Growing in size and complexity

Living things, masses of atoms, DNA, protein

Dancing a pattern ever more intricate.

Out of the cradle onto the dry land

Here it is standing

Atoms with consciousness

Matter with curiosity.

Stands at the sea

Wonders at wondering

I

A universe of atoms

An atom in the universe.

RICHARD P. FEYNMAN (1918 – 1988)
From *"The Value of Science"* in Edward Hutchings, Jr. (ed.), *Frontiers of Science: A Survey.* New York: Basic Books, 1958.

*C*hapter 6

Proteins: Secondary, Tertiary, and Quaternary Structure

Like the Greek sea god Proteus, who could assume different forms, proteins act through changes in conformation. Proteins (from the Greek proteios, *meaning "primary") are the primary agents of biological function. ("Proteus, Old Man of the Sea, Roman period mosaic, from Thessalonika, 1st century A.D. National Archaeological Museum, Athens/Ancient Art and Architecture Collection Ltd./Bridgeman Art Library, London/New York.)*

Nearly all biological processes involve the specialized functions of one or more protein molecules. Proteins function to produce other proteins, control all aspects of cellular metabolism, regulate the movement of various molecular and ionic species across membranes, convert and store cellular energy, and carry out many other activities. Essentially all of the information required to initiate, conduct, and regulate each of these functions must be contained in

the structure of the protein itself. The previous chapter described the details of primary protein structure. However, proteins do not normally exist as fully extended polypeptide chains but rather as compact, folded structures, and the function of a given protein is rarely if ever dependent only on the amino acid sequence. Instead, the ability of a particular protein to carry out its function in nature is normally determined by its overall three-dimensional shape or *conformation*. This native, folded structure of the protein is dictated by several factors: (1) interactions with solvent molecules (normally water), (2) the pH and ionic composition of the solvent, and most important, (3) the sequence of the protein. The first two of these effects are intuitively reasonable, but the third, the role of the amino acid sequence, may not be. In ways that are just now beginning to be understood, the primary structure facilitates the development of short-range interactions among adjacent parts of the sequence and also long-range interactions among distant parts of the sequence. Although the resulting overall structure of the complete protein molecule may at first look like a disorganized and random arrangement, it is in nearly all cases a delicate and sophisticated balance of numerous forces that combine to determine the protein's unique conformation. This chapter considers the details of protein structure and the forces that maintain these structures.

6.1 • Forces Influencing Protein Structure

Several different kinds of noncovalent interactions are of vital importance in protein structure. Hydrogen bonds, hydrophobic interactions, electrostatic bonds, and van der Waals forces are all noncovalent in nature, yet are extremely important influences on protein conformations. The stabilization free energies afforded by each of these interactions may be highly dependent on the local environment within the protein, but certain generalizations can still be made.

Hydrogen Bonds

Hydrogen bonds are generally made wherever possible within a given protein structure. In most protein structures that have been examined to date, component atoms of the peptide backbone tend to form hydrogen bonds with one another. Furthermore, side chains capable of forming H bonds are usually located on the protein surface and form such bonds primarily with the water solvent. Although each hydrogen bond may contribute an average of only about 12 kJ/mol in stabilization energy for the protein structure, the number of H-bonds formed in the typical protein is very large. For example, in α-helices, the C=O and N—H groups of every residue participate in H bonds. The importance of H bonds in protein structure cannot be overstated.

Hydrophobic Interactions

Hydrophobic "bonds," or, more accurately, *interactions,* form because nonpolar side chains of amino acids and other nonpolar solutes prefer to cluster in a nonpolar environment rather than to intercalate in a polar solvent such as water. The forming of hydrophobic bonds minimizes the interaction of nonpolar residues with water and is therefore highly favorable. Such clustering is entropically driven. The side chains of the amino acids in the interior or core of the protein structure are almost exclusively hydrophobic. Polar amino acids are almost never found in the interior of a protein, but the protein surface may consist of both polar and nonpolar residues.

