.
10. Glycoproteins are proteins containing covalently bound oligosaccharides. The oligosaccharide chains of most glycoproteins are either *O*-linked to serine or threonine residues or *N*-linked to asparagine residues and exhibit great variety in structure and sugar composition.

Problems

1. Identify each of the following:
   (a) Two aldoses whose configuration at carbons 3, 4, and 5 matches that of D-fructose.
   (b) The enantiomer of D-galactose.
   (c) An epimer of D-galactose that is also an epimer of D-mannose.
   (d) A ketose that has no chiral centers.
   (e) A ketose that has only one chiral center.
   (f) Monosaccharide residues of cellulose, amylose, and glycogen.
   (g) Monosaccharide residues of chitin.
2. Draw Fischer projections for (a) L-mannose, (b) L-fucose (6-deoxy-L-galactose), (c) D-xylitol, and (d) D-iduronate.
3. Describe the general structural features of glycosaminoglycans.
4. Honey is an emulsion of microcrystalline D-fructose and D-glucose. Although D-fructose in polysaccharides exists mainly in the furanose form, solution or crystalline D-fructose (as in honey) is a mixture of several forms with β-D-fructopyranose (67%) and β-D-fructofuranose (25%) predominating. Draw the Fischer projection for D-fructose and show how it can cyclize to form both of the cyclized forms above.
5. Sialic acid (*N*-acetyl-α-D-neuraminic acid) is often found in *N*-linked oligosaccharides that are involved in cell-cell interactions. Cancer cells synthesize much greater amounts of sialic acid than normal cells. Derivatives of sialic acid have been proposed as anticancer agents to block cell-surface interactions between normal and cancerous cells. Answer the following questions about the structure of sialic acid.
   (a) Is it an α or a β anomeric form?
   (b) Will sialic acid mutorotate between α and β anomeric forms?
   (c) Is this a "deoxy" sugar?
   (d) Will the open-chain form of sialic acid be an aldehyde or a ketone?
   (e) How many chiral carbons are there in the sugar ring?



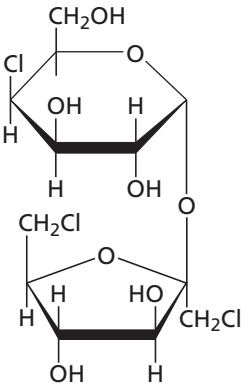
Sialic acid

6. How many stereoisomers are possible for glucopyranose and for fructofuranose? How many are D sugars in each case, and how many are L sugars?
7. Draw the structure of each of the following molecules and label each chiral carbon with an asterisk:
   (a) α-D-Glucose 1-phosphate.
   (b) 2-Deoxy-β-D-ribose 5-phosphate.
   (c) D-Glyceraldehyde 3-phosphate.
   (d) L-Glucuronate.
8. In aqueous solution, almost all D-glucose molecules (>99%) are in the pyranose form. Other aldoses have a greater proportion of molecules in the open-chain form. D-Glucose may have evolved to be the predominant hexose because it is less likely than its isomers to react with and damage cellular proteins. Explain why D-glucose reacts less than other aldoses with the amino groups of proteins.
9. Why is the β-D-glucopyranose form of glucose more abundant than α-D-glucopyranose in aqueous solution?
10. The relative orientations of substituents on ribose rings are determined by the conformation of the ring itself. If the ribose is part of a polymeric molecule, then ring conformation will affect overall polymer structure. For example, the orientation of ribose phosphate substituents connecting monomeric nucleoside units is important in determining the overall structure of nucleic acid molecules. In one major form of DNA (B-DNA), the ribofuranose rings adopt an envelope conformation in which C-2′ carbon is above the plane defined by C-1, C-3, C-4, and the ring oxygen (C-2′ endo conformation). Draw the envelope structure of D-ribose 5-phosphate with a nucleoside base (B) attached in a β-anomeric position at the C-1 carbon.
11. In a procedure for testing blood glucose, a drop of blood is placed on a paper strip impregnated with the enzyme glucose oxidase and all the reagents necessary for the reaction

$$\beta\text{-}D\text{-Glucose} + O_2 \longrightarrow D\text{-Gluconolactone} + H_2O_2$$

The H_2O_2 produced causes a color change on the paper that indicates how much glucose is present. Since glucose oxidase is specific for the β anomer of glucose, why can the total blood glucose be measured?
12. Sucralose (registered under the brand name Splenda®) is a nonnutritive (noncaloric) sweetener that is approximately 600 times sweeter than sugar. Since sucralose is heat stable, it can be used in cooking and baking. The structure of sucralose is shown below.

Name the disaccharide that is used as a starting substrate for the synthesis of sucralose. What chemical modifications have been made to the starting disaccharide?



13. Draw Haworth projections for the following glycosides:

(a) Isomaltose [α-D-glucopyranosyl-(1 → 6)-α-D-glucopyranose].
(b) Amygdalin, a compound in the pits of certain fruits, which has a —CH(CN)C_6H_5 group attached to C-1 of β-D-glucopyranosyl-(1 → 6)-β-D-glucopyranose.
(c) The *O*-linked oligosaccharide in collagen (β-D-galactose attached to a 5-hydroxylysine residue)

14. Keratan sulfate is a glycosaminoglycan composed primarily of the following repeating disaccharide unit:—Gal β(1 → 4) GlcNAc6S β(1 → 3)—. The acetylated sugar has a sulfate ester on C-6. Keratan sulfate is found in cornea, bone, and cartilage aggregated with other glycosaminoglycans such as chondroitin sulfate. Draw a Haworth projection of the repeating disaccharide unit found in keratan sulfate.

15. A number of diseases result from hereditary deficiencies in specific glycosidases. In these diseases, certain glycoproteins are incompletely degraded and oligosaccharides accumulate in tissues. Which of the *N*-linked oligosaccharides in Figure 8.35 would be affected by deficiencies of the following enzymes?

(a) *N*-Acetyl-β-glucosaminyl asparagine amidase
(b) β-Galactosidase
(c) Sialidase
(d) Fucosidase

16. A carbohydrate–amino acid polymer that is a potent inhibitor of influenza virus has been synthesized. The virus is thought to be inactivated when multiple sialyl groups bind to viral surface proteins. Draw the chemical structure of the carbohydrate portion of this polymer (below, where X represents the rest of the polymer).

NeuNAc α-(2 → 3) Gal β-(1 → 4) Glu β-(1 →)-X

17. Imagine that you could take a pill containing β-glucosidase. If, after taking this pill, you ate this textbook, what would it taste like? Would it taste any different if you could marinate it overnight in a solution containing β-glucosidase? Should publishers use flavored ink in order to encourage students to eat their textbooks?

Selected Readings

General

Collins, P. M., ed. (1987). *Carbohydrates* (London and New York: Chapman and Hall).

El Khadem, H. S. (1988). *Carbohydrate Chemistry: Monosaccharides and Their Derivatives* (Orlando, FL: Academic Press).

Li, X., Glaser, D., Li, W., Johnson, W. E., O'Brien, S. J., Beauchamp, G. K., and Brand, J. G. (2009). Analyses of sweet receptor gene (Tas1r2) and preference for sweet stimuli in species of Carnivora. *J. Hered.* 100(Supplement 1):S90–S100.

Li, X., Li, W., Wang, H., Cao, J., Maehashi, K., Huang, L., Bachmanov, A. A., Reed, D. R., Legrand-Defretin, V., Beauchamp, G. K., and Brand, J. G. (2005). Pseudogenization of a sweet-receptor gene accounts for cats' indifference toward sugar. *PloS Genet.* 1(1): e3. DOI:10.1371/journal.pgen.0010003

Nodulation Factors

Dénarié, J., and Debellé, F. (1996). Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu. Rev. Biochem.* 65:503–535.

Madsen, L. H., Tirichine, L., Jurkiewicz, A., Sullivan, J. T., Heckmann, A. B., Bek, A. S., Ronson, C. W., James, E. K., and Stougaard, J. (2010). The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nature Communications.* DOI:10.1038/ncomms1009

Mergaert, P., Van Montagu, M., and Holsters, M. (1997). Molecular mechanisms of Nod factor diversity. *Mol. Microbiol.* 25:811–817.

Thoden, J. B., Kim, J., Raushel, F. M., and Holden, H. M. (2002). Structural and kinetic studies of sugar binding to galactose mutarotase from *Lactococcus lactis*. *J. Biol. Chem.* 277:45458–45465.

Proteoglycans

Heinegård, D., and Oldberg, Å. (1989). Structure and biology of cartilage and bone matrix noncollagenous macromolecules. *FASEB J.* 3:2042–2051.

Iozzo, R. V. (1999). The biology of the small leucine-rich proteoglycans: functional network of interactive proteins. *J. Biol. Chem.* 274:18843–18846.

Iozzo, R. V., and Murdoch, A. D. (1996). Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function. *FASEB J.* 10:598–614.

Kjellén, L., and Lindahl, U. (1991). Proteoglycans: structures and interactions. *Annu. Rev. Biochem.* 60:443–475.

Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* 75:39–68.

Glycoproteins

Drickamer, K., and Taylor, M. E. (1998). Evolving views of protein glycosylation. *Trends Biochem. Sci.* 23:321–324.

Dwek, R. A., Edge, C. J., Harvey, D. J., Wormald, M. R., and Parekh, R. B. (1993). Analysis of glycoprotein-associated oligosaccharides. *Annu. Rev. Biochem.* 62:65–100.

Fudge, D. S., Levy, N., Chiu, S., and Gosline, J. M. (2005). Composition, morphology and mechanics of hagfish slime. *J. Exp. Biol.* 208:4613–4625.

Lairson, L. L., Henrissat, B., Davies, G., and Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem.* 77:521–555.

Lechner, J., and Wieland, F. (1989). Structure and biosynthesis of prokaryotic glycoproteins. *Annu. Rev. Biochem.* 58:173–194.

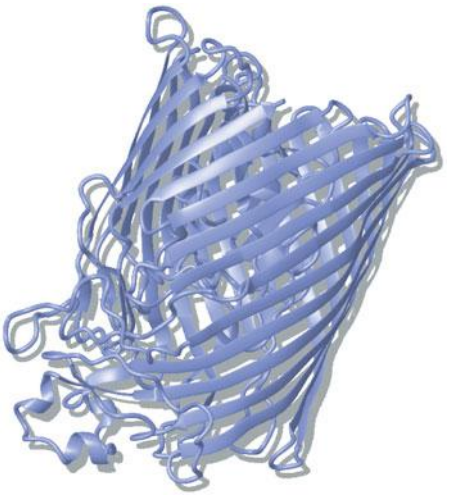
Marionneau, S., Caileau-Thomas, A., Rocher, J., Le Moullac-Vaidye, B. Ruvoën, N., Clément, M., and Le Pendu, J. (2001). ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie.* 83:565–573.

Patenaude, S. I., Seto, N. O. L., Borisova, S. N., Szpacenko, A., Marcus, S. L., Palcic, M. M., and Evans, S. V. (2002). The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nat. Struct. Biol.* 9:685–690.

Rademacher, T. W., Parekh, R. B., and Dwek, R. A. (1988). Glycobiology. *Annu. Rev. Biochem.* 57:785–838

Rudd, P. M., and Dwek, R. A. (1997). Glycosylation: heterogeneity and the 3D structure of proteins. *Crit. Rev. Biochem. Mol. Biol.* 32:1–100.

Strous, G. J., and Dekker, J. (1992). Mucin-type glycoproteins. *Crit. Rev. Biochem. Mol. Biol.* 27:57–92.



Lipids and Membranes

In this chapter, we consider **lipids**, (*lipo*-, fat) a third major class of biomolecules. Lipids—like proteins and carbohydrates—are essential components of all living organisms. However, unlike these other types of biomolecules, lipids have widely varied structures. They are often defined as water-insoluble (or only sparingly soluble) organic compounds found in biological systems but that's a very broad definition. Lipids are very soluble in nonpolar organic solvents. They are either hydrophobic (nonpolar) or amphipathic (containing both nonpolar and polar regions).

We begin this chapter with a discussion of the structures and functions of the different classes of lipids. In the second part of the chapter, we examine the structures of biological membranes whose properties as cellular barriers depend on the properties of their lipids. Finally, we describe the principles of membrane transport and transmembrane signaling pathways.

In this article, we therefore present and discuss a fluid mosaic model of membrane structure, and propose that it is applicable to most biological membranes, such as plasmalemmal and intracellular membranes, including the membranes of different cell organelles such as mitochondria and chloroplasts.

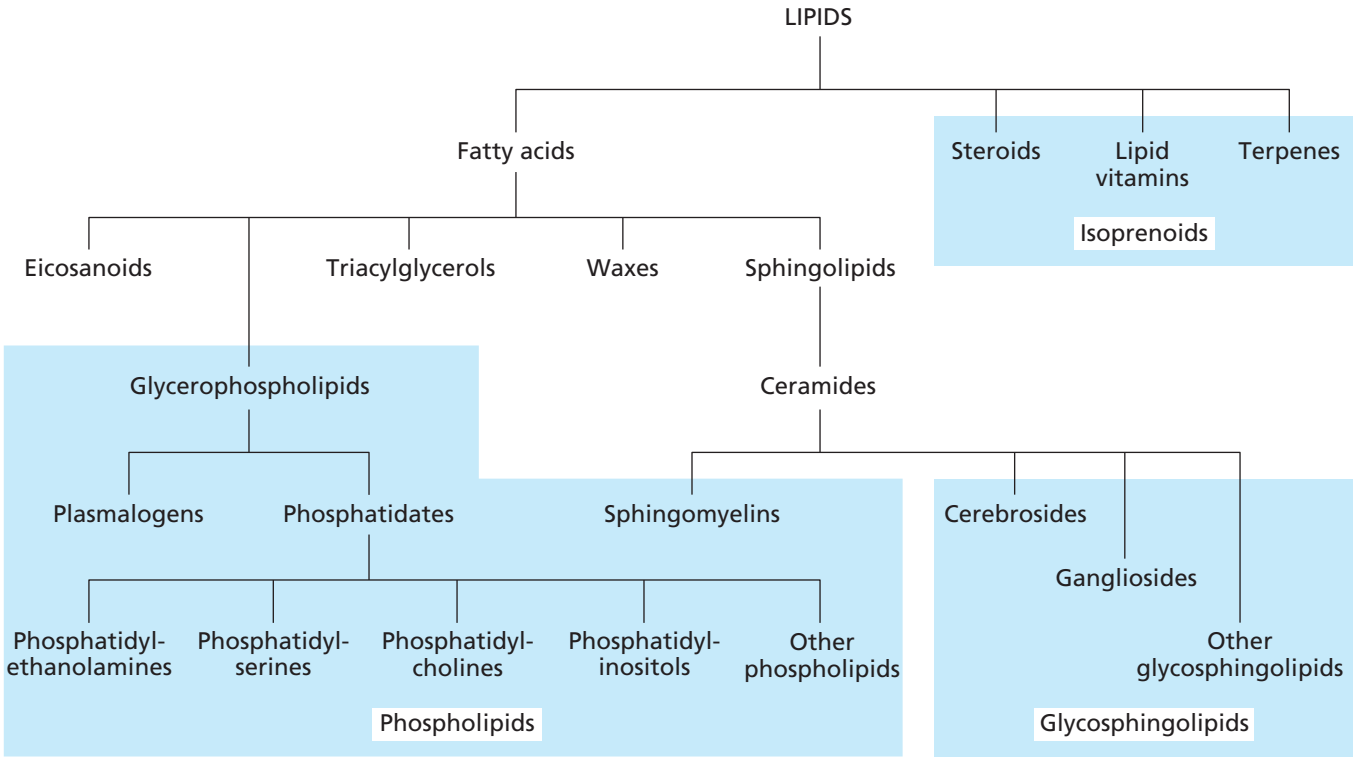
—S.J. Singer and G.L. Nicholson (1972)

9.1 Structural and Functional Diversity of Lipids

Figure 9.1 shows the major types of lipids and their structural relationships to one another. The simplest lipids are the **fatty acids** that have the general formula R—COOH, where R represents a hydrocarbon chain composed of various lengths of —CH_2-(methylene) units. Fatty acids are components of many more complex types of lipids, including triacylglycerols, glycerophospholipids, and sphingolipids. Lipids containing phosphate groups are called **phospholipids** and lipids containing both sphingosine and carbohydrate groups are called **glycosphingolipids**. Steroids, lipid vitamins, and terpenes are related to the five-carbon molecule isoprene and are therefore called **isoprenoids**. The name *terpenes* has been applied to all isoprenoids but usually is restricted to those that occur in plants.

Lipids have diverse biological functions as well as diverse structures. Biological membranes contain a variety of amphipathic lipids including glycerophospholipids and

Top: Ribbon structure of the transmembrane portion of porin FhuA from *Escherichia coli* (see Figure 9.28).



▲ **Figure 9.1**
Structural relationships of the major classes of lipids. Fatty acids are the simplest lipids. Many other types of lipids either contain or are derived from fatty acids. Glycerophospholipids and sphingomyelins contain phosphate and are classified as phospholipids. Cerebrosides and gangliosides contain sphingosine and carbohydrate and are classified as glycosphingolipids. Steroids, lipid vitamins, and terpenes are called isoprenoids because they are related to the five-carbon molecule isoprene rather than to fatty acids.

sphingolipids. In some organisms, triacylglycerols (fats and oils) function as intracellular storage molecules for metabolic energy. Fats also provide animals with thermal insulation and padding. Waxes in cell walls, exoskeletons, and skins protect the surfaces of some organisms. Some lipids have highly specialized functions. For example, steroid hormones regulate and integrate a host of metabolic activities in animals and eicosanoids participate in the regulation of blood pressure, body temperature, and smooth-muscle contraction in mammals. Gangliosides and other glycosphingolipids are located at the cell surface and can participate in cellular recognition.

9.2 Fatty Acids

Fatty acid biosynthesis is discussed in Chapter 16.

More than 100 different fatty acids have been identified in various species. Fatty acids differ from one another in the length of their hydrocarbon tails, the number of carbon–carbon double bonds, the positions of the double bonds in the chains, and the number of branches. Some of the fatty acids commonly found in mammals are shown in Table 9.1.

All fatty acids have a carboxyl group (—COOH) at their "head." This is why they are acids. The pK_a of this group is about 4.5 to 5.0 so it is ionized at physiological pH (—COO^-). Fatty acids are a form of detergent because they have a long hydrophobic tail and a polar head (Section 2.4). As expected, the concentration of *free* fatty acid in cells is quite low because high concentrations of free fatty acids could disrupt membranes. Most fatty acids are components of more complex lipids. They are joined to other molecules by an ester linkage at the terminal carboxyl group.

Fatty acids can be referred to by either International Union of Pure and Applied Chemistry (IUPAC) names or common names. Common names are used for the most frequently encountered fatty acids.

The number of carbon atoms in the most abundant fatty acids ranges from 12 to 20 and is almost always an even number since fatty acids are synthesized by the sequential addition of two-carbon units. In IUPAC nomenclature, the carboxyl carbon is labeled C-1 and the remaining carbon atoms are numbered sequentially. In common

Table 9.1 Some common fatty acids (anionic forms)

Number of carbons	Number of double bonds	Common name	IUPAC name	Molecular formula	Melting point, °C
12	0	Laurate	Dodecanoate	$CH_3(CH_2)_{10}COO^{\ominus}$	44
14	0	Myristate	Tetradecanoate	$CH_31(CH_21)_{12}COO^{\ominus}$	52
16	0	Palmitate	Hexadecanoate	$CH_31(CH_21)_{14}COO^{\ominus}$	63
18	0	Stearate	Octadecanoate	$CH_3(CH_2)_{16}COO^{\ominus}$	70
20	0	Arachidate	Eicosanoate	$CH_3(CH_2)_{18}COO^{\ominus}$	75
22	0	Behenate	Docosanoate	$CH_3(CH_2)_{20}COO^{\ominus}$	81
24	0	Lignocerate	Tetracosanoate	$CH_3(CH_2)_{22}COO^{\ominus}$	84
16	1	Palmitoleate	*cis*-Δ^9-Hexadecenoate	$CH_3(CH_2)_5CH{=}CH(CH_2)_7COO^{\ominus}$	−0.5
18	1	Oleate	*cis*-Δ^9-Octadecenoate	$CH_3(CH_2)_7CH{=}CH(CH_2)_7COO^{\ominus}$	13
18	2	Linoleate	cis, *cis*-$\Delta^{9,12}$-Octadecadienoate	$CH_3(CH_2)_4(CH{=}CHCH_2)_2(CH_2)_6COO^{\ominus}$	−9
18	3	Linolenate	all *cis*-$\Delta^{9,12,15}$-Octadecatrienoate	$CH_3CH_2(CH{=}CHCH_2)_3(CH_2)_6COO^{\ominus}$	−17
20	4	Arachidonate	all *cis*-$\Delta^{5,8,11,14}$-Eicosatetraenoate	$CH_3(CH_2)_4(CH{=}CHCH_2)_4(CH_2)_2COO^{\ominus}$	−49

nomenclature, Greek letters are used to identify the carbon atoms. The carbon adjacent to the carboxyl carbon (C-2 in IUPAC nomenclature) is designated α, and the other carbons are lettered β, γ, δ, and ε and so on (Figure 9.2). The Greek letter ω (omega) specifies the carbon atom farthest from the carboxyl group, whatever the length of the hydrocarbon tail. (ω is the last letter in the Greek alphabet.)

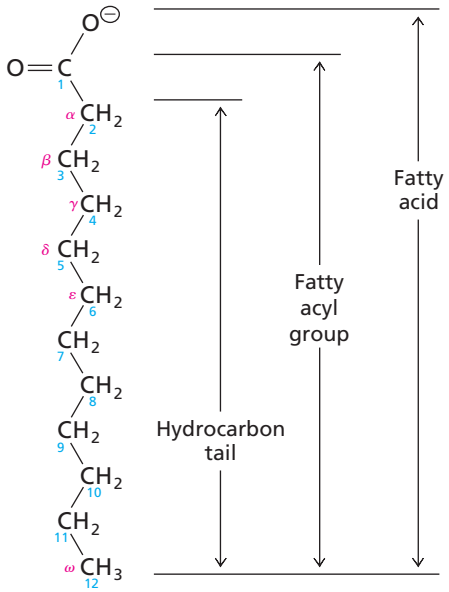
Fatty acids without a carbon–carbon double bond are classified as **saturated**, whereas those with at least one carbon–carbon double bond are classified as **unsaturated**. Unsaturated fatty acids with only one carbon–carbon double bond are called **monounsaturated** and those with two or more are called **polyunsaturated**. The configuration of the double bonds in unsaturated fatty acids can be either *cis* or *trans* . The configuration is usually *cis* in naturally occurring fatty acids (see Box. 9.2).

The positions of double bonds are indicated by the symbol Δ^n in IUPAC nomenclature, where the superscript n indicates the lower-numbered carbon atom of each

BOX 9.1 COMMON NAMES OF THE FATTY ACIDS

Laurate present in oil from the laurel plant (*Laurus nobilis*) (1873)
Myristate oil from nutmeg (*Myristica fragrans*) (1848)
Palmitate from palm oil (1857)
Stearate from French *stéarique* referring to fat from steers, or tallow (1831)
Arachidate present in oil from peanuts (*Arachis hypogaea*) (1866)
Behenate a corruption of "ben" from ben-nut = seeds of the Horseradish tree (1873)
Lignocerate probably from Latin *lignum* ("wood") (~1900)
Oleate from Latin *oleum* ("oil") (1899)
Linoleate found in linseed oil (*lin* + *oleate*) (1857)

▲ **The African oil palm tree, *Elaeis guineensis*.** Palm oil is a complex mixture of saturated and unsaturated fatty acids but palmitate makes up 44% of the total. The presence of such a large amount of saturated fatty acid means that palm oil is a semisolid at room temperature. It can never be "virgin" or "extra virgin" (see Box 16.6).



◀ **Figure 9.2**
Structure and nomenclature of fatty acids. Fatty acids consist of a long hydrocarbon tail terminating with a carboxyl group. Since the pK_a of the carboxyl group is approximately 4.5 to 5.0, fatty acids are anionic at physiological pH. In IUPAC nomenclature, carbons are numbered beginning with the carboxyl carbon. In common nomenclature, the carbon atom adjacent to the carboxyl carbon is designated α, and the remaining carbons are lettered β, γ, δ, and so on. The carbon atom farthest from the carboxyl carbon is designated the ω carbon, whatever the length of the tail. The fatty acid shown, laurate (or dodecanoate), has 12 carbon atoms and contains no carbon–carbon double bonds.

double-bonded pair (Table 9.1). The double bonds of most polyunsaturated fatty acids are separated by a methylene group and are therefore not conjugated.

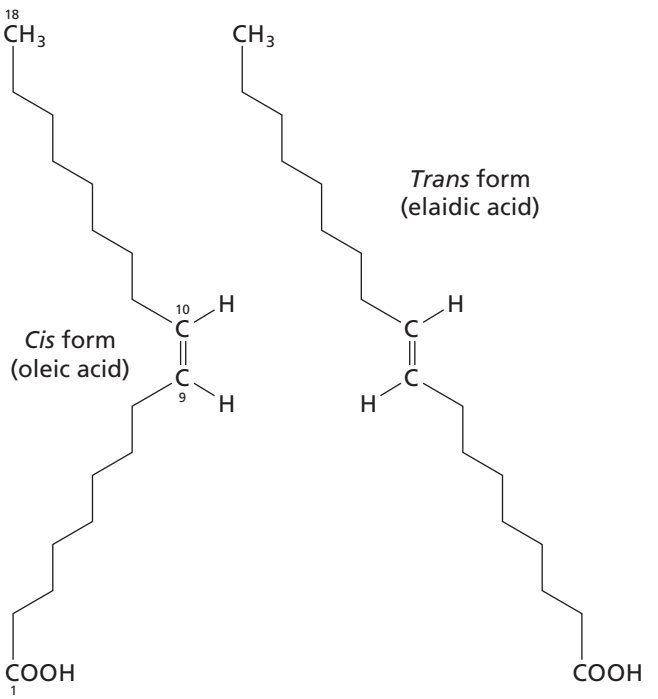
A shorthand notation for identifying fatty acids uses two numbers separated by a colon—the first refers to the number of carbon atoms in the fatty acid and the second refers to the number of carbon–carbon double bonds, with their positions indicated as superscripts following a Greek symbol, Δ. In this notation, palmitate is written as 16:0, oleate as 18:1 Δ^9, and arachidonate as 20:4 $\Delta^{5,8,11,14}$. Unsaturated fatty acids can

BOX 9.2 *TRANS* FATTY ACIDS AND MARGARINE

The configuration of most double bonds in unsaturated fatty acids is *cis* but some fatty acids in the human diet have the *trans* configuration. *Trans* fatty acids can come from animal sources such as dairy products and ruminant meats. However, most of the edible *trans* fatty acids consumed in Western industrialized countries are present as hydrogenated vegetable oils in some margarines or shortenings. Dietary *trans* monounsaturated fatty acids can increase plasma levels of cholesterol and triglycerides and their ingestion may increase the risk of cardiovascular disease. More work is required to establish the exact level of risk.

