CHAPTER 4

The Three-Dimensional Structure of Proteins

Red blood cells contain hemoglobin, a classic example of protein structure.

Amino acids joined together form a protein (polypeptide) chain. The repeating units are amide planes containing peptide bonds. These amide planes can twist about their connecting carbon atoms to create the three-dimensional conformations of proteins. More than 50 years ago, Linus Pauling predicted that linked amino acids could form an α-helix. Years later, his prediction was confirmed when myoglobin, an oxygen-binding protein, was found to be made from Pauling's α-helices. This type of local folding of the protein chain is called secondary structure, the linear sequence being the primary structure. The conformation of a complete protein chain is its tertiary structure. Myoglobin, a molecule that binds oxygen tightly, has a single protein chain. Hemoglobin, a protein with four myoglobin-like subunits fitted together, has a quaternary structure. This allows it to change from the oxy conformation, when it binds oxygen in the lungs, to the deoxy form, when it releases oxygen to working tissues. The discovery of structure–function relationships in hemoglobin led to an understanding of the way complex multisubunit enzymes regulate metabolic pathways.

Critical Questions

BiochemistryNow™
Test yourself on these Critical Questions at the BiochemistryNow website at **http://now.brookscole.com/campbell5**

4.1 How Does the Structure of Proteins Determine Their Function?

Levels of Structure in Proteins

Biologically active proteins are polymers consisting of amino acids linked by covalent peptide bonds. Many different conformations (three-dimensional structures) are possible for a molecule as large as a protein. Of these many structures, one or (at most) a few have biological activity; these are called the **native conformations.** Many proteins have no obvious regular repeating structure. As a consequence, these proteins are frequently described as having large segments of "random structure" (also referred to as random coil). The term "random" is really a misnomer, since the same nonrepeating structure is found in the native conformation of all molecules of a given protein, and this conformation is needed for its proper function. Because proteins are complex, they are defined in terms of four levels of structure.

Primary structure is the order in which the amino acids are covalently linked together. The peptide Leu—Gly—Thr—Val—Arg—Asp—His (recall that the N-terminal amino acid is listed first) has a different primary structure from the peptide Val—His—Asp—Leu—Gly—Arg—Thr, even though both have the same number and kinds of amino acids. Note that the order of amino acids can be written on one line. The primary structure is the one-dimensional first step in specifying the three-dimensional structure of a protein. Some biochemists define primary structure to include all covalent interactions, including the disulfide bonds that can be formed by cysteines; however, we shall consider the disulfide bonds to be part of the tertiary structure, which will be considered later.

Two three-dimensional aspects of a single polypeptide chain, called the secondary and tertiary structure, can be considered separately. **Secondary structure** is the arrangement in space of the atoms in the peptide backbone. The

α-helix and β-pleated sheet arrangements are two different types of secondary structure. Secondary structures have repetitive interactions resulting from hydrogen bonding between the amide N—H and the carbonyl groups of the peptide backbone. The conformations of the side chains of the amino acids are not part of the secondary structure. In many proteins, the folding of parts of the chain can occur independently of the folding of other parts. Such independently folded portions of proteins are referred to as **domains** or **supersecondary structure.**

Tertiary structure includes the three-dimensional arrangement of all the atoms in the protein, including those in the side chains and in any **prosthetic groups** (groups of atoms other than amino acids).

A protein can consist of multiple polypeptide chains called **subunits.** The arrangement of subunits with respect to one another is the **quaternary structure.** Interaction between subunits is mediated by noncovalent interactions, such as hydrogen bonds, electrostatic attractions, and hydrophobic interactions.

We shall discuss secondary structure in more detail in Section 4.3, tertiary structure in Section 4.5, and quaternary structure in Section 4.7.

4.2 What Is the Primary Structure of Proteins?

The amino acid sequence (the primary structure) of a protein determines its three-dimensional structure, which, in turn, determines its properties. In every protein, the correct three-dimensional structure is needed for correct functioning.

One of the most striking demonstrations of the importance of primary structure is found in the hemoglobin associated with *sickle-cell anemia.* In this genetic disease, red blood cells cannot bind oxygen efficiently. The red blood cells also assume a characteristic sickle shape, giving the disease its name. The sickled cells tend to become trapped in small blood vessels, cutting off circulation and thereby causing organ damage. These drastic consequences stem from a change in one amino acid residue in the sequence of the primary structure.

Considerable research is being done to determine the effects of changes in primary structure on the functions of proteins. Using molecular-biology techniques, such as site-directed mutagenesis (Section 14.7), it is possible to replace any chosen amino acid residue in a protein with another specific amino acid residue. The conformation of the altered protein, as well as its biological activity, can then be determined. The results of such amino acid substitutions range from negligible effects to complete loss of activity, depending on the protein and the nature of the altered residue.

Determining the sequence of amino acids in a protein is a routine, but not trivial, operation in classical biochemistry. It consists of several steps, which must be carried out carefully to obtain accurate results (Section 5.4).

The following Biochemical Connections box describes an important practical aspect of the amino acid composition of proteins. This property can differ markedly, depending on the source of the protein (plant or animal), with important consequences for human nutrition.

Essential Information

The primary structure of a protein is the sequence of amino acids. Determination of the sequence involves cleaving the protein to smaller peptides, determining the sequence of the individual peptides, and combining the peptide sequences to obtain that of the protein.

4.3 What Is the Secondary Structure of Proteins?

The secondary structure of proteins is the hydrogen-bonded arrangement of the backbone of the protein, the polypeptide chain. The nature of the bonds in the peptide backbone plays an important role here. Within each amino acid residue are two bonds with reasonably free rotation. They are (1) the

Biochemical Connections

Complete Proteins and Nutrition

A **complete protein** is one that provides all essential amino acids (Section 23.5) in appropriate amounts for human survival. These amino acids cannot be synthesized by humans, but they are needed for the biosynthesis of proteins. Lysine and methionine are two essential amino acids that are frequently in short supply in plant proteins.

Because grains such as rice and corn are usually poor in lysine, and because beans are usually poor in methionine, vegetarians are at risk for malnutrition unless they eat grains and beans together. This leads to the concept of *complementary proteins,* mixtures that provide all the essential amino acids—for example, corn and beans in *succotash,* or a bean burrito made with a corn tortilla. The specific recommended dietary allowances for adult males follow. Adult females who are neither pregnant nor lactating need 20% less than the amounts indicated for adult males.

	RDA		RDA
Arg*	Unknown	Met	0.70 g
His*	Unknown	Phe	1.12 g (includes Tyr)
Ile	0.84 g	Thr	0.56 g
Leu	1.12 g	Trp	0.21 g
Lys	0.84 g	Val	0.96 g

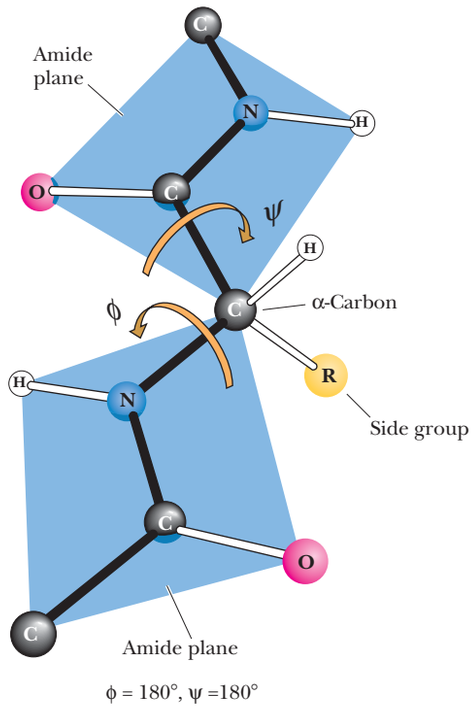
*The inclusion of His and Arg is controversial. They appear to be required only by growing children and for the repair of injured tissue. Arg is required to maintain fertility in males.

The *p*rotein *e*fficiency *r*atio (PER) describes how well a protein supplies essential amino acids. This parameter is useful for deciding how much of a food you need to eat. Most college-age, nonpregnant females require 46 g (or about 1.6 oz) of complete protein, and males require 58 g (or about 2 oz) of complete protein per day. If one chooses to pick only a *single* source of protein for the diet, eggs are perhaps the best choice because they contain high-quality protein. For a female, the need for 1.6 oz of complete protein could be met with 10.7 oz of eggs, or about four whole extra-large eggs. For a male, 13.6 oz of eggs, or a little more than five eggs, would be needed. The same requirement could be met with a lean beef steak, but it would require 345 g, or about 0.75 lb, for a female (or 431 g, or nearly a full pound, for a male) because beef steak has a lower PER. If one ate only corn, it would require 1600 g/day for women and 2000 g/day for men (1600 g is about 3.6 pounds of fresh corn kernels—something in excess of 160 eight-inch ears per day). However, if you simply combine a small amount of beans or peas with the corn, it complements the low amount of lysine in the corn, and the protein is now complete. This can easily be done with normal food portions.

Protein	PER	% Protein
Whole egg	100	15
Beef muscle	84	16
Cow's milk	66	4 (largely H_2O)
Peanuts	45	28
Corn	32	9
Wheat	26	12

In an attempt to increase the nutritional value of certain crops that are grown as food for livestock, scientists have used genetic techniques to create strains of corn that are much higher in lysine than the wild-type corn. This has proven effective in increasing growth rates in pigs. Many vegetable crops are now being produced using biotechnology to increase shelf life, decrease spoilage, and give crops defenses against insects. These genetically modified foods are currently a hot spot of debate and controversy.

bond between the α-carbon and the amino nitrogen of that residue and (2) the bond between the α-carbon and the carboxyl carbon of that residue. The combination of the planar peptide group and the two freely rotating bonds has important implications for the three-dimensional conformations of peptides and proteins. A peptide-chain backbone can be visualized as a series of playing cards, each card representing a planar peptide group. The cards are linked at opposite corners by swivels, representing the bonds about which there is considerable freedom of rotation (Figure 4.1). The side chains also play a vital role in determining the three-dimensional shape of a protein, but only the backbone is considered in the secondary structure. The angles ϕ (phi) and ψ (psi), frequently called Ramachandran angles (after their originator, G. N. Ramachandran), are used to designate rotations around the C—N and C—C bonds, respectively. The conformation of a protein backbone can be described by specifying the values of ϕ and ψ for each residue (−180° to 180°). Two kinds of secondary structures that occur frequently in proteins are the repeating **α-helix** and **β-pleated sheet** (or β-sheet) hydrogen-



◀ **FIGURE 4.1** Definition of the angles that determine the conformation of a polypeptide chain. The rigid planar peptide groups (called "playing cards" in the text) are shaded. The angle of rotation around the C^{α}—N bond is designated ϕ (phi), and the angle of rotation around the C^{α}—C bond is designated ψ(psi). These two bonds are the ones around which there is freedom of rotation. (*Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.*)

bonded structures. The ϕ and ψ angles repeat themselves in contiguous amino acids in regular secondary structures. The α-helix and β-pleated sheet are not the only possible secondary structures, but they are by far the most important and deserve a closer look.

Essential Information

Two of the most important structural motifs in proteins are the α-helix and β-pleated sheet.

Periodic Structures in Protein Backbones

The α-helix and β-pleated sheet are periodic structures; their features repeat at regular intervals. The α-helix is rodlike and involves only one polypeptide chain. The β-pleated sheet structure can give a two-dimensional array and can involve one or more polypeptide chains.

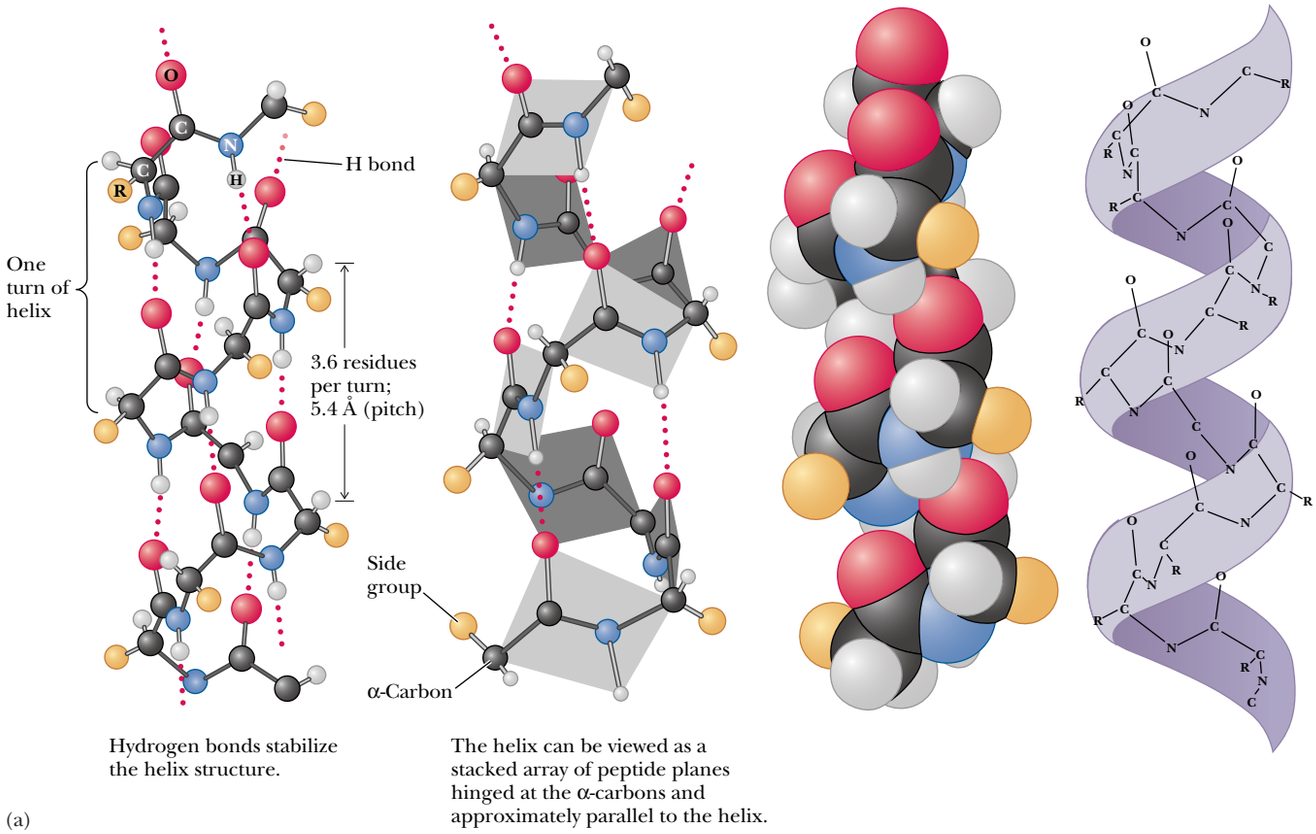
The α-Helix

The α-helix is stabilized by hydrogen bonds parallel to the helix axis within the backbone of a single polypeptide chain. Counting from the N-terminal end, the C—O group of each amino acid residue is hydrogen bonded to the N—H group of the amino acid four residues away from it in the covalently bonded sequence. The helical conformation allows a linear arrangement of the atoms involved in the hydrogen bonds, which gives the bonds maximum strength and thus makes the helical conformation very stable (Section 2.2). There are 3.6 residues for each turn of the helix, and the *pitch* of the helix (the linear distance between corresponding points on successive turns) is 5.4 Å (Figure 4.2).

The angstrom unit, 1 Å = 10^{-8} cm = 10^{-10} m, is convenient for interatomic distances in molecules, but it is not a Système International [SI] unit. Nanometers (1 nm = 10^{-9} m) and picometers (1 pm = 10^{-12} m) are the SI units used for interatomic distances. In SI units, the pitch of the α-helix is 0.54 nm or 540 pm.). Figure 4.3 shows the structures of two proteins with a high degree of α-helical content.

Biochemistry Now™

Go to BiochemistryNow and click on Biochemistry Interactive to explore the anatomy of the α-helix.



Jane and David Richardson, Dept. of Biochem., Duke Medical Center, NC

▲ **FIGURE 4.2** The α-helix. (a) From left to right, ball-and-stick model of the α-helix, showing terminology; ball-and-stick model with planar peptide groups shaded; computer-generated space-filling model of the α-helix; outline of the α-helix. (b) Model of the protein hemoglobin, showing the helical regions. (*Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.*)

Proteins have varying amounts of α-helical structures, varying from a few percent to nearly 100%. Several factors can disrupt the α-helix. The amino acid proline creates a bend in the backbone because of its *cyclic* structure. It cannot fit into the α-helix because (1) rotation around the bond between the nitrogen and the α-carbon is severely restricted, and (2) proline's α-amino group cannot participate in intrachain hydrogen bonding. Other localized factors involving the side chains include strong electrostatic repulsion owing to the proximity of several charged groups of the same sign, such as groups of positively charged lysine and arginine residues or groups of negatively

β-Hemoglobin subunit

Myohemerythrin

Biochemistry Now™ ANIMATED FIGURE 4.3
The three-dimensional structure of two proteins with substantial amounts of α-helix in their structures. The helices are represented by the regularly coiled sections of the ribbon diagram. Myohemerythrin is an oxygen-carrying protein in invertebrates. **See this figure animated at http://now.brookscole.com/campbell5** (*Jane Richardson.*)

charged glutamate and aspartate residues. Another possibility is crowding (steric repulsion) caused by the proximity of several bulky side chains. In the α-helical conformation, all the side chains lie outside the helix; there is not enough room for them in the interior. The β-carbon is just outside the helix, and crowding can occur if it is bonded to two atoms other than hydrogen, as is the case with valine, isoleucine, and threonine.