intercalate • to insert between

FIGURE 6.1 • An electrostatic interaction between the ε-amino group of a lysine and the γ-carboxyl group of a glutamate residue.
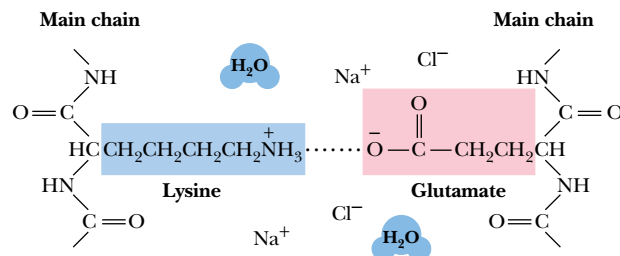
Electrostatic Interactions

Ionic interactions arise either as electrostatic attractions between opposite charges or repulsions between like charges. Chapter 4 discusses the ionization behavior of amino acids. Amino acid side chains can carry positive charges, as in the case of lysine, arginine, and histidine, or negative charges, as in aspartate and glutamate. In addition, the NH_2-terminal and COOH-terminal residues of a protein or peptide chain usually exist in ionized states and carry positive or negative charges, respectively. All of these may experience electrostatic interactions in a protein structure. Charged residues are normally located on the protein surface, where they may interact optimally with the water solvent. It is energetically unfavorable for an ionized residue to be located in the hydrophobic core of the protein. Electrostatic interactions between charged groups on a protein surface are often complicated by the presence of salts in the solution. For example, the ability of a positively charged lysine to attract a nearby negative glutamate may be weakened by dissolved NaCl (Figure 6.1). The Na^+ and Cl^- ions are highly mobile, compact units of charge, compared to the amino acid side chains, and thus compete effectively for charged sites on the protein. In this manner, electrostatic interactions among amino acid residues on protein surfaces may be damped out by high concentrations of salts. Nevertheless, these interactions are important for protein stability.

Van der Waals Interactions

Both attractive forces and repulsive forces are included in van der Waals interactions. The attractive forces are due primarily to instantaneous dipole-induced dipole interactions that arise because of fluctuations in the electron charge distributions of adjacent nonbonded atoms. Individual van der Waals interactions are weak ones (with stabilization energies of 4.0 to 1.2 kJ/mol), but many such interactions occur in a typical protein, and, by sheer force of numbers, they can represent a significant contribution to the stability of a protein. Peter Privalov and George Makhatadze have shown that, for pancreatic ribonuclease A, hen egg white lysozyme, horse heart cytochrome *c,* and sperm whale myoglobin, van der Waals interactions between tightly packed groups in the interior of the protein are a major contribution to protein stability.

6.2 • Role of the Amino Acid Sequence in Protein Structure

It can be inferred from the first section of this chapter that many different forces work together in a delicate balance to determine the overall three-dimensional structure of a protein. These forces operate both within the protein structure itself and between the protein and the water solvent. How, then, does nature dictate the manner of protein folding to generate the three-dimensional

structure that optimizes and balances these many forces? *All of the information necessary for folding the peptide chain into its "native" structure is contained in the amino acid sequence of the peptide.* This principle was first appreciated by C. B. Anfinsen and F. White, whose work in the early 1960s dealt with the chemical denaturation and subsequent renaturation of bovine pancreatic ribonuclease. Ribonuclease was first denatured with urea and mercaptoethanol, a treatment that cleaved the four covalent disulfide (S—S) cross-bridges in the protein. Subsequent air oxidation permitted random formation of disulfide cross-bridges, most of which were incorrect. Thus, the air-oxidized material showed little enzymatic activity. However, treatment of these inactive preparations with small amounts of mercaptoethanol allowed a reshuffling of the disulfide bonds and permitted formation of significant amounts of active native enzyme. In such experiments, the only road map for the protein, that is, the only "instructions" it has, are those directed by its primary structure, the linear sequence of its amino acid residues.

Just how proteins recognize and interpret the information that is stored in the polypeptide sequence is not well understood yet. It may be assumed that certain loci along the peptide chain act as nucleation points, which initiate folding processes that eventually lead to the correct structures. Regardless of how this process operates, it must take the protein correctly to the final native structure, without getting trapped in a local energy-minimum state which, although stable, may be different from the native state itself. A long-range goal of many researchers in the protein structure field is the prediction of three-dimensional conformation from the amino acid sequence. As the details of secondary and tertiary structure are described in this chapter, the complexity and immensity of such a prediction will be more fully appreciated. This area is perhaps the greatest uncharted frontier remaining in molecular biology.