Plant oils such as corn oil and sunflower oil can be converted to "spreadable" semisolid substances known as margarines. Margarines can be produced by the partial or complete hydrogenation of double bonds in plant oils. The hydrogenation process itself not only saturates the carbon–carbon double bonds of fatty acid esters but can also change the configuration of the remaining double bonds from *cis* to *trans*. The physical properties of these *trans* fatty acids are similar to those of saturated fatty acids.

In order to reduce consumption of *trans* fatty acids, many margarines are now produced from plant oils without hydrogenation by adding other edible components such as skim milk powder.



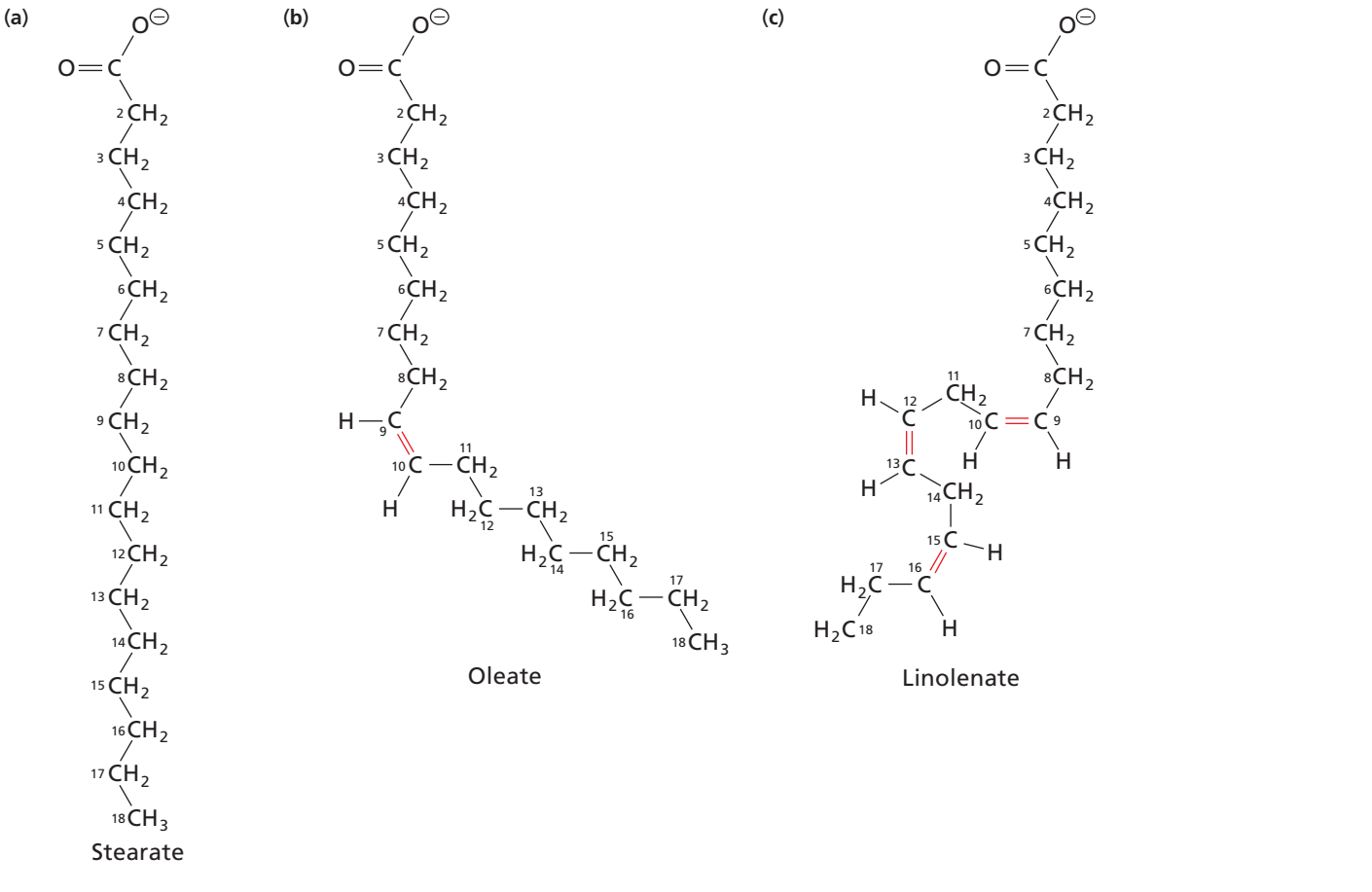
▲ ***Cis* and *trans* forms of Δ^9-octadecanoate.** (Left) Oleate (*cis*-Δ^9-octadecanoate). (Right) the *trans* configuration after hydrogenation.

also be described by the location of the last double bond in the chain. This double bond is usually found three, six, or nine carbon atoms from the end of the chain. Such fatty acids are called ω - 3 (e.g., 18:3 $\Delta^{9,12,15}$), ω - 6 (e.g., 18:2 $\Delta^{9,12}$), or ω - 9 (e.g., 18:1 Δ^{9}).

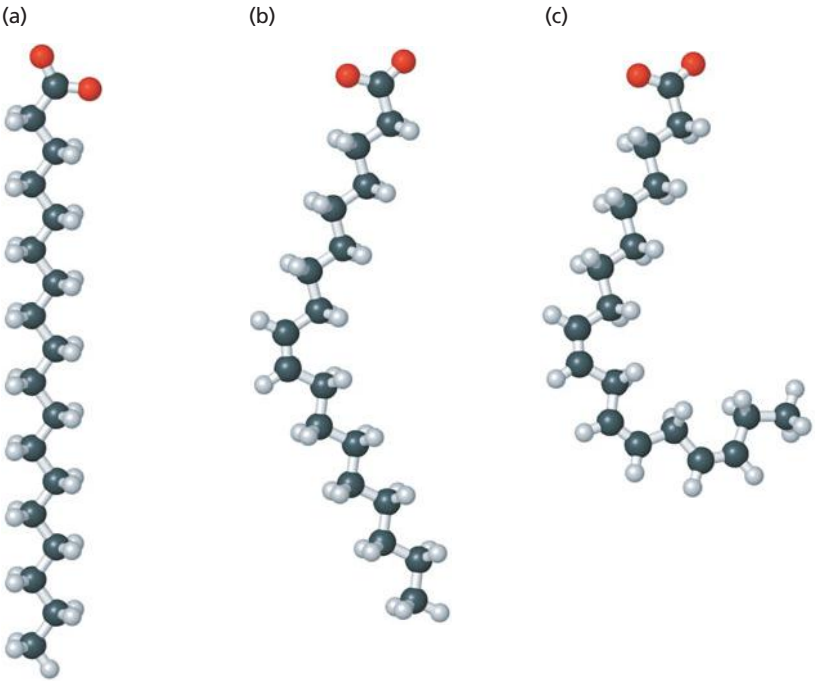
The physical properties of saturated and unsaturated fatty acids differ considerably. Typically, saturated fatty acids are waxy solids at room temperature (22°C) whereas unsaturated fatty acids are liquids at this temperature. The length of the hydrocarbon chain of a fatty acid and its degree of unsaturation influence the melting point. Compare the melting points listed in Table 9.1 for the saturated fatty acids laurate (12:0), myristate (14:0), and palmitate (16:0). As the lengths of the hydrocarbon tails increase, the melting points of the saturated fatty acids also increase. The number of van der Waals interactions among neighboring hydrocarbon tails increases as the tails get longer so more energy is required to disrupt the interactions.

Compare the structures of stearate (18:0), oleate (18:1), and linolenate (18:3) in Figures 9.3 and 9.4. The saturated hydrocarbon tail of stearate is flexible since rotation can occur around every carbon–carbon bond. In a crystal of stearic acid, the hydrocarbon chains are extended and pack together closely. The presence of *cis* double bonds in oleate and linolenate produces pronounced bends in the hydrocarbon chains since rotation around double bonds is hindered. These bends prevent close packing and extensive van der Waals interactions among the hydrocarbon chains. Consequently, *cis* unsaturated fatty acids have lower melting points than saturated fatty acids. As the degree of unsaturation increases, fatty acids become more fluid. Note that stearic acid (melting point 70°C) is a solid at body temperature but oleic acid (melting point 13°C) and linolenic acid (melting point −17°C) are both liquids.

As mentioned earlier, free fatty acids occur only in trace amounts in living cells. Most fatty acids are esterified to glycerol or other backbone compounds to form more complex lipid molecules. In esters and other derivatives of carboxylic acids, the RC═O moiety contributed by the acid is called the acyl group. In common nomenclature,



▼ **Figure 9.3**
Structures of three C_{18} fatty acids. (a) Stearate (octadecanoate), a saturated fatty acid. **(b)** Oleate (*cis*-Δ^9-octadecenoate) a monounsaturated fatty acid. **(c)** Linolenate (all-*cis*-$\Delta^{9,12,15}$-octadecatrienoate) a polyunsaturated fatty acid. The *cis* double bonds produce kinks in the tails of the unsaturated fatty acids. Linolenate is a very flexible molecule that can assume a variety of conformations.



◀ **Figure 9.4**
Stearate (left), oleate (center), and linolenate (right). Color key: carbon, grey; hydrogen, white; oxygen, red.

complex lipids that contain specific fatty acyl groups are named after the parent fatty acid. For example, esters of the fatty acid laurate are called lauroyl esters, and esters of linoleate are called linoleoyl esters. (A lauryl group is the alcohol analog of the lauroyl acyl group). The relative abundance of particular fatty acids varies with the type of organism, type of organ (in multicellular organisms), and food source. The most abundant fatty acids in animals are usually oleate (18:1), palmitate (16:0), and stearate (18:0).

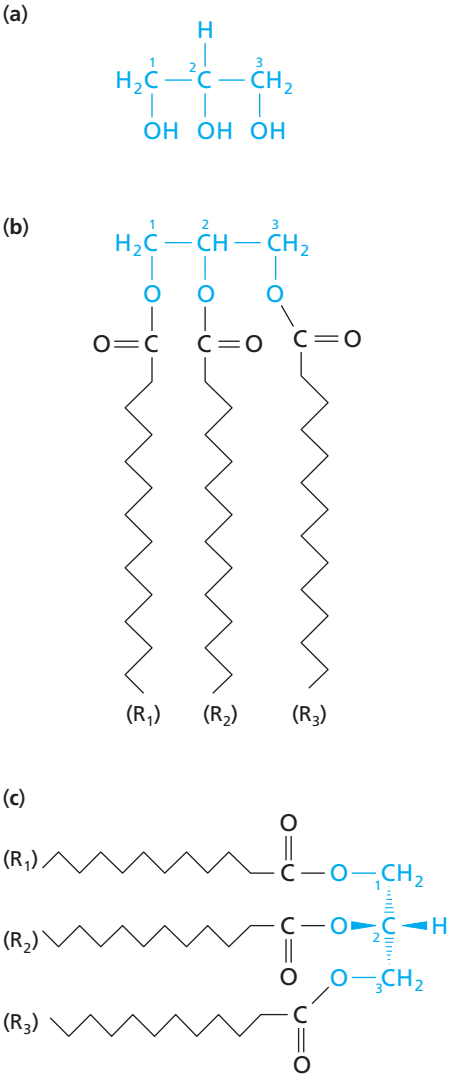
Mammals require certain dietary polyunsaturated fatty acids that they cannot synthesize, such as linoleate (18:2 $\Delta^{9,12}$) and linolenate (18:3 $\Delta^{9,12,15}$). These fatty acids are called essential fatty acids. Mammals can synthesize other polyunsaturated fatty acids from an adequate supply of linoleate and linolenate. (Recall that many vitamins are also essential components of the mammalian diet because mammals cannot synthesize them. In addition to vitamins and essential fatty acids, we will see in Chapter 17 that many amino acids cannot be synthesized in mammals.)

Linolenate is an omega-3 (ω - 3) fatty acid since the last double bond is three carbon atoms from the tail end of the molecule. Omega-3 fatty acids are very popular dietary supplements. They are enriched in fish oils, which is why many people recommend that you include fish and fish oils in your diet. Linolenate is an essential fatty acid so your diet must include an adequate supply of this omega-3 fatty acid. This adequate amount is readily supplied in the typical diet of people all over the world, which is why essential fatty acid deficiency is rare. The market for supplemental omega-3 fatty acids is driven by other factors. The most important benefit is protection against cardiovascular disease. The scientific evidence indicates that extra amounts of omega-3 fatty acids provide a small benefit in terms of reducing the risk of heart attacks, particularly a second heart attack. None of the other claims are based on reproducible double-blind test results after controlling for other factors. Eating fish, for example, will not make you smarter.

Many fatty acids besides those listed in Table 9.1 are present in nature. For example, fatty acids containing cyclopropane rings are found in bacteria. Branched-chain fatty acids are common components of bacterial membranes and also occur on the feathers of ducks. Many other fatty acids are rare and have highly specialized functions.

9.3 Triacylglycerols

As their name implies, **triacylglycerols** (historically called triglycerides) are composed of three fatty acyl residues esterified to glycerol, a three-carbon sugar alcohol (Figure 9.5). Triacylglycerols are very hydrophobic.



▲ **Figure 9.5**
Structure of a triacylglycerol. Glycerol **(a)** is the backbone to which three fatty acyl residues are esterified **(b)**. Although glycerol is not chiral, C-2 of a triacylglycerol is chiral when the acyl groups bound to C-1 and C-3 (R_1 and R_3) differ. The general structure of a triacylglycerol is shown in **(c)**, oriented for comparison with the structure of L-glyceraldehyde (Figure 8.1). This orientation allows stereospecific numbering of glycerol derivatives with C-1 at the top and C-3 at the bottom.

▲ **Figure 9.6**
Adipocytes. This is a colorized scanning electron micrograph of clusters of adipocytes. A fat droplet occupies most of the volume of each adipocyte.

The structures and functions of lipoproteins are discussed in Section 16.1B.

Fats and oils are mixtures of triacylglycerols. They can be solids (fats) or liquids (oils), depending on their fatty acid compositions and on the temperature. Triacylglycerols containing only saturated long chain fatty acyl groups tend to be solids at body temperature and those containing unsaturated or short chain fatty acyl groups tend to be liquids. A sample of naturally occurring triacylglycerols can contain as many as 20 to 30 molecular species that differ in their fatty acid constituents. Tripalmitin, found in animal fat, contains three residues of palmitic acid. Triolein, which contains three oleic acid residues, is the principal triacylglycerol in olive oil.

In most cells, triacylglycerols coalesce as fat droplets. These droplets are sometimes seen near mitochondria in cells that rely on fatty acids for metabolic energy. In mammals, most fat is stored in adipose tissue that is composed of specialized cells known as adipocytes. Each adipocyte contains a large fat droplet that accounts for nearly the entire volume of the cell (Figure 9.6). Although distributed throughout the bodies of mammals, most adipose tissue occurs just under the skin and in the abdominal cavity. Extensive subcutaneous fat serves both as a storage depot for energy and as thermal insulation and is especially pronounced in aquatic mammals.

9.4 Glycerophospholipids

Triacylglycerols are not found in biological membranes. The most abundant lipids in most membranes are **glycerophospholipids** (also called phosphoglycerides). Glycerophospholipids, like triacylglycerols, have a glycerol backbone. The simplest glycerophospholipids are the, phosphatidates—they consist of two fatty acyl groups esterified to C-1 and C-2 of glycerol 3-phosphate (Table 9.2). Note that there are three fatty acyl groups esterified to glycerol in triacylglycerols whereas there are only two fatty acyl groups (R_1 and R_2) in the glycerophospholipids. The distinguishing feature of the glycerophospholipids is the presence of a phosphate group on C-3 of the glycerol backbone. The structures of glycerophospholipids can be drawn as derivatives of L-glycerol 3-phosphate with the C-2 substituent on the left in a Fischer projection, as in Table 9.2. For simplicity, we usually show these compounds as stereochemically uncommitted structures.

KEY CONCEPT

Glycerophospholipids have polar heads and long, hydrophobic fatty acid tails.

Phosphatidates are present in small amounts as intermediates in the biosynthesis or breakdown of more complex glycerophospholipids. In most glycerophospholipids, the phosphate group is esterified to both glycerol and another compound bearing an —OH group. Table 9.2 lists some common types of glycerophospholipids. Note that glycerophospholipids are amphipathic molecules with a polar head and long, nonpolar tails. The structures of three types of glycerophospholipids—phosphatidylethanolamine, phosphatidylserine, and phosphatidylcholine—are shown in Figure 9.7.

Each type of glycerophospholipid consists of a family of molecules with the same polar head group and different fatty acyl chains. For example, human red blood cell membranes contain at least 21 different species of phosphatidylcholine that differ from one another in the fatty acyl chains esterified at C-1 and C-2 of the glycerol backbone. In general, glycerophospholipids have saturated fatty acids esterified to C-1 and unsaturated fatty acids esterified to C-2. The major membrane glycerophospholipids in *Escherichia coli* are phosphatidylethanolamine and phosphatidylglycerol.

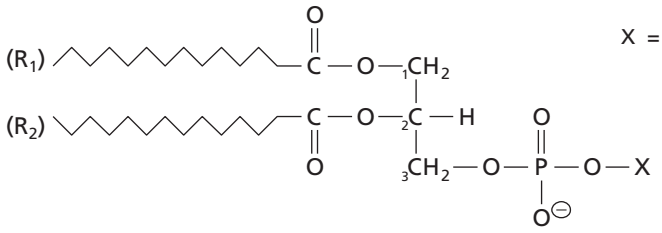
KEY CONCEPT

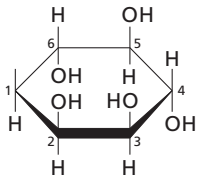
Many important lipids are derivatives of glycerol (see Box 16.1).

▲ **Yellow jacket wasp.** The venom of wasps, bees, and snakes contains phospholipases.

A variety of phospholipases can be used to dissect glycerophospholipid structures and determine the identities of their individual fatty acids. The specific positions of fatty acids in glycerophospholipids can be determined by using phospholipase A_1 and phospholipase A_2 that specifically catalyze the hydrolysis of the ester bonds at C-1 and C-2, respectively (Figure 9.8). Phospholipase A_2 is the major phospholipase in pancreatic juice and it is responsible for the digestion of membrane phospholipids in the diet. It is also present in snake, bee, and wasp venom. High concentrations of the products of the action of phospholipase A_2 can disrupt cell membranes. Thus, injection of snake venom into the blood can result in life-threatening lysis of the membranes of red blood cells. Phospholipase C catalyzes hydrolysis of the P—O bond between glycerol and

Table 9.2 Some common types of glycerophospholipids



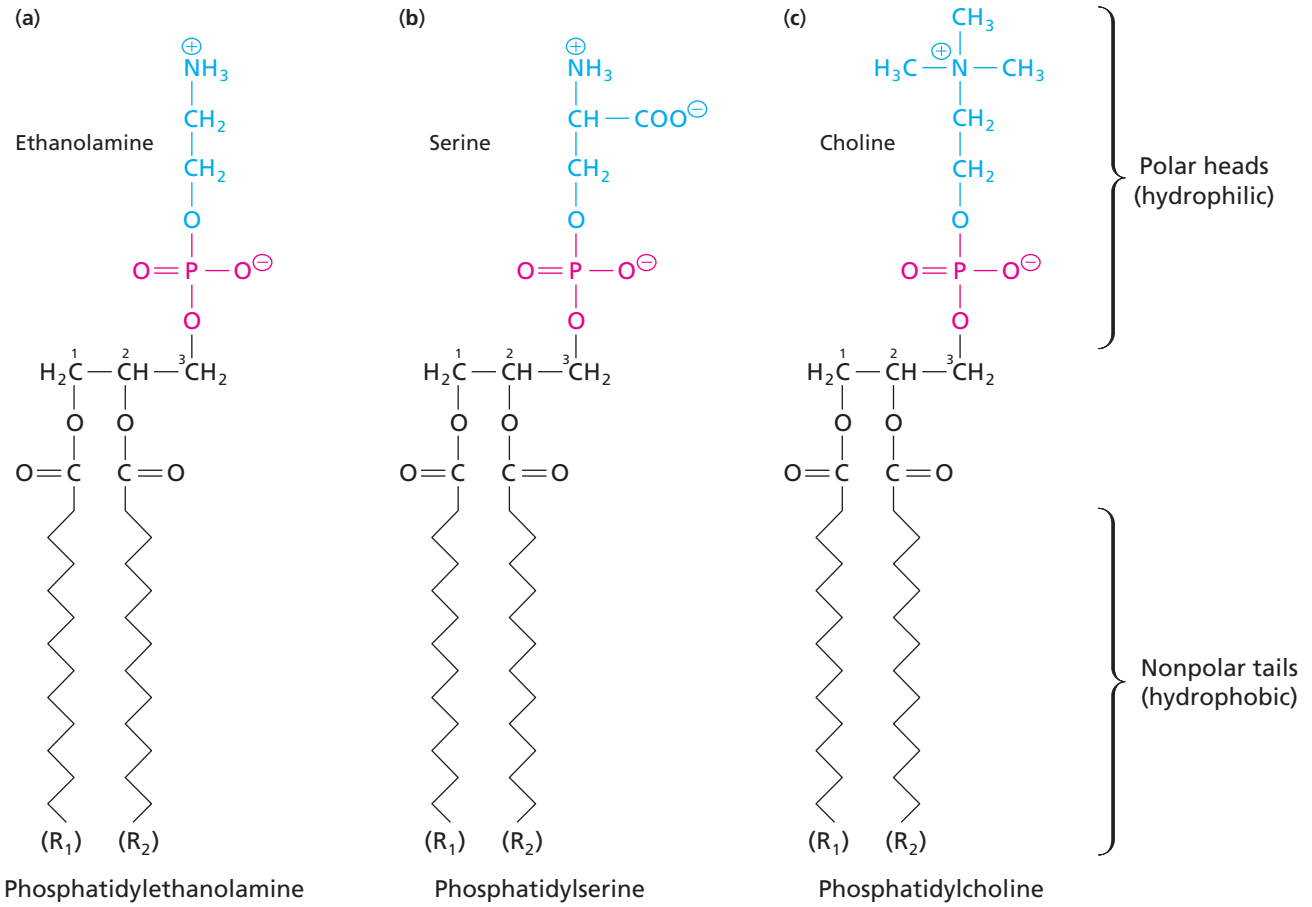
Precursor of X (HO—X)	Formulas of —O—X	Name of resulting glycerophospholipid
Water	—H	Phosphatidate
Choline	$-CH_2CH_2\overset{\oplus}{N}(CH_3)_3$	Phosphatidylcholine
Ethanolamine	$-CH_2CH_2\overset{\oplus}{N}H_3$	Phosphatidylethanolamine
Serine	$-CH_2-CH(\overset{\oplus}{N}H_3)-COO^{\ominus}$	Phosphatidylserine
Glycerol	$-CH_2CH(OH)-CH_2OH$	Phosphatidylglycerol
Phosphatidyl-glycerol	$-CH_2CH(OH)-CH_2-O-P(O^{\ominus})(=O)-O-CH_2-CH(OCOCH_2R_4)-CH_2OCOR_3$	Diphosphatidylglycerol (Cardiolipin)
myo-Inositol		Phosphatidylinositol

phosphate to liberate diacylglycerol. Phospholipase D converts glycerophospholipids to phosphatidates.

Plasmalogens are the other major type of glycerophospholipids. They differ from phosphatidates because the hydrocarbon substituent on the C-1 hydroxyl group of glycerol is attached by a vinyl ether linkage rather than an ester linkage (Figure 9.9). Ethanolamine or choline is commonly esterified to the phosphate group of plasmalogens. Plasmalogens account for about 23% of the glycerophospholipids in the human central nervous system and are also found in the membranes of peripheral nerve and muscle tissue.

9.5 Sphingolipids

Sphingolipids are the second most abundant lipids in plant and animal membranes. In mammals, sphingolipids are particularly abundant in tissues of the central nervous system. Most bacteria do not have sphingolipids. The structural backbone of sphingolipids is sphingosine (*trans*-4-sphingenine), an unbranched C_{18} alcohol with a *trans* double

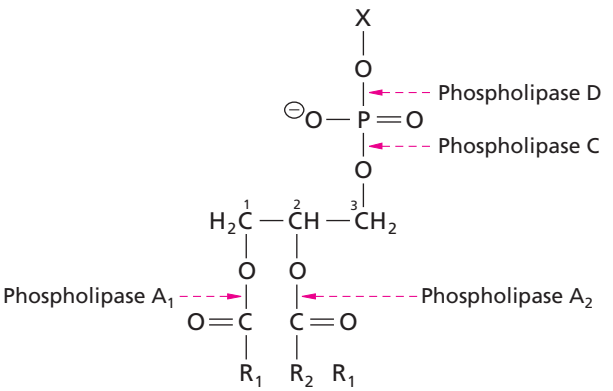


▲ **Figure 9.7**
Structures of (a) phosphatidylethanolamine, (b) phosphatidylserine, and (c) phosphatidylcholine. Functional groups derived from esterified alcohols are shown in blue. Since each of these lipids can contain many combinations of fatty acyl groups, the general name refers to a family of compounds, not to a single molecule.

bond between C-4 and C-5, an amino group at C-2, and hydroxyl groups at C-1 and C-3 (Figure 9.10a). **Ceramide** consists of a fatty acyl group linked to the C-2 amino group of sphingosine by an amide bond (Figure 9.10b). Ceramides are the metabolic precursors of all sphingolipids. The three major families of sphingolipids are the sphingomyelins, the cerebrosides, and the gangliosides. Of these, only sphingomyelins contain phosphate and are classified as phospholipids; cerebrosides and gangliosides contain carbohydrate residues and are classified as glycosphingolipids (Figure 9.1).