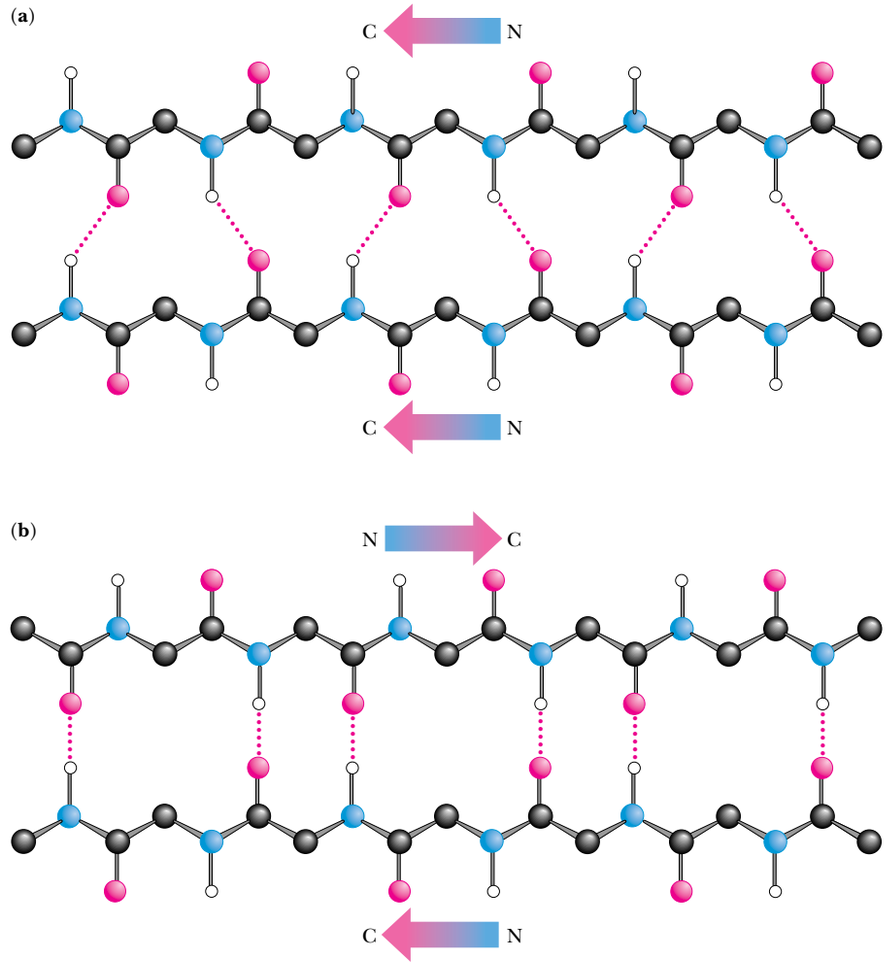
The β-Sheet

The arrangement of atoms in the β-pleated sheet conformation differs markedly from that in the α-helix. The peptide backbone in the β-sheet is almost completely extended. Hydrogen bonds can be formed between different parts of a single chain that is doubled back on itself *(intrachain bonds)* or between different chains *(interchain bonds)*. If the peptide chains run in the same direction (i.e., if they are all aligned in terms of their N-terminal and C-terminal ends), a *parallel pleated sheet* is formed. When alternating chains run in opposite directions, an *antiparallel* pleated sheet is formed (Figure 4.4). The hydrogen bonding between peptide chains in the β-pleated sheet gives rise to a repeated zigzag structure; hence, the name "pleated sheet" (Figure 4.5). Note that the hydrogen bonds are perpendicular to the direction of the protein chain, not parallel to it as in the α-helix.

Biochemistry Now™
Go to BiochemistryNow and click on Biochemistry Interactive to explore β-sheets, one of the principal types of secondary structure in proteins.

Irregularities in Regular Structures

Other helical structures are found in proteins. These are often found in shorter stretches than with the α-helix, and they sometimes break up the regular nature of the α-helix. The most common is the 3_{10} helix, which has three residues per turn and ten atoms in the ring formed by making the hydrogen bond. Other common helices are designated 2_7 and 4.4_{16}, following the same nomenclature as the 3_{10} helix.



▶ **FIGURE 4.4** The arrangement of hydrogen bonds in (a) parallel and (b) antiparallel β-pleated sheets.

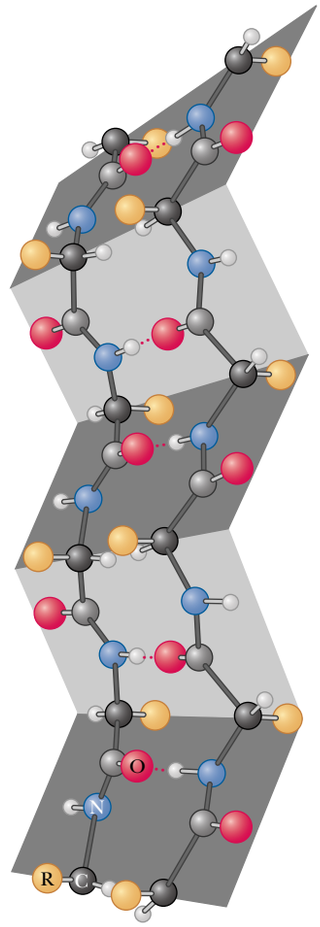
Biochemistry Now™
Go to BiochemistryNow and click on Biochemistry Interactive to explore discover the features of β-turns and how they change the direction of a polypeptide strand.

A **β-bulge** is a common nonrepetitive irregularity found in antiparallel β-sheets. It occurs between two normal β-structure hydrogen bonds and involves two residues on one strand and one on the other. Figure 4.6 shows typical β-bulges.

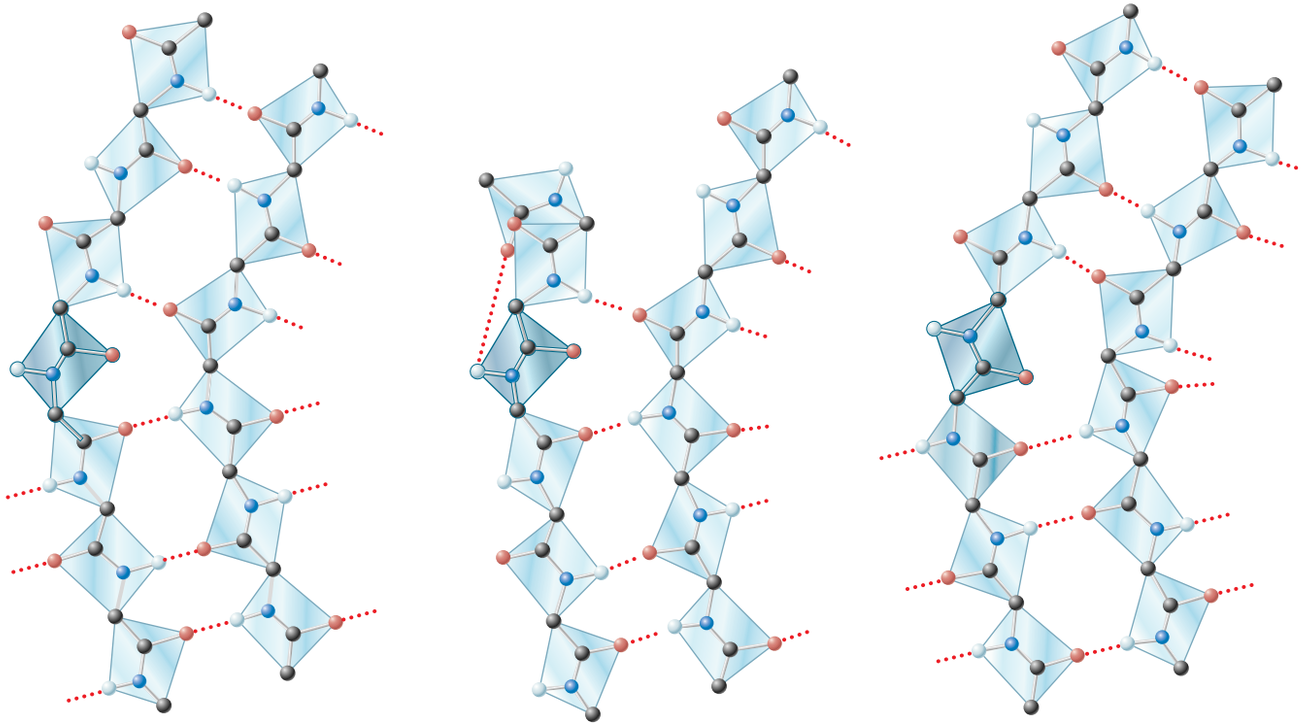
Protein folding requires that the peptide backbones and the secondary structures be able to change directions. Often a reverse turn marks a transition between one secondary structure and another. For steric (spatial) reasons, glycine is frequently encountered in **reverse turns,** at which the polypeptide chain changes direction; the single hydrogen of the side chain prevents crowding (Figures 4.7a and 4.7b). Because the cyclic structure of proline has the correct geometry for a reverse turn, this amino acid is also frequently encountered in such turns (Figure 4.7c).

Supersecondary Structures and Domains

The α-helix, β-pleated sheet, and other secondary structures are combined in many ways as the polypeptide chain folds back on itself in a protein. The combination of α- and β-strands produces various kinds of supersecondary structures in proteins. The most common feature of this sort is the βαβ *unit,* in which two parallel strands of β-sheet are connected by a stretch of α-helix (Figure 4.8a). An αα *unit* (helix-turn-helix) consists of two antiparallel α-helices

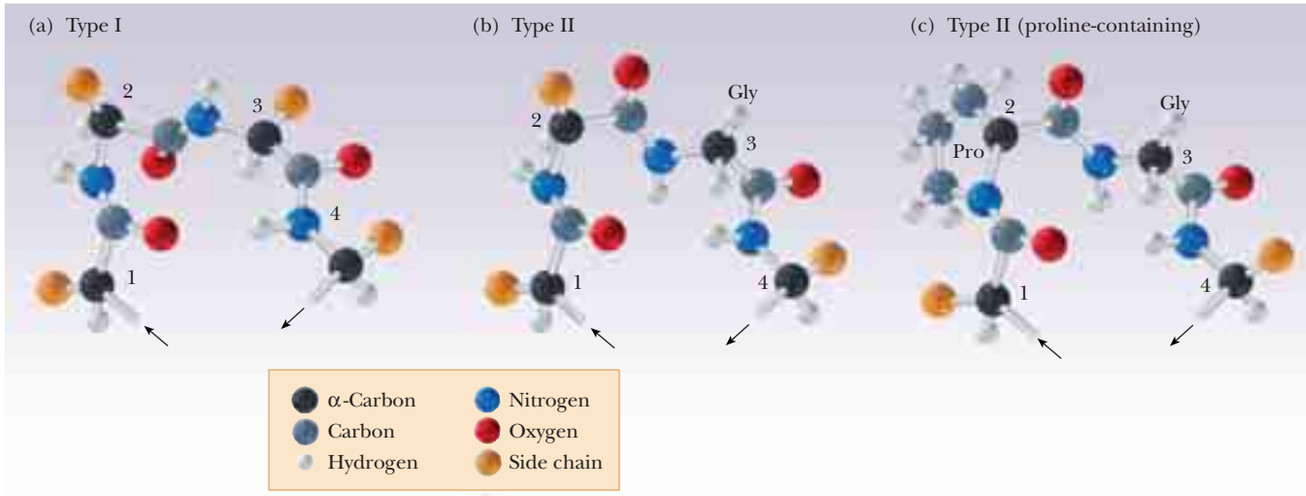


◀ **FIGURE 4.5** The three-dimensional form of the antiparallel β-pleated sheet arrangement. The chains do not fold back on each other but are in a fully extended conformation. (*Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.*)

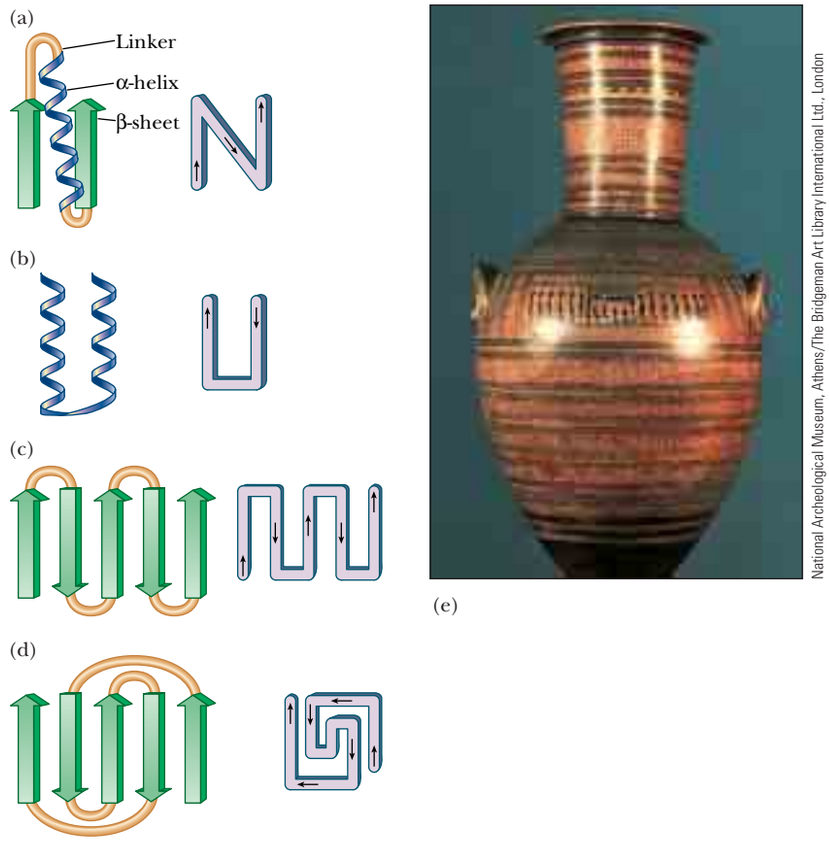


Classic bulge G-1 bulge Wide bulge

▲ **FIGURE 4.6** Three different β-bulge structures. Hydrogen bonds are shown as red dots.



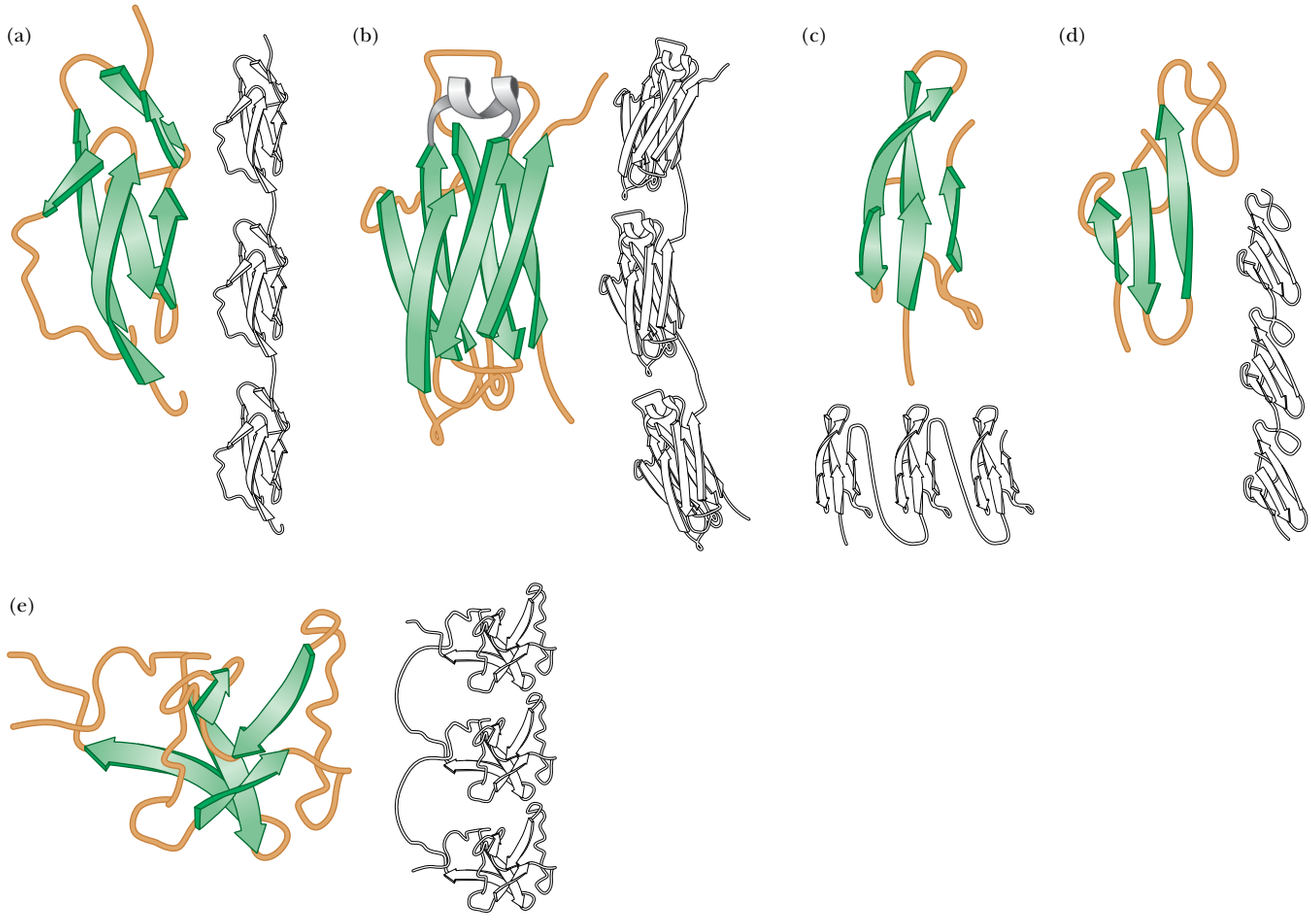
▲ **FIGURE 4.7** Structures of reverse turns. Arrows indicate the directions of the polypeptide chains. (a) A type I reverse turn. In residue 3, the side chain (gold) lies outside the loop, and any amino acid can occupy this position. (b) A type II reverse turn. The side chain of residue 3 has been rotated 180° from the position in the type I turn and is now on the inside of the loop. Only the hydrogen side chain of glycine can fit into the space available, so glycine must be the third residue in a type II reverse turn. (c) The five-membered ring of proline has the correct geometry for a reverse turn; this residue normally occurs as the second residue of a reverse turn. The turn shown here is type II, with glycine as the third residue.