6.3 • Secondary Structure in Proteins

Any discussion of protein folding and structure must begin with the *peptide bond,* the fundamental structural unit in all proteins. As we saw in Chapter 5, the resonance structures experienced by a peptide bond constrain the oxygen, carbon, nitrogen, and hydrogen atoms of the peptide group, as well as the adjacent α-carbons, to all lie in a plane. The resonance stabilization energy of this planar structure is approximately 88 kJ/mol, and substantial energy is required to twist the structure about the C—N bond. A twist of θ degrees involves a twist energy of $88 \sin^2\theta$ kJ/mol.

Consequences of the Amide Plane

The planarity of the peptide bond means that there are only two degrees of freedom per residue for the peptide chain. Rotation is allowed about the bond linking the α-carbon and the carbon of the peptide bond and also about the bond linking the nitrogen of the peptide bond and the adjacent α-carbon. As shown in Figure 6.2, each α-carbon is the joining point for two planes defined by peptide bonds. The angle about the C_α—N bond is denoted by the Greek letter ϕ (phi) and that about the C_α—C_o is denoted by ψ (psi). For either of these bond angles, a value of 0° corresponds to an orientation with the amide plane bisecting the H—C_α—R (sidechain) plane and a *cis* configuration of the main chain around the rotating bond in question (Figure 6.3). In any case, the entire path of the peptide backbone in a protein is known if the ϕ and ψ rotation angles are all specified. Some values of ϕ and ψ are not allowed due to steric interference between nonbonded atoms. As shown in Figure 6.4, values
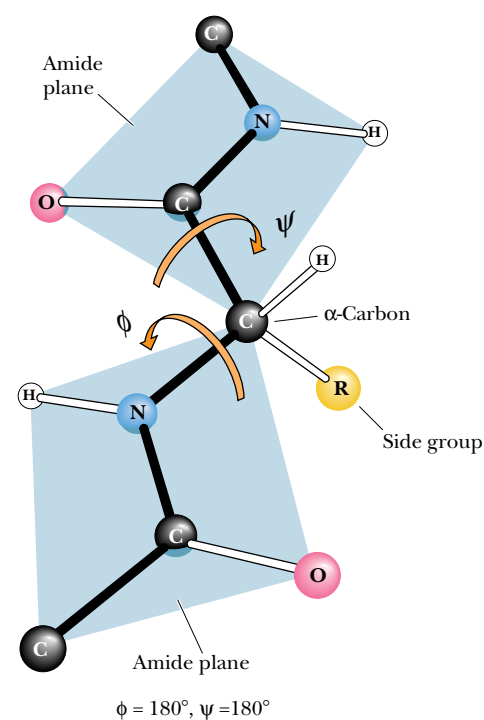