In **sphingomyelins**, phosphocholine is attached to the C-1 hydroxyl group of a ceramide (Figure 9.10c). Note the resemblance of sphingomyelin to phosphatidylcholine (Figure 9.7c)—both molecules are zwitterions containing choline, phosphate, and two long hydrophobic tails. Sphingomyelins are present in the plasma membranes of most mammalian cells and are a major component of the myelin sheaths that surround certain nerve cells.

Figure 9.8 ►
Action of four phospholipases. Phospholipases A_1, A_2, C, and D can be used to dissect glycerophospholipid structure. Phospholipases catalyze the selective removal of fatty acids from C-1 or C-2 or convert glycerophospholipids to diacylglycerols or phosphatidates.

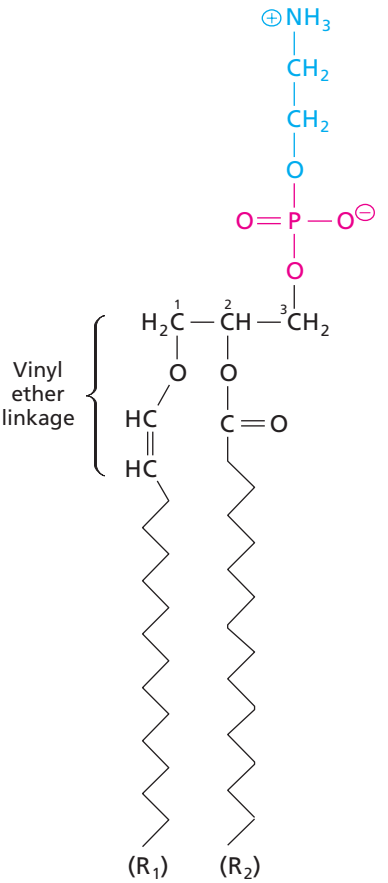


Cerebrosides are glycosphingolipids that contain one monosaccharide residue attached by a β-glycosidic linkage to C-1 of a ceramide. Galactocerebrosides, also known as galactosylceramides, have a single β-D-galactosyl residue as a polar head group (Figure 9.11). Galactocerebrosides are abundant in nerve tissue and account for about 15% of the lipids of myelin sheaths. Many other mammalian tissues contain glucocerebrosides, ceramides with a β-D-glucosyl head group. In some glycosphingolipids, a linear chain of up to three more monosaccharide residues is attached to the galactosyl or glucosyl moiety of a cerebroside.

Gangliosides are more complex glycosphingolipids in which oligosaccharide chains containing *N*-acetylneuraminic acid (NeuNAc) are attached to a ceramide. NeuNAc (Figure 8.15), an acetylated derivative of neuraminic acid, makes the head groups of gangliosides anionic. The structure of a representative ganglioside, G_{M2}, is shown in Figure 9.12. The M in G_{M2} stands for monosialo (i.e., one NeuNAc residue); G_{M2} was the second monosialo ganglioside characterized, thus the subscript 2.

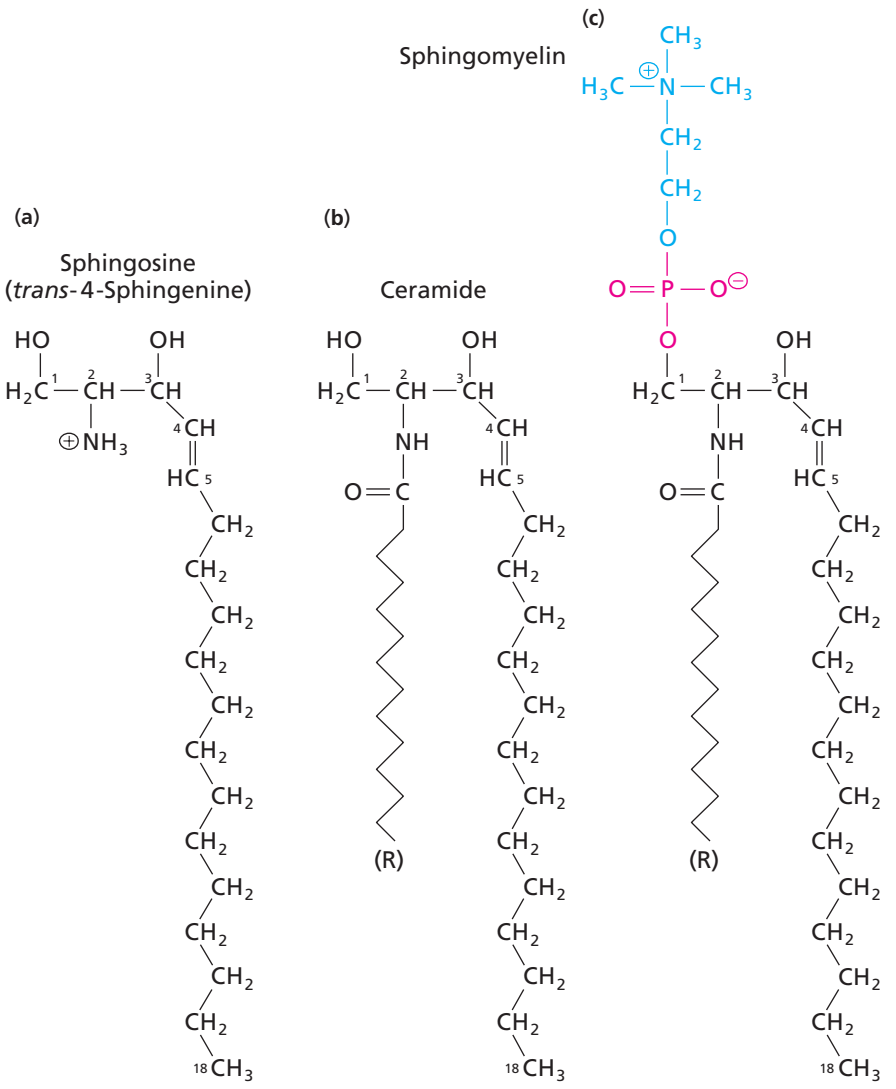
More than 60 varieties of gangliosides have been characterized. Their structural diversity results from variations in the composition and sequence of sugar residues. Ganglioside G_{M1}, for example, is similar to ganglioside G_{M2} shown in Figure 9.12 except that it has an additional β-D-galactose residue attached to the terminal *N*-acetyl-β-D-galactosamine residue via a β-(1 → 4) linkage. In all gangliosides, the ceramide is linked through its C-1 to a β-glucosyl residue, which in turn is bound to a β-galactosyl residue.

Gangliosides are present on cell surfaces with the two hydrocarbon chains of the ceramide moiety embedded in the plasma membrane and the oligosaccharides on the

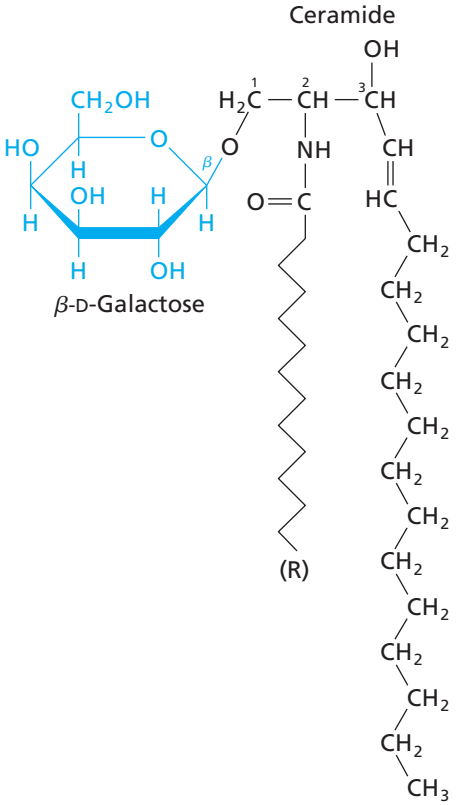


▲ **Figure 9.9**
Structure of an ethanolamine plasmalogen. A hydrocarbon is linked to the C-1 hydroxyl group of glycerol to form a vinyl ether.

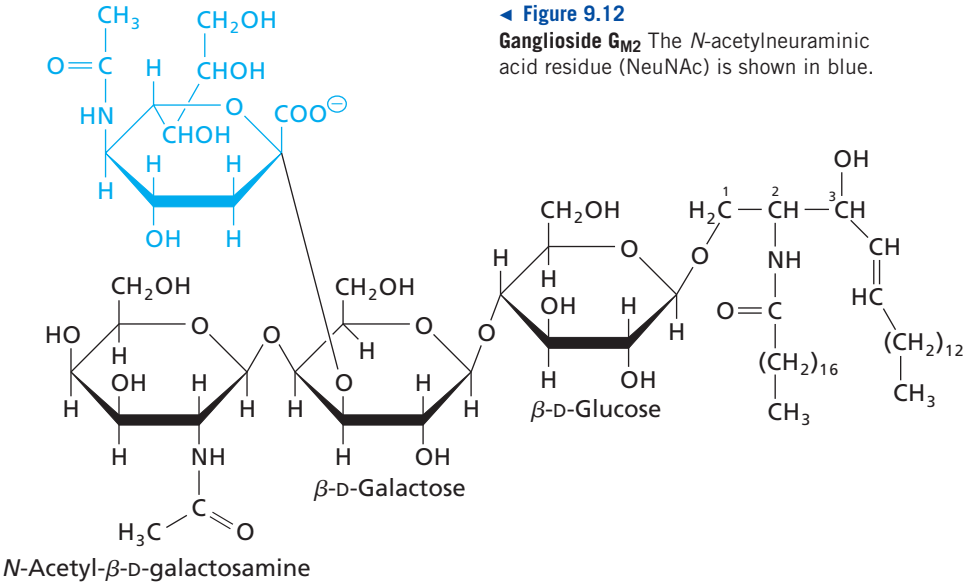
Genetic defects associated with lipid metabolism are described in Chapter 16.



◀ **Figure 9.10**
Structures of sphingosine, ceramide, and sphingomyelin. (a) Sphingosine, the backbone for sphingolipids, is a long-chain alcohol with an amino group at C-2. **(b)** Ceramides have a long-chain fatty acyl group attached to the amino group of sphingosine. **(c)** Sphingomyelins have a phosphate group (red) attached to the C-1 hydroxyl group of a ceramide and a choline group (blue) attached to the phosphate.



▲ **Figure 9.11**
Structure of a galactocerebroside. β-D-Galactose (blue) is attached to the C-1 hydroxyl group of a ceramide (black).

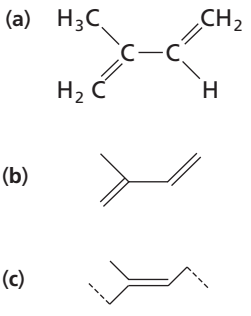


◄ **Figure 9.12**
Ganglioside G_{M2} The *N*-acetylneuraminic acid residue (NeuNAc) is shown in blue.

extracellular surface. Gangliosides and other glycosphingolipids are part of the cell surface repertoire of diverse oligosaccharide chains along with glycoproteins. Collectively, these markers provide cells with distinguishing surface markers that can serve in cellular recognition and cell-to-cell communication. Structures similar to the ABO blood group antigens on the surface of human cells (Box. 8.3) can be oligosaccharide components of glycosphingolipids in addition to being linked to proteins to form glycoproteins.

Genetically inherited defects in ganglioside metabolism are responsible for a number of debilitating and often lethal diseases, such as Tay-Sachs disease and generalized gangliosidosis. Certain rare genetic defects lead to deficiencies of enzymes responsible for the degradation of sphingolipids in the lysosomes of cells. In Tay-Sachs disease, there is a deficiency of a hydrolase that catalyzes removal of *N*-acetylgalactosamine from G_{M2}. Accumulation of G_{M2} causes lysosomes to swell leading to tissue enlargement. In the central nervous tissue, where there is little room for expansion, nerve cells die causing blindness, mental retardation, and death.

The exposed carbohydrates on the cell surface also provide convenient receptors for bacteria, viruses, and toxins. For example, cholera toxin, produced by the bacterium *Vibrio cholerae*, binds to the ganglioside G_{M1} of intestinal epithelial cells. Binding stimulates entry of the toxin into the cells where it interferes with normal signaling pathways leading to massive efflux of fluid into the intestine. This often produces death by dehydration.

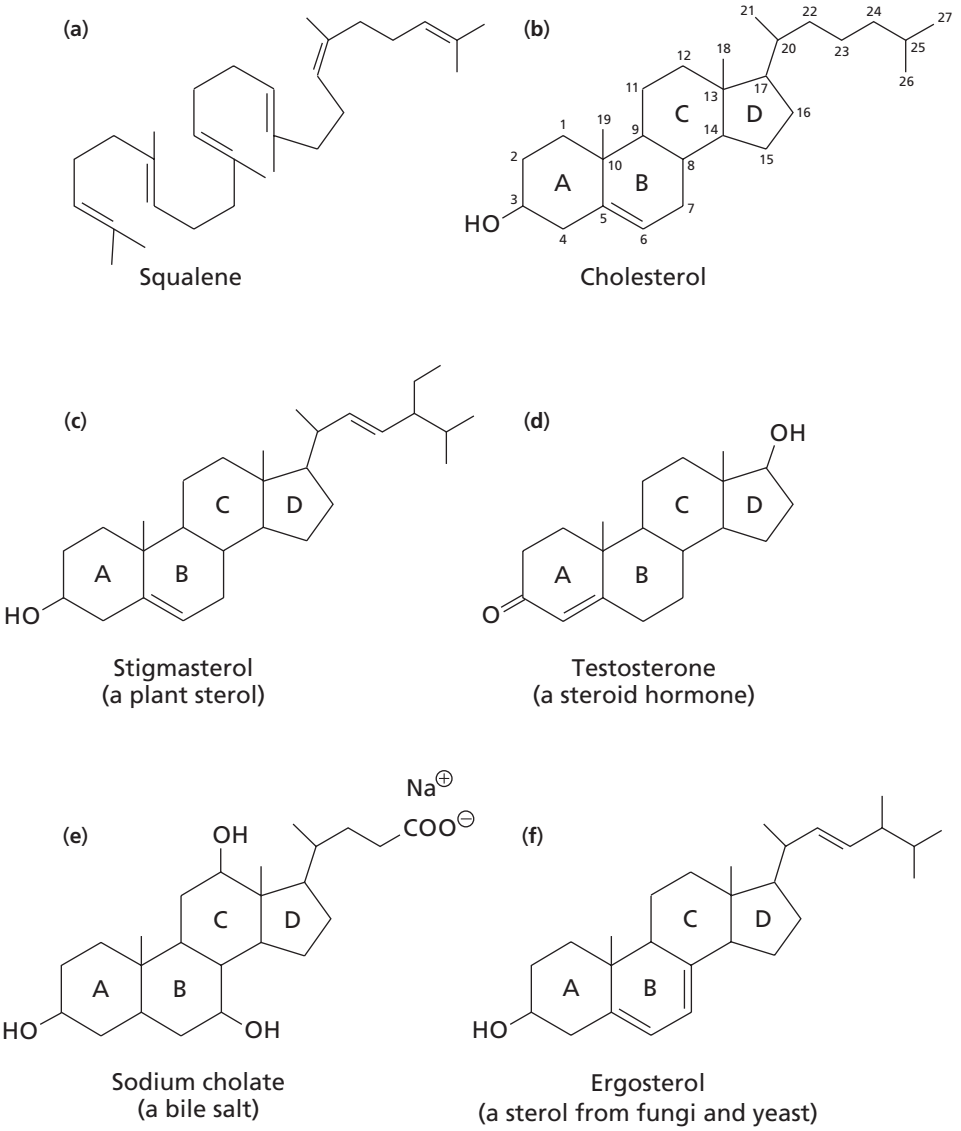


▲ **Figure 9.13**
Isoprene (2-methyl-1,3-butadiene), the basic structural unit of isoprenoids. (a) Chemical structure. **(b)** Carbon backbone. **(c)** Isoprene unit where dashed lines represent covalent bonds to a adjacent units.

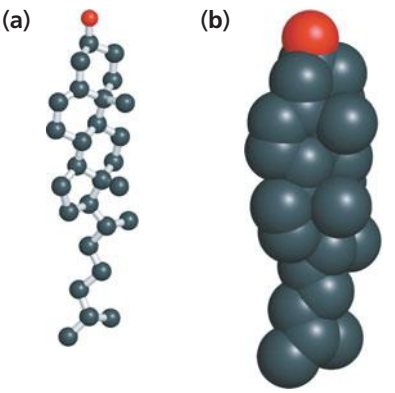
9.6 Steroids

Steroids are a third class of lipids found in the membranes of eukaryotes and, very rarely, in bacteria. Steroids, along with lipid vitamins and terpenes, are classified as isoprenoids because their structures are related to the five-carbon molecule isoprene (Figure 9.13). Steroids contain four fused rings: three six-carbon rings designated A, B, and C and a five-carbon D ring. The characteristic ring structure is derived from squalene (Figure 9.14a). Substituents of the nearly planar ring system can point either down (the α configuration) or up (the β configuration). The structures of several steroids are shown in Figure 9.14.

The steroid cholesterol is an important component of animal plasma membranes but is less common in plants and absent from prokaryotes, protists, and fungi. These species have other steroids (e.g., stigmasterol, ergosterol) that are very similar to cholesterol. Cholesterol is actually a **sterol** because it has a hydroxyl group at C-3. Other steroids include the sterols of plants, fungi, and yeast (which also have a hydroxyl group at C-3); mammalian steroid hormones (such as estrogens, androgens, progestins, and



◀ **Figure 9.14**
Structures of several steroids. Squalene **(a)** is the precursor of most steroids. Steroids contain four fused rings (lettered A, B, C, and D). **(b)** Cholesterol. **(c)** Stigmasterol, a common component of plant membranes. **(d)** Testosterone, a steroid hormone involved in male development in animals. **(e)** Sodium cholate, a bile salt, which aids in the digestion of lipids. **(f)** Ergosterol, a compound from fungi and yeast.



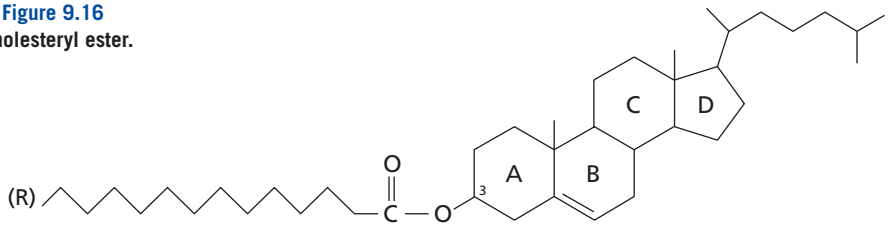
▲ **Figure 9.15**
Cholesterol. (a) Ball-and-stick model with the oxygen atom (red) at the top. Hydrogen atoms are not shown. The fused ring system of cholestrol is almost planar. **(b)** Space-filling model.

adrenal corticosteroids); and bile salts. These steroids differ in the length of the side chain attached to C-17 and in the number and placement of methyl groups, double bonds, hydroxyl groups, and in some cases, keto groups. Prokaryotes use squalene and some related nonsteroid lipids that do not have the complete ring structure of the steroids.

Cholesterol plays an essential role in mammalian biochemistry. It is not only a component of certain membranes but is also a precursor of steroid hormones and bile salts. The fused ring system of cholesterol, shown from the side in Figure 9.15, makes it less flexible than most other lipids. As a result, cholesterol modulates the fluidity of mammalian cell membranes, as we will see later in this chapter.

Steroids are far more hydrophobic than glycerophospholipids and sphingolipids. For example, free cholesterol's maximal concentration in water is only 10^{-8} M. Esterification of a fatty acid to the C-3 hydroxyl group forms a cholesteryl ester (Figure 9.16).

▼ **Figure 9.16**
Cholesteryl ester.



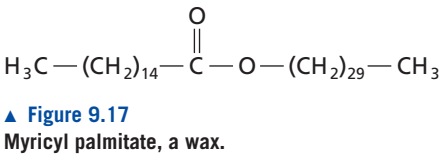
Because the 3-acyl group of the ester is nonpolar, a cholesteryl ester is even more hydrophobic than cholesterol itself. Cholesterol is converted to cholesteryl esters for storage in cells or for transport through the bloodstream. Because they are essentially insoluble in water, cholesterol and its esters must be complexed with phospholipids and amphipathic proteins in lipoproteins for transport (Section 16.1B).

9.7 Other Biologically Important Lipids

There are many kinds of lipids not found in membranes. These include diverse compounds such as waxes, eicosanoids, and some isoprenoids. Non-membrane lipids have a variety of specialized functions—some of which we have already encountered (e.g., lipid vitamins).

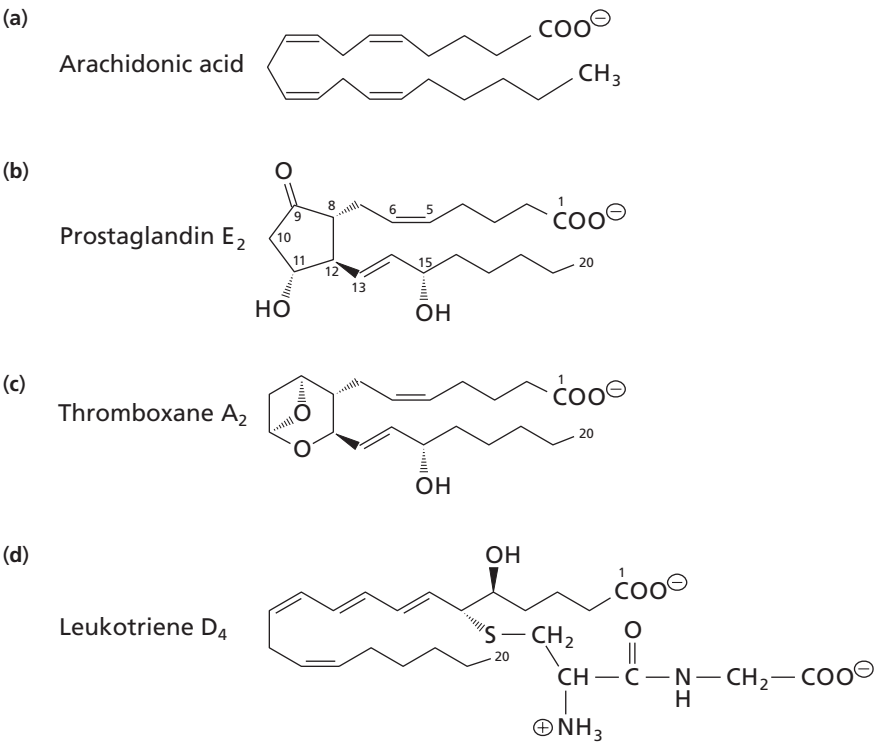
Waxes are nonpolar esters of long-chain fatty acids and long chain monohydroxylic alcohols. For example, myricyl palmitate, a major component of beeswax, is the ester of palmitate (16:0) and the 30-carbon myricyl alcohol (Figure 9.17). The hydrophobicity of myricyl palmitate makes beeswax very insoluble and its high melting point (due to the long, saturated hydrocarbon chains) makes beeswax hard and solid at typical outdoor temperatures. Waxes are widely distributed in nature. They provide protective waterproof coatings on the leaves and fruits of certain plants and on animal skin, fur, feathers, and exoskeletons. Ear wax, also known as cerumen (from the Latin word *cera*, "wax"), is secreted by cells lining the auditory canal. It serves to lubricate the canal and trap particles that could damage the eardrum. Ear wax is a complex mixture made up mostly of long chain fatty acids, cholesterol, and ceramides. It also contains squalene, triacylglycerols, and true waxes (about 10% of the weight).

Eicosanoids are oxygenated derivatives of C_{20} polyunsaturated fatty acids such as arachidonic acid. Some examples of eicosanoids are shown in Figure 9.18. Eicosanoids participate in a variety of physiological processes and can also mediate many potentially pathological responses. **Prostaglandins** are eicosanoids that have a cyclopentane ring.



▲ **Figure 9.17**
Myricyl palmitate, a wax.

▲ **Earwax and beeswax are two examples of naturally occurring waxes.**



▲ **Figure 9.18**
Structures of arachidonic acid (a) and three eicosanoids derived from it. Arachidonate is a C_{20} polyunsaturated fatty acid with four *cis* double bonds.

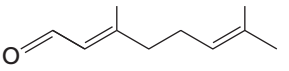
Prostaglandin E_2 can cause constriction of blood vessels, and thromboxane A_2 is involved in the formation of blood clots that in some cases can block the flow of blood to the heart or brain. Leukotriene D_4, a mediator of smooth-muscle contraction, also provokes the bronchial constriction seen in asthmatics. Aspirin (acetylsalicylic acid) alleviates pain, fever, swelling, and inflammation by inhibiting the synthesis of prostaglandins (Box. 16.1).

Some nonmembrane lipids are related to isoprene (Figure 9.13) but they are not steroids. We encountered several of these lipids in Chapter 7. The lipid vitamins A, E, and K are isoprenoids that contain long hydrocarbon chains or fused rings (Section 7.14). Vitamin D is an isoprenoid derivative of cholesterol. There are several carotenes related to retinol (vitamin A). The hydrophobic chain of ubiquinone contains 6–10 isoprenoid units (Section 7.15).

Simple isoprenoids are often called terpenes. They have structures that reveal their formation from isoprene units. Citral is a good example: it is present in many plants and imparts a strong lemon odor (Figure 19.19a). Other isoprenoids are bactoprenol (undecaprenyl alcohol) (Figure 9.19b) and juvenile hormone I (Figure 9.19c) that regulates the expression of genes required for development in insects. Isoprenoids similar to bactoprenol are important lipids in archaebacteria, where they replace fatty acids in most membrane phospholipids (see Box 9.5).

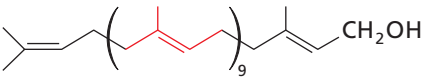
Terpenes can be extensively modified to form a more complex class of lipid called terpenoids. Many of these are cyclic compounds like limonene, which is responsible for the smell of oranges (Figure 19.19d). Gibberellins are multi-ring terpenoids that function as growth hormones in plants (Figure 19.19e).