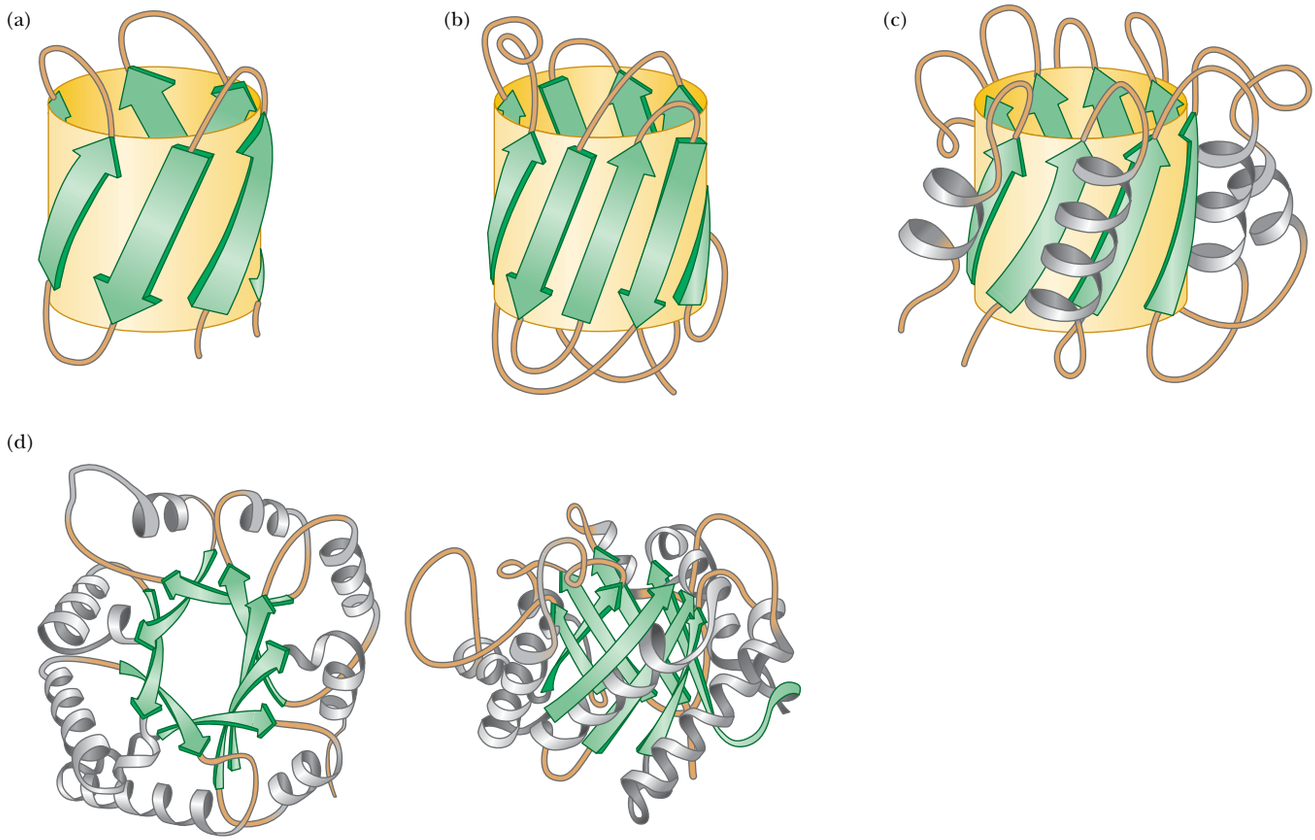
▲ **FIGURE 4.8** Schematic diagrams of supersecondary structures. Arrows indicate the directions of the polypeptide chains. (a) A βαβ unit, (b) an αα unit, (c) a β-meander, and (d) the Greek key. (e) The Greek key motif in protein structure resembles the geometric patterns on this ancient Greek vase, giving rise to the name.

(Figure 4.8b). In such an arrangement, energetically favorable contacts exist between the side chains in the two stretches of helix. In a β-*meander,* an antiparallel sheet is formed by a series of tight reverse turns connecting stretches of the polypeptide chain (Figure 4.8c). Another kind of antiparallel sheet is formed when the polypeptide chain doubles back on itself in a pattern known as the *Greek key,* named for a decorative design found on pottery from the classical period (Figure 4.8e). A **motif** is a repetitive supersecondary structure. Some of the common smaller motifs are shown in Figure 4.9. These smaller motifs can often be repeated and organized into larger motifs. Protein sequences that allow for a β-meander or Greek key can often be found arranged into a β-barrel in the tertiary structure of the protein (Figure 4.10). Motifs are important and tell us much about the folding of proteins. However, these motifs do not allow us to predict anything about the biological function of the protein because they are found in proteins and enzymes with very dissimilar functions.

Many proteins that have the same type of function have similar protein sequences; consequently, domains with similar conformations are associated with the particular function. Many types of domains have been identified, including three different types of domains by which proteins bind to DNA.



▲ **FIGURE 4.9** Motifs are repeated supersecondary structures, sometimes called modules. (a) The complement-control protein module. (b) The immunoglobulin module. (c) The fibronectin type I module. (d) The growth-factor module. (e) The kringle module. All of these have a particular secondary structure that is repeated in the protein. (*Reprinted from "Protein Modules,"* Trends in Biochemical Sciences, *Vol. 16, p. 13–17, Copyright © 1991, with permission from Elsevier.*)



▲ **FIGURE 4.10** Some β-barrel arrangements. (a) A linked series of β-meanders. This arrangement occurs in the protein rubredoxin from *Clostridium pasteurianum*. (b) The Greek key pattern occurs in human prealbumin. (c) A β-barrel involving alternating βαβ units. This arrangement occurs in triose phosphate isomerase from chicken muscle. (d) Top and side views of the polypeptide backbone arrangement in triose phosphate isomerase. Note that the α-helical sections lie outside the actual β-barrel.

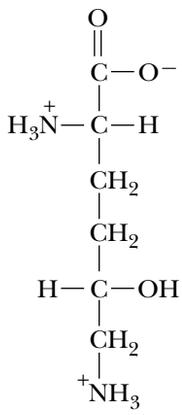
In addition, short polypeptide sequences within a protein direct the post-translational modification and subcellular localization. For example, several sequences play a role in the formation of glycoproteins (ones that contain sugars in addition to the polypeptide chain). Other specific sequences indicate that a protein is to be bound to a membrane or secreted from the cell. Still other specific sequences mark a protein for phosphorylation by a specific enzyme.

The Collagen Triple Helix

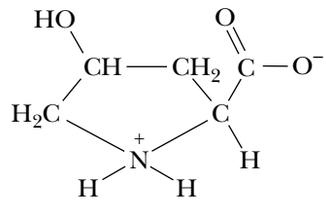
Collagen, a component of bone and connective tissue, is the most abundant protein in vertebrates. It is organized in water-insoluble fibers of great strength. A collagen fiber consists of three polypeptide chains wrapped around each other in a ropelike twist, or triple helix. Each of the three chains has, within limits, a repeating sequence of three amino acid residues, X—Pro—Gly or X—Hyp—Gly, where Hyp stands for hydroxyproline, and any amino acid can occupy the first position, designated by X.

Proline and hydroxyproline can constitute up to 30% of the residues in collagen. Hydroxyproline is formed from proline by a specific hydroxylating enzyme after the amino acids are linked together. Hydroxylysine also occurs in collagen. In the amino acid sequence of collagen, every third position must

be occupied by glycine. The triple helix is arranged so that every third residue on each chain is inside the helix. Only glycine is small enough to fit into the space available (Figure 4.11).



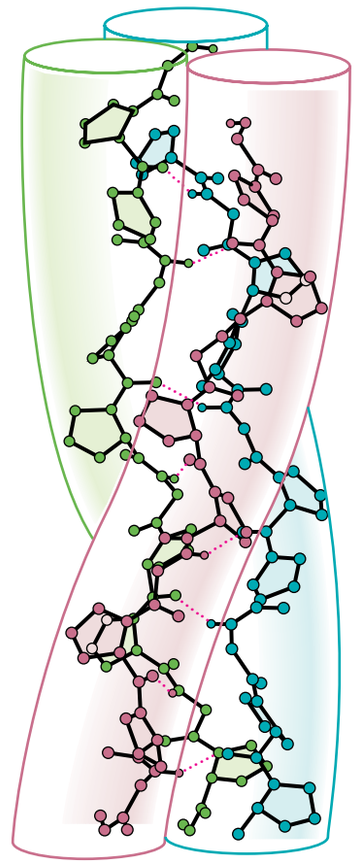
Hydroxylysine



Hydroxyproline

The three individual collagen chains are themselves helices that differ from the α-helix. They are twisted around each other in a superhelical arrangement to form a stiff rod. This triple helical molecule is called *tropocollagen;* it is 300 nm (3000 Å) long and 1.5 nm (15 Å) in diameter. The three strands are held together by hydrogen bonds involving the hydroxyproline and hydroxylysine residues. The molecular weight of the triple-stranded array is about 300,000; each strand contains about 800 amino acid residues. Collagen is both intramolecularly and intermolecularly linked by covalent bonds formed by reactions of lysine and histidine residues. The amount of cross-linking in a tissue increases with age. That is why meat from older animals is tougher than meat from younger animals.

Collagen in which the proline is not hydroxylated to hydroxyproline to the usual extent is less stable than normal collagen. Symptoms of scurvy, such as bleeding gums and skin discoloration, are the results of fragile collagen. The enzyme that hydroxylates proline and thus maintains the normal state of collagen requires ascorbic acid (vitamin C) to remain active. Scurvy is ultimately caused by a dietary deficiency of vitamin C. See the Biochemical Connections box in Chapter 16.

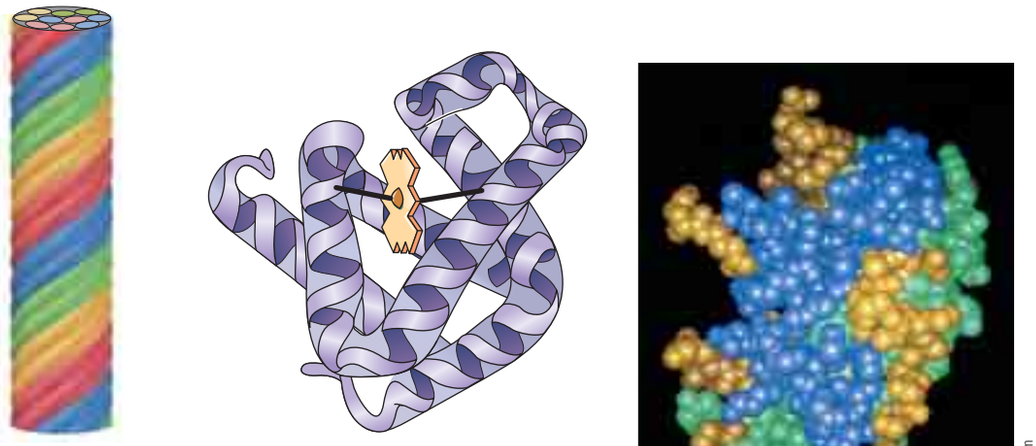


Biochemistry Now™ ACTIVE FIGURE 4.11
Poly (Gly—Pro—Pro), a collagen-like right-handed triple helix composed of three left-handed helical chains. (*Adapted from M. H. Miller and H. A. Scheraga, 1976, Calculation of the structures of collagen models. Role of interchain interactions in determining the triple-helical coiled-coil conformations. I. Poly(glycyl-prolyl-prolyl).* Journal of Polymer Science Symposium ***54:****171–200. © 1976 John Wiley & Sons, Inc. Reprinted by permission.*) **Watch this Active Figure at http://now.brookscole.com/campbell5**

Two Types of Protein Conformations: Fibrous and Globular

It is difficult to draw a clear separation between secondary and tertiary structures. The nature of the side chains in a protein (part of the tertiary structure) can influence the folding of the backbone (the secondary structure). Comparing collagen with silk and wool fibers can be illuminating. Silk fibers consist largely of the protein fibroin, which, like collagen, has a fibrous structure, but which, unlike collagen, consists largely of β-sheets. Fibers of wool consist largely of the protein keratin, which is largely α-helical. The amino acids of which collagen, fibroin, and keratin are composed determine which conformation they will adopt, but all are **fibrous proteins** (Figure 4.12a).

In other proteins, the backbone folds back on itself to produce a more or less spherical shape. These are called **globular proteins** (Figure 4.12b), and we shall see many examples of them. Their helical and pleated-sheet sections can be arranged so as to bring the ends of the sequence close to each other in three dimensions. Globular proteins, unlike fibrous proteins, are water-soluble and have compact structures; their tertiary and quaternary structures can be quite complex.



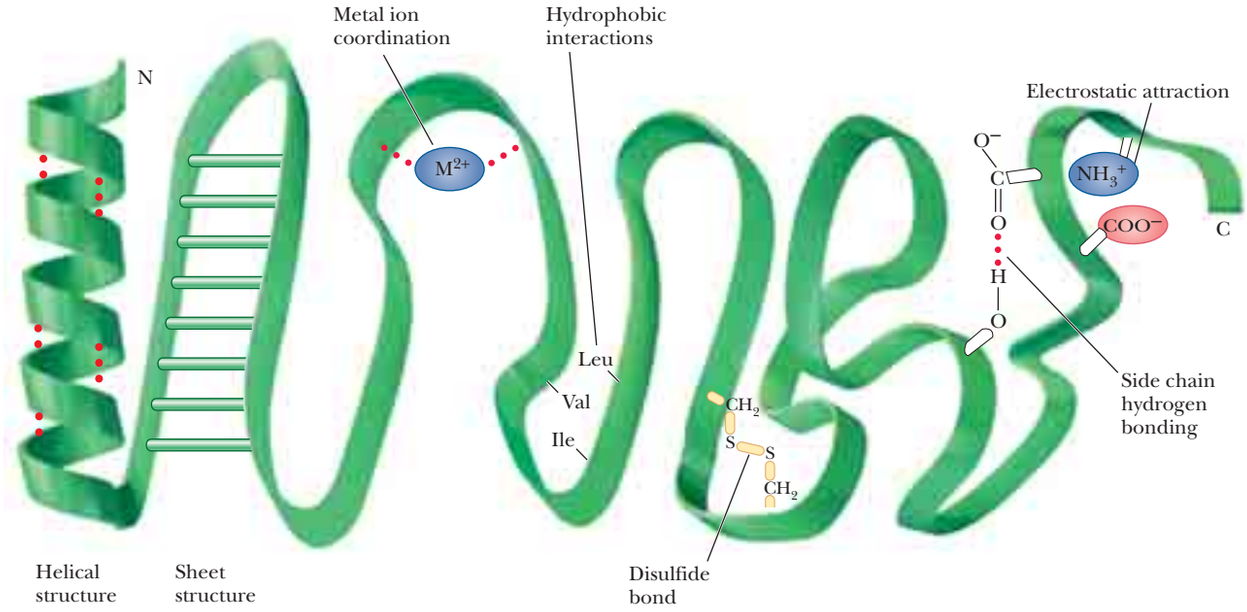
▲ **FIGURE 4.12** A comparison of the shapes of fibrous and globular proteins. (a) Schematic diagrams of a portion of a fibrous protein and of a globular protein. (b) Computer-generated model of a globular protein. The color-coding in this model differs from that of models of smaller molecules. The carbons are represented by light blue spheres, and the yellow spheres represent sulfur.