$\phi = 180°, \psi = 180°$

FIGURE 6.2 • The amide or peptide bond planes are joined by the tetrahedral bonds of the α-carbon. The rotation parameters are ϕ and ψ. The conformation shown corresponds to $\phi = 180°$ and $\psi = 180°$. Note that positive values of ϕ and ψ correspond to clockwise rotation as viewed from C_α. Starting from 0°, a rotation of 180° in the clockwise direction $(+180°)$ is equivalent to a rotation of 180° in the counterclockwise direction $(-180°)$. *(Irving Geis)*
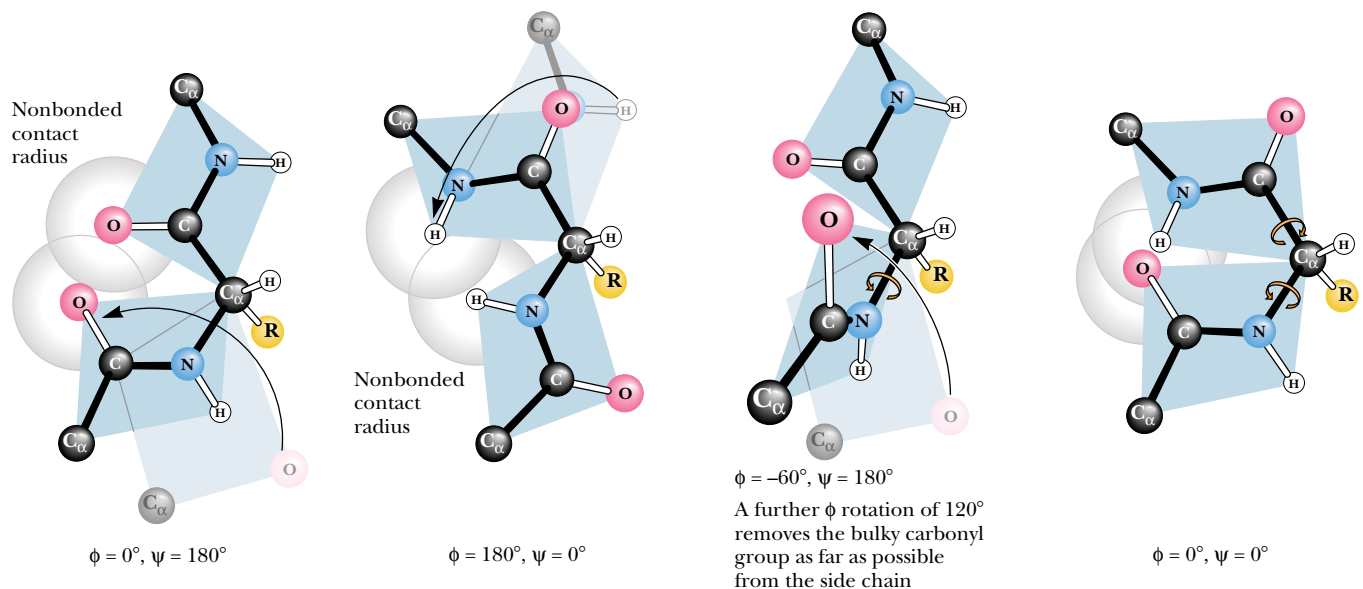
Nonbonded contact radius

$\phi = 0°$, $\psi = 180°$

Nonbonded contact radius

$\phi = 180°$, $\psi = 0°$

$\phi = -60°$, $\psi = 180°$

A further ϕ rotation of 120° removes the bulky carbonyl group as far as possible from the side chain

$\phi = 0°$, $\psi = 0°$

FIGURE 6.3 • Many of the possible conformations about an α-carbon between two peptide planes are forbidden because of steric crowding. Several noteworthy examples are shown here.

Note: The formal IUPAC-IUB Commission on Biochemical Nomenclature convention for the definition of the torsion angles ϕ and ψ in a polypeptide chain (*Biochemistry* **9:**3471–3479, 1970) is different from that used here, where the C_α atom serves as the point of reference for both rotations, but the result is the same. *(Irving Geis)*

of $\phi = 180°$ and $\psi = 0°$ are not allowed because of the forbidden overlap of the N—H hydrogens. Similarly, $\phi = 0°$ and $\psi = 180°$ are forbidden because of unfavorable overlap between the carbonyl oxygens.

G. N. Ramachandran and his coworkers in Madras, India, first showed that it was convenient to plot ϕ values against ψ values to show the distribution of allowed values in a protein or in a family of proteins. A typical **Ramachandran plot** is shown in Figure 6.4. Note the clustering of ϕ and ψ values in a few regions of the plot. Most combinations of ϕ and ψ are sterically forbidden, and the corresponding regions of the Ramachandran plot are sparsely populated. The combinations that are sterically allowed represent the subclasses of structure described in the remainder of this section.