(a)



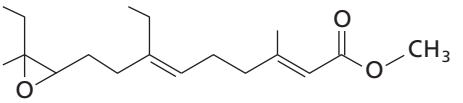
Citral

(b)



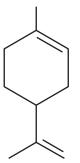
Bactoprenol
(Undecaprenyl alcohol)

(c)



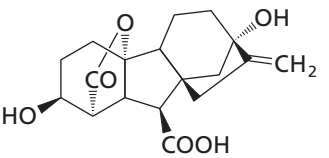
Juvenile hormone I

(d)



Limonene

(e)



Gibberellin GA1

▲ **Figure 9.19**
Some isoprenoids. Note the isoprene unit (red) in bactoprenol.

9.8 Biological Membranes

Biological membranes define the external boundaries of cells and separate compartments within cells. They are essential components of all living cells. A typical membrane consists of two layers of lipid molecules and many embedded proteins.

Biological membranes are not merely passive barriers to diffusion. They have a wide variety of complex functions. Some membrane proteins serve as selective pumps controlling the transport of ions and small molecules into and out of the cell. Membranes are also responsible for generating and maintaining the proton concentration gradients essential for the production of ATP. Receptors in membranes recognize extracellular signals and communicate them to the cell interior.

Many cells have membranes with specialized structures. For example, many bacteria have double membranes: an outer membrane and an inner plasma membrane. The liquid in the periplasmic space between these two membranes contains proteins that carry specific solutes to transport proteins in the inner membrane. The solutes then pass through the inner membrane by an ATP-dependent process. A mitochondrion's smooth outer membrane has proteins that form aqueous channels while its convoluted inner membrane is selectively permeable and has many membrane-bound enzymes. The nucleus also has a double membrane—nuclear contents interact with the cytosol through nuclear pores. The single membrane of the endoplasmic reticulum is highly convoluted. Its extensive network in eukaryotic cells is involved in the synthesis of transmembrane and secreted proteins and of lipids for many membranes.

In this section, we explore the structure of biological membranes. In the remaining sections of this chapter, we discuss the properties and functions of biological membranes.

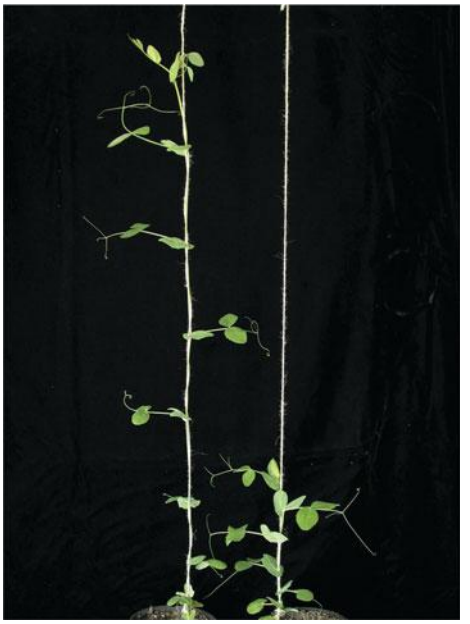
A. Lipid Bilayers

We saw earlier that detergents in aqueous solutions can spontaneously form monolayers or micelles (Section 2.4). Like detergents, amphipathic glycerophospholipids and glycosphingolipids can form monolayers under some conditions. In cells, these lipids do not pack well into micelles but rather tend to form **lipid bilayers** (Figure 9.20). Lipid bilayers are the main structural component of all biological membranes, including plasma membranes and the internal membranes of eukaryotic cells. The noncovalent

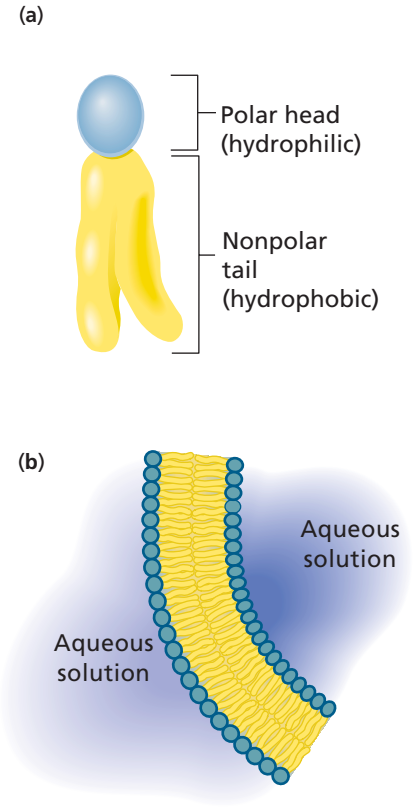
BOX 9.3 GREGOR MENDEL AND GIBBERELLINS

Gregor Mendel studied seven traits in order to come up with the basic laws of heredity. One of the traits was stem length (*Le/le*). The *Le* gene has been cloned and sequenced (Lester et al., 1997). It encodes the enzyme gibberellin 3β-hydroxylase, an enzyme required for the synthesis of the terpenoid gibberellin GA1. The production of gibbberellin GA1 by the normal gene stimulates growth producing a tall pea plant. The mutant gene produces a less active enzyme that synthesizes less hormone and plants homozygous for the mutant allele (*le*) are short.

The mutation is a single nucleotide substitution that converts an alanine codon into a threonine codon (A229T). Another one of Mendel's seven traits is described in Box. 15.3.



► **The stem length mutation.** Tall plants (left) are normal. Mutations in the stem length gene (*Le*) produce short plants (right).



▲ **Figure 9.20**
Membrane lipid and bilayer. (a) An amphipathic membrane lipid. **(b)** Cross-section of a lipid bilayer. The hydrophilic head groups (blue) of each leaflet face the aqueous medium, and the hydrophobic tails (yellow) pack together in the interior of the bilayer.

interactions among lipid molecules in bilayers make membranes flexible and allow them to self-seal. Triacylglycerols, which are very hydrophobic rather than amphipathic, cannot form bilayers and cholesterol, although slightly amphipathic, does not form bilayers by itself.

A lipid bilayer is typically about 5 to 6 nm thick and consists of two sheets, or monolayers (also called leaflets). In each sheet, the polar head groups of amphipathic lipids are in contact with the aqueous medium and the nonpolar hydrocarbon tails point toward the interior of the bilayer (Figure 9.20).

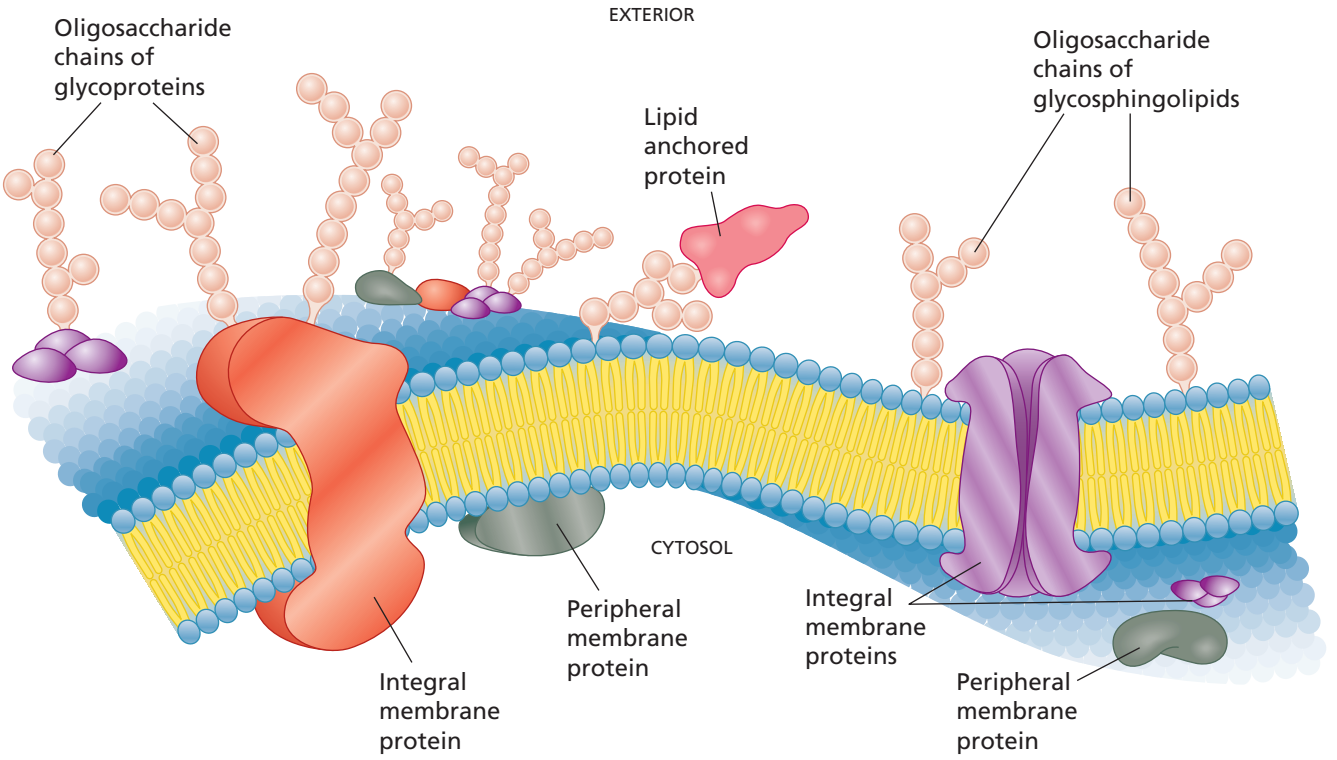
The spontaneous formation of lipid bilayers is driven by the hydrophobic interactions (Section 2.5D). When lipid molecules associate, the entropy of the solvent molecules increases and this favors formation of the lipid bilayer.

B. Three Classes of Membrane Proteins

Cellular and intracellular membranes contain specialized membrane-bound proteins. These proteins are divided into three classes based on their mode of association with the lipid bilayer: integral membrane proteins, peripheral membrane proteins, and lipid anchored membrane proteins (Figure 9.21).

Integral membrane proteins, also referred to as transmembrane proteins, contain hydrophobic regions embedded in the hydrophobic core of the lipid bilayer. Integral membrane proteins usually span the bilayer completely, with one part of the protein exposed on the outer surface and one part exposed on the inner surface. Some integral membrane proteins are anchored by only a single membrane-spanning portion of the polypeptide chain, whereas other membrane proteins have several transmembrane segments connected by loops at the membrane surface. The membrane-spanning segment is often an α helix containing approximately 20 amino acid residues.

One of the best characterized integral membrane proteins is bacteriorhodopsin (Figure 9.22a). This protein is found in the cytoplasmic membrane of the halophilic (salt-loving) bacterium *Halobacterium halobium*, where it helps harness light energy used in the synthesis of ATP. Bacteriorhodopsin consists of a bundle of seven α helices. The exterior surface of the helical bundle is hydrophobic and interacts directly with lipid molecules in the membrane. The interior surface contains charged amino acid side chains that bind the pigment molecule. Bacteriorhodopsin is one of several α-helical membrane proteins whose structures are known in detail. These α-helix bundle

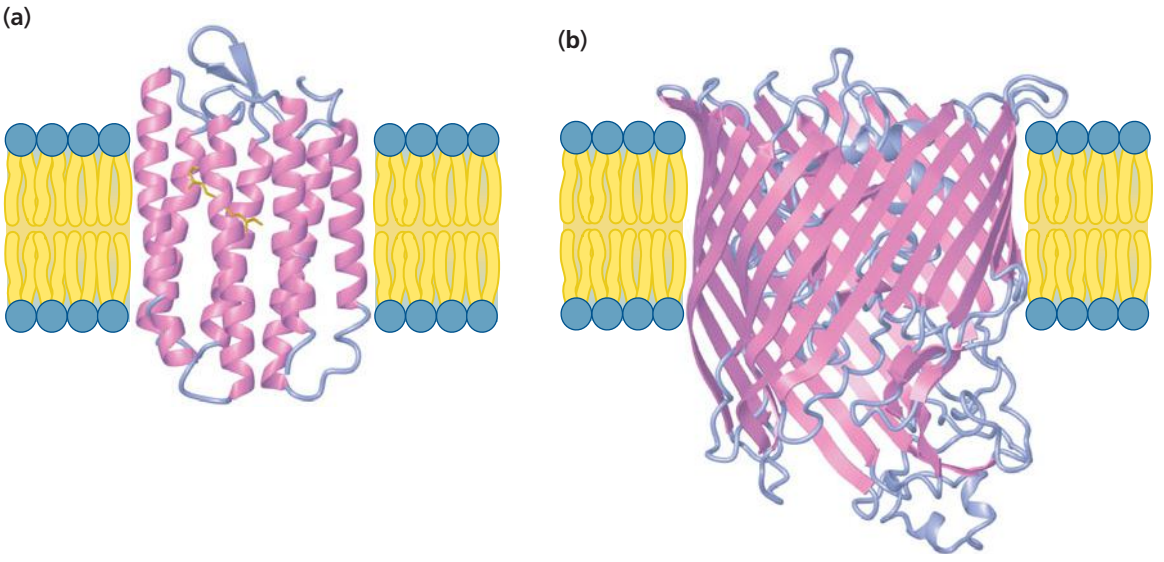


▲ **Figure 9.21**
Structure of a typical eukaryotic plasma membrane. A lipid bilayer forms the basic matrix of biological membranes, and proteins (some of which are glycoproteins) are associated with it in various ways. The oligosaccharides of glycoproteins and glycolipids face the extracellular space.

proteins make up one of the two major classes of integral membrane proteins. The other class is the β-barrel proteins (see below).

In the absence of data on three-dimensional structure, the presence of transmembrane α-helical regions in membrane proteins can often be predicted by searching amino acid sequences for regions that are hydrophobic (i.e., that have high hydropathy values) (Section 3.2G) and a tendency to be present in α-helices (Section 4.4). Various prediction algorithms have been developed over the years and they are currently able to detect 70% of known transmembrane α-helices. These predictions are important because it is still very difficult to crystallize membrane proteins in order to determine their true structure.

▼ **Figure 9.22**
Integral membrane proteins. (a) Bacteriorhodopsin: seven membrane-spanning α helices, connected by loops, form a bundle that spans the bilayer. The light-harvesting prosthetic group is shown in yellow. [PDB 1FBB]. **(b)** Porin FhuA from *Escherichia coli*: this porin forms a channel for the passage of protein-bound iron into the bacterium. The channel is formed from 22 antiparallel β strands that form a β-barrel. [PDB 1BY3].



Protein folding is another example of an entropically driven assembly reaction (Section 4.11A).

We consider the functions of some of these membrane proteins later in this chapter. We will also encounter membrane proteins in other chapters, including those on membrane-associated electron transport (Chapter 14), photosynthesis (Chapter 15), and protein synthesis (Chapter 22).

The function of bacteriorhodopsin is described in Section 15.2.

Some prenyl-decorated proteins will be encountered in the discussion of signal transduction (Section 9.12).

Many integral membrane proteins have a β barrel fold (Figure 4.23b). The exterior surface of the β strands contacts the membrane lipids and the center of the barrel often serves as a pore or channel for passing molecules from one side of the membrane to the other. The *E. coli* porin, FhuA, is a typical example of this type of integral membrane protein (Figure 9.22b).

Peripheral membrane proteins are associated with one face of the membrane through charge–charge interactions and hydrogen bonding with integral membrane proteins or with the polar head groups of membrane lipids. Peripheral membrane proteins are more readily dissociated from membranes by changes in pH or ionic strength.

Lipid anchored membrane proteins are tethered to a membrane through a covalent bond to a lipid anchor. In the simplest lipid anchored membrane proteins, an amino acid side chain is linked by an amide or ester bond to a fatty acyl group, often from myristate or palmitate. The fatty acid is inserted into the cytoplasmic leaflet of the bilayer, anchoring the protein to the membrane (Figure 9.23a). Proteins of this type are found in viruses and eukaryotic cells.

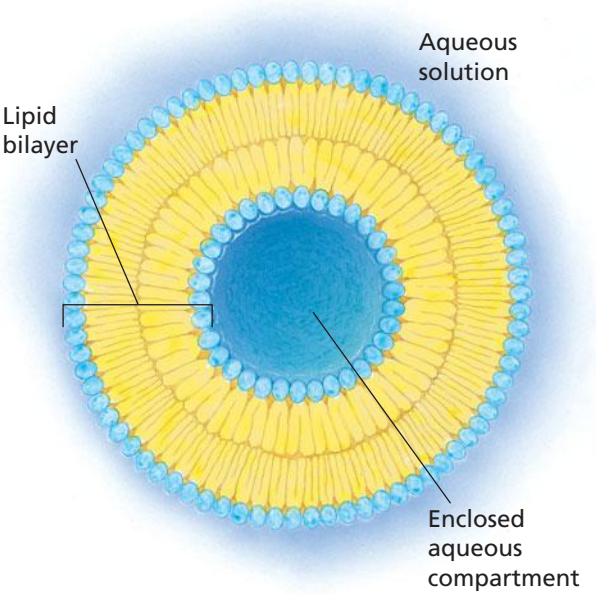
Other lipid anchored membrane proteins are covalently linked to an isoprenoid chain (either 15- or 20-carbon) through the sulfur atom of a cysteine residue at or near the C-terminus of the protein (Figure 9.23b). These *prenylated proteins* are found on the cytoplasmic face of both plasma membranes and intracellular membranes.

Many eukaryotic lipid anchored proteins are linked to a molecule of glycosylphosphatidylinositol (Figure 9.23c). The membrane anchor is the 1,2-diacylglycerol portion of the glycosylphosphatidylinositol. A glycan of varied composition is attached to the inositol by a glucosamine residue, a mannose residue links the glycan to a phosphoethanolamine residue, and the C-terminal α-carboxyl group of the protein is linked to the ethanolamine by an amide bond. Over 100 different proteins are known to be associated with membranes by a glycosylphosphatidylinositol anchor. These proteins have a variety of functions and they are present only in the outer monolayer of the plasma membrane. They are found in the cholesterol-sphingolipid rafts described in Section 9.9.

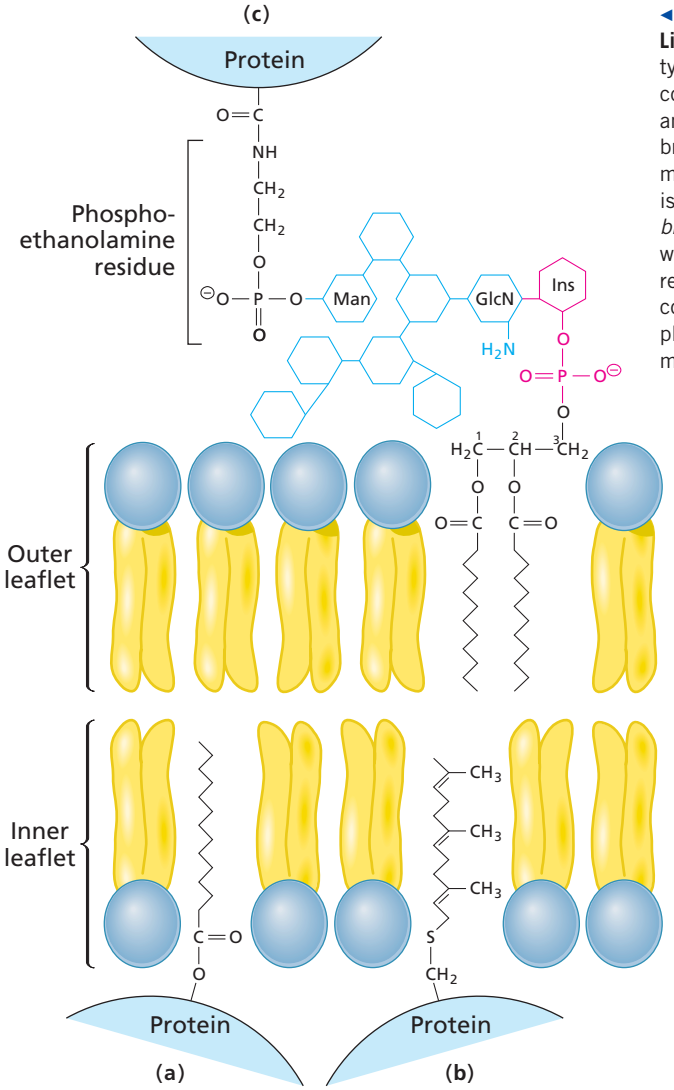
All three types of lipid anchors are covalently linked to amino acid residues post-translationally, that is, after the protein has been synthesized. Like integral membrane proteins, most lipid anchored proteins are permanently associated with the membrane, although the proteins themselves do not interact with the membrane. Once released by treatment with phospholipases, the proteins behave like soluble proteins.

BOX 9.4 NEW LIPID VESICLES, OR LIPOSOMES

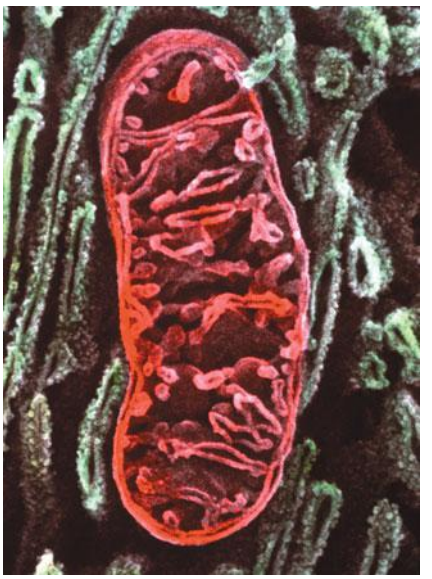
Synthetic vesicles (often called liposomes) consisting of phospholipid bilayers that enclose an aqueous compartment can be formed in the laboratory. In order to minimize unfavorable contact between the hydrophobic edge of the bilayer and the aqueous solution, lipid bilayers tend to close up to form these spherical structures. The vesicles are generally quite stable and impermeable to many substances. Liposomes whose aqueous inner compartment contains drug molecules can be used to deliver drugs to particular tissues in the body, provided that specific targeting proteins are present in the liposome membrane. Synthetic bilayers are an important experimental tool in the investigation of cellular membranes. An example of such an experiment is described in Box. 15.3.

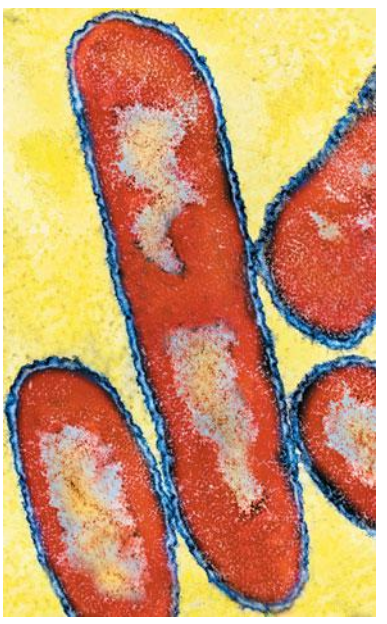


► **Schematic cross-section of a lipid vesicle, or liposome.** The bilayer is made up of two leaflets. In each leaflet, the polar head groups of the amphipathic lipids extend into the aqueous medium and the nonpolar hydrocarbon tails point inward and are in van der Waals contact with each other.



◀ **Figure 9.23**
Lipid anchored membrane proteins attached to the plasma membrane. The three types of anchors can be found in the same membrane, but they do not form a complex as shown here. **(a)** A fatty acyl anchored protein. **(b)** A prenyl anchored membrane protein. Note that fatty acyl and prenyl anchored membrane proteins can also occur on the cytoplasmic (outer) leaflet of intracellular membranes. **(c)** Protein anchored by glycosylphosphatidylinositol. Shown here is the variant surface glycoprotein of the parasitic protozoan *Trypanosoma brucei*. The protein is covalently bound to a phosphoethanolamine residue, which in turn is bound to a glycan. The glycan (blue) includes a mannose residue to which the phosphoethanolamine residue is attached and a glucosamine residue that is attached to the phosphoinositol group (red) of phosphatidylinositol. Abbreviations: GlcN, glucosamine; Ins, inositol; Man, mannose.





The total number of membrane proteins in a typical cell isn't known for certain but they are likely to represent a significant fraction of the proteome. In *E. coli*, for example, there appear to be roughly 1000 membrane proteins of all types. Since the total number of proteins is about 4000 (Chapter 4), membrane proteins account for about 25% of the total. This fraction is probably higher in multicellular eukaryotes because there are many more membrane proteins involved in cell-cell interactions and intracellular signaling.

Different membranes have different proteins (and lipids). In some cases a cell or compartment is enclosed by a double membrane consisting of two separate lipid bilayers (Figure 9.24). In the case of mitochondria and *E. coli*, the inner membranes have many more membrane proteins than the outer membranes.

Figure 9.24 ▶
Double membrane of mitochondria and many bacteria. The plasma membrane of most eukaryotic cells is a single lipid bilayer. Within eukaryotic cells the nucleus and major organelles such as mitochondria (top right) are bounded by double membranes. In bacteria, the gram-negative bacteria have a double membrane consisting of an inner and outer lipid bilayer as shown for *E. coli* (bottom right). It's not surprising that mitochondria (and chloroplasts) have a double membrane since they are derived from gram-negative bacteria that use the double membrane as part of the energy-producing mechanism of electron transport and ATP synthesis (Chapter 14).

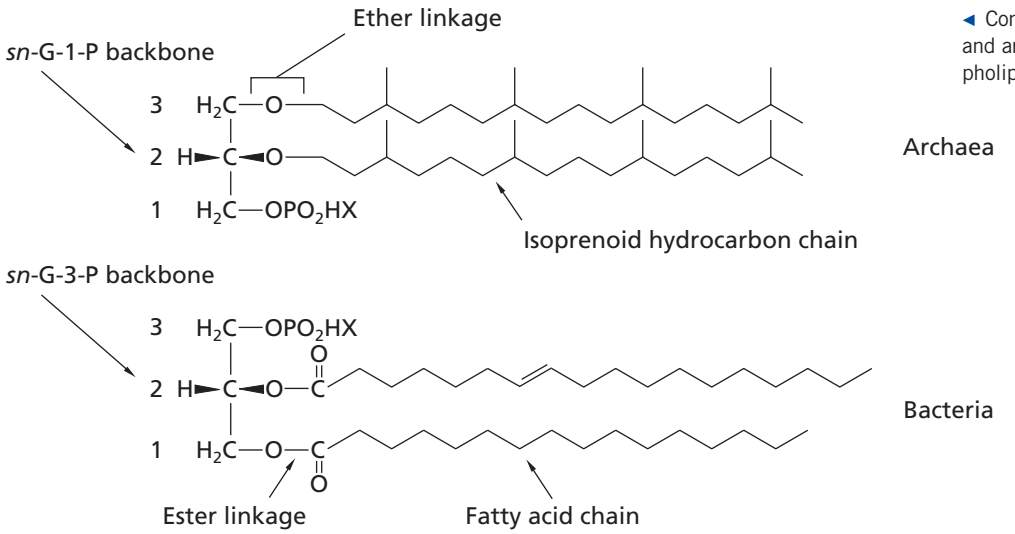
BOX 9.5 SOME SPECIES HAVE UNUSUAL LIPIDS IN THEIR MEMBRANES

Many species have unusual lipids in some of their membranes. The unusual lipids are sometimes confined to genera or families and sometimes entire orders share some distinctive lipid compositions. Within the eukaryotes, there are some lipids found only in some classes of animals and not others or in some classes of plants and not others. There are even distinctive lipid compositions in some entire kingdoms such as plants, animals, or fungi.