4.4 What Can We Say about the Thermodynamics of Protein Folding?

The primary structure of a protein—the order of amino acids in the polypeptide chain—depends on the formation of peptide bonds, which are covalent. Higher-order levels of structure, such as the conformation of the backbone (secondary structure) and the positions of all the atoms in the protein (tertiary structure), depend on noncovalent interactions; if the protein consists of several subunits, the interaction of the subunits (quaternary structure) also depends on noncovalent interactions. Noncovalent stabilizing forces contribute to the most stable structure for a given protein, the one with the lowest energy.

Several types of hydrogen bonding occur in proteins. *Backbone* hydrogen bonding is a major determinant of secondary structure; hydrogen bonds *between the side chains of amino acids* are also possible in proteins. Nonpolar residues tend to cluster together in the interior of protein molecules as a result of *hydrophobic* interactions. *Electrostatic* attraction between oppositely charged groups, which frequently occurs on the surface of the molecule, results in such groups being close to one another. Several side chains can be *complexed* to a single metal ion. (Metal ions also occur in some prosthetic groups.)

In addition to these noncovalent interactions, *disulfide bonds* form covalent links between the side chains of cysteines. When such bonds form, they restrict the folding patterns available to polypeptide chains. There are specialized laboratory methods for determining the number and positions of disulfide links in a given protein. Information about the locations of disulfide links can then be combined with knowledge of the primary structure to give the *complete covalent structure* of the protein. Note the subtle difference here: The primary structure is the order of amino acids, whereas the complete covalent structure also specifies the positions of the disulfide bonds (Figure 4.13).



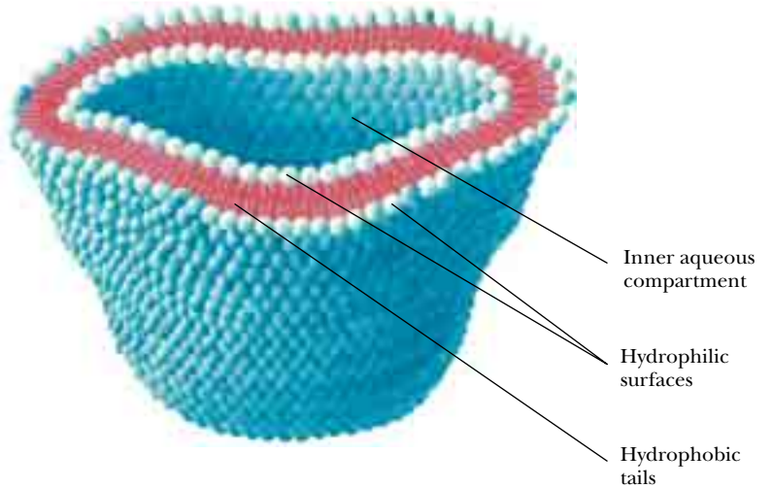
▲ **FIGURE 4.13** Forces that stabilize the tertiary structure of proteins. Note that the helical structure and sheet structure are two kinds of backbone hydrogen bonding. Although backbone hydrogen bonding is part of secondary structure, the conformation of the backbone puts constraints on the possible arrangement of the side chains.

Recall that, as a result of this assortment of stabilizing forces, residues that are far apart in the primary sequence can be close to each other in the three-dimensional structure produced by the folding of the protein. When a polypeptide chain folds back on itself, it can assume a compact globular shape. A different polypeptide chain (or the same chain under different conditions) can assume a rodlike fibrous form.

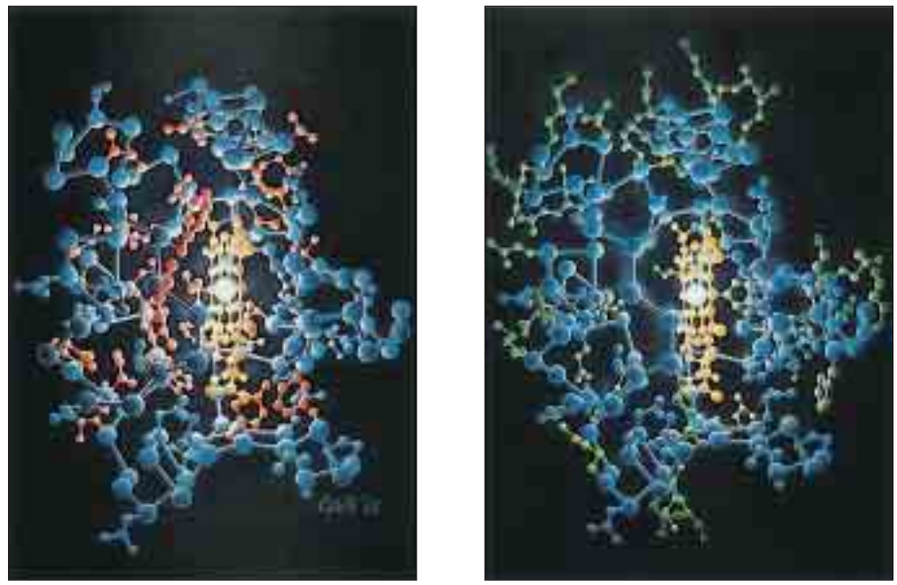
The most stable form of the protein is the one with the lowest energy, representing a complex interplay of all the forces involved. Many of these forces involve bond formation, frequently the formation of a large number of weak, noncovalent bonds. Of these, hydrophobic interactions are a special case in the sense that the concept of entropy plays a large role in describing them. This is a good place to take a detailed look at hydrophobic interactions.

Hydrophobic Interactions: A Case Study in Thermodynamics

Hydrophobic interactions have important consequences in biochemistry. Large arrays of molecules can take on definite structures as a result of hydrophobic interactions. We have already seen the way in which phospholipid bilayers can form one such array. Recall (Chapter 2, Section 2.1) that phospholipids are molecules that have polar head groups and long nonpolar tails of hydrocarbon chains. These bilayers are less complex than a folded protein, but the interactions that lead to their formation also play a vital role in protein folding. Under suitable conditions, a double-layer arrangement is formed so that the polar head groups of many molecules face the aqueous environment, while the nonpolar tails are in contact with each other and are kept away from the aqueous environment. These bilayers form three-dimensional structures called **liposomes** (Figure 4.14). Such structures are useful model systems for biological membranes, which consist of similar bilayers with proteins embedded in them. The interactions between the bilayer and the



▶ **FIGURE 4.14** Schematic diagram of a liposome. This three-dimensional structure is arranged so that hydrophilic head groups of lipids are in contact with the aqueous environment. The hydrophobic tails are in contact with each other and are kept away from the aqueous environment.



(a) (b)

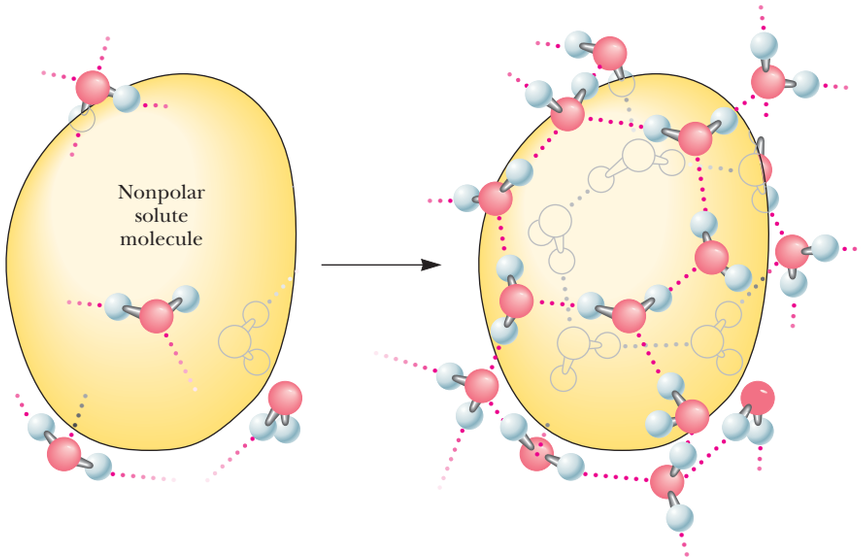
▶ **FIGURE 4.15** The three-dimensional structure of the protein cytochrome c. (a) The hydrophobic side chains (shown in red) are found in the interior of the molecule. (b) The hydrophilic side chains (shown in green) are found on the exterior of the molecule. (*Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.*)

embedded proteins are also examples of hydrophobic interactions. The very existence of membranes depends on hydrophobic interactions. The same hydrophobic interactions play a crucial role in protein folding.

Hydrophobic interactions are a major factor in the folding of proteins into the specific three-dimensional structures required for their functioning as enzymes, oxygen carriers, or structural elements. The order of amino acids (i.e., the nature of the side chains) automatically determines the three-dimensional structure of the protein. It is known experimentally that proteins tend to be folded so that the nonpolar hydrophobic side chains are sequestered from water in the interior of the protein, while the polar hydrophilic side chains lie on the exterior of the molecule and are accessible to the aqueous environment (Figure 4.15). What makes hydrophobic interactions favorable?

Hydrophobic interactions are spontaneous processes. The entropy of the universe increases when hydrophobic interactions occur.

$$\Delta S_{\text{universe}} > 0$$



BiochemistryNow™ ANIMATED FIGURE 4.16
A "cage" of water molecules forms around a nonpolar solute. **See this figure animated at http://now.brookscole.com/campbell5**

As an example, let us assume that we have tried to mix the liquid hydrocarbon hexane (C_6H_{14}) with water and have obtained not a solution but a two-layer system, one layer of hexane and one of water. Formation of a mixed solution is nonspontaneous, and the formation of two layers is spontaneous. Unfavorable entropy terms enter into the picture if solution formation requires the creation of ordered arrays of solvent, in this case water (Figure 4.16). The water molecules surrounding the nonpolar molecules can hydrogen bond with each other, but they have fewer possible orientations than if they were surrounded by other water molecules on all sides. This introduces a higher degree of order, preventing the dispersion of energy, more like the lattice of ice than liquid water, and thus a lower entropy. The required entropy decrease is too large for the process to take place. Therefore, nonpolar substances do not dissolve in water; rather, nonpolar molecules associate with one another by hydrophobic interactions and are excluded from water.

4.5 What Is the Tertiary Structure of Proteins?

The tertiary structure of a protein is the three-dimensional arrangement of all the atoms in the molecule. The conformations of the side chains and the positions of any prosthetic groups are parts of the tertiary structure, as is the arrangement of helical and pleated-sheet sections with respect to one another. In a fibrous protein, the overall shape of which is a long rod, the secondary structure also provides much of the information about the tertiary structure. The helical backbone of the protein does not fold back on itself, and the only important aspect of the tertiary structure that is not specified by the secondary structure is the arrangement of the atoms of the side chains.

For a globular protein, considerably more information is needed. It is necessary to determine the way in which the helical and pleated-sheet sections fold back on each other, in addition to the positions of the side-chain atoms and any prosthetic groups. The interactions between the side chains play an important role in the folding of proteins. The folding pattern frequently brings residues that are separated in the amino acid sequence into proximity in the tertiary structure of the native protein.

Not every protein necessarily exhibits all possible structural features of the kinds we described in Section 4.4. For instance, there are no disulfide bridges in myoglobin and hemoglobin, which are oxygen-storage and transport proteins and classic examples of protein structure, but they both contain Fe(II) ions as part of a prosthetic group. In contrast, the enzymes trypsin and chymotrypsin do not contain complexed metal ions, but they do have disulfide bridges. Hydrogen bonds, electrostatic interactions, and hydrophobic interactions occur in most proteins.

The three-dimensional conformation of a protein is the result of the interplay of all the stabilizing forces. It is known, for example, that proline does not fit into an α-helix and that its presence can cause a polypeptide chain to turn a corner, ending an α-helical segment. The presence of proline is not, however, a *requirement* for a turn in a polypeptide chain. Other residues are routinely encountered at bends in polypeptide chains. The segments of proteins at bends in the polypeptide chain and in other portions of the protein that are not involved in helical or pleated-sheet structures are frequently referred to as "random" or "random coil." In reality, the forces that stabilize each protein are responsible for its conformation.

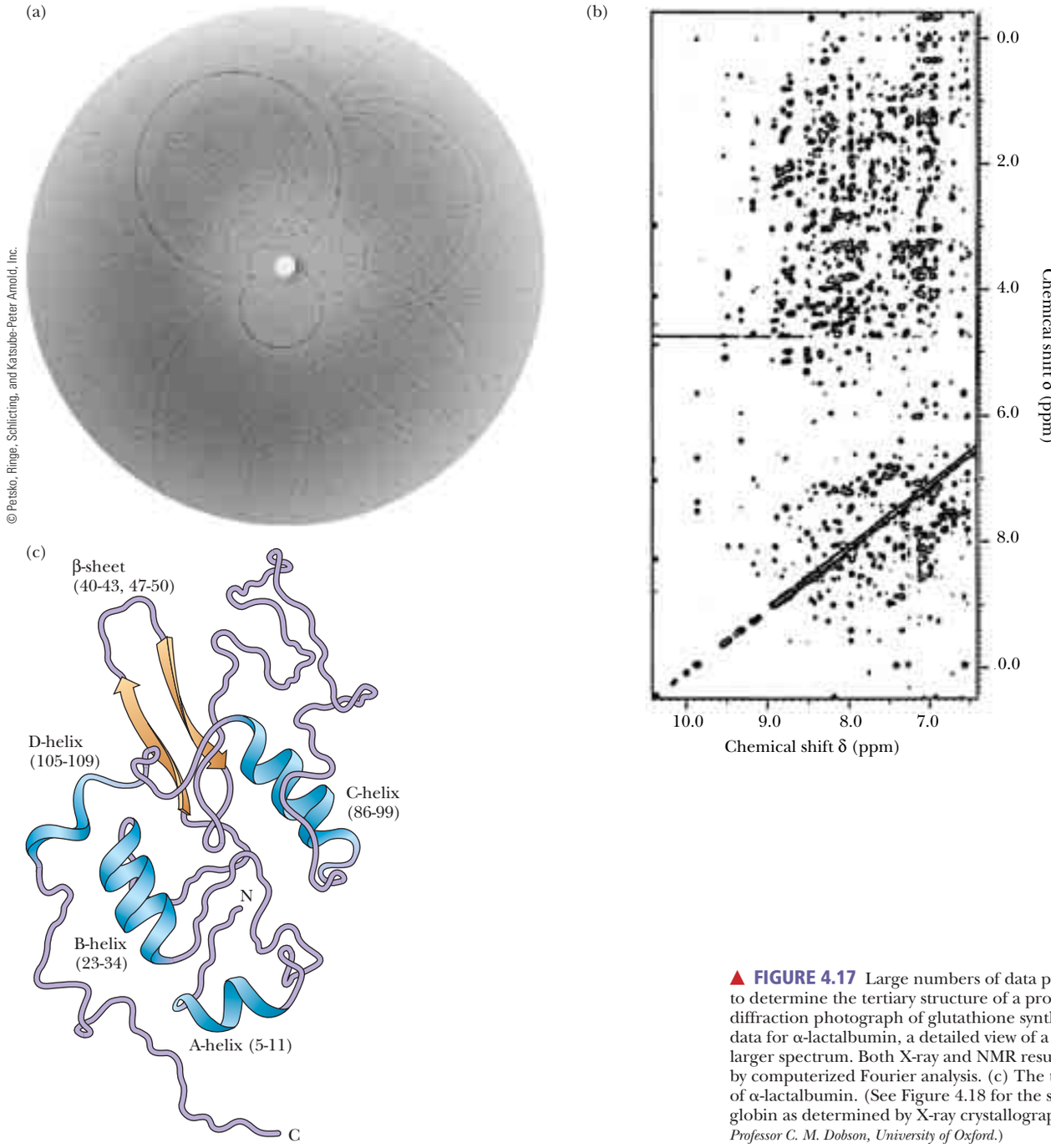
The experimental technique used to determine the tertiary structure of a protein is **X-ray crystallography.** Perfect crystals of some proteins can be grown under carefully controlled conditions. In such a crystal, all the individual protein molecules have the same three-dimensional conformation and the same orientation. Crystals of this quality can be formed only from proteins of very high purity, and it is not possible to obtain a structure if the protein cannot be crystallized.

When a suitably pure crystal is exposed to a beam of X rays, a *diffraction pattern* is produced on a photographic plate (Figure 4.17a) or a radiation counter. The pattern is produced when the electrons in each atom in the molecule scatter the X rays. The number of electrons in the atom determines the intensity of its scattering of X rays; heavier atoms scatter more effectively than lighter atoms. The scattered X rays from the individual atoms can reinforce each other or cancel each other (set up constructive or destructive interference), giving rise to the characteristic pattern for each type of molecule. A series of diffraction patterns taken from several angles contains the information needed to determine the tertiary structure. The information is extracted from the diffraction patterns through a mathematical analysis known as a *Fourier series*. Many thousands of such calculations are required to determine the structure of a protein, and even though they are performed by computer, the process is a fairly long one. Improving the calculation procedure is a subject of active research. The articles by Hauptmann and by Karle listed in the bibliography at the end of this chapter outline some of the accomplishments in the field.