The Alpha-Helix

The discussion of hydrogen bonding in Section 6.1 pointed out that the carbonyl oxygen and amide hydrogen of the peptide bond could participate in H bonds either with water molecules in the solvent or with other H-bonding groups in the peptide chain. In nearly all proteins, the carbonyl oxygens and the amide protons of many peptide bonds participate in H bonds that link one peptide group to another, as shown in Figure 6.5. These structures tend to form in cooperative fashion and involve substantial portions of the peptide chain. Structures resulting from these interactions constitute **secondary structure** for proteins (see Chapter 5). When a number of hydrogen bonds form between portions of the peptide chain in this manner, two basic types of structures can result: α-*helices* and β-*pleated sheets*.
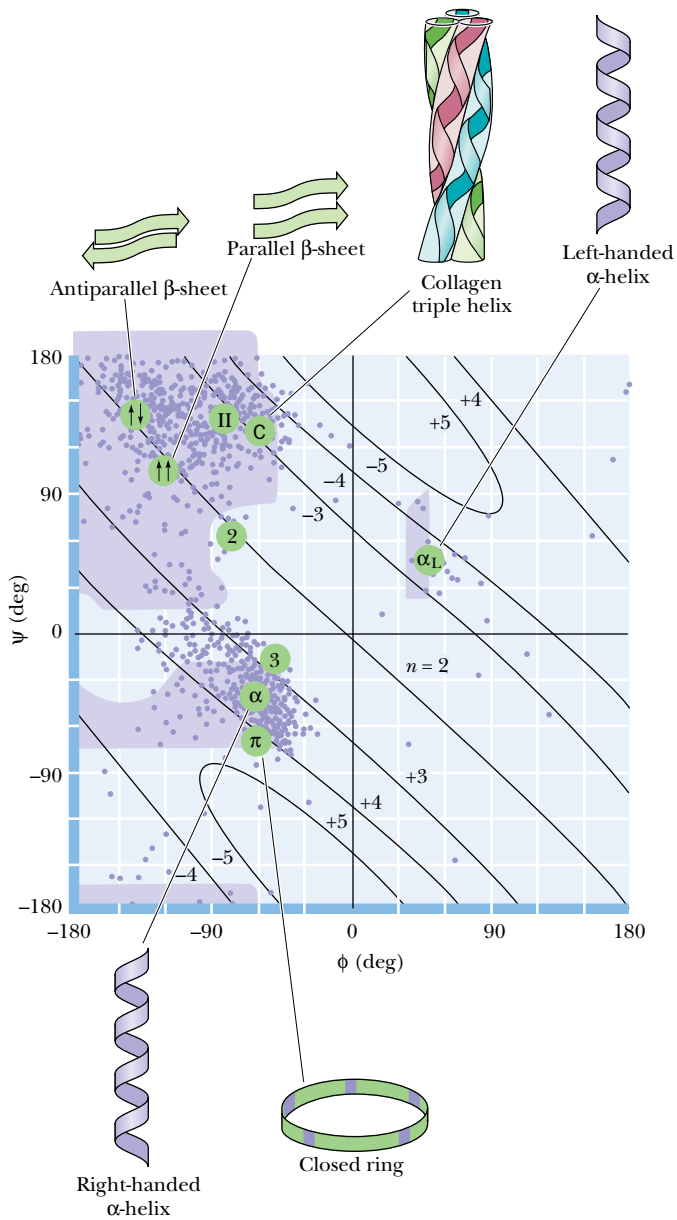
FIGURE 6.4 • A Ramachandran diagram showing the sterically reasonable values of the angles ϕ and ψ. The shaded regions indicate particularly favorable values of these angles. Dots in purple indicate actual angles measured for 1000 residues (excluding glycine, for which a wider range of angles is permitted) in eight proteins. The lines running across the diagram (numbered +5 through 2 and −5 through −3) signify the number of amino acid residues per turn of the helix; "+" means right-handed helices; "−" means left-handed helices. *(After Richardson, J. S., 1981, Advances in Protein Chemistry **34**:167–339.)*

FIGURE 6.5 • A hydrogen bond between the amide proton and carbonyl oxygen of adjacent peptide groups.
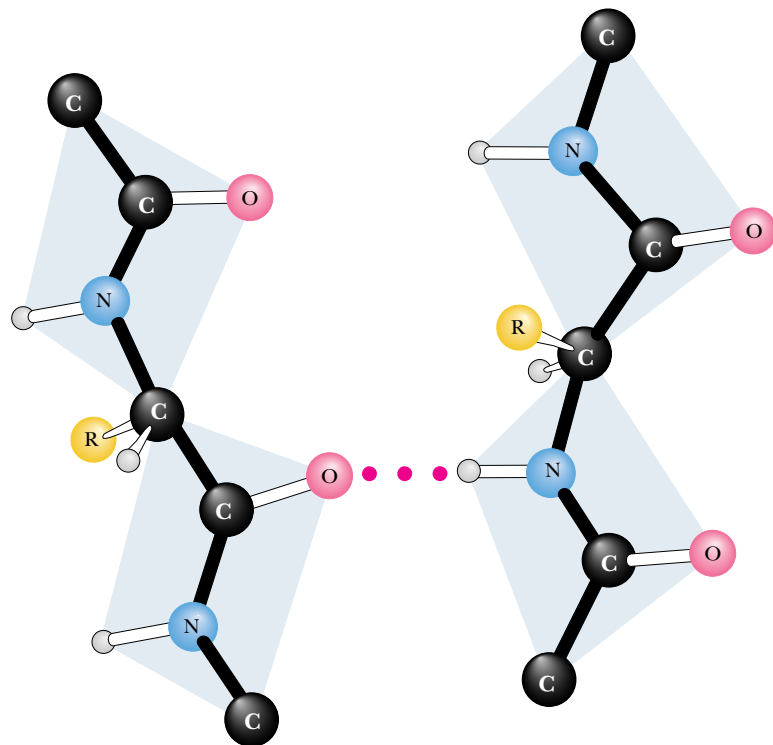
𝒜 D E E P E R   L O O K