Prokaryotes are a very diverse group with many varieties of lipids. Major groups such as cyanobacteria, mycoplasma, and gram positive bacteria, can have quite characteristic lipid compositions in their membranes.

The archaebacteria (or Archaea) have glycerophospholipids that are quite unusual and distinctive. The glycerol phosphate backbone in archaebacterial glycerophospholipids is *sn*-glycerol-1-phosphate, a stereoisomer of the one found in other species (*sn*-glycerol-3-phosphate). (see Box 16.1) The hydrocarbon chains are attached to the glycerol backbone via ether linkages, not ester linkages, and the hydrocarbon chains in archaebacteria are often isoprenoid derivatives, not fatty acid derivatives.

There are a few species of gram-negative bacteria that have mixtures of ether and ester linkages in their lipids but unusual lipid composition of archaebacteria argues strongly in favor of classifying them as a distinctive monophyletic group. As mentioned earlier (Section 1.5), some scientists argue that the distinctiveness of archaebacteria justifies creating a third domain of life but the current view favors a more complex web of life perspective.



◄ Comparison of typical bacterial and archaebacterial glycero phospholipids.

C. The Fluid Mosaic Model of Biological Membranes

A typical biological membrane contains about 25% to 50% lipid and 50% to 75% protein by mass. Carbohydrates are present as components of glycolipids and glycoproteins. The lipids are a complex mixture of phospholipids, glycosphingolipids (in animals), and cholesterol (in some eukaryotes). Cholesterol and some other lipids that do not form bilayers by themselves (about 30% of the total) are stabilized in a bilayer arrangement by the other 70% of lipids in the membrane (see next section).

The compositions of biological membranes vary considerably among species and even among different cell types in multicellular organisms. For example, the myelin membrane that insulates nerve fibers contains relatively little protein. In contrast, the inner mitochondrial membrane is rich in proteins reflecting its high level of metabolic activity. The plasma membrane of red blood cells is also exceptionally rich in proteins.

Each biological membrane has a characteristic lipid composition, in addition to having a characteristic lipid to protein ratio. Membranes in brain tissue, for example, have a relatively high content of phosphatidylserines whereas membranes in heart and lung cells have high levels of phosphatidylglycerols and sphingomyelins, respectively. Phosphatidylethanolamines constitute nearly 70% of the inner membrane lipids of *E. coli* cells. The outer membranes of gram-negative bacteria contain lipopolysaccharides.

In addition to being distributed differentially among different tissues, phospholipids are also distributed asymmetrically between the inner and outer monolayers of a single biological membrane. In mammalian cells, for example, 90% of the sphingomyelin molecules are in the outer surface of the plasma membrane. Phosphatidylserines are also asymmetrically distributed in many cells, with 90% of the molecules in the cytoplasmic monolayer.

A biological membrane is thicker than a lipid bilayer—typically 6 to 10 nm thick. The **fluid mosaic model** proposed in 1972 by S. Jonathan Singer and Garth L. Nicolson is still generally valid for describing the arrangement of lipid and protein within a membrane. According to the fluid mosaic model, the membrane is a dynamic structure in which both proteins and lipids can rapidly and randomly diffuse laterally or rotate within the bilayer. Membrane proteins are visualized as icebergs floating in a highly fluid lipid bilayer sea (Figure 9.21). (Actually, some proteins are immobile and some lipids have restricted movement.)

KEY CONCEPT

Membranes consist of a lipid bilayer and embedded proteins. Lipids and proteins can diffuse rapidly within the membrane.

9.9 Membranes Are Dynamic Structures

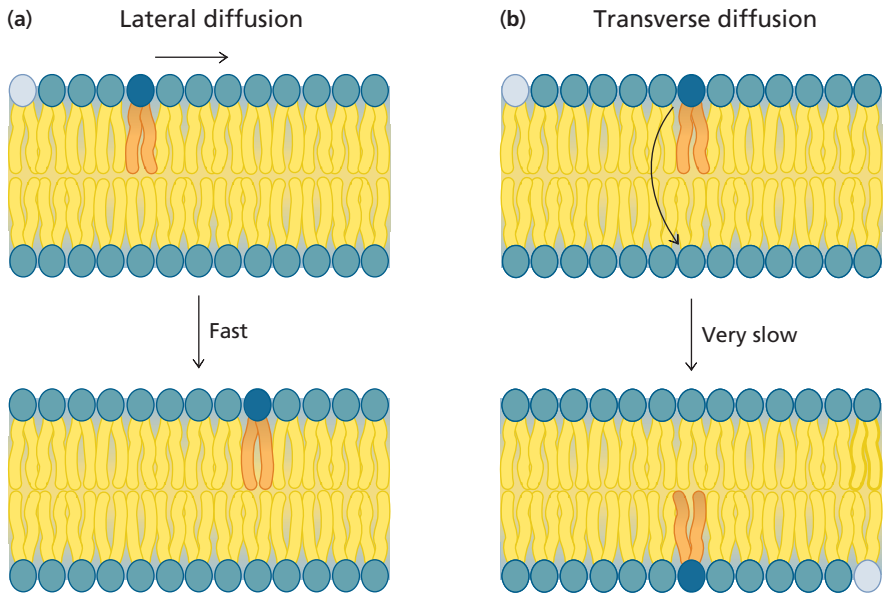
The lipids in a bilayer are in constant motion giving lipid bilayers many of the properties of fluids. A lipid bilayer can therefore be regarded as a two-dimensional solution. Lipids undergo several types of molecular motion within bilayers. The rapid movement of lipids within the plane of one monolayer is an example of two-dimensional lateral diffusion. A phospholipid molecule can diffuse from one end of a bacterial cell to the other (a distance of about 2 μm) in about 1 second at 37°C.

In contrast, transverse diffusion (or flip-flop) is the passage of lipids from one monolayer of the bilayer to the other. Transverse diffusion is much slower than lateral diffusion (Figure 9.25). The polar head of a phospholipid molecule is highly solvated and must shed its solvation sphere and penetrate the hydrocarbon interior of the bilayer in order to move from one leaflet to the other. The energy barrier associated with this movement is so high that transverse diffusion of phospholipids in a bilayer occurs at about one-billionth the rate of lateral diffusion. The very slow rate of transverse diffusion of membrane lipids is what allows the inner and outer layers of biological membranes to maintain different lipid compositions.

All cells synthesize new membrane by adding lipids and protein to preexisting membranes. As the plasma membrane is extended, the cell increases in size. Eventually the cell will divide and each daughter cell will inherit a portion (usually half) of the parental membranes. Internal membranes are extended and divide in the same manner.

You might have inherited lipid molecules from your grandmother! (see Problem 18).

In bacteria, lipid molecules are usually added to the cytoplasmic side of the lipid bilayer. Lipid asymmetry is generated by preferentially adding newly synthesized lipids to



◀ **Figure 9.25**
Diffusion of lipids within a bilayer. (a) Lateral diffusion of lipids is relatively rapid. **(b)** Transverse diffusion, or flip-flop, of lipids is very slow.

only one of the monolayers. Since transverse diffusion is so slow, these newly synthesized molecules will not spread to the outer layer of the plasma membrane. This accounts for the enrichment of some types of lipids in the inner layer. Lipid asymmetry can also be generated and maintained by the activity of membrane-bound flipases and flopases—enzymes that use the energy of ATP to move specific phospholipids from one monolayer to the other. The activity of these enzymes accounts for the enrichment of certain types of phospholipid in the outer layer. Eukaryotic cells make their membrane lipids in an asymmetric arrangement in the endoplasmic reticulum or the Golgi apparatus. The membrane fragments flow from these organelles—retaining the asymmetry—to other membranes.

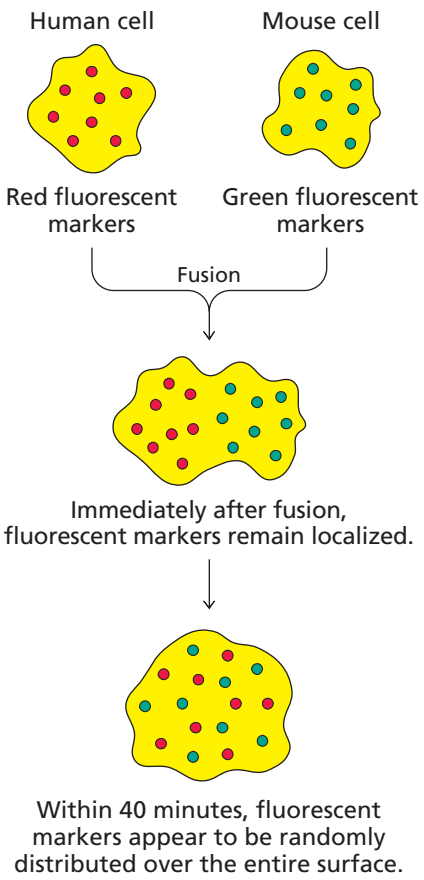
In 1970, L. D. Frye and Michael A. Edidin devised an elegant experiment to test whether membrane proteins diffuse within the lipid bilayer. Frye and Edidin fused mouse cells with human cells to form heterokaryons (hybrid cells). By using red fluorescence-labeled antibodies that specifically bind to certain proteins in human plasma membranes and green fluorescence-labeled antibodies that specifically bind to certain proteins in mouse plasma membranes, they observed the changes in the distribution of membrane proteins over time by immunofluorescence microscopy. The labeled proteins were intermixed within 40 minutes after cell fusion (Figure 9.26). This experiment demonstrated that at least some membrane proteins diffuse freely within biological membranes.

A few membrane proteins move laterally very rapidly but the majority of membrane proteins diffuse about 100 to 500 times more slowly than membrane lipids. The diffusion of some proteins is severely restricted by aggregation or by attachment to the cytoskeleton just beneath the membrane surface. Relatively immobile membrane proteins may act as fences or cages, restricting the movement of other proteins. The limited diffusion of membrane proteins produces protein patches, or domains—areas of membrane whose composition differs from that of the surrounding membrane.

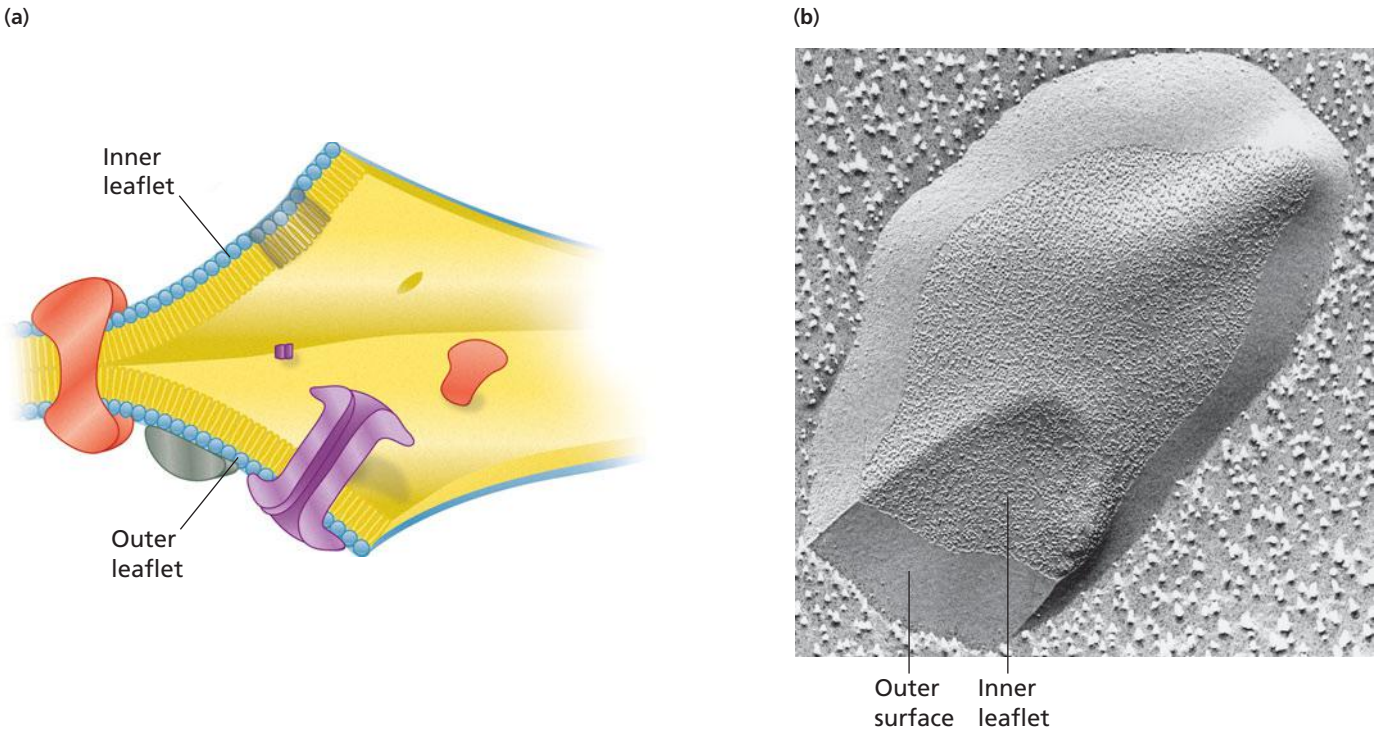
The distribution of membrane proteins can be visualized by *freeze-fracture electron microscopy*. In this technique, a membrane sample is rapidly frozen to the temperature of liquid nitrogen and then fractured with a knife. The membrane splits between the leaflets of the lipid bilayer where the intermolecular interactions are weakest (Figure 9.27a). Ice is evaporated in a vacuum and the exposed internal surface of the membrane is then coated with a thin film of platinum to make a metal replica for examination in an electron microscope. Membranes that are rich in membrane proteins contain pits and bumps indicating the presence of proteins. In contrast, membranes that contain no proteins are smooth. Figure 9.27b shows the bumpy surface of the inner monolayer of a red blood cell membrane exposed by removal of the outer layer.

The fluid properties of lipid bilayers depend on the flexibility of their fatty acyl chains. Saturated acyl chains are fully extended at low temperatures forming a crystalline array with maximal van der Waals contact between the chains. When the lipid bilayer is heated, a phase transition analogous to the melting of a crystalline solid occurs. The acyl chains of lipids in the resulting liquid crystalline phase are relatively disordered and loosely packed. During the phase transition, the thickness of the bilayer decreases by about 15% as the hydrocarbon tails become less extended because of rotation around C—C bonds (Figure 9.28). Bilayers composed of a single type of lipid undergo phase transition at a distinct temperature called the phase-transition temperature. When the lipids contain unsaturated acyl chains, the hydrophobic core of the bilayer is fluid well below room temperature (23°C). Biological membranes, which contain a heterogeneous mixture of lipids, change gradually from the gel to the liquid crystalline phase, typically over a temperature range of 10° to 40°C. Phase transitions in biological membranes can be localized so fluid- and gel-phase regions can coexist at certain temperatures.

The structure of a phospholipid has dramatic effects on its fluidity and phase-transition temperature. As we saw in Section 9.2, the hydrocarbon chain of a fatty acid with a *cis* double bond has a kink that disrupts packing and increases fluidity. Incorporating an unsaturated fatty acyl group into a phospholipid lowers the phase-transition temperature. Changes in membrane fluidity affect the membrane transport and catalytic functions of membrane proteins so many organisms maintain membrane fluidity under different conditions by adjusting the ratio of unsaturated to saturated fatty acyl groups in membrane



▲ **Figure 9.26**
Diffusion of membrane proteins. Human cells whose membrane proteins had been labeled with a red fluorescent marker were fused with mouse cells whose membrane proteins had been labeled with a green fluorescent marker. The initially localized markers became dispersed over the entire surface of the fused cell within 40 minutes.

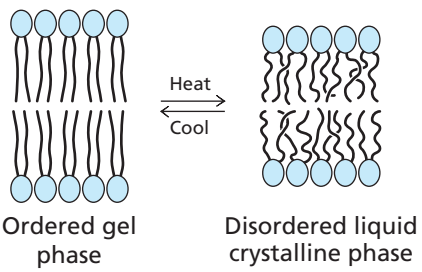


▲ **Figure 9.27**
Freeze fracturing a biological membrane.
(a) Splitting the lipid bilayer along the interface of the two leaflets. A platinum replica of the exposed internal surface is examined in an electron microscope. Membrane proteins appear as protrusions or cavities in the replica. **(b)** Electron micrograph of a freeze-fractured erythrocyte membrane. The bumps on the inner membrane surface show the locations of membrane proteins.

lipids. For example, when bacteria are grown at low temperatures, the proportion of unsaturated fatty acyl groups in membranes increases. Goldfish adapt to the temperature of the water in which they swim: as the environmental temperature drops, there is a rise in unsaturated fatty acids in goldfish intestinal membranes and whole brain. The lower melting point and greater fluidity of unsaturated fatty acyl groups preserve membrane fluidity allowing membrane processes to continue at colder temperatures.

Cholesterol accounts for 20% to 25% of the mass of lipids in a typical mammalian plasma membrane and significantly affects membrane fluidity. When the rigid cholesterol molecules intercalate between the hydrocarbon chains of the membrane lipids, the mobility of fatty acyl chains in the membrane is restricted and fluidity decreases at high temperatures (Figure 9.29). Cholesterol disrupts the ordered packing of the extended fatty acyl chains and thereby increases fluidity at low temperatures. Cholesterol in animal cell membranes thus helps maintain fairly constant fluidity despite fluctuations in temperature or degree of fatty acid saturation.

Cholesterol tends to associate with sphingolipids because they have long saturated fatty acid chains. The unsaturated chains of most glycerophospholipids produce kinks that don't easily accommodate cholesterol molecules in the membrane. Because of this preferential association, mammalian membranes consist of patches of cholesterol/sphingolipids regions surrounded by regions that have very little cholesterol. These patches are called **lipid rafts**. Certain membrane proteins may preferentially associate with lipid rafts. Thus, some membrane proteins may also have a patch-like distribution on the cell surface. Membrane proteins are thought to play an important role in maintaining the integrity of lipid rafts.



▲ **Figure 9.28**
Phase transition of a lipid bilayer. In the ordered gel state, the hydrocarbon chains are extended. Above the phase-transition temperature, rotation around C—C bonds disorders the chains in the liquid crystalline phase.

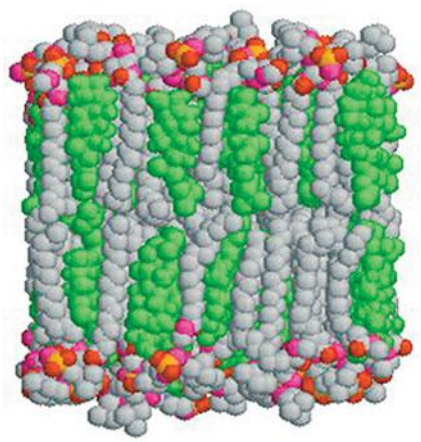
9.10 Membrane Transport

Plasma membranes physically separate a living cell from its environment. In addition, within both prokaryotic and eukaryotic cells there are membrane-bound compartments. The nucleus and mitochondria are obvious examples in eukaryotes.

(a)

(b)

▲ **Goldfish adapt to water temperature. (a)** These goldfish (carp, *Carassius auratus*) have adapted to the water temperature in Kyoto, Japan, by adjusting the lipid composition of their membranes. **(b)** These Goldfish® do not adapt well to any water temperature.



▲ **Figure 9.29**
Model of a lipid membrane. Cholesterol molecules (green) are packed between phospholipid fatty acid chains (grey).

Membranes are selectively permeable barriers that restrict the free passage of most molecules. As a general rule, the permeability of molecules is related to their hydrophobicity and their tendency to dissolve in organic solvents. Thus, hexanoic acid, acetic acid, and ethanol are able to move across membranes quite readily. They have high permeability coefficients (Figure 9.30). Water, despite its strong polar character, is able to diffuse freely across lipid bilayers although, as the permeability coefficient indicates, its movement is still greatly restricted compared to organic solvents like hexanoic acid.

Small ions like Na^+, K^+, and Cl^- have very low permeability coefficients. They are unable to diffuse across a membrane because the hydrophobic core of the lipid bilayer presents an almost impenetrable barrier to most polar or charged species. H^+ ions have a much higher permeability coefficient although membranes still act as an effective barrier to protons.

As mentioned above, very hydrophobic molecules and some small uncharged molecules can move through biological membranes. Water, oxygen, and other small molecules must also be able to enter all cells and move freely between compartments inside eukaryotic cells even if they are not able to diffuse as quickly across membranes. Larger molecules, such as proteins and nucleic acids, have to be transported across membranes, including the membranes between compartments. Living cells move molecules across membranes using transport proteins (sometimes called pores, carriers, permeases, or pumps) and they transport macromolecules by endocytosis or exocytosis.

Nonpolar gases, such as O_2 and CO_2, and hydrophobic molecules, such as steroid hormones, lipid vitamins, and some drugs, enter and leave the cell by diffusing through the membrane moving from the side with the higher concentration to the side with the lower concentration. The rate of movement depends on the difference in concentrations, or the concentration gradient, between the two sides. Diffusion down a concentration gradient (i.e., downhill diffusion) is a spontaneous process driven by an increase in entropy and therefore a decrease in free energy (see below).

The traffic of other molecules and ions across membranes is mediated by three types of integral membrane proteins: channels and pores, passive transporters, and active transporters. These transport systems differ in their kinetic properties and energy requirements. For example, the rate of solute movement through pores and channels may increase with increasing solute concentration but the rate of movement through passive and active transporters may approach a maximum as the solute concentration increases (i.e., the transport protein becomes saturated). Some types of transport require a source of energy (Section C). The characteristics of membrane transport are summarized in Table 9.3. In this section, we describe the different membrane transport systems, as well as endocytosis and exocytosis.

A. Thermodynamics of Membrane Transport

Recall from Chapter 1 (Section 1.4C) that the actual Gibbs free energy change of a reaction is related to the standard Gibbs free energy change by the equation

$$\Delta G_{\text{reaction}} = \Delta G^{\circ\prime}_{\text{reaction}} + RT \ln \frac{[C][D]}{[A][B]} \qquad \textbf{(9.1)}$$

where $\Delta G^{\circ\prime}_{\text{reaction}}$ represents the standard Gibbs free energy change for the reaction, [C][D] represents the concentrations of the products, and [A][B] represents the concentration of the reactants. The Gibbs free energy change associated with membrane transport depends only on the concentrations of the molecules on either side of the membrane.

For any molecule, A, the concentration on the inside of the membrane is $[A_{in}]$ and the concentration outside is $[A_{out}]$. The Gibbs free energy change associated with transporting molecules of A is

$$\Delta G_{\text{transport}} = RT \ln \frac{[A_{in}]}{[A_{out}]} = 2.303\ RT \log \frac{[A_{in}]}{[A_{out}]} \qquad \textbf{(9.2)}$$

Table 9.3 Characteristics of different types of membrane transport

	Protein carrier	Saturable with substrate	Movement relative to concentration gradient	Energy input required
Simple diffusion	No	No	Down	No
Channels and pores	Yes	No	Down	No
Passive transport	Yes	Yes	Down	No
Active transport				
Primary	Yes	Yes	Up	Yes (direct source)
Secondary	Yes	Yes	Up	Yes (ion gradient)

If the concentration of A inside the cell is much less than the concentration of A outside the cell then $\Delta G_{transport}$ will be negative and the flow of A into the cell will be thermodynamically favored. For exmple, if $[A_{in}] = 1$ mM and $[A_{out}] = 100$ mM, then at 25°C

$$\Delta G_{transport} = 2.303\ RT \log \frac{[A_{in}]}{[A_{out}]} = 2.303 \times 8.325 \times 298 \times (-2) = -11.4\ \text{kJ mol}^{-1} \quad (9.3)$$

Under these conditions, molecules of solute A will tend to flow into the cell in order to reduce the concentration gradient. Flow in the opposite direction is thermodynamically unfavorable since it is associated with a positive Gibbs free energy change ($\Delta G_{transport}$ = +11.4 kJ mol^{-1} for molecules moving from the inside of the cell to the outside).

Equation 9.2 only applies to uncharged molecules. In the case of ions, the Gibbs free energy change has to include a factor that takes into account the charge difference across a biological membrane. Most cells selectively export cations so the inside of a cell is negatively charged with respect to the outside. The charge difference across the membrane is

$$\Delta\Psi = \Psi_{in} - \Psi_{out} \quad (9.4)$$

where $\Delta\Psi$ is called the membrane potential (in volts). The Gibbs free energy change due to this electric potential is

$$\Delta G = zF\Delta\Psi \quad (9.5)$$

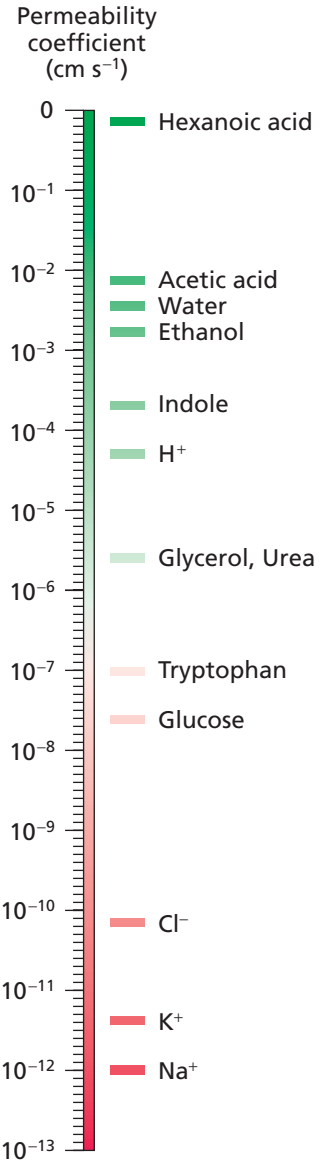
where z is the charge on the molecule being transported (e.g., +1, −1, +2, −2, etc.) and F is Faradays's constant (96,485 JV^{-1}mol^{-1}). Since the inside of the cell is negatively charged, the import of cations such as $Na^{\oplus}$ and $K^{\oplus}$ is thermodynamically favored by the membrane potential. The export of cations must be coupled to an energy-producing reaction since it is associated with a positive Gibbs free energy change.