Another technique that supplements the results of X-ray diffraction has come into wide use in recent years. It is a form of **nuclear magnetic resonance (NMR) spectroscopy.** In this particular application of NMR, called *2-D* (two-dimensional) *NMR*, large collections of data points are subjected to computer analysis (Figure 4.17b). Like X-ray diffraction, this method uses a Fourier series to analyze results. It is similar to X-ray diffraction in other ways: It is a long process, and it requires considerable amounts of computing power and milligram quantities of protein. One way in which 2-D NMR differs from X-ray diffraction is that it uses protein samples in aqueous solution rather than crystals. This environment is closer to that of proteins in cells, and thus it is one of the main advantages of the method. The NMR method most widely used in the determination of protein structure ultimately depends on the distances between hydrogen atoms, giving results independent of those obtained by X-ray crystallography. The NMR method is undergoing constant improvement and is being applied to larger proteins as these improvements progress.

Essential Information

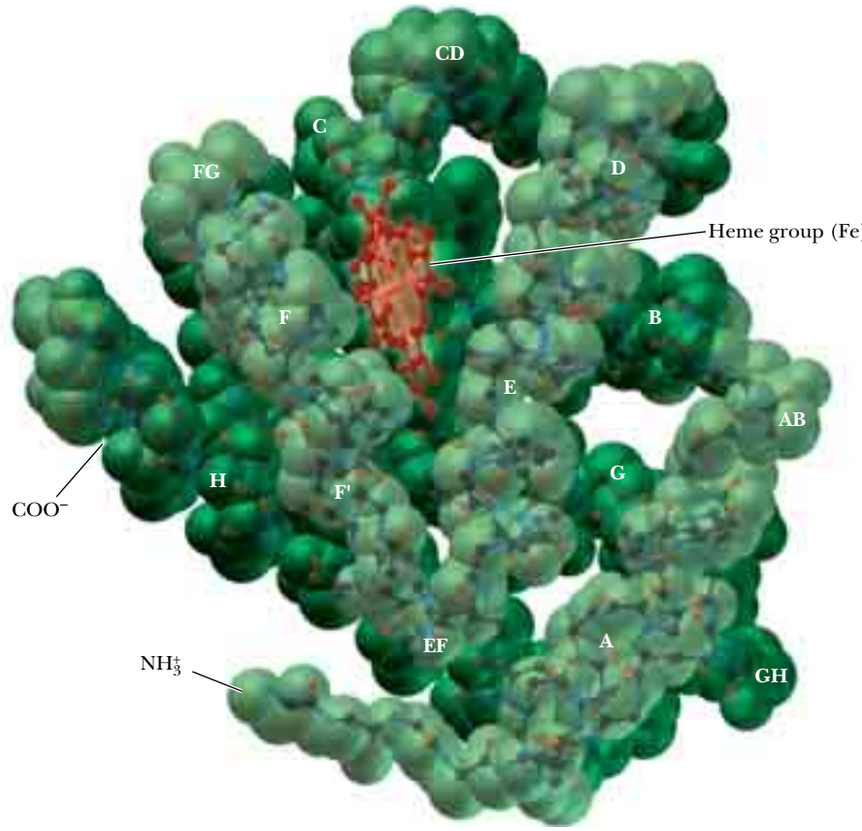
The tertiary structure of a protein is the three-dimensional arrangement of all atoms in a protein chain. The secondary and tertiary structures of a protein can be determined simultaneously.



▲ **FIGURE 4.17** Large numbers of data points are needed to determine the tertiary structure of a protein. (a) X-ray diffraction photograph of glutathione synthetase. (b) NMR data for α-lactalbumin, a detailed view of a key part of a larger spectrum. Both X-ray and NMR results are processed by computerized Fourier analysis. (c) The tertiary structure of α-lactalbumin. (See Figure 4.18 for the structure of myoglobin as determined by X-ray crystallography.) (*b, courtesy of Professor C. M. Dobson, University of Oxford.*)

Myoglobin: An Example of Protein Structure

In many ways, myoglobin is the classic example of a globular protein. We shall use it here as a case study in tertiary structure. (We shall see the tertiary structures of many other proteins in context when we discuss their roles in biochemistry.) Myoglobin was the first protein for which the complete tertiary structure (Figure 4.18) was determined by X-ray crystallography. The complete myoglobin molecule consists of a single polypeptide chain of 153 amino acid residues and includes a prosthetic group, the **heme** group, which also occurs in hemoglobin. The myoglobin molecule (including the heme group) has a compact structure, with the interior atoms very close to each other. This

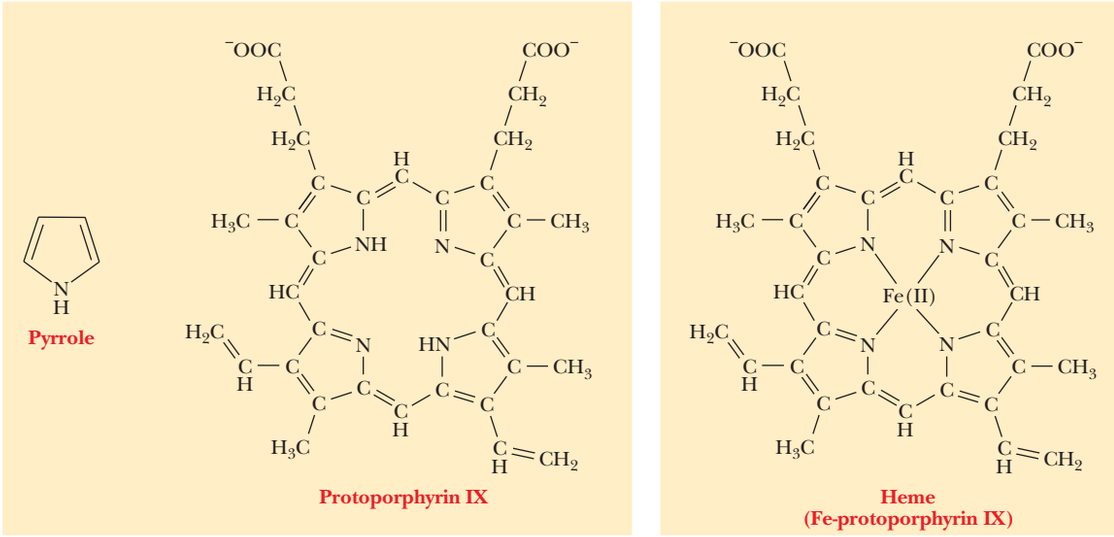


▶ **FIGURE 4.18** The structure of the myoglobin molecule, showing the peptide backbone and the heme group. The helical segments are designated by the letters A through H. The terms NH_3 and COO^- indicate the N-terminal and C-terminal ends, respectively.

structure provides examples of many of the forces responsible for the three-dimensional shapes of proteins.

In myoglobin, there are eight α-helical regions and no β-pleated sheet regions. Approximately 75% of the residues in myoglobin are found in these helical regions, which are designated by the letters A through H. Hydrogen bonding in the polypeptide backbone stabilizes the α-helical regions; amino acid side chains are also involved in hydrogen bonds. The polar residues are on the exterior of the molecule. The interior of the protein contains almost exclusively nonpolar amino acid residues. Two polar histidine residues are found in the interior; they are involved in interactions with the heme group and bound oxygen, and thus play an important role in the function of the molecule. The planar heme group fits into a hydrophobic pocket in the protein portion of the molecule and is held in position by hydrophobic attractions between heme's porphyrin ring and the nonpolar side chains of the protein. The presence of the heme group drastically affects the conformation of the polypeptide: The apoprotein (the polypeptide chain alone, without the prosthetic heme group) is not as tightly folded as the complete molecule.

The heme group consists of a metal ion, Fe(II), and an organic part, protoporphyrin IX (Figure 4.19). (The notation Fe(II) is preferred to Fe^{2+} when metal ions occur in complexes.) The porphyrin part consists of four five-membered rings based on the pyrrole structure; these four rings are linked by bridging methine (—CH═) groups to form a square planar structure. The Fe(II) ion has six coordination sites, and it forms six metal–ion complexation bonds. Four of the six sites are occupied by the nitrogen atoms of the four pyrrole-type rings of the porphyrin to give the complete heme group. The presence of the heme group is required for myoglobin to bind oxygen.



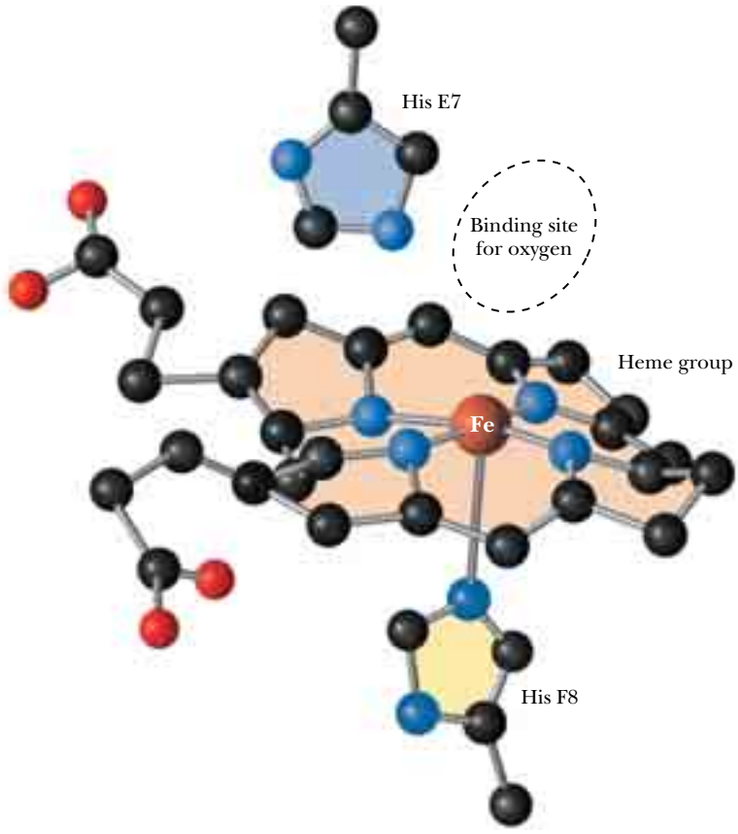
▲ **FIGURE 4.19** The structure of the heme group. Four pyrrole rings are linked by bridging groups to form a planar porphyrin ring. Several isomeric porphyrin rings are possible, depending on the nature and arrangement of the side chains. The porphyrin isomer found in heme is protoporphyrin IX. Addition of iron to protoporphyrin IX produces the heme group.

The fifth coordination site of the Fe(II) ion is occupied by one of the nitrogen atoms of the imidazole side chain of histidine residue F8 (the eighth residue in helical segment F). This histidine residue is one of the two in the interior of the molecule. The oxygen is bound at the sixth coordination site of the iron. The fifth and sixth coordination sites lie perpendicular to, and on opposite sides of, the plane of the porphyrin ring. The other histidine residue in the interior of the molecule, residue E7 (the seventh residue in helical segment E), lies on the same side of the heme group as the bound oxygen (Figure 4.20). This second histidine is not bound to the iron, or to any part of the heme group, but it acts as a gate that opens and closes as oxygen enters the hydrophobic pocket to bind to the heme. The E7 histidine sterically inhibits oxygen from binding perpendicularly to the heme plane, with biologically important ramifications. The affinity of free heme for carbon monoxide (CO) is 25,000 times greater than its affinity for oxygen. When carbon monoxide is forced to bind at an angle in myoglobin due to the steric block by His E7, its advantage over oxygen drops by two orders of magnitude (Figure 4.21). This guards against the possibility that traces of CO produced during metabolism would occupy all the oxygen-binding sites on the hemes. Nevertheless, CO is a potent poison in larger quantities because of its effect both on oxygen binding to hemoglobin and on the final step of the electron transport chain (Section 20.5).

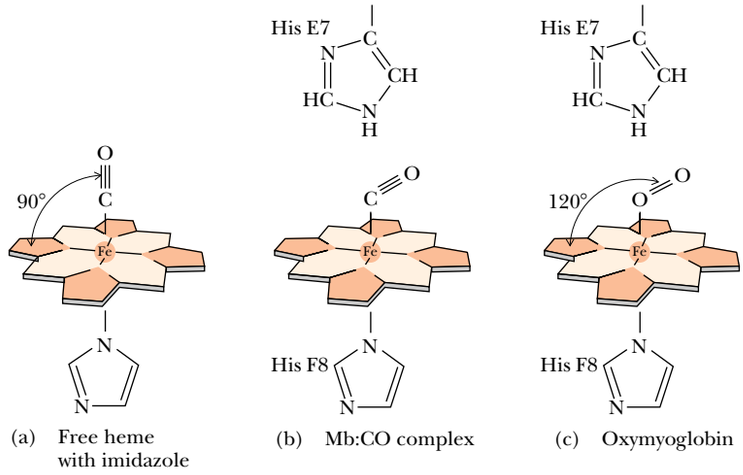
In the absence of the protein, the iron of the heme group can be oxidized to Fe(III); the oxidized heme will not bind oxygen. Thus, the combination of both heme and protein is needed to bind O_2 for oxygen storage.

Denaturation and Refolding

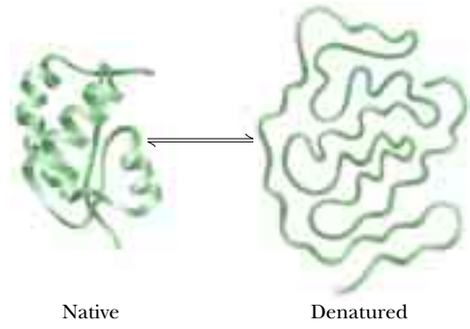
The noncovalent interactions that maintain the three-dimensional structure of a protein are weak, and it is not surprising that they can be disrupted easily. The unfolding of a protein is called **denaturation.** Reduction of disulfide bonds leads (Section 3.5) to even more extensive unraveling of the tertiary



▶ **FIGURE 4.20** The oxygen-binding site of myoglobin. The porphyrin ring occupies four of the six coordination sites of the Fe(II). Histidine F8 (His F8) occupies the fifth coordination site of the iron (see text). Oxygen is bound at the sixth coordination site of the iron, and histidine E7 lies close to the oxygen. (*Leonard Lessin/Waldo Feng/Mt. Sinai CORE.*)



▶ **FIGURE 4.21** Oxygen and carbon monoxide binding to the heme group of myoglobin. The presence of the E7 histidine forces a 120° angle to the oxygen or CO.



Biochemistry Now™ ANIMATED FIGURE 4.22 Denaturation of a protein. The native conformation can be recovered when denaturing conditions are removed. **See this figure animated at http://now.brookscole.com/campbell5**

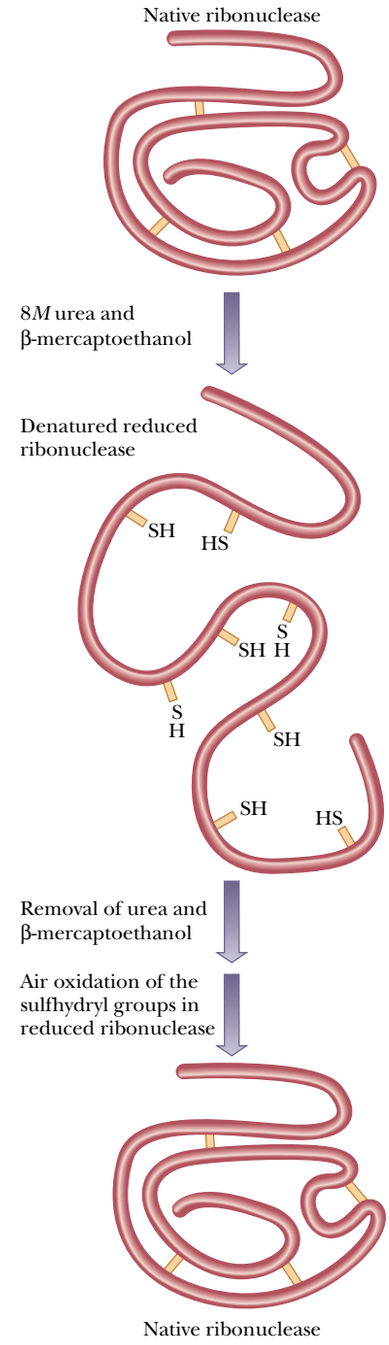
structure. Denaturation and reduction of disulfide bonds are frequently combined when complete disruption of the tertiary structure of proteins is desired. Under proper experimental conditions, the disrupted structure can then be completely recovered. This process of denaturation and refolding is a dramatic demonstration of the relationship between the primary structure of the protein and the forces that determine the tertiary structure. For many proteins, various other factors are needed for complete refolding, but the important point is that the primary structure determines the tertiary structure.