Knowing What the Right Hand and Left Hand Are Doing

Certain conventions related to peptide bond angles and the "handedness" of biological structures are useful in any discussion of protein structure. To determine the ϕ and ψ angles between peptide planes, viewers should imagine themselves at the C_α carbon looking outward and should imagine starting from the $\phi = 0°$, $\psi = 0°$ conformation. From this perspective, positive values of ϕ correspond to clockwise rotations about the C_α—N bond of the plane that includes the adjacent N—H group. Similarly, positive values of ψ correspond to clockwise

rotations about the C_α—C bond of the plane that includes the adjacent C=O group.

Biological structures are often said to exhibit "right-hand" or "left-hand" twists. For all such structures, the sense of the twist can be ascertained by holding the structure in front of you and looking along the polymer backbone. If the twist is clockwise as one proceeds outward and through the structure, it is said to be right-handed. If the twist is counterclockwise, it is said to be left-handed.

Evidence for helical structures in proteins was first obtained in the 1930s in studies of fibrous proteins. However, there was little agreement at that time about the exact structure of these helices, primarily because there was also lack of agreement about interatomic distances and bond angles in peptides. In 1951, Linus Pauling, Robert Corey, and their colleagues at the California Institute of Technology summarized a large volume of crystallographic data in a set of dimensions for polypeptide chains. (A summary of data similar to what they reported is shown in Figure 5.2.) With these data in hand, Pauling, Corey, and their colleagues proposed a new model for a helical structure in proteins, which they called the **α-helix.** The report from Caltech was of particular interest to Max Perutz in Cambridge, England, a crystallographer who was also interested in protein structure. By taking into account a critical but previously ignored feature of the X-ray data, Perutz realized that the α-helix existed in keratin, a protein from hair, and also in several other proteins. Since then, the α-helix has proved to be a fundamentally important peptide structure. Several representations of the α-helix are shown in Figure 6.6. One turn of the helix represents 3.6 amino acid residues. (A single turn of the α-helix involves 13 atoms from the O to the H of the H bond. For this reason, the α-helix is sometimes referred to as the 3.6_{13} helix.) This is in fact the feature that most confused crystallographers before the Pauling and Corey α-helix model. Crystallographers were so accustomed to finding twofold, threefold, sixfold, and similar integral axes in simpler molecules that the notion of a nonintegral number of units per turn was never taken seriously before Pauling and Corey's work.

Each amino acid residue extends **1.5 Å (0.15 nm)** along the helix axis. With **3.6 residues per turn,** this amounts to 3.6×1.5 Å or **5.4 Å (0.54 nm)** of travel along the helix axis per turn. This is referred to as the translation distance or the **pitch of the helix.** If one ignores side chains, the helix is about 6 Å in diameter. The side chains, extending outward from the core structure of the helix, are removed from steric interference with the polypeptide backbone. As can be seen in Figure 6.6, *each peptide carbonyl is hydrogen bonded to the peptide N—H group four residues farther up the chain.* Note that all of the H bonds lie parallel to the helix axis and that all of the carbonyl groups are pointing in one direction along the helix axis while the N—H groups are pointing in the opposite direction. Recall that the entire path of the peptide backbone can be known if the ϕ and ψ twist angles are specified for each residue. The α-helix is formed if the values of ϕ are approximately $-60°$ and the values of ψ are in the range of -45 to $-50°$. Figure 6.7 shows the structures of two proteins that contain α-helical segments. The number of residues involved in a given α-helix

(a)
Hydrogen bonds stabilize
the helix structure.