Both the chemical (concentration) and electric (charge) effects have to be considered, for any transport process involving charged molecules. Thus,

$$\Delta G_{transport} = 2.303\ RT \log \frac{[A_{in}]}{[A_{out}]} + zF\Delta\Psi \quad (9.6)$$

B. Pores and Channels

Pores and *channels* are transmembrane proteins with a central passage for ions and small molecules. (Usually, the term *pore* is used for bacteria and *channel* for animals.) Solutes of the appropriate size, charge, and molecular structure can move rapidly

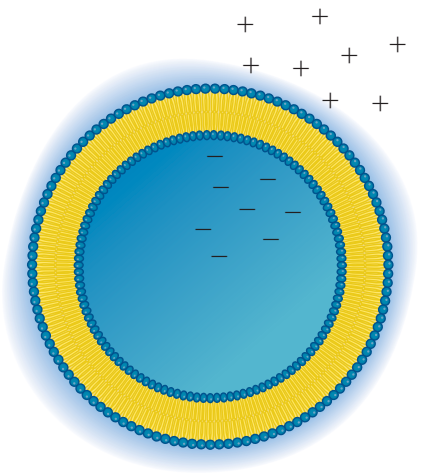


▲ **Figure 9.30**
Permeability coefficients of various molecules. Molecules with high permeability coefficients (top) are able to diffuse unaided across a membrane.

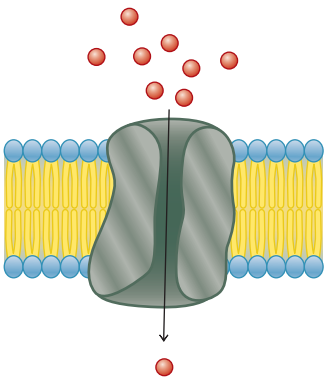
KEY CONCEPT

For a given solute, the Gibbs free energy change of transport depends on both the membrane potential and solute concentrations on either side of the membrane.

The importance of Equation 9.6 will become apparent when we describe chemiosmotic theory (Section 14.3).



▲ **Membrane potential.** In most cases the inside of a cell or membrane compartment is negative with respect to the outside and the membrane potential ($\Delta\psi$) is negative.



▲ **Figure 9.31**
Membrane transport through a pore or channel. A central passage allows molecules and ions of the appropriate size, charge, and geometry to traverse the membrane in either direction.

through the passage in either direction by diffusing down a concentration gradient (Figure 9.31). This process requires no energy. In general, the rate of movement of solute through a pore or channel is not saturable at high concentrations. For some channels, the rate may approach the diffusion controlled limit.

The outer membranes of some bacteria are rich in porins, a family of pore proteins that allow ions and many small molecules to gain access to specific transporters in the plasma membrane. Similar channels are found in the outer membranes of mitochondria. Porins are usually only weakly solute-selective. They can act as sieves that are permanently open or they can be regulated by the concentration of solutes. In contrast, plasma membranes also contain many channel proteins that are highly specific for certain ions and they open or close in response to a specific signal.

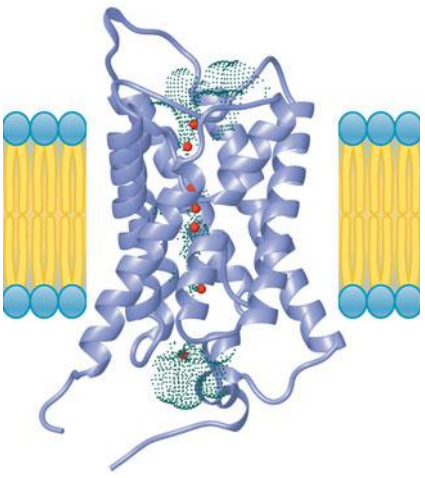
Aquaporin is an integral membrane protein that acts as a pore for water molecules. The channel through the middle of the protein will allow for passage of water molecules and other small uncharged molecules but it blocks passage of any charged molecules or large molecules. This channel is larger on the outside surface but narrows to a much smaller channel on the cytoplasmic side as shown for yeast aquaporin in Figure 9.32. Aquaporins are common in all species. They are required in cells where the rapid uptake of water is necessary because the rate of diffusion of water across the membrane is too slow. This is an example of a simple, somewhat specific, porin. It was discovered by Peter Agre, who received the Nobel Prize in Chemistry in 2003.

CorA is the primary Mg^{2+} pump in prokaryotic cells. It is highly selective for Mg^{2+} and permits the import of Mg^{2+} against a concentration gradient in response to the membrane potential. Positively charged ions "want" to flow into cells and the CorA pore allows passage of Mg^{2+} but not other ions. Mg^{2+} is essential for many cell functions. The rate of influx is regulated by the large cytoplasmic domain of CorA (Figure 9.33). It binds Mg^{2+} ions and when a sufficient number have bound, the pore is closed. Thus, influx of Mg^{2+} is controlled by the cytoplasmic concentration.

Membranes of nerve tissues have gated (i.e., controlled) potassium channels that selectively allow rapid outward transport of potassium ions. These channels permit $K^{\oplus}$ ions to pass through the membrane at least 10,000 times faster than the smaller $Na^{\oplus}$ ions. Crystallographic studies have shown that the potassium channel has a wide mouth (like a funnel) containing negatively charged amino acids to attract cations and repel anions. Hydrated cations are directed electrostatically to an electrically neutral constriction of the pore called the selectivity filter. Potassium ions rapidly lose some of their water of hydration and pass through the selectivity filter. Sodium ions apparently retain more water of hydration and therefore transit the filter much more slowly. The remainder of the channel has a hydrophobic lining. Based on comparisons of amino acid sequences, the general structural properties of the potassium channel seem to also apply to other types of channels and pores. Roderick MacKinnon shared the 2003 Nobel Prize in Chemistry with Peter Agre. MacKinnnon's work focused mainly on potassium channels.

C. Passive Transport and Facilitated Diffusion

Pore and channel proteins are examples of **passive transport** where the Gibbs free energy change for transport is negative and transport from one side of the membrane to the other is a spontaneous process. In *active transport* (see below), the solute moves against a concentration gradient and/or a charge difference. Active transport must be coupled to an energy-producing reaction in order to overcome the unfavorable Gibbs free energy change for unassisted transport. The simplest membrane transporters—whether active or passive—carry out **uniport**; that is, they carry only a single type of solute across the membrane (Figure 9.34a). Many transporters carry out the simultaneous transport of two different solute molecules. The process is called **symport** if both solutes are



◀ **Figure 9.32**
Fungal aquaporin. Aquaporin is an integral membrane protein with an α-helix bundle domain. The water channel (green dots) is open on the exterior surface and narrows to a tiny passage on the cytoplasmic side. *[Pichia pastoris* PDB 2W2E]

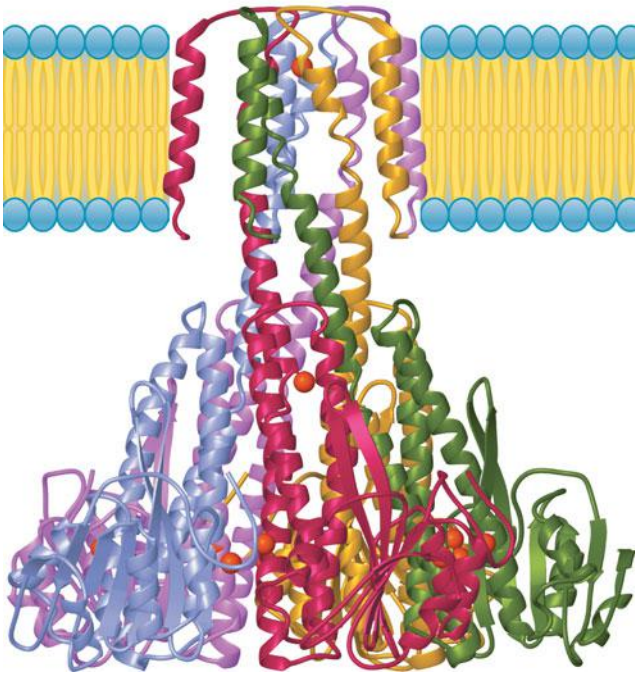


Figure 9.33 ▲
CorA, a magnesium pump. CorA is the prokaryotic magnesium pump. Mg^{2+} ions bind on the exterior surface and are transported through a highly selective channel in response to the membrane potential. The cytoplasmic domain binds Mg^{2+} ions and closes the pore in response to high internal concentrations of Mg^{2+}. This is the *Thermotoga maritima* version with each of the fire subunits in a different color. [PDB 2HN2]

transported in the same direction, (Figure 9.34b). If they are transported in opposite directions, the process is **antiport** (Figure 9.34c).

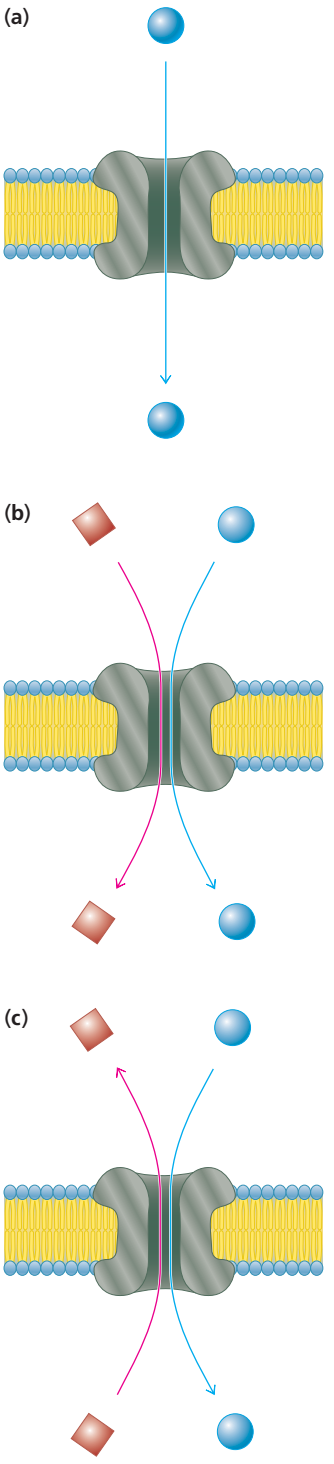
Passive transport includes simple diffusion across a membrane. When pores, channels, and transporters are involved, we call the process **facilitated diffusion**. Facilitated diffusion is still an example of passive transport since it does not require an energy source. The transport protein accelerates the movement of solute down its concentration gradient, or charge gradient, a process that would occur very slowly by diffusion alone. In this case, transport proteins are similar to enzymes because they increase the rate of a process that is thermodynamically favorable. For a simple passive uniport system, the initial rate of inward transport, like the initial rate of an enzyme-catalyzed reaction, depends on the external concentration of substrate. The equation describing this dependence is analogous to the Michaelis–Menten equation for enzyme catalysis (Equation 5.14).

$$v_0 = \frac{V_{max}[S]_{out}}{K_{tr} + [S]_{out}} \qquad \textbf{(9.7)}$$

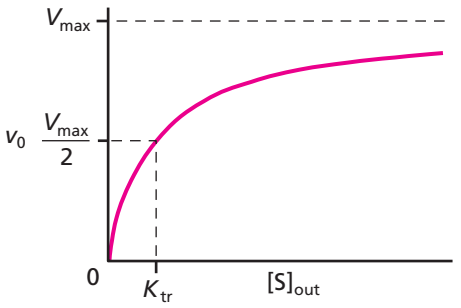
where v_0 is the initial rate of inward transport of the substrate at an external concentration $[S]_{out}$, V_{max} is the maximum rate of transport of the substrate, and K_{tr} is a constant analogous to the Michaelis constant (K_m) (i.e., K_{tr} is the substrate concentration at which the transporter is half-saturated). The lower the value of K_{tr}, the higher the affinity of the transporter for the substrate. The rate of transport is saturable, approaching a maximum value at a high substrate concentration (Figure 9.35).

As substrate accumulates inside the cell, the rate of outward transport increases until it equals the rate of inward transport, and $[S]_{in}$ equals $[S]_{out}$. At this point, there is no net change in the concentration of substrate on either side of the membrane, although substrate continues to move across the membrane in both directions.

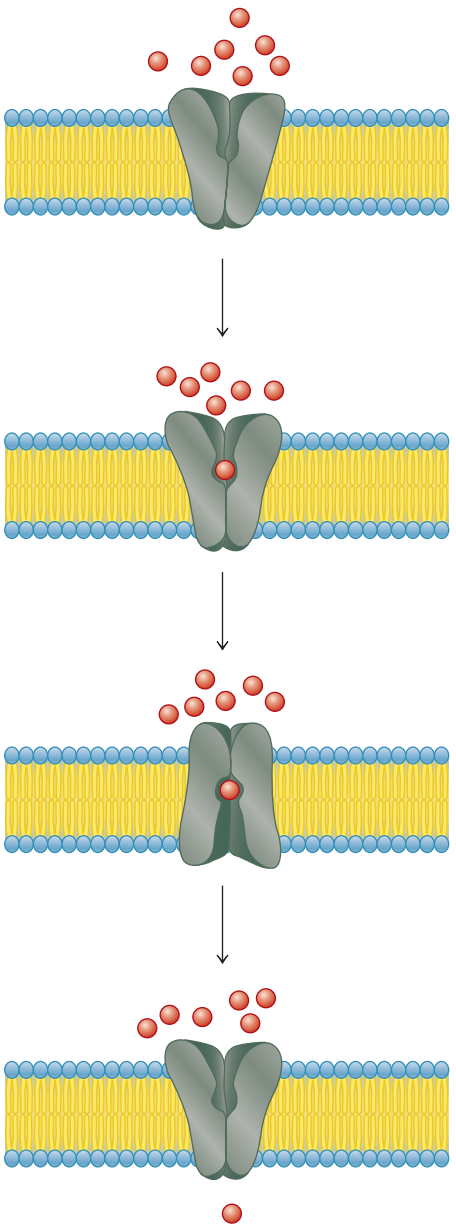
Models of transport protein operation suggest that some transporters undergo a conformational change after they bind their substrates. This conformational change allows the substrate to be released on the other side of the membrane; the transporter



▲ Figure 9.34
Types of passive and active transport. Although the transport proteins are depicted as having an open central pore, passive and active transporters actually undergo conformational changes when transporting their solutes. **(a)** Uniport. **(b)** Symport. **(c)** Antiport.



▲ **Figure 9.35**
Kinetics of passive transport. The initial rate of transport increases with substrate concentration until a maximum is reached. K_{tr} is the concentration of substrate at which the rate of transport is half-maximal.



◄ **Figure 9.36**
Passive and active transport protein function. The protein binds its specific substrate and then undergoes a conformational change, allowing the molecule or ion to be released on the other side of the membrane. Cotransporters have specific binding sites for each transported species.

then reverts to its original state (Figure 9.36). The conformational change in the transporter is often triggered by binding of the transported species, as in the induced fit of certain enzymes to their substrates (Section 6.9). In active transport, the conformational change can be driven by ATP or other sources of energy. Like enzymes, transport proteins can be susceptible to reversible and irreversible inhibition.

D. Active Transport

Active transport resembles passive transport in overall mechanism and kinetic properties. However, active transport requires energy to move a solute up its concentration gradient. In some cases, active transport of charged molecules or ions also results in a charge gradient across the membrane and active transport moves ions against the membrane potential.

Active transporters use a variety of energy sources, most commonly ATP. Ion-transporting ATPases are found in all organisms. These active transporters, which include $Na^{\oplus}$-$K^{\oplus}$ ATPase, and Ca^{2+} ATPase, create and maintain ion concentration gradients across the plasma membrane and across the membranes of internal organelles.

Primary active transport is powered by a direct source of energy such as ATP or light. For example, bacteriorhodopsin (Figure 9.22) uses light energy to generate a transmembrane proton concentration gradient that can be used for ATP formation. One primary active transport protein, P-glycoprotein, appears to play a major role in the resistance of tumor cells to multiple chemotherapeutic drugs. Multidrug resistance is a leading cause of failure in the clinical treatment of human cancers. P-Glycoprotein is an integral membrane glycoprotein (M_r 170,000) that is abundant in the plasma membrane of drug-resistant cells. Using ATP as an energy source, P-glycoprotein pumps a large variety of structurally unrelated nonpolar compounds, such as drugs, out of the cell up a concentration gradient. In this way, the cytosolic drug concentration is maintained at a level low enough to avoid cell death. The normal physiological function of P-glycoprotein appears to be removal of toxic hydrophobic compounds in the diet.

Secondary active transport is driven by an ion concentration gradient. The active uphill transport of one solute is coupled to the downhill transport of a second solute that was concentrated by primary active transport. For example, in *E. coli*, electron flow through a series of membrane-bound oxidation–reduction enzymes generates a higher extracellular concentration of protons. As protons flow back into the cell down their concentration gradient, lactose is also transported in, against its concentration gradient (Figure 9.37). The energy of the proton concentration gradient drives the secondary active transport of lactose. The symport of $H^{\oplus}$ and lactose is mediated by the transmembrane protein lactose permease.

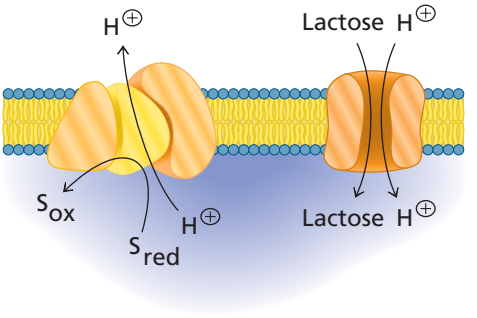
In large multicellular animals, secondary active transport is often powered by a sodium ion gradient. Most cells maintain an intracellular potassium ion concentration of about 140 mM in the presence of an extracellular concentration of about 5 mM. The cytosolic concentration of sodium ions is maintained at about 5 to 15 mM in the presence of an extracellular concentration of about 145 mM. These ion concentration gradients are maintained by $Na^{\oplus}$–$K^{\oplus}$ ATPase, an ATP-driven antiport system that pumps two $K^{\oplus}$ into the cell and ejects three $Na^{\oplus}$ for every molecule of ATP hydrolyzed (Figure 9.38). Each $Na^{\oplus}$–$K^{\oplus}$ ATPase can catalyze the hydrolysis of about 100 molecules of ATP per minute, a significant portion (up to one-third) of the total energy consumption of a typical animal cell. The $Na^{\oplus}$ gradient that is generated by $Na^{\oplus}$–$K^{\oplus}$ ATPase is the major source of energy for secondary active transport of glucose in intestinal cells. One glucose molecule is imported with each sodium ion that enters the cell. The energy released by the downhill movement of $Na^{\oplus}$ powers the uphill transport of glucose.

E. Endocytosis and Exocytosis

The transport we have discussed so far occurs by the flow of molecules or ions across an intact membrane. Cells also need to import and export molecules too large to be transported via pores, channels, or transport proteins. Prokaryotes possess specialized multicomponent export systems in their plasma and outer membranes that allow them to secrete certain proteins (often toxins or enzymes) into the extracellular medium. In eukaryotic cells, many—but not all—proteins (and certain other large substances) are moved into and out of the cell by **endocytosis** and **exocytosis**, respectively. In both cases, transport involves formation of a specialized type of lipid vesicle.

Endocytosis is the process by which macromolecules are engulfed by the plasma membrane and brought into the cell inside a lipid vesicle. Receptor-mediated endocytosis begins with the binding of macromolecules to specific receptor proteins in the plasma membrane of the cell. The membrane then invaginates, forming a vesicle that contains the bound molecules. As shown in Figure 9.39, the inside of such a membrane vesicle is equivalent to the outside of a cell; thus, substances inside the vesicle have not actually crossed the plasma membrane. Once inside the cell, the vesicle can fuse with an endosome (another type of vesicle) and then with a lysosome. Inside a lysosome, the endocytosed material and the receptor itself can be degraded. Alternatively, the ligand, the receptor, or both, can be recycled from the endosome back to the plasma membrane.

Exocytosis is similar to endocytosis except that the direction of transport is reversed. During exocytosis, materials destined for secretion from the cell are enclosed in vesicles by the Golgi apparatus (Section 1.8B). The vesicles then fuse with the plasma membrane releasing the vesicle contents into the extracellular space. The zymogens of digestive enzymes are exported from pancreatic cells in this manner (Section 6.7A).



▲ **Figure 9.37**
Secondary active transport in *Escherichia coli*. The oxidation of reduced substrates (S_{red}) generates a transmembrane proton concentration gradient. The energy released by protons moving down their concentration gradient drives the transport of lactose into the cell by lactose permease.

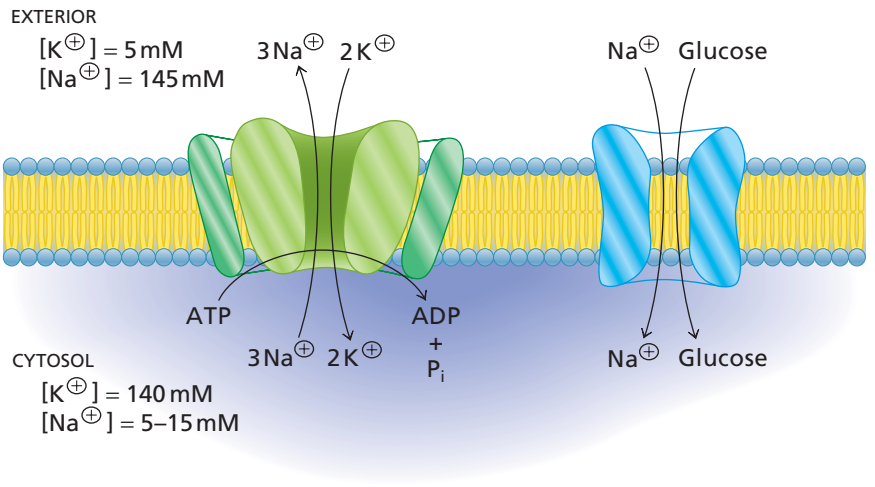
The secretory pathway in eukaryotic cells is described in Section 22.10.

9.11 Transduction of Extracellular Signals

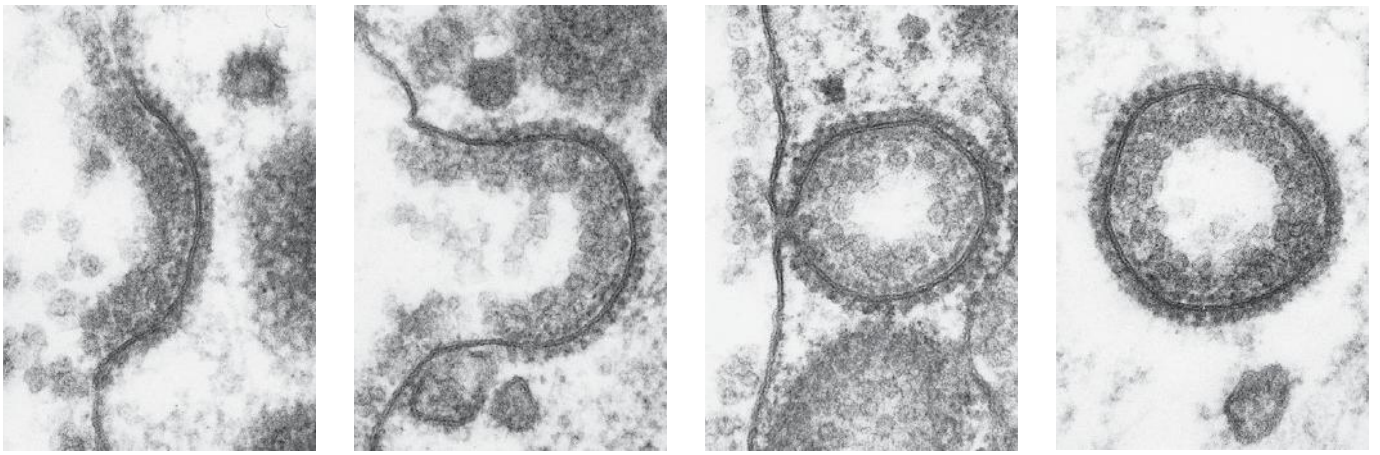
In order for a cell to interact with its external environment, it must detect molecules outside of the plasma membrane and convey that information to the inside of the cell. This process is called **signal transduction** and it is a very active field of research. In this section we'll cover the basic mechanism of the most common signaling pathways. As you learn more biochemistry, you'll encounter many variations of these themes.

A. Receptors

The plasma membranes of all cells contain specific receptors that allow the cell to respond to external chemical stimuli that cannot cross the membrane. For example,



◀ **Figure 9.38**
Secondary active transport in animals. The $Na^{\oplus}$–$K^{\oplus}$ ATPase generates a sodium ion gradient that drives secondary active transport of glucose in intestinal cells.



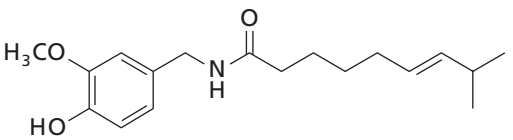
▲ **Figure 9.39**
Electron micrographs of endocytosis. Endocytosis begins with the binding of macromolecules to the plasma membrane of the cell. The membrane then invaginates forming a vesicle that contains the bound molecules. The inside of the vesicle is topologically equivalent to the outside of the cell.

bacteria can detect certain chemicals in their environment. A signal is passed via a cell surface receptor to the flagella, causing the bacterium to swim toward a potential food source. This is called positive **chemotaxis**. In negative chemotaxis, the bacteria swim away from toxic chemicals.

In multicellular organisms, stimuli such as *hormones*, *neurotransmitters* (substances that transmit nerve messages at synapses), and *growth factors* (proteins that regulate cell proliferation) are produced by specialized cells. These ligands can travel to other tissues where they bind to and produce specific responses in cells with the appropriate receptors on their surfaces. In this section, we see how the binding of water-soluble ligands to receptors elicits intracellular responses in mammals. These signal transduction pathways involve adenylyl cyclase, inositol phospholipids, and receptor tyrosine kinases.