Proteins can be denatured in several ways. One is *heat.* An increase in temperature favors vibrations within the molecule, and the energy of these vibrations can become great enough to disrupt the tertiary structure. At either high or low *extremes of pH,* at least some of the charges on the protein are missing, and so the electrostatic interactions that would normally stabilize the native, active form of the protein are drastically reduced. This leads to denaturation. The binding of *detergents,* such as sodium dodecyl sulfate (SDS), also denatures proteins. Detergents tend to disrupt hydrophobic interactions. If a detergent is charged, it can also disrupt electrostatic interactions within the protein. Other reagents, such as *urea* and *guanidine hydrochloride,* form hydrogen bonds with the protein that are stronger than those within the protein itself. These two reagents can also disrupt hydrophobic interactions in much the same way as detergents (Figure 4.22).

β-*Mercaptoethanol* ($HS—CH_2—CH_2—OH$) is frequently used to reduce disulfide bridges to two sulfhydryl groups. Urea is usually added to the reaction mixture to facilitate unfolding of the protein and to increase the accessibility of the disulfides to the reducing agent. If experimental conditions are properly chosen, the native conformation of the protein can be recovered when both mercaptoethanol and urea are removed (Figure 4.23). Experiments of this type provide some of the strongest evidence that the amino acid sequence of the protein contains all the information required to produce the complete three-dimensional structure. Protein researchers are pursuing with some interest the conditions under which a protein can be denatured—including reduction of disulfides—and its native conformation later recovered.

Essential Information

The primary structure of a protein contains all the information needed to specify the tertiary structure.

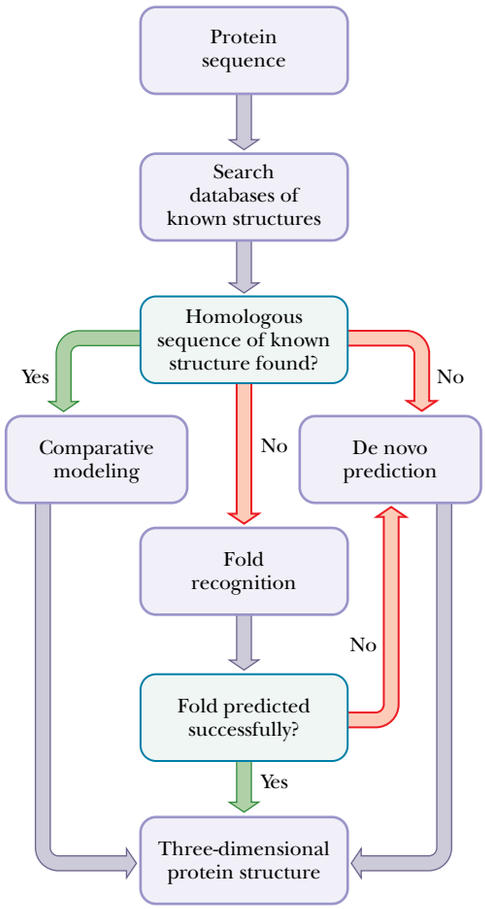


▲ **FIGURE 4.23** Denaturation and refolding in ribonuclease. The protein ribonuclease can be completely denatured by the actions of urea and mercaptoethanol. When denaturing conditions are removed, activity is recovered.

4.6 Can We Predict Protein Folding from Sequence?

Since the sequence of amino acids determines the three-dimensional structure of a protein, a question that arises naturally is, "Can we predict the tertiary structure of a protein if we know its amino acid sequence?" The answer is that we can, within limits. Modern computing techniques greatly facilitate the operation, which requires processing large amounts of information. The encounter of biochemistry and computing has given rise to the burgeoning field of **bioinformatics.** Prediction of protein structure is one of the principal applications of bioinformatics. Another important application is the comparison of base sequences in nucleic acids, a topic we shall discuss in Chapter 14, along with other methods for working with nucleic acids.

The first step in predicting protein architecture is a search of databases of known structures for *sequence homology* between the protein whose structure is to be determined and proteins of known architecture, where the term **homology** refers to similarity of two or more sequences. If the sequence of the known protein is similar enough to that of the protein being studied, the known protein's structure becomes the point of departure for *comparative modeling.* Use of modeling algorithms that compare the protein being studied with known structures leads to a structure prediction. This method is most



▲ **FIGURE 4.24** A flow chart showing the use of existing information from databases to predict protein conformation. (*Courtesy of Rob Russell, EMBL.*)

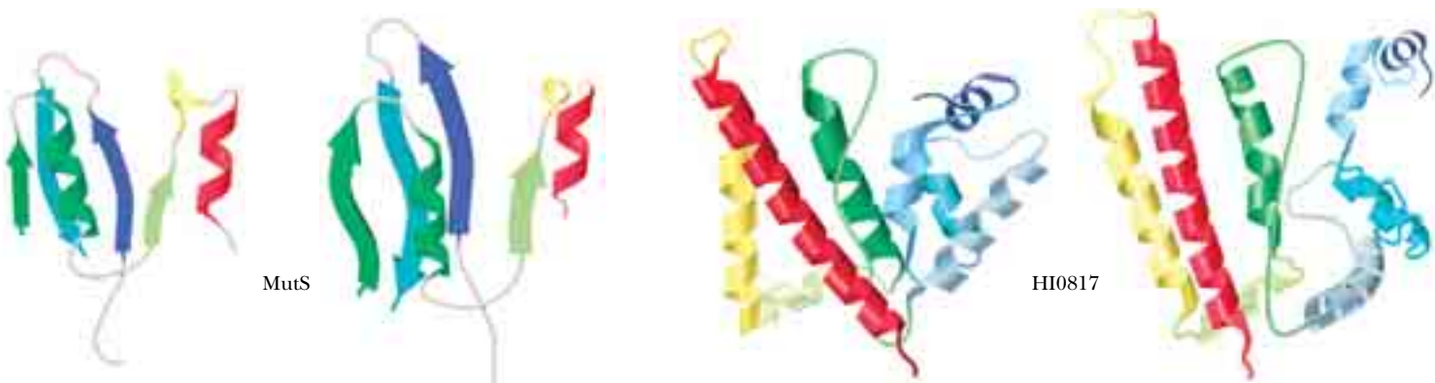
useful when the sequence homology is greater than 25–30%. If the sequence homology is less than 25–30%, other approaches are more useful. *Fold recognition* algorithms allow comparison with known folding motifs common to many secondary structures. We saw a number of these motifs in Section 4.3. Here is an application of that information. Yet another method is *de novo prediction,* based on first principles from chemistry, biology, and physics. This method too can give rise to structures subsequently confirmed by X-ray crystallography. The flow chart in Figure 4.24 shows how prediction techniques use existing information from databases. Figure 4.25 shows a comparison of the predicted structures of two proteins (right side) for the DNA repair protein MutS and the bacterial protein HI0817. The crystal structures of the two proteins are shown on the left.

A considerable amount of information about protein sequences and architecture is available on the World Wide Web. One of the most important resources is the Protein Data Bank operated under the auspices of the Research Collaboratory for Structural Bioinformatics (RCSB). Its URL is http://www.rcsb.org/pdb. This site, which has a number of mirror sites around the world, is the single repository of structural information about large molecules. It includes material about nucleic acids as well as proteins. Its home page has a button with links specifically geared to educational applications.

Results of structure prediction using the methods discussed in this section are available on the Web as well. One of the most useful URLs is http://predictioncenter.llnl.gov/casp5. Other excellent sources of information are available through the National Institutes of Health (http://pubmedcentral.nih.gov/tocrender.fcgi?iid=1005, and http://www.ncbi.nlm.nih.gov), and through the ExPASy (Expert Protein Analysis System) server (http://us.expasy.org).

Protein-Folding Chaperones

The primary structure conveys all the information necessary to produce the correct tertiary structure, but the folding process in vivo can be a bit trickier. In the protein-dense environment of the cell, proteins may begin to fold incorrectly as they are produced, or they may begin to associate with other proteins before completing their folding process. In eukaryotes, proteins may need to remain unfolded long enough to be transported across the membrane of a subcellular organelle. Special proteins called **chaperones** aid in the correct and timely folding of many other proteins (see the Biochemical Connections box in Chapter 12). The first such proteins discovered were a family called



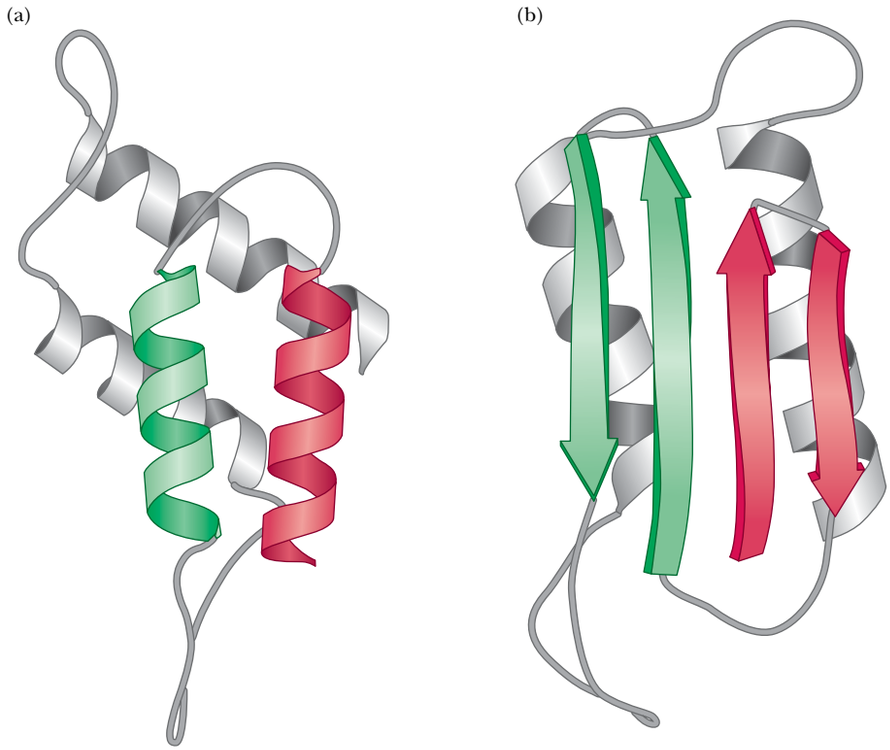
▲ **FIGURE 4.25** A comparison of the predicted structures of two proteins (right side) for the DNA repair protein MutS and the bacterial protein HI0817. The crystal structures of the two proteins are shown on the left. (*Courtesy of University of Washington, Seattle.*)

Biochemical Connections

Prions

It has been established that the causative agent of mad-cow disease, as well as the related diseases scrapie in sheep and spongiform encephalopathy (kuru and Creutzfeldt-Jakob disease) in humans, is a small (28-kD) protein called a **prion.** Prions are glycoproteins found in the cell membranes of nerve tissue. The diseases come about when the normal form of the prion protein, PrP (Figure a), folds into an incorrect form called PrP^{sc} (Figure b). The abnormal form of the prion protein is able to convert other, normal forms into abnormal forms. As recently discovered, this change can be propagated in nervous tissue. Scrapie had been known for years, but it had not been known to cross species barriers. Then an outbreak of mad-cow disease was shown to have followed the inclusion of sheep remains in cattle feed. It is now known that eating tainted beef from animals with mad-cow disease can cause spongiform encephalopathy, now known as new variant Creutzfeldt-Jakob disease, in humans. The normal prions have a large percentage of α-helix, but the abnormal forms have more β-pleated sheets. Notice that in this case the same protein (a single, well-defined sequence) can exist in alternative forms. These β-pleated sheets in the abnormal proteins interact between protein molecules and form insoluble plaques, a fate also seen in Alzheimer's disease. Ingested abnormal prions use macrophages from the immune system to travel in the body until they come in contact with nerve tissue. They can then propagate up the nerves until they reach the brain.

This mechanism was a subject of considerable controversy when it was first proposed. A number of scientists expected that a slow-acting virus would be found to be the ultimate cause of these neurological diseases. A susceptibility to these diseases can be inherited, so some involvement of DNA (or RNA) was also expected. Some went so far as to talk about "heresy" when Stanley Prusiner received the 1997 Nobel Prize in medicine for his discovery of prions. It now appears that genes for susceptibility to the incorrect form exist in all vertebrates, giving rise to the observed pattern of disease transmission, but many individuals with the genetic susceptibility never develop the disease if they do not come in contact with abnormal prions from another source. See the articles by Ferguson and Peretz in the bibliography of this chapter.



◀ (a) Normal prion structure (PrP).
(b) Abnormal prion (PrP^{sc}).

hsp70 (for 70,000 MW Heat-Shock Protein), which are proteins produced in *E. coli* grown above optimal temperatures. Chaperones exist in organisms from prokaryotes through humans, and their mechanisms of action are currently being studied. (See the article by Helfand in the bibliography of this chapter.) In recent years, it has become evident that protein-folding dynamics is crucial

to protein function in vivo. The Biochemical Connections box on the previous page describes a particularly striking example of the importance of protein folding.

4.7 What Is the Quaternary Structure of Proteins?

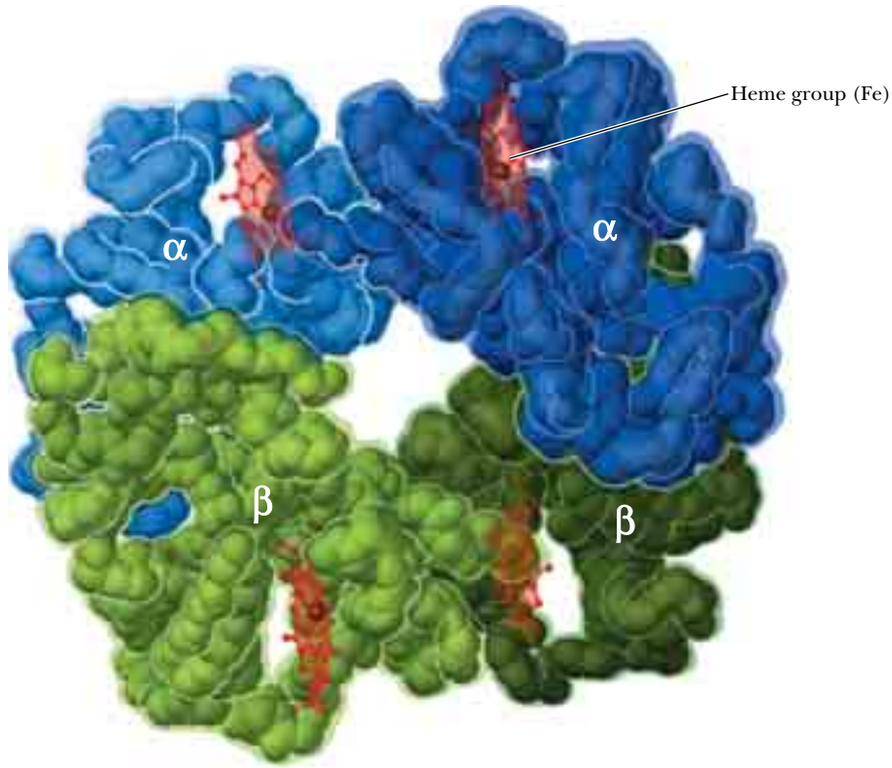
Quaternary structure is a property of proteins that consist of more than one polypeptide chain. Each chain is called a subunit. The number of chains can range from two to more than a dozen, and the chains may be identical or different. Commonly occurring examples are **dimers, trimers,** and **tetramers,** consisting of two, three, and four polypeptide chains, respectively. (The generic term for such a molecule, made up of a small number of subunits, is **oligomer.**) The chains interact with one another noncovalently via electrostatic attractions, hydrogen bonds, and hydrophobic interactions.

As a result of these noncovalent interactions, subtle changes in structure at one site on a protein molecule may cause drastic changes in properties at a distant site. Proteins that exhibit this property are called **allosteric.** Not all multisubunit proteins exhibit allosteric effects, but many do.

A classic illustration of the quaternary structure of proteins and its effect on properties is a comparison of hemoglobin, an allosteric protein, with myoglobin, which consists of a single polypeptide chain.

Hemoglobin

Hemoglobin is a tetramer, consisting of four polypeptide chains, two α-chains and two β-chains (Figure 4.26). (In oligomeric proteins, the types of polypep-



▶ **FIGURE 4.26** The structure of hemoglobin. Hemoglobin ($\alpha_2\beta_2$) is a tetramer consisting of four polypeptide chains (two α-chains and two β-chains).

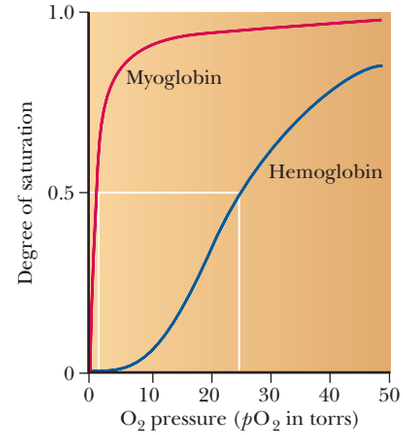
tide chains are designated with Greek letters.) The two α-chains of hemoglobin are identical, as are the two β-chains. The overall structure of hemoglobin is $\alpha_2\beta_2$ in Greek-letter notation. Both the α- and β-chains of hemoglobin are very similar to the myoglobin chain. The α-chain is 141 residues long, and the β-chain is 146 residues long; for comparison, the myoglobin chain is 153 residues long. Many of the amino acids of the α-chain, the β-chain, and myoglobin are *homologous;* that is, the same amino acid residues are in the same positions. The heme group is the same in myoglobin and hemoglobin.