(b)
The helix can be viewed
as a stacked array of
peptide planes hinged
at the α-carbons and
approximately parallel
to the helix.

α-Carbon

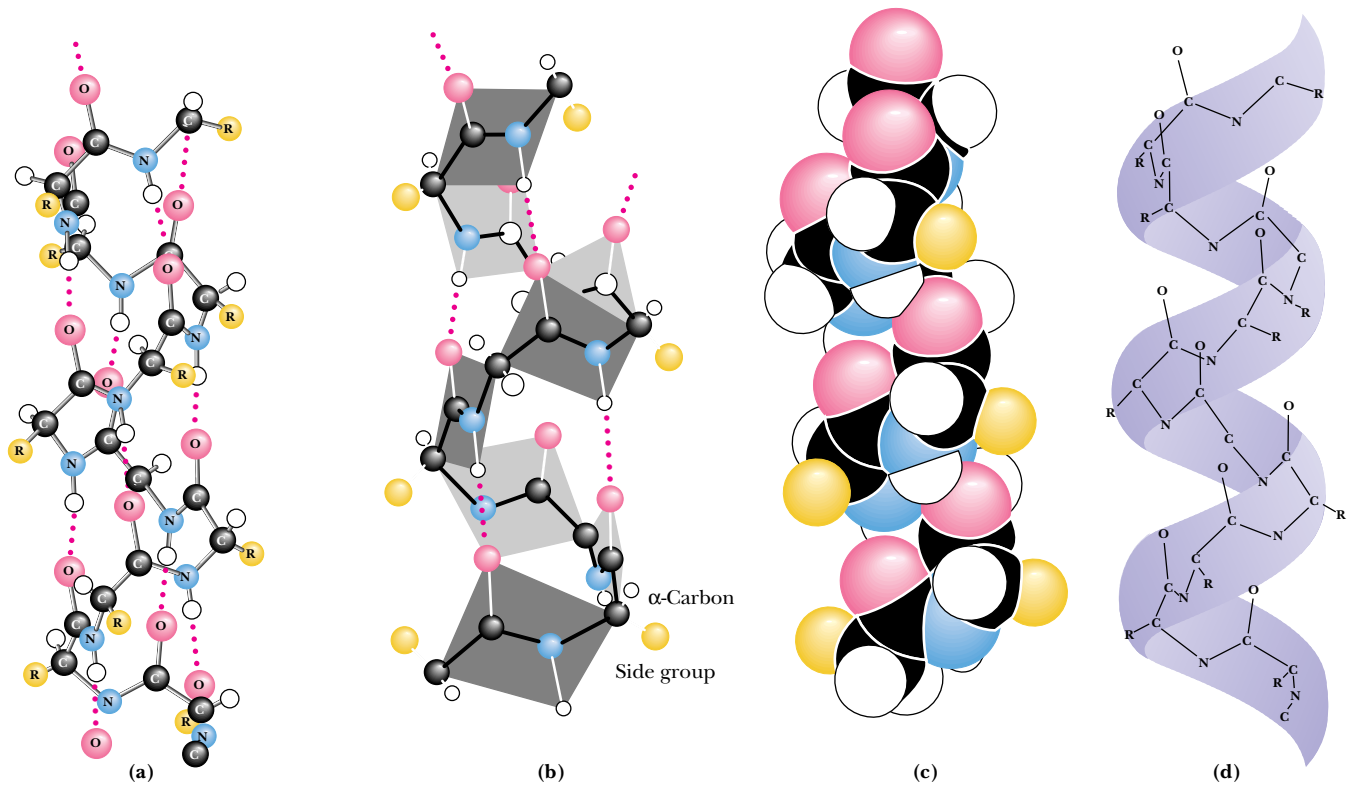Side group

(c)

(d)

FIGURE 6.6 • Four different graphic repre-
sentations of the α-helix. (a) As it originally
appeared in Pauling's 1960 *The Nature of the
Chemical Bond*. (b) Showing the arrangement of
peptide planes in the helix. (c) A space-filling
computer graphic presentation. (d) A "ribbon
structure" with an inlaid stick figure, showing