BOX 9.6 THE HOT SPICE OF CHILI PEPPERS

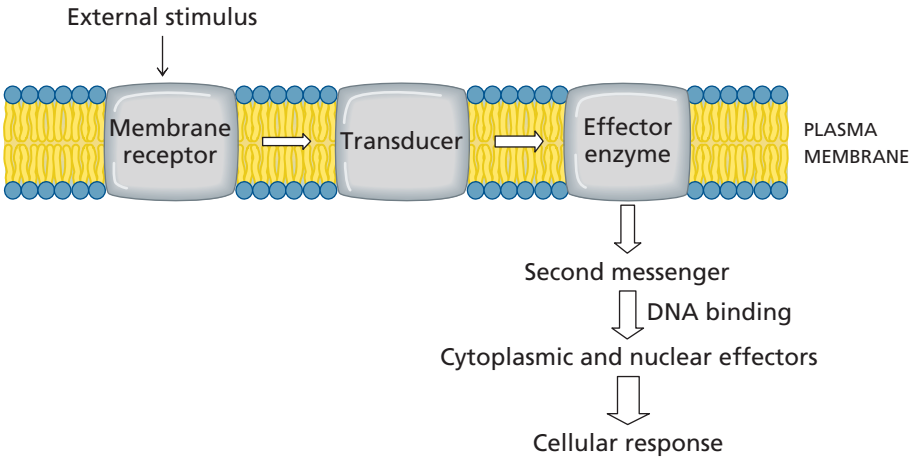
Biochemists now know the mechanism by which spice from "hot" peppers exerts its action, causing a burning pain. The active factor in capsaicin peppers is a lipophilic vanilloid compound called capsaicin.



Capsaicin

A nerve cell protein receptor that responds to capsaicin has been identified and characterized. It is an ion channel and its amino acid sequence suggests that it has six transmembrane domains. Activation of the receptor by capsaicin causes the channel to open so that calcium and sodium ions can flow into the nerve cell and send an impulse to the brain. The receptor is activated not only by vanilloid spices but also by rapid increases in temperature. In fact, the main function of the receptor is detection of heat.

Chili peppers ►



◀ **Figure 9.40**
General mechanism of signal transduction across the plasma membrane of a cell.

A general mechanism for signal transduction is shown in Figure 9.40. A ligand binds to its specific receptor on the surface of the target cell. This interaction generates a signal that is passed through a membrane protein **transducer** to a membrane-bound **effector enzyme**. The action of the effector enzyme generates an intracellular **second messenger** that is usually a small molecule or ion. The diffusible second messenger carries the signal to its ultimate destination which may be in the nucleus, an intracellular compartment, or the cytosol. Ligand binding to a cell-surface receptor almost invariably results in the activation of protein kinases. These enzymes catalyze the transfer of a phosphoryl group from ATP to various protein substrates, many of which help regulate metabolism, cell growth, and cell division. Some proteins are activated by phosphorylation, whereas others are inactivated. A vast diversity of ligands, receptors, and transducers exists but only a few second messengers and types of effector enzymes are known.

Kinases were introduced in Section 6.9.

Receptor tyrosine kinases have a simpler mechanism for signal transduction. With these enzymes, the membrane receptor, transducer, and effector enzyme are combined in one enzyme. A receptor domain on the extracellular side of the membrane is connected to the cytosolic active site by a transmembrane segment. The active site catalyzes phosphorylation of its target proteins.

KEY CONCEPT

Membrane receptors are the primary step in carrying information across a membrane.

Amplification is an important feature of signaling pathways. A single ligand receptor complex can interact with a number of transducer molecules, each of which can activate several molecules of effector enzyme. Similarly, the production of many second messenger molecules can activate many kinase molecules that catalyze the phosphorylation of many target proteins. This series of amplification events is called a **cascade**. The cascade mechanism means that small amounts of an extracellular compound can affect large numbers of intracellular enzymes without crossing the plasma membrane or binding to each target protein.

The actions of the hormones insulin, glucagon, and epinephrine and the roles of transmembrane signaling pathways in the regulation of carbohydrate and lipid metabolism are described in Sections 11.5, 13.3, 13.7, 13.10, 16.1C, 16.4 (Box), and 16.7.

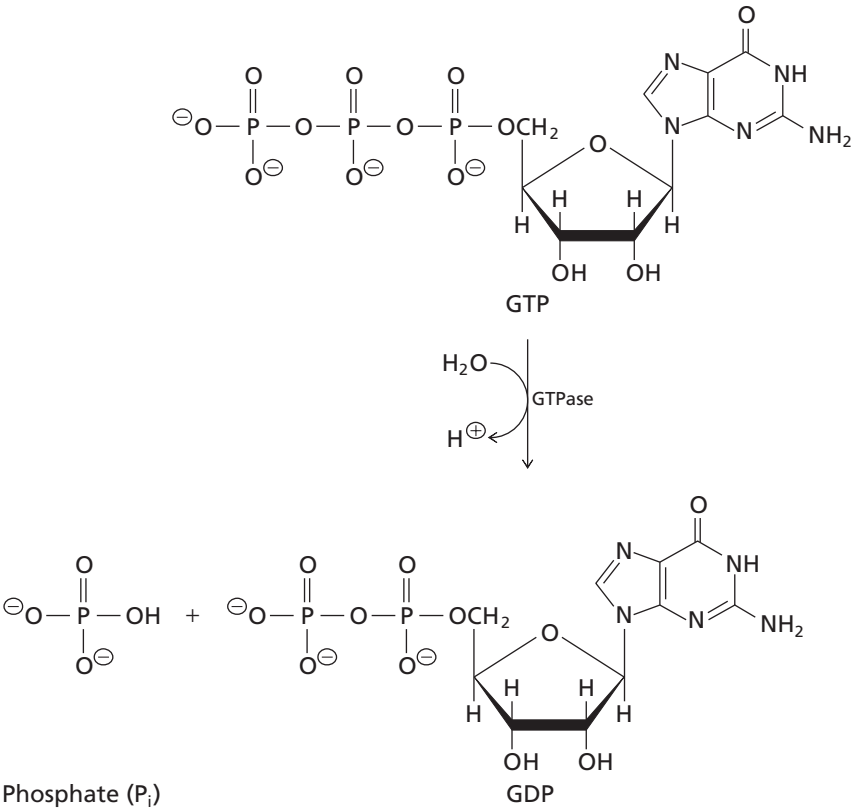
Not all chemical stimuli follow the general mechanism of signal transduction shown in Figure 9.40. For example, because steroid hormones are hydrophobic, they can diffuse across the plasma membrane into the cell where they can bind to specific receptor proteins in the cytoplasm. The steroid receptor complexes are then transferred to the nucleus. The complexes bind to specific regions of DNA called hormone response elements and thereby enhance or suppress the expression of adjacent genes.

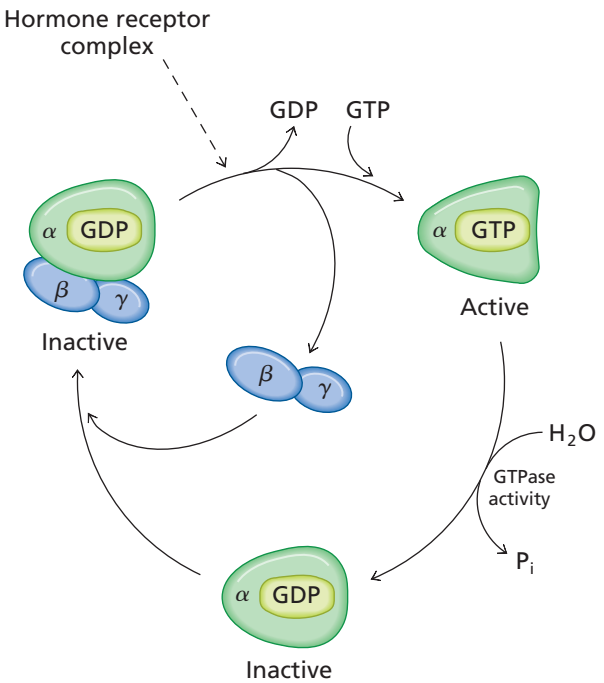
B. Signal Transducers

There are many kinds of receptors and many different transducers. Bacterial transducers are different than eukaryotic ones. There are some eukaryotic transducers found in most species. In this section, we'll concentrate on those general transducers.

Many membrane receptors interact with a family of guanine nucleotide binding proteins called **G proteins**. G proteins act as transducers—the agents that transmit external

Figure 9.41 ►
Hydrolysis of guanosine 5′-triphosphate (GTP) to guanosine 5′-diphosphate (GDP) and phosphate (P_i).





▲ Figure 9.42
G-protein cycle. G proteins undergo activation after binding to a receptor ligand complex and are slowly inactivated by their own GTPase activity. Both G_α–GTP/GDP and $G_{\beta\gamma}$ are membrane-bound.

stimuli to effector enzymes. G proteins have GTPase activity; that is, they slowly catalyze hydrolysis of bound guanosine 5′-triphosphate (GTP, the guanine analog of ATP) to guanosine 5′-diphosphate (GDP) (Figure 9.41). When GTP is bound to G protein it is active in signal tranduction and when G protein is bound to GDP it is inactive. The cyclic activation and deactivation of G proteins is shown in Figure 9.42. The G proteins involved in signaling by hormone receptors are peripheral membrane proteins located on the inner surface of the plasma membrane. Each protein consists of an α, a β, and a γ subunit. The α and γ subunits are lipid anchored membrane proteins; the α subunit is a fatty acyl anchored protein and the γ subunit is a prenyl anchored protein. The complex of $G_{\alpha\beta\gamma}$ and GDP is inactive.

When a hormone receptor complex diffusing laterally in the membrane encounters and binds $G_{\alpha\beta\gamma}$, it induces the G protein to change to an active conformation. Bound GDP is rapidly exchanged for GTP promoting the dissociation of G_α–GTP from $G_{\beta\gamma}$. Activated G_α–GTP then interacts with the effector enzyme. The GTPase activity of the G protein acts as a built-in timer since G proteins slowly catalyze the hydrolysis of GTP to GDP. When GTP is hydrolyzed the G_α–GDP complex reassociates with $G_{\beta\gamma}$ and the $G_{\alpha\beta\gamma}$–GDP complex is regenerated. G proteins have evolved into good switches because they are very slow catalysts, typically having a k_{cat} of only about 3 min^{-1}.

G proteins are found in dozens of signaling pathways including the adenylyl cyclase and the inositol-phospholipid pathways discussed below. An effector enzyme can respond to stimulatory G proteins (Gs) or inhibitory G proteins (Gi). The α subunits of different G proteins are distinct providing varying specificity but the β and γ subunits are similar and often interchangeable. Humans have two dozen α proteins, five β proteins, and six γ proteins.

C. The Adenylyl Cyclase Signaling Pathway

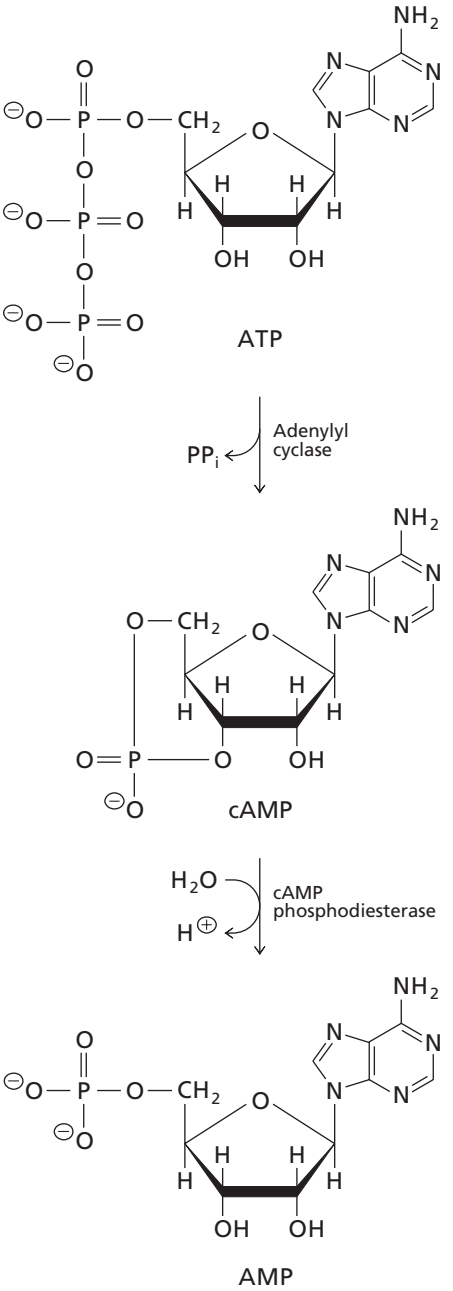
The cyclic nucleotides 3′,5′-cyclic adenosine monophosphate (cAMP) and its guanine analog, 3′,5′-cyclic guanosine monophosphate (cGMP), are second messengers that help transmit signals from external sources to intracellular enzymes. cAMP is produced from ATP by the action of adenylyl cyclase (Figure 9.43) and cGMP is formed from GTP in a similar reaction.

Many hormones that regulate intracellular metabolism exert their effects on target cells by activating the adenylyl cyclase signaling pathway. Binding of a hormone to a stimulatory receptor causes the conformation of the receptor to change promoting interaction between the receptor and a stimulatory G protein, G_s. The receptor ligand complex activates G_s that, in turn, binds the effector enzyme adenylyl cyclase and activates it by allosterically inducing a conformational change at its active site.

Adenylyl cyclase is an integral membrane enzyme whose active site faces the cytosol. It catalyzes the formation of cAMP from ATP. cAMP then diffuses from the membrane surface through the cytosol and activates an enzyme known as protein kinase A. This kinase is made up of a dimeric regulatory subunit and two catalytic subunits and is inactive in its fully assembled state. When the cytosolic concentration of cAMP increases as a result of signal transduction through adenylyl cyclase, four molecules of cAMP bind to the regulatory subunit of the kinase releasing the two catalytic subunits, which are enzymatically active (Figure 9.44). Protein kinase A, a serine-threonine protein kinase, catalyzes phosphorylation of the hydroxyl groups of specific serine and threonine residues in target enzymes. Phosphorylation of amino acid side chains on the target enzymes is reversed by the action of protein phosphatases that catalyze hydrolytic removal of the phosphoryl groups.

The ability to turn off a signal transduction pathway is an essential element of all signaling processes. For example, the cAMP concentration in the cytosol increases only transiently. A soluble cAMP phosphodiesterase catalyzes the hydrolysis of cAMP to AMP (Figure 9.43) limiting the lifetime of the second messenger. At high concentrations, the methylated purines caffeine and theophylline (Figure 9.45) inhibit cAMP phosphodiesterase, thereby decreasing the rate of conversion of cAMP to AMP. These inhibitors prolong and intensify the effects of cAMP and hence the activating effects of the stimulatory hormones.

Hormones that bind to stimulatory receptors activate adenylyl cyclase and raise intracellular cAMP levels. Hormones that bind to inhibitory receptors inhibit adenylyl cyclase activity via receptor interaction with the transducer G_i. The ultimate response of a cell to a hormone depends on the type of receptors present and the type of G protein to which they are coupled. The main features of the adenylyl cyclase signaling pathway, including G proteins, are summarized in Figure 9.46.



▲ **Figure 9.43**
Production and inactivation of cAMP. ATP is converted to cAMP by the transmembrane enzyme adenylyl cyclase. The second messenger is subsequently converted to 5′-AMP by the action of a cytosolic cAMP phosphodiesterase.

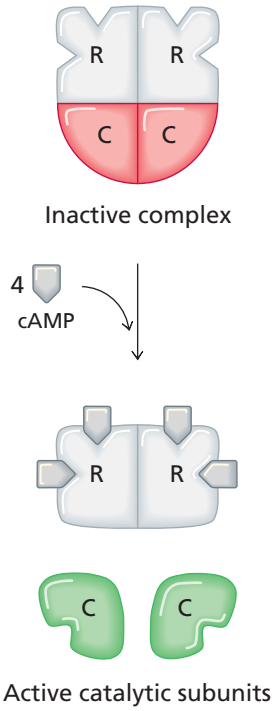
The response of *E. coli* to changes in glucose concentrations, modulated by cAMP, is described in Section 21.7B.

D. The Inositol–Phospholipid Signaling Pathway

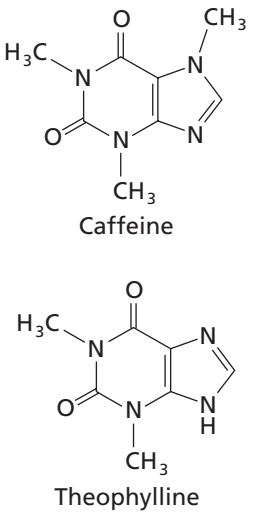
Another major signal transduction pathway produces two different second messengers, both derived from a plasma membrane phospholipid called phosphatidylinositol 4,5-*bis*phosphate (PIP_2) (Figure 9.47). PIP_2 is a minor component of plasma membranes located in the inner monolayer. It is synthesized from phosphatidylinositol by two successive phosphorylation steps catalyzed by ATP-dependent kinases.

Following binding of a ligand to a specific receptor, the signal is transduced through the G protein G_q. The active GTP-bound form of G_q activates the effector enzyme phosphoinositide-specific phospholipase C that is bound to the cytoplasmic face of the plasma membrane. Phospholipase C catalyzes the hydrolysis of PIP_2 to inositol 1,4,5-*tris*phosphate (IP_3) and diacylglycerol (Figure 9.47). Both IP_3 and diacylglycerol are second messengers that transmit the original signal to the interior of the cell.

IP_3 diffuses through the cytosol and binds to a calcium channel in the membrane of the endoplasmic reticulum. This causes the calcium channel to open for a short time, releasing Ca^{2+} from the lumen of the endoplasmic reticulum into the cytosol. Calcium is also an intracellular messenger because it activates calcium-dependent protein



▲ **Figure 9.44**
Activation of protein kinase A. The assembled complex is inactive. When four molecules of cAMP bind to the regulatory subunit (R) dimer, the catalytic subunits (C) are released.



▲ **Figure 9.45**
Caffeine and theophylline.

kinases that catalyze phosphorylation of various protein targets. The calcium signal is short-lived since Ca^{2+} is pumped back into the lumen of the endoplasmic reticulum when the channel closes.

The other product of PIP_2 hydrolysis, diacylglycerol, remains in the plasma membrane. Protein kinase C, which exists in equilibrium between a soluble cytosolic form and a peripheral membrane form, moves to the inner face of the plasma membrane where it binds transiently and is activated by diacylglycerol and Ca^{2+}. Protein kinase C catalyzes phosphorylation of many target proteins altering their catalytic activity. Several protein kinase C isozymes exist, each with different catalytic properties and tissue distribution. They are members of the serine–threonine kinase family.

Signaling via the inositol–phospholipid pathway is turned off in several ways. First, when GTP is hydrolyzed, G_q returns to its inactive form and no longer stimulates phospholipase C. The activities of IP_3 and diacylglycerol are also transient. IP_3 is rapidly hydrolyzed to other inositol phosphates (which can also be second messengers) and inositol. Diacylglycerol is rapidly converted to phosphatidate. Both inositol and phosphatidate are recycled back to phosphatidylinositol. The main features of the inositol–phospholipid signaling pathway are summarized in Figure 9.48.

Phosphatidylinositol is not the only membrane lipid that gives rise to second messengers. Some extracellular signals lead to the activation of hydrolases that catalyze the conversion of membrane sphingolipids to sphingosine, sphingosine 1-phosphate, or ceramide. Sphingosine inhibits protein kinase C, and ceramide activates a protein kinase and a protein phosphatase. Sphingosine 1-phosphate can activate phospholipase

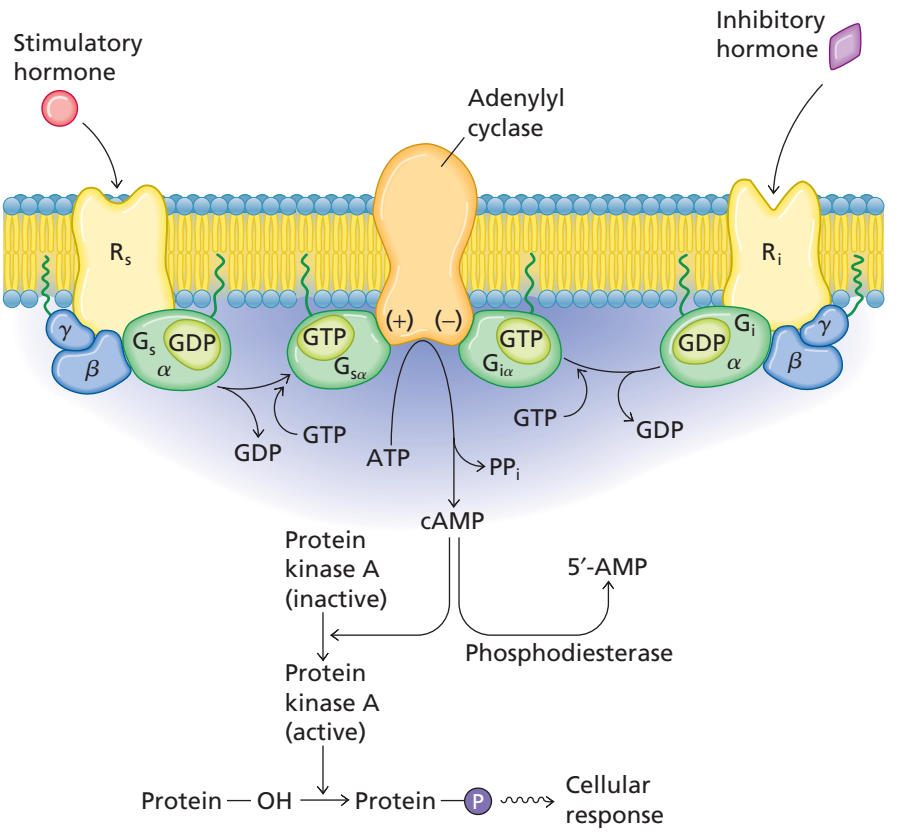
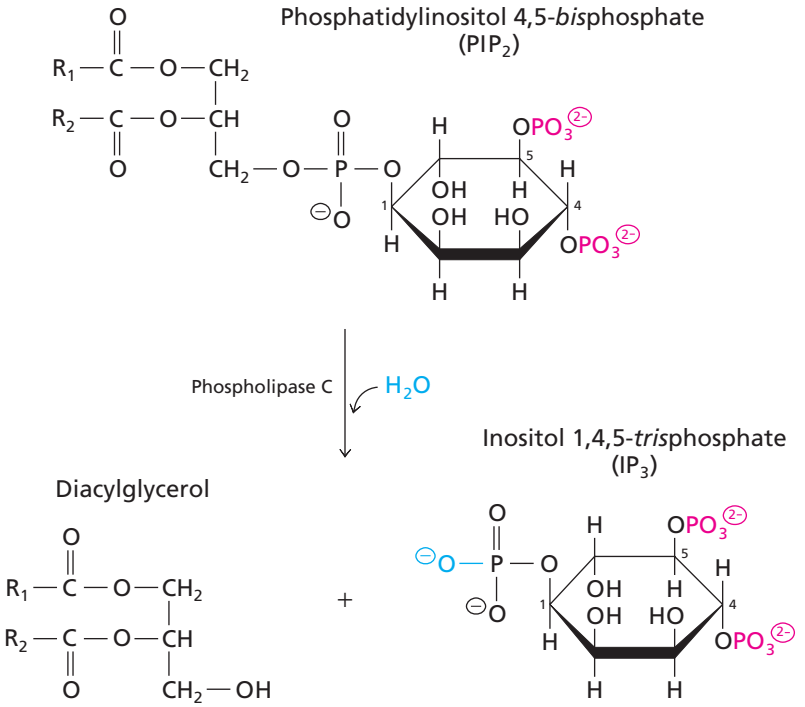
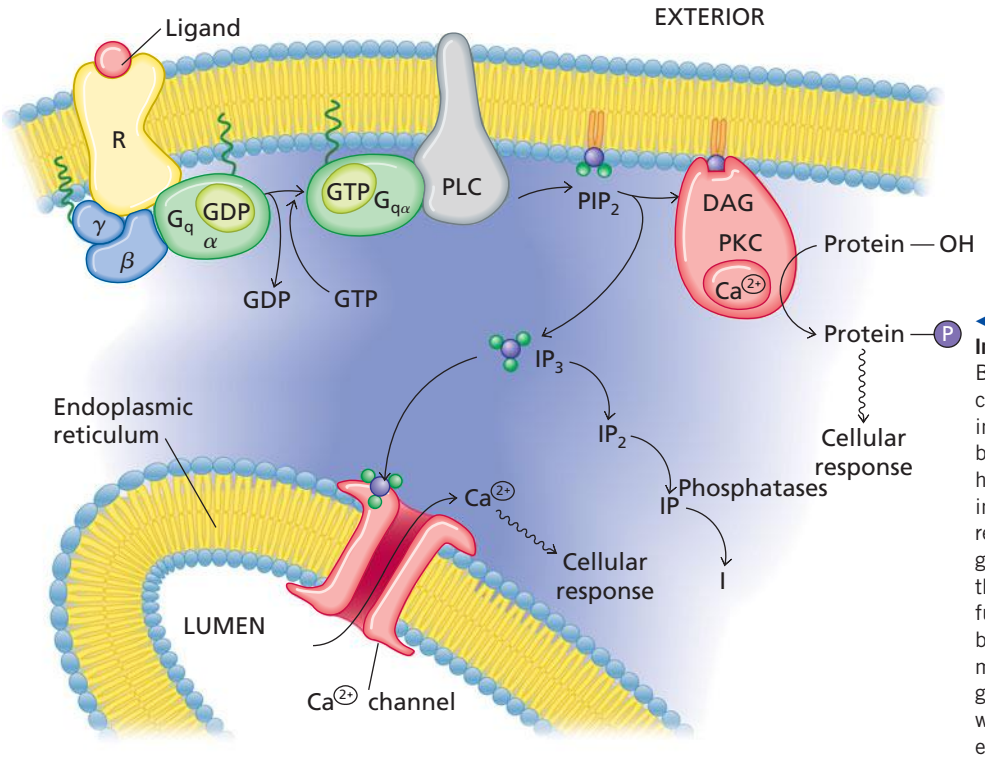


Figure 9.46 ▲
Summary of the adenylyl cyclase signaling pathway. Binding of a hormone to a stimulatory transmembrane receptor (R_s) leads to activation of the stimulatory G protein (G_s) on the inside of the membrane. Other hormones can bind to inhibitory receptors (R_i) that are coupled to adenylyl cyclase by the inhibitory G protein G_i. G_s activates the integral membrane enzyme adenylyl cyclase whereas G_i inhibits it. cAMP activates protein kinase A resulting in the phosphorylation of cellular proteins.