We have already seen that one molecule of myoglobin binds one oxygen molecule. Four molecules of oxygen can therefore bind to one hemoglobin molecule. Both hemoglobin and myoglobin bind oxygen reversibly, but the binding of oxygen to hemoglobin exhibits **positive cooperativity,** whereas oxygen binding to myoglobin does not. Positive cooperativity means that when one oxygen molecule is bound, it becomes easier for the next to bind. A graph of the oxygen-binding properties of hemoglobin and myoglobin is one of the best ways to illustrate this point (Figure 4.27).

When the degree of saturation of myoglobin with oxygen is plotted against oxygen pressure, a steady rise is observed until complete saturation is approached and the curve levels off. The oxygen-binding curve of myoglobin is thus said to be **hyperbolic.** In contrast, the shape of the oxygen-binding curve for hemoglobin is **sigmoidal.** This shape indicates that the binding of the first oxygen molecule facilitates the binding of the second oxygen, which facilitates the binding of the third, which in turn facilitates the binding of the fourth. This is precisely what is meant by the term "cooperative binding." However, note that even though cooperative binding means that binding of each subsequent oxygen is easier than the previous one, the binding curve is still lower than that of myoglobin at any oxygen pressure. In other words, at any oxygen pressure, myoglobin will have a higher percentage of saturation than hemoglobin.

The two types of behavior are also related to the functions of these proteins. Myoglobin has the function of oxygen *storage* in muscle. It must bind strongly to oxygen at very low pressures, and it is 50% saturated at 1 torr partial pressure of oxygen. (The **torr** is a widely used unit of pressure, but it is not an SI unit. One torr is the pressure exerted by a column of mercury 1 mm high at 0°C. One atmosphere is equal to 760 torr.) The function of hemoglobin is oxygen *transport,* and it must be able both to bind strongly to oxygen and to release oxygen easily, depending upon conditions. In the alveoli of lungs (where hemoglobin must bind oxygen for transport to the tissues), the oxygen pressure is 100 torr. At this pressure, hemoglobin is 100% saturated with oxygen. In the capillaries of active muscles, the pressure of oxygen is 20 torr, corresponding to less than 50% saturation of hemoglobin, which occurs at 26 torr. In other words, hemoglobin gives up oxygen easily in capillaries, where the need for oxygen is great.

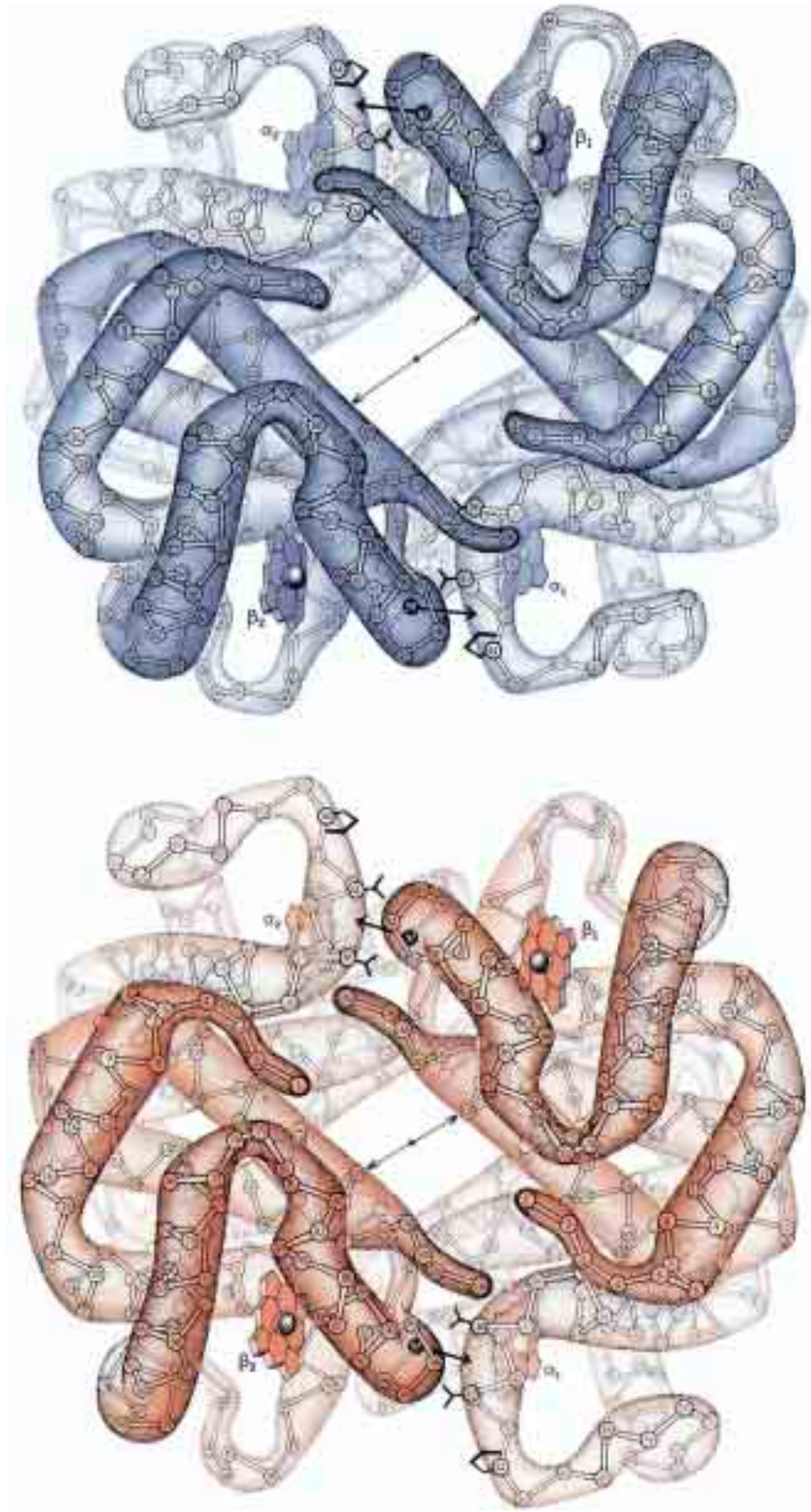
Structural changes during binding of small molecules are characteristic of allosteric proteins such as hemoglobin. Hemoglobin has different quaternary structures in the bound (oxygenated) and unbound (deoxygenated) forms. The two β-chains are much closer to each other in oxygenated hemoglobin than in deoxygenated hemoglobin. The change is so marked that the two forms of hemoglobin have different crystal structures (Figure 4.28).



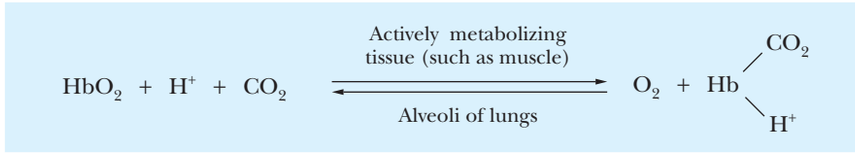
▲ **FIGURE 4.27** A comparison of the oxygen-binding behavior of myoglobin and hemoglobin. The oxygen-binding curve of myoglobin is hyperbolic, whereas that of hemoglobin is sigmoidal. Myoglobin is 50% saturated with oxygen at 1 torr partial pressure; hemoglobin does not reach 50% saturation until the partial pressure of oxygen reaches 26 torr.

Conformational Changes That Accompany Hemoglobin Function

Other ligands are involved in cooperative effects when oxygen binds to hemoglobin. Both H^+ and CO_2, which themselves bind to hemoglobin, affect the affinity of hemoglobin for oxygen by altering the protein's three-dimensional



▶ **FIGURE 4.28** The structures of (a) deoxyhemoglobin and (b) oxyhemoglobin. Note the motions of subunits with respect to one another. There is much less room at the center of oxyhemoglobin. (*Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.*)

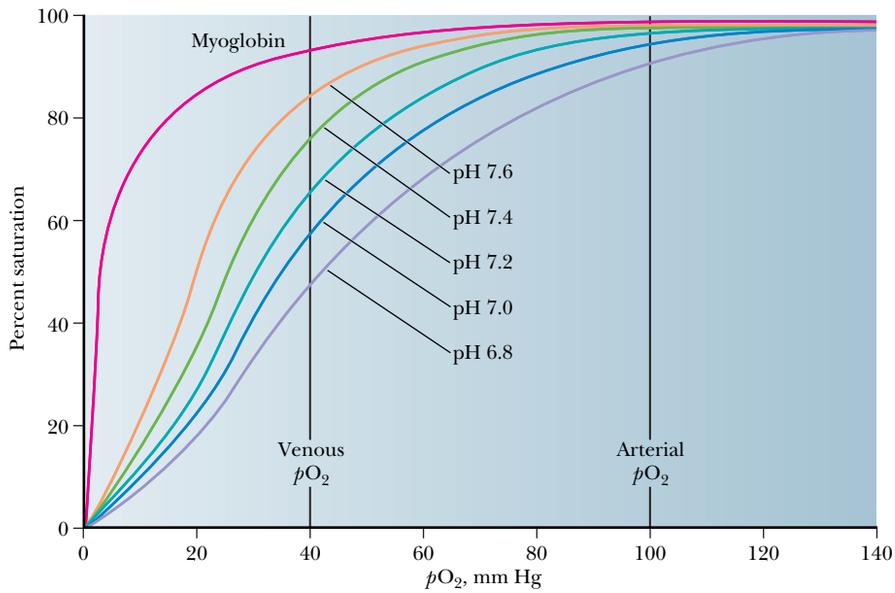


◀ **FIGURE 4.29** The general features of the Bohr effect. In actively metabolizing tissue, hemoglobin releases oxygen and binds both CO_2 and H^+. In the lungs, hemoglobin releases both CO_2 and H^+ and binds oxygen.

structure in subtle but important ways. The effect of H^+ (Figure 4.29) is called the *Bohr effect,* after its discoverer, Christian Bohr (the father of physicist Niels Bohr). The oxygen-binding ability of myoglobin is not affected by the presence of H^+ or of CO_2.

An increase in the concentration of H^+ (i.e., a lowering of the pH) reduces the oxygen affinity of hemoglobin. Increasing H^+ causes the protonation of key amino acids, including the N-terminals of the α-chains and His^{146} of the β-chains. The protonated histidine is attracted to, and stabilized by, a salt bridge to Asp^{94}. This favors the deoxygenated form of hemoglobin. Actively metabolizing tissue, which requires oxygen, releases H^+, thus acidifying its local environment. Hemoglobin has a lower affinity for oxygen under these conditions, and it releases oxygen where it is needed (Figure 4.30). Hemoglobin's acid–base properties affect, and are affected by, its oxygen-binding properties. The oxygenated form of hemoglobin is a stronger acid (has a lower pK_a) than the deoxygenated form. In other words, deoxygenated hemoglobin has a higher affinity for H^+ than does the oxygenated form. Thus, changes in the quaternary structure of hemoglobin can modulate the buffering of blood through the hemoglobin molecule itself.

Table 4.1 summarizes the important features of the Bohr effect.

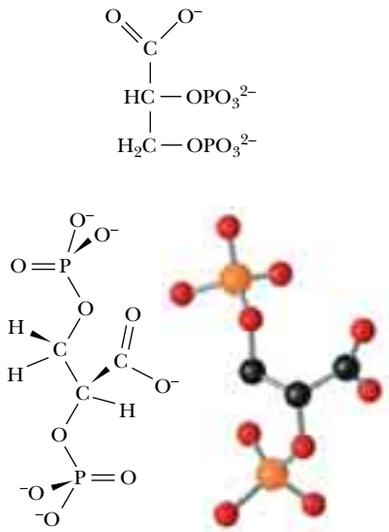


◀ **FIGURE 4.30** The oxygen saturation curves for myoglobin and for hemoglobin at five different pH values.

Table 4.1

A Summary of the Bohr Effect

Lungs	Actively Metabolizing Muscle
Higher pH than actively metabolizing tissue	Lower pH due to production of H^+
Hemoglobin binds O_2	Hemoglobin releases O_2
Hemoglobin releases H^+	Hemoglobin binds H^+



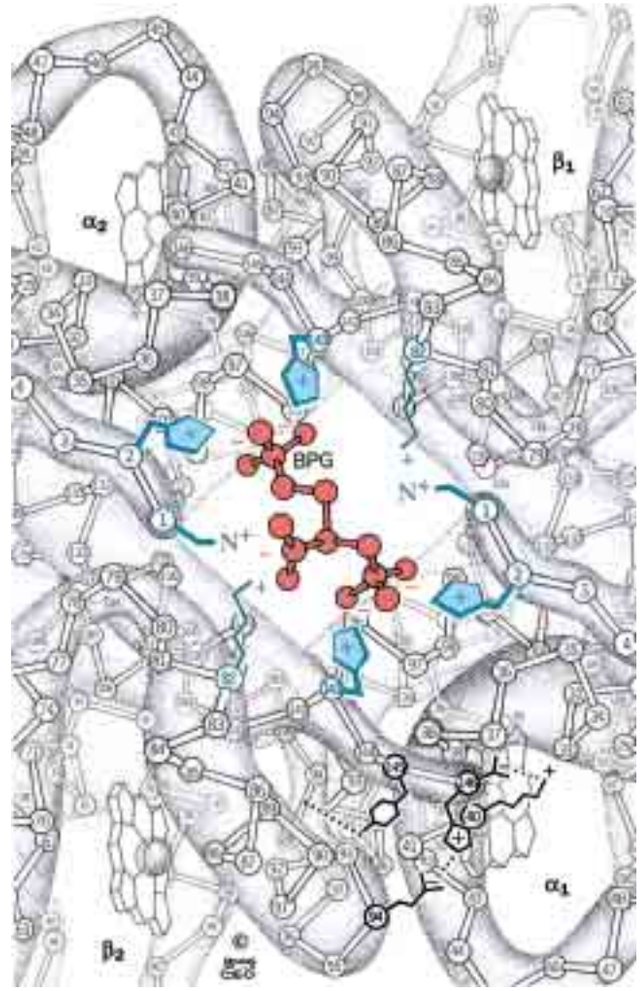
▲ **FIGURE 4.31** The structure of BPG (2,3-*bis*phosphoglycerate), an important allosteric effector of hemoglobin.

Large amounts of CO_2 are produced by metabolism. The CO_2, in turn, forms carbonic acid, H_2CO_3. The pK_a of H_2CO_3 is 6.35; the normal pH of blood is 7.4. As a result, about 90% of dissolved CO_2 will be present as the bicarbonate ion, HCO_3^-, releasing H^+. (The Henderson–Hasselbalch equation can be used to confirm this point.) The in vivo buffer system involving H_2CO_3 and HCO_3^- in blood was discussed in Section 2.6. The presence of larger amounts of H^+ as a result of CO_2 production favors the quaternary structure that is characteristic of deoxygenated hemoglobin. Hence, the affinity of hemoglobin for oxygen is lowered. The HCO_3^- is transported to the lungs, where it combines with H^+ released when hemoglobin is oxygenated, producing H_2CO_3. In turn, H_2CO_3 liberates CO_2, which is then exhaled. Hemoglobin also transports some CO_2 directly. When the CO_2 concentration is high, it combines with the free α-amino groups to form carbamate:

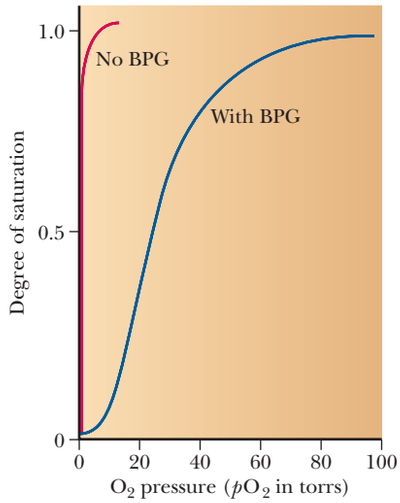
$$R{-}NH_2 + CO_2 \rightleftharpoons R{-}NH{-}COO^- + H^+$$

This reaction turns the α-amino terminals into anions, which can then interact with the α-chain Arg^{141}, also stabilizing the deoxygenated form.

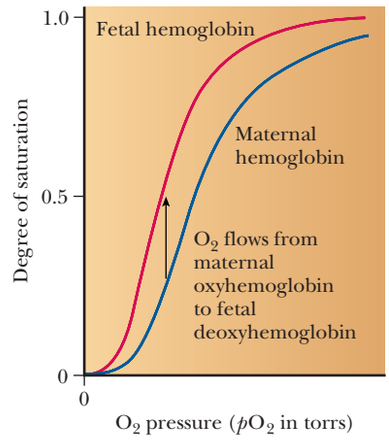
In the presence of large amounts of H^+ and CO_2, as in respiring tissue, hemoglobin releases oxygen. The presence of large amounts of oxygen in the lungs reverses the process, causing hemoglobin to bind O_2. The oxygenated



▶ **FIGURE 4.32** The binding of BPG to deoxyhemoglobin. Note the electrostatic interactions between the BPG and the protein. (*Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.*)



▲ **FIGURE 4.33** A comparison of the oxygen-binding properties of hemoglobin in the presence and absence of BPG. Note that the presence of the BPG markedly decreases the affinity of hemoglobin for oxygen.



▲ **FIGURE 4.34** A comparison of the oxygen-binding capacity of fetal and maternal hemoglobins. Fetal hemoglobin binds less strongly to BPG and, consequently, has a greater affinity for oxygen than does maternal hemoglobin.

hemoglobin can then transport oxygen to the tissues. The process is complex, but it allows for fine tuning of pH as well as levels of CO_2 and O_2.

Hemoglobin in blood is also bound to another ligand, **2,3-*bis*phosphoglycerate (BPG)** (Figure 4.31), with drastic effects on its oxygen-binding capacity. The binding of BPG to hemoglobin is electrostatic; specific interactions take place between the negative charges on BPG and the positive charges on the protein (Figure 4.32). In the presence of BPG, the partial pressure at which 50% of hemoglobin is bound to oxygen is 26 torr. If BPG were not present in blood, the oxygen-binding capacity of hemoglobin would be much higher (50% of hemoglobin bound to oxygen at about 1 torr), and little oxygen would be released in the capillaries. "Stripped" hemoglobin, which is isolated from blood and from which the endogenous BPG has been removed, displays this behavior (Figure 4.33).

BPG also plays a role in supplying a growing fetus with oxygen. The fetus obtains oxygen from the mother's bloodstream via the placenta. Fetal hemoglobin (Hb F) has a higher affinity for oxygen than does maternal hemoglobin, allowing for efficient transfer of oxygen from the mother to the fetus (Figure 4.34). Two features of fetal hemoglobin contribute to this higher oxygen-binding capacity. One is the presence of two different polypeptide chains. The subunit structure of Hb F is $\alpha_2\gamma_2$, where the β-chains of adult hemoglobin (Hb A), the usual hemoglobin, have been replaced by the γ-chains, which are similar but not identical in structure. The second feature is that Hb F binds less strongly to BPG than does Hb A. In the β-chain of adult hemoglobin, His^{143} makes a salt bridge to BPG. In the fetal hemoglobin, the γ-chain has an amino acid substitution of a serine for His^{143}. This change of a positively charged amino acid for a neutral one diminishes the number of contacts between the hemoglobin and the BPG, effectively reducing the allosteric effect enough to give fetal hemoglobin a higher binding curve than adult hemoglobin.

Summary

4.1 How Does the Structure of Proteins Determine Their Function? The structure of proteins is complex, with few obvious regular structures. Many three-dimensional conformations are possible for proteins, but only one, or at most a few, have biological activity; these are called the native conformations. To facilitate structure determination, it is customary to define four levels of organization.

4.2 What Is the Primary Structure of Proteins? Primary structure is the order in which the amino acids are covalently linked. The primary structure of a protein can be determined by chemical methods. The amino acid sequence (the primary structure) of a protein determines its three-dimensional structure, which in turn determines its properties. A striking example of the importance of primary structure is sickle-cell anemia, a disease caused by a change in one amino acid in each of two of the four chains of hemoglobin.

4.3 What Is the Secondary Structure of Proteins? Secondary structure is the hydrogen-bonded arrangement in space of the backbone, the polypeptide chain. Some of the most important backbone arrangements are the α-helix, the β-sheet, and the β-turn. They can be combined in a number of ways to produce structural motifs that occur in many proteins.

4.4 What Can We Say about the Thermodynamics of Protein Folding? The higher-order (secondary and tertiary) levels of structure depend primarily on noncovalent interactions, including hydrogen bonds, hydrophobic interactions, electrostatic interactions, and complexation of metal ions. Hydrophobic interactions, which depend on the unfavorable entropy of the water of hydration surrounding nonpolar solutes, are particularly important determinants of protein folding.

4.5 What Is the Tertiary Structure of Proteins? Tertiary structure includes the three-dimensional arrangement of *all* the atoms in the protein. The three-dimensional structures of proteins can be completely disrupted and, under proper experimental conditions, completely recovered. This process of denaturation and refolding is a dramatic example of the relationship between the primary structure of the protein and the forces that determine the tertiary structure. The secondary and tertiary structures of a protein can be determined simultaneously by X-ray crystallography. The oxygen-storage protein myoglobin was the first protein for which the complete tertiary structure was determined by crystallography.

4.6 Can We Predict Protein Folding from Sequence? It is possible, to some extent, to predict the three-dimensional structure of a protein from its amino acid sequence. Computer algorithms are based on two approaches, one of which is based on comparison of sequences with those of proteins whose folding pattern is known. Another one is based on the folding motifs that occur in many proteins.

4.7 What Is the Quaternary Structure of Proteins? Quaternary structure is the arrangement of subunits in multisubunit proteins. The individual polypeptide chains of multisubunit proteins interact with one another noncovalently. As a result, subtle changes in structure at one site on the molecule can cause drastic changes in properties at a distant site. Proteins that exhibit this property are referred to as allosteric. The properties of the allosteric protein hemoglobin can be contrasted with those of myoglobin, which is not allosteric. In hemoglobin, an oxygen-transport protein, the binding of oxygen is cooperative (as each oxygen is bound, it becomes easier for the next one to bind) and is modulated by such ligands as H^+, CO_2, and BPG. The binding of oxygen to myoglobin is not cooperative.

Critical Questions to Review

4.1 How Does the Structure of Proteins Determine Their Function?

1. **Fact Check** Match the following statements about protein structure with the proper levels of organization.

(a) Primary structure	(1) Three-dimensional arrangement of all atoms
(b) Secondary structure	(2) The order of amino acid residues in the polypeptide chain
(c) Tertiary structure	(3) The interaction between subunits in proteins that consist of more than one polypeptide chain
(d) Quaternary structure	(4) The hydrogen-bonded arrangement of the polypeptide backbone

2. **Fact Check** Define denaturation in terms of the effects of secondary, tertiary, and quaternary structure.
3. **Fact Check** What is the nature of "random" structure in proteins?

4.2 What Is the Primary Structure of Proteins?

4. **Thought Question** Suggest an explanation for the observation that, when proteins are chemically modified so that specific side chains have a different chemical nature, these proteins cannot be denatured reversibly.
5. **Thought Question** Rationalize the following observations.
   (a) Serine is the amino acid residue that can be replaced with the least effect on protein structure and function.
   (b) Replacement of tryptophan causes the greatest effect on protein structure and function.
   (c) Replacements such as Lys → Arg and Leu → Ile usually have very little effect on protein structure and function.
6. **Thought Question** Glycine is a highly conserved amino acid residue in proteins (i.e., it is found in the same position in the primary structure of related proteins). Suggest a reason why this might occur.
7. **Thought Question** A mutation that changes an alanine residue in a protein to an isoleucine leads to a loss of activity. Activity is regained when a further mutation at the same site changes the isoleucine to a glycine. Why?
8. **Thought Question** A biochemistry student characterizes the process of cooking meat as an exercise in denaturing proteins. Comment on the validity of this remark.
9. **Biochemical Connections** Severe combined immunodeficiency disease (SCID) is characterized by the complete lack of an immune system. Strains of mice have been developed that have SCID. When SCID mice that carry genetic predisposition to prion diseases are infected with PrP^{sc}, they do not develop prion diseases. How do these facts relate to the transmission of prion diseases?
10. **Biochemical Connections** An isolated strain of sheep was found in New Zealand. Most of these sheep carried the gene for predisposition to scrapie, yet none of them ever came down with the disease. How do these facts relate to the transmission of prion diseases?

4.3 What Is the Secondary Structure of Proteins?

11. **Fact Check** List three major differences between fibrous and globular proteins.
12. **Biochemical Connections** What is a protein efficiency ratio?
13. **Biochemical Connections** Which food has the highest PER?
14. **Biochemical Connections** What are the essential amino acids?
15. **Biochemical Connections** Why are scientists currently trying to create genetically modified foods?
16. **Fact Check** What are Ramachandran angles?
17. **Fact Check** What is a β-bulge?
18. **Fact Check** What is a reverse turn? Draw two types of reverse turns.
19. **Fact Check** List some of the differences between the α-helix and β-sheet forms of secondary structure.
20. **Fact Check** List some of the possible combinations of α-helices and β-sheets in supersecondary structures.
21. **Fact Check** Why is proline frequently encountered at the places in the myoglobin and hemoglobin molecules where the polypeptide chain turns a corner?
22. **Fact Check** Why must glycine be found at regular intervals in the collagen triple helix?
23. **Thought Question** You hear the comment that the difference between wool and silk is the difference between helical and pleated-sheet structures. Do you consider this a valid point of view? Why or why not?
24. **Thought Question** Woolen clothing shrinks when washed in hot water, but items made of silk do not. Suggest a reason, based on information from this chapter.

4.4 What Can We Say about the Thermodynamics of Protein Folding?

25. **Fact Check** List five forces that are responsible for maintaining the correct three-dimensional shapes of proteins. Specify which groups on the protein are involved in each type of interaction.
26. **Thought Question** Comment on the energetics of protein folding in light of the information in this chapter.

4.5 What Is the Tertiary Structure of Proteins?

27. **Fact Check** Draw two hydrogen bonds, one that is part of a secondary structure and another that is part of a tertiary structure.
28. **Fact Check** Draw a possible electrostatic interaction between two amino acids in a polypeptide chain.
29. **Fact Check** Draw a disulfide bridge between two cysteines in a polypeptide chain.
30. **Fact Check** Draw a region of a polypeptide chain showing a hydrophobic pocket containing nonpolar side chains.
31. **Fact Check** What is a chaperone?
32. **Thought Question** The terms *configuration* and *conformation* appear in descriptions of molecular structure. How do they differ?
33. **Thought Question** Theoretically, a protein could assume a virtually infinite number of configurations and conformations. Suggest several features of proteins that drastically limit the actual number.
34. **Thought Question** What is the highest level of protein structure found in collagen?

4.6 Can We Predict Protein Folding from Sequence?

35. **Thought Question** You have discovered a new protein, one whose sequence has about 25% homology with ribonuclease A. How would you go about predicting, rather than experimentally determining, its tertiary structure?
36. **Thought Question** Go to the RCSB site for the Protein Data Bank (http://www.rcsb.org/pdb). Give a brief description of the molecule prefoldin, which can be found under *chaperones*.

4.7 What Is the Quaternary Structure of Proteins?

37. **Biochemical Connections** What is a prion?
38. **Biochemical Connections** What are the known diseases caused by abnormal prions?
39. **Biochemical Connections** What are the protein secondary structures that differ between a normal prion and an infectious one?
40. **Fact Check** List two similarities and two differences between hemoglobin and myoglobin.
41. **Fact Check** What are the two critical amino acids near the heme group in both myoglobin and hemoglobin?
42. **Fact Check** What is the highest level of organization in myoglobin? In hemoglobin?
43. **Fact Check** Suggest a way in which the difference between the functions of hemoglobin and myoglobin is reflected in the shapes of their respective oxygen-binding curves.
44. **Fact Check** Describe the Bohr effect.
45. **Fact Check** Describe the effect of 2,3-*bis*phosphoglycerate on the binding of oxygen by hemoglobin.
46. **Fact Check** How does the oxygen-binding curve of fetal hemoglobin differ from that of adult hemoglobin?
47. **Fact Check** What is the critical amino acid difference between the β-chain and the γ-chain of hemoglobin?
48. **Thought Question** In oxygenated hemoglobin, pK_a = 6.6 for the histidines at position 146 on the β-chain. In deoxygenated hemoglobin, the pK_a of these residues is 8.2. How can this piece of information be correlated with the Bohr effect?
49. **Thought Question** You are studying with a friend who is in the process of describing the Bohr effect. She tells you that, in the lungs, hemoglobin binds oxygen and releases hydrogen ion; as a result, the pH increases. She goes on to say that, in actively metabolizing muscle tissue, hemoglobin releases oxygen and binds hydrogen ion and, as a result, the pH decreases. Do you agree with her reasoning? Why or why not?
50. **Thought Question** How does the difference between the β-chain and the γ-chain of hemoglobin explain the differences in oxygen binding between Hb A and Hb F?
51. **Thought Question** Suggest a reason for the observation that persons with sickle-cell trait sometimes have breathing problems during high-altitude flights.
52. **Thought Question** Does a fetus homozygous for Hb S have normal Hb F?
53. **Thought Question** Why is fetal Hb essential for the survival of placental animals?
54. **Thought Question** Why might you expect to find some Hb F in adults who are afflicted with sickle-cell anemia?
55. **Thought Question** When deoxyhemoglobin was first isolated in crystalline form, the researcher who did so noted that the crystals changed color from purple to red and also changed shape as he observed them under a microscope. What is happening on the molecular level? *Hint:* The crystals were mounted on a microscope slide with a *loosely* fitting cover slip.

BiochemistryNow™
Assess your understanding of this chapter's topics with additional quizzing and tutorials at **http://now.brookscole.com/campbell5**

Annotated Bibliography

Ferguson, N. M., A. C. Ghan, C. A. Donnelly, T. J. Hagenaars, and R. M. Anderson. Estimating the Human Health Risk from Possible BSE Infection of the British Sheep Flock. *Nature* **415,** 420–424 (2002). [The title says it all.]

Gibbons, A., and M. Hoffman. New 3-D Protein Structures Revealed. *Science* **253,** 382–383 (1991). [Examples of the use of X-ray crystallography to determine protein structure.]

Gierasch, L. M., and J. King, eds. *Protein Folding: Deciphering the Second Half of the Genetic Code*. Waldorf, Md.: AAAS Books, 1990. [A collection of articles on recent discoveries about the processes involved in protein folding. Experimental methods for studying protein folding are emphasized.]

Hall, S. Protein Images Update Natural History. *Science* **267,** 620–624 (1995). [Combining X-ray crystallography and computer software to produce images of protein structure.]

Hauptmann, H. The Direct Methods of X-ray Crystallography. *Science* **233,** 178–183 (1986). [A discussion of improvements in methods of doing the calculations involved in determining protein structure; based on a Nobel Prize address. This article should be read in connection with the one by Karle, and it provides an interesting contrast with the articles by Perutz, both of which describe early milestones in protein crystallography.]

Helfand, S. L. Chaperones Take Flight. *Science* **295,** 809–810 (2002). [An article about using chaperones to combat Parkinson's disease.]

Holm, L., and C. Sander. Mapping the Protein Universe. *Science* **273,** 595–602 (1996). [An article on searching databases on protein structure to predict the three-dimensional structure of proteins. Part of a series of articles on computers in biology.]

Karle, J. Phase Information from Intensity Data. *Science* **232,** 837–843 (1986). [A Nobel Prize address on the subject of X-ray crystallography. See remarks on the article by Hauptmann.]

Kasha, K. J. Biotechnology and the World Food Supply. *Genome* **42** (4), 642–645 (1999). [Proteins are frequently in short supply in the diet of many people in the world, but biotechnology can help improve the situation.]

Mitten, D. D., R. MacDonald, and D. Klonus. Regulation of Foods Derived from Genetically Engineered Crops. *Curr. Opin. Biotechnol.* **10,** 298–302 (1999). [How genetic engineering can affect the food supply, especially that of proteins.]

O'Quinn, P. R., J. L. Nelssen, R. D. Goodband, D. A. Knabe, J. C. Woodworth, M. D. Tokach, and T. T. Lohrmann. Nutritional Value of a Genetically Improved High-Lysine, High-Oil Corn for Young Pigs. *J. Anim. Sci.* **78** (8), 2144–2149 (2000). [The availability of amino acids affects the proteins formed.]

Peretz, D., R. A. Williamson, K. Kaneko, J. Vergara, E. Leclerc, G. Schmitt-Ulms, I. R. Mehlhorn, G. Legname, M. R. Wormald, P. M. Rudd, R. A. Dwek, D. R. Burton, and S. B. Prusiner. Antibodies Inhibit Prion Propagation and Clear Cell Cultures of Prion Infectivity. *Nature* **412,** 739–742 (2001). [Description of a possible treatment for prion diseases.]

Perutz, M. The Hemoglobin Molecule. *Sci. Amer.* **211** (5), 64–76 (1964). [A description of work that led to a Nobel Prize.]

Perutz, M. The Hemoglobin Molecule and Respiratory Transport. *Sci. Amer.* **239** (6), 92–125 (1978). [The relationship between molecular structure and cooperative binding of oxygen.]

Ruibal-Mendieta, N. L., and F. A. Lints. Novel and Transgenic Food Crops: Overview of Scientific versus Public Perception. *Transgenic Res.* **7** (5), 379–386 (1998). [A practical application of protein structure research.]

Yam, P. Mad Cow Disease's Human Toll. *Sci. Amer.* **284** (5), 12–13 (2001). [An overview of mad-cow disease and how it has crossed over to infect people.]