◀ **Figure 9.47**
Phosphatidylinositol 4,5-*bis*phosphate (PIP_2). Phosphatidylinositol 4,5-*bis*phosphate (PIP_2) produces two second messengers, inositol 1,4,5-*tris*phosphate (IP_3) and diacylglycerol. PIP_2 is synthesized by the addition of two phosphoryl groups (red) to phosphatidylinositol and hydrolyzed to IP_3 and diacylglycerol by the action of a phosphoinositide-specific phospholipase C.

D, which specifically catalyzes hydrolysis of phosphatidylcholine. The phosphatidate and the diacylglycerol formed by this hydrolysis appear to be second messengers. The full significance of the wide variety of second messengers generated from membrane lipids (each with its own specific fatty acyl groups) has not yet been determined.



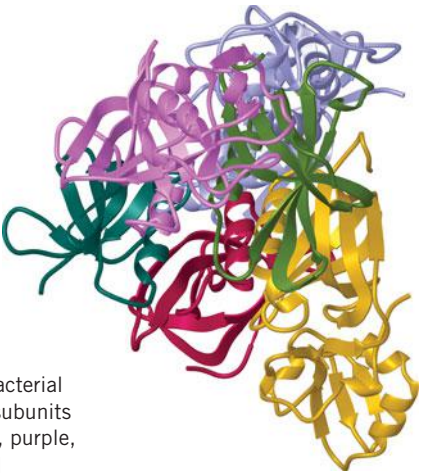
◀ **Figure 9.48**
Inositol–phospholipid signaling pathway. Binding of a ligand to its transmembrane receptor (R) activates the G protein (G_q). This in turn stimulates a specific membrane-bound phospholipase C (PLC) that catalyzes hydrolysis of the phospholipid PIP_2 in the inner leaflet of the plasma membrane. The resulting second messengers, IP_3 and diacylglycerol (DAG), are responsible for carrying the signal to the interior of the cell. IP_3 diffuses to the endoplasmic reticulum where it binds to and opens a Ca^{2+} channel in the membrane releasing stored Ca^{2+}. Diacylglycerol remains in the plasma membrane where it—along with Ca^{2+}—activates the enzyme protein kinase C (PKC).

BOX 9.7 BACTERIAL TOXINS AND G PROTEINS

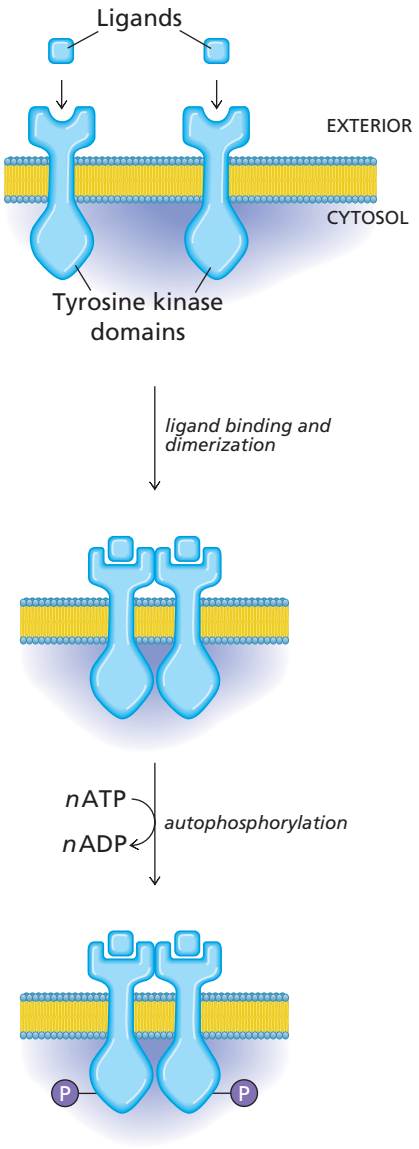
G proteins are the biological targets of cholera and pertussis (whooping cough) toxins that are secreted by the disease-producing bacteria *Vibrio cholerae* and *Bordetella pertussis*, respectively. Both diseases involve overproduction of cAMP.

Cholera toxin binds to ganglioside G_{M1} on the cell surface (Section 9.5) and a subunit of it crosses the plasma membrane and enters the cytosol. This subunit catalyzes covalent modification of the α subunit of the G protein G_s inactivating its GTPase activity. The adenylyl cyclase of these cells remains activated and cAMP levels stay high. In people infected with *V. cholerae*, cAMP stimulates certain transporters in the plasma membrane of the intestinal cells leading to a massive secretion of ions and water into the gut. The dehydration resulting from diarrhea can be fatal unless fluids are replenished.

Pertussis toxin binds to a glycolipid called lactosylceramide found on the cell surface of epithelial cells in the lung. It is taken up by endocytosis. The toxin catalyzes covalent modification of G_i. In this case, the modified G protein is unable to replace GDP with GTP and therefore adenylyl cyclase activity cannot be reduced via inhibitory receptors. The resulting increase in cAMP levels produces the symptoms of whooping cough.



► **Pertussis toxin.** The bacterial toxin has five different subunits colored red, green, blue, purple, and yellow. [PDB 1BCP]



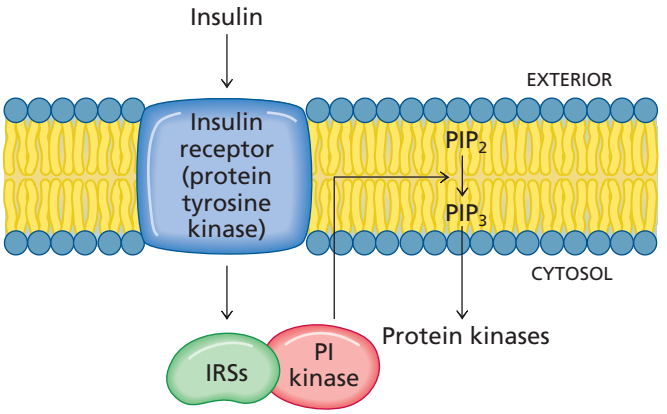
E. Receptor Tyrosine Kinases

Many growth factors operate by a signaling pathway that includes a multifunctional transmembrane protein called a receptor tyrosine kinase. As shown in Figure 9.49, the receptor, transducer, and effector functions are all found in a single membrane protein. In one type of activation, a ligand binds to the extracellular domain of the receptor, activating tyrosine kinase catalytic activity in the intracellular domain by dimerization of the receptor. When two receptor molecules associate, each tyrosine kinase domain catalyzes the phosphorylation of specific tyrosine residues of its partner, a process called *autophosphorylation*. The activated tyrosine kinase then catalyzes phosphorylation of certain cytosolic proteins, setting off a cascade of events in the cell.

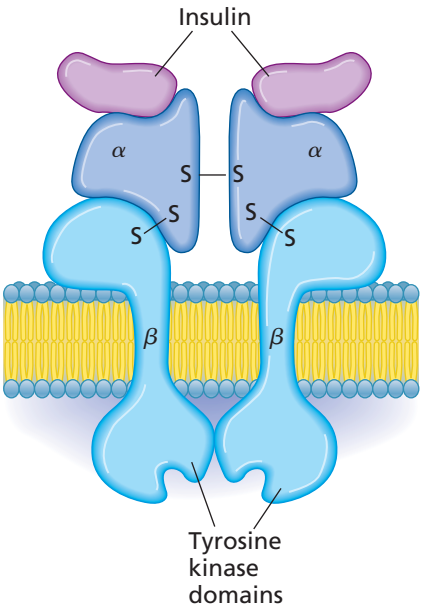
The insulin receptor is an $\alpha_2\beta_2$ tetramer (Figure 9.50). When insulin binds to the α subunit, it induces a conformational change that brings the tyrosine kinase domains of the β subunits together. Each tyrosine kinase domain in the tetramer catalyzes the phosphorylation of the other kinase domain. The activated tyrosine kinase also catalyzes the phosphorylation of tyrosine residues in other proteins that help regulate nutrient utilization.

Recent research has found that many of the signaling actions of insulin are mediated through PIP_2 (Section 9.12C and Figure 9.51). Rather than causing hydrolysis of PIP_2, insulin (via proteins called insulin receptor substrates, IRSs) activates phosphotidylinositol 3-kinase, an enzyme that catalyzes the phosphorylation of PIP_2 to phosphatidylinositol 3,4,5-*tris*phosphate (PIP_3). PIP_3 is a second messenger that transiently activates a series of target proteins, including a specific phosphoinositide-dependent protein kinase. In this way, phosphotidylinositol 3-kinase is the molecular switch that regulates several serine–threonine protein kinase cascades.

◄ **Figure 9.49**
Activation of receptor tyrosine kinases. Activation occurs as a result of ligand induced receptor dimerization. Each kinase domain catalyzes phosphorylation of its partner. The phosphorylated dimer can catalyze phosphorylation of various target proteins.



▲ **Figure 9.51**
Insulin-stimulated formation of phosphatidylinositol 3,4,5-*tris*phosphate (PIP_3). Binding of insulin to its receptor activates the protein tyrosine kinase activity of the receptor leading to the phosphorylation of insulin receptor substrates (IRSs). The phosphorylated IRSs interact with phosphotidylinositiol 3-kinase (PI kinase) at the plasma membrane where the enzyme catalyzes the phosphorylation of PIP_2 to PIP_3. PIP_3 acts as a second messenger carrying the message from extracellular insulin to certain intracellular protein kinases.



▲ **Figure 9.50**
Insulin receptor. Two extracellular α chains, each with an insulin binding site, are linked to two transmembrane β chains, each with a cytosolic tyrosine kinase domain. Following insulin binding to the α chains, the tyrosine kinase domain of each β chain catalyzes autophosphorylation of tyrosine residues in the adjacent kinase domain. The tyrosine kinase domains also catalyze the phosphorylation of proteins called insulin receptor substrates (IRSs).

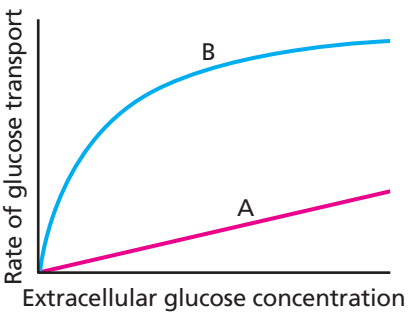
Phosphoryl groups are removed from both the growth factor receptors and their protein targets by the action of protein tyrosine phosphatases. Although only a few of these enzymes have been studied, they appear to play an important role in regulating the tyrosine kinase signaling pathway. One means of regulation appears to be the localized assembly and separation of enzyme complexes.

Summary

1. Lipids are a diverse group of water-insoluble organic compounds.
2. Fatty acids are monocarboxylic acids, usually with an even number of carbon atoms ranging from 12 to 20.
3. Fatty acids are generally stored as triacylglycerols (fats and oils), which are neutral and nonpolar.
4. Glycerophospholipids have a polar head group and nonpolar fatty acyl tails linked to a glycerol backbone.
5. Sphingolipids, which occur in plant and animal membranes, contain a sphingosine backbone. The major classes of sphingolipids are sphingomyelins, cerebrosides, and gangliosides.
6. Steroids are isoprenoids containing four fused rings.
7. Other biologically important lipids are waxes, eicosanoids, lipid vitamins, and terpenes.
8. The structural basis for all biological membranes is the lipid bilayer that includes amphipathic lipids such as glycerophospholipids, sphingolipids, and sometimes cholesterol. Lipids can diffuse rapidly within a leaflet of the bilayer.
9. A biological membrane contains proteins embedded in or associated with a lipid bilayer. The proteins can diffuse laterally within the membrane.
10. Most integral membrane proteins span the hydrophobic interior of the bilayer, but peripheral membrane proteins are more loosely associated with the membrane surface. Lipid anchored membrane proteins are covalently linked to lipids in the bilayer.
11. Some small or hydrophobic molecules can diffuse across the bilayer. Channels, pores, and passive and active transporters mediate the movement of ions and polar molecules across membranes. Macromolecules can be moved into and out of the cell by endocytosis and exocytosis, respectively.
12. Extracellular chemical stimuli transmit their signals to the cell interior by binding to receptors. A transducer passes the signal to an effector enzyme, which generates a second messenger. Signal transduction pathways often include G proteins and protein kinases. The adenylyl cyclase signaling pathway leads to activation of the cAMP-dependent protein kinase A. The inositol-phospholipid signaling pathway generates two second messengers and leads to the activation of protein kinase C and an increase in the cytosolic Ca^{2+} concentration. In receptor tyrosine kinases, the kinase is part of the receptor protein.

Problems

1. Write the molecular formulas for the following fatty acids: (a) nervonic acid (*cis*-Δ^{15}-tetracosenoate; 24 carbons); (b) vaccenic acid (*cis*-Δ^{11}-octadecenoate); and (c) EPA (all *cis*-$\Delta^{5,8,11,14,17}$-eicosapentaenoate).
2. Write the molecular formulas for the following modified fatty acids:
   (a) 10-(Propoxy) decanoate, a synthetic fatty acid with antiparasitic activity used to treat African sleeping sickness, a disease caused by the protozoan *T. brucei* (the propoxy group is $—O—CH_2CH_2CH_3$)
   (b) Phytanic acid (3,7,11,15-tetramethylhexadecanoate), found in dairy products
   (c) Lactobacillic acid (*cis*-11,12-methyleneoctadecanoate), found in various microorganisms
3. Fish ois are rich sources of omega-3 and polyunsaturated fatty acids and omega-6 fatty acids are relatively abundant in corn and sunflower oils. Classify the following fatty acids as omega-3, omega-6, or neither: (a) linolenate, (b) linoleate, (c) arachidonate, (d) oleate, (e) $\Delta^{8,11,14}$-eicosatrienoate.
4. Mammalian platelet activating factor (PAF), a messenger in signal transduction, is a glycerophospholipid with an ether linkage at C-1. PAF is a potent mediator of allergic responses, inflammation, and the toxic-shock syndrome. Draw the structure of PAF (1-alkyl-2-acetyl-phosphatidyl-choline), where the 1-alkyl group is a C_{16} chain.
5. Docosahexaenoic acid, 22:6 $\Delta^{4,7,10,13,16,19}$, is the predominate fatty acyl group in the C-2 position of glycerol-3-phosphate in phosphatidylethanolamine and phosphatidylcholine in many types of fish.
   (a) Draw the structure of docosahexaenoic acid (all double bonds are *cis*).
   (b) Classify docosahexaenoic acid as an omega-3, omega-6, or omega-9 fatty acid.
6. Many snake venoms contain phospholipase A_2 that catalyzes the degradation of glycerophospholipids into a fatty acid and a "lysolecithin." The amphipathic nature of lysolecithins allows them to act as detergents in disrupting the membrane structure of red blood cells, causing them to rupture. Draw the structures of phosphatidyl serine (PS) and the products (including a lysolecithin) that result from the reaction of PS with phospholipase A_2.
7. Draw the structures of the following membrane lipids:
   (a) 1-stearoyl-2-oleoyl-3-phosphatidylethanolamine
   (b) palmitoylsphingomyelin
   (c) myristoyl-β-D-glucocerebroside.
8. (a) The steroid cortisol participates in the control of carbohydrate, protein, and lipid metabolism. Cortisol is derived from cholesterol and possesses the same four-membered fused ring system but with: (1) a C-3 keto group, (2) C-4-C-5 double bond (instead of the C-5-C-6 as in cholesterol), (3) a C-11 hydroxyl, and (4) a hydroxyl group and a $—C(O)CH_2OH$ group at C-17. Draw the structure of cortisol.
   (b) Ouabain is a member of the cardiac glycoside family found in plants and animals. This steroid inhibits $Na^{\oplus}$–$K^{\oplus}$ ATPase and ion transport and may be involved in hypertension and high blood pressure in humans. Ouabain possesses a four-membered fused ring system similar to cholesterol but has the following structural features: (1) no double bonds in the rings, (2) hydroxy groups on C-1, C-5, C-11, and C-14, (3) $—CH_2OH$ on C-19, (4) 2-3 unsaturated five-membered lactone ring on C-17 (attached to C-3 of lactone ring), and (5) 6-deoxymannose attached β-1 to the C-3 oxygen. Draw the structure of ouabain.
9. A consistent response in many organisms to changing environmental temperatures is the restructuring of cellular membranes. In some fish, phosphatidylethanolamine (PE) in the liver microsomal lipid membrane contains predominantly docosahexaenoic acid, 22:6 $\Delta^{4,7,10,13,16,19}$ at C-2 of the glycerol-3-phosphate backbone and then either a saturated or monounsaturated fatty acyl group at C-1. The percentage of the PE containing saturated or monounsaturated fatty acyl groups was determined in fish acclimated at 10°C or 30°C. At 10°C, 61% of the PE molecules contained saturated fatty acyl groups at C-1, and 39% of the PE molecules contained monounsaturated fatty acyl groups at C-1. When fish were acclimated to 30°C, 86% of the PE lipids contained saturated fatty acyl groups at C-1, while 14% of the PE molecules had monounsaturated acyl groups at C-1 [Brooks, S., Clark, G.T., Wright, S.M., Trueman, R.J., Postle, A.D., Cossins, A.R., and Maclean, N.M. (2002). Electrospray ionisation mass spectrometric analysis of lipid restructuring in the carp (*Cyprinus carpio* L.) during cold acclimation. *J. Exp. Biol.* 205:3989–3997]. Explain the purpose of the membrane restructuring observed with the change in environmental temperature.
10. A mutant gene (*ras*) is found in as many as one-third of all human cancers including lung, colon, and pancreas, and may be partly responsible for the altered metabolism in tumor cells. The *ras* protein coded for by the *ras* gene is involved in cell signaling pathways that regulate cell growth and division. Since the *ras* protein must be converted to a lipid anchored membrane protein in order to have cell-signaling activity, the enzyme farnesyl transferase (FT) has been selected as a potential chemotherapy target for inhibition. Suggest why FT might be a reasonable target.
11. Glucose enters some cells by simple diffusion through channels or pores, but glucose enters red blood cells by passive transport. On the plot below, indicate which line represents diffusion through a channel or pore and which represents passive transport. Why do the rates of the two processes differ?



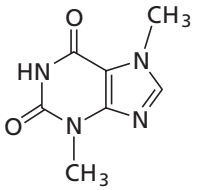
12. The pH gradient between the stomach (pH 0.8–1.0) and the gastric mucosal cells lining the stomach (pH 7.4) is maintained by an $H^{\oplus}$–$K^{\oplus}$ ATPase transport system that is similar to the ATP-driven $Na^{\oplus}$–$K^{\oplus}$ ATPase transport system (Figure 9.38). The $H^{\oplus}$–$K^{\oplus}$ ATPase antiport system uses the energy of ATP to pump $H^{\oplus}$ out of the mucosal cells (mc) into the stomach (st) in exchange for $K^{\oplus}$ ions. The $K^{\oplus}$ ions that are transported into the mucosal cells are then cotransported back into the stomach along

with $Cl^{\ominus}$ ions. The net transport is the movement of HCl into the stomach.

$$K^{\oplus}_{(mc)} + Cl^{\ominus}_{(mc)} + H^{\oplus}_{(mc)} + K^{\oplus}_{(st)} + ATP \rightleftharpoons$$
$$K^{\oplus}_{(st)} + Cl^{\ominus}_{(st)} + H^{\oplus}_{(st)} + K^{\oplus}_{(mc)} + ADP + P_i$$

Draw a diagram of this $H^{\oplus}$–$K^{\oplus}$ ATPase system.

13. Chocolate contains the compound theobromine, which is structurally related to caffeine and theophylline. Chocolate products may be toxic or lethal to dogs because these animals metabolize theobromine more slowly than humans. The heart, central nervous system, and kidneys are affected. Early signs of theobromine poisoning in dogs include nausea and vomiting, restlessness, diarrhea, muscle tremors, and increased urination or incontinence. Comment on the mechanism of toxicity of theobromine in dogs.



Theobromine

14. In the inositol signaling pathway, both IP_3 and diacylglycerol (DAG) are hormonal second messengers. If certain protein kinases in cells are activated by binding Ca^{2+}, how do IP_3 and DAG act in a complementary fashion to elicit cellular responses inside cells?

15. In some forms of diabetes, a mutation in the β subunit of the insulin receptor abolishes the enzymatic activity of that subunit. How does the mutation affect the cell's response to insulin? Can additional insulin (e.g., from injections) overcome the defect?

16. The *ras* protein (described in Problem 10) is a mutated G protein that lacks GTPase activity. How does the absence of this activity affect the adenylyl cyclase signaling pathway?

17. At the momentof fertilization a female egg is about 100μm in diameter. Assuming that each lipid molecule in the plasma membrane has a suface area of 10^{-14} cm^2, how many lipid molecules are there in the egg plasma membrane if 25% of the surface is protein?

18. Each fertilized egg cell (zygote) divides 30 times to produce all the eggs that a female child will need in her lifetime. One of these eggs will be fertilized giving rise to a new generation. If lipid molecles are never degraded, how many lipid molecules have you inherited that were synthesized in your grandmother?

Selected Readings

General

Gurr, M. I., and Harwood, J. L. (1991). *Lipid Biochemistry: An Introduction*, 4th ed. (London: Chapman and Hall).

Lester, D. R., Ross, J. J., Davies, P. J., and Reid, J. B. (1997). Mendel's stem length gene (*Le*) encodes a gibberellin 3 beta-hydroxylase. *Plant Cell.* 9:1435–1443.

Vance, D. E., and Vance, J. E., eds. (2008). *Biochemistry of Lipids, Lipoproteins, and Membranes*, 5th ed. (New York: Elsevier).

Membranes

Dowhan, W. (1997). Molecular basis for membrane phospholipid diversity: why are there so many lipids? *Annu. Rev. Biochem.* 66:199–232.

Jacobson, K., Sheets, E. D., and Simson, R. (1995). Revisiting the fluid mosaic model of membranes. *Science* 268:1441–1442.

Koga, Y., and Morii, H. (2007). Biosynthesis of ether-type polar lipids in Archaea and evolutionary considerations. *Microbiol. and Molec. Biol. Rev.* 71: 97–120.

Lai, E.C. (2003) Lipid rafts make for slippery platforms. *J. Cell Biol.* 162:365–370.

Lingwood, D., and Simons, K. (2010). Lipid rafts as a membrane-organizing principle. *Science.* 327:46–50.

Simons, K., and Ikonen, E. (1997). Functional rafts in cell membranes. *Nature.* 387:569–572.

Singer, S. J. (1992). The structure and function of membranes: a personal memoir. *J. Membr. Biol.* 129:3–12.

Singer, S. J. (2004) Some early history of membrane molecular biology. *Annu. Rev. Physiol.* 66:1–27.

Singer, S. J., and Nicholson, G. L. (1972). The fluid mosaic model of the structure of cell membranes. *Science* 175:720–731.

Membrane Proteins

Casey, P. J., and Seabra, M. C. (1996). Protein prenyltransferases. *J. Biol. Chem.* 271:5289–5292.

Bijlmakers, M-J., and Marsh, M. (2003). The on-off story of protein palmitoylation. *Trends in Cell Biol.* 13:32–42.

Elofsson, A., and von Heijne, G. (2007). Membrane protein structure: prediction versus reality. *Annu. Rev. Biochem.* 76:125–140.

Membrane Transport

Borst, P., and Elferink, R. O. (2002). Mammalian ABC transporters in health and disease. *Annu. Rev. Biochem.* 71:537–592.

Caterina, M. J., Schumacher, M. A., Tominaga, M., Rosen, T. A., Levine, J. D., and Julius, D. (1997). The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* 389:816–824.

Clapham, D. (1997). Some like it hot: spicing up ion channels. *Nature* 389:783–784.

Costanzo, M. et. al. (2010). The genetic landscape of a cell. *Science* 327:425–432.

Doherty, G. J. and McMahon, H. T. (2009). Mechanisms of endocytosis. *Annu. Rev. Biochem.* 78:857–902.

Doyle, D. A., Cabral, J. M., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and McKinnon, R. (1998). The structure of the potassium channel: molecular basis of $K^{\oplus}$ conduction and selectivity. *Science* 280:69–75.

Jahn, R., and Südhof, T. C. (1999). Membrane fusion and exocytosis. *Annu. Rev. Biochem.* 68:863–911.

Kaplan, J. H. (2002). Biochemistry of Na, K-AT-Pase. *Annu. Rev. Biochem.* 71:511–535.

Loo, T. W., and Clarke, D. M. (1999). Molecular dissection of the human multidrug resistance P-glycoprotein. *Biochem. Cell Biol.* 77:11–23.

Signal Transduction

Fantl, W. J., Johnson, D. E., and Williams, L. T. (1993). Signalling by receptor tyrosine kinases. *Annu. Rev. Biochem.* 62:453–481.

Hamm, H. E. (1998). The many faces of G protein signaling. *J. Biol. Chem.* 273:669–672.

Hodgkin, M. N., Pettitt, T. R., Martin, A., Michell, R. H., Pemberton, A. J., and Wakelam, M. J. O. (1998). Diacylglycerols and phosphatidates: which molecular species are intracellular messengers? *Trends Biochem. Sci.* 23:200–205.

Hurley, J. H. (1999). Structure, mechanism, and regulation of mammalian adenylyl cyclase. *J. Biol. Chem.* 274:7599–7602.

Luberto, C., and Hannun, Y. A. (1999). Sphingolipid metabolism in the regulation of bioactive molecules. *Lipids* 34 (Suppl.):S5–S11.

Prescott, S. M. (1999). A thematic series on kinases and phosphatases that regulate lipid signaling. *J. Biol. Chem.* 274:8345.

Shepherd, P. R., Withers, D. J., and Siddle, K. (1998). Phosphoinositide 3-kinase: the key switch mechanism in insulin signalling. *Biochem. J.* 333:471–490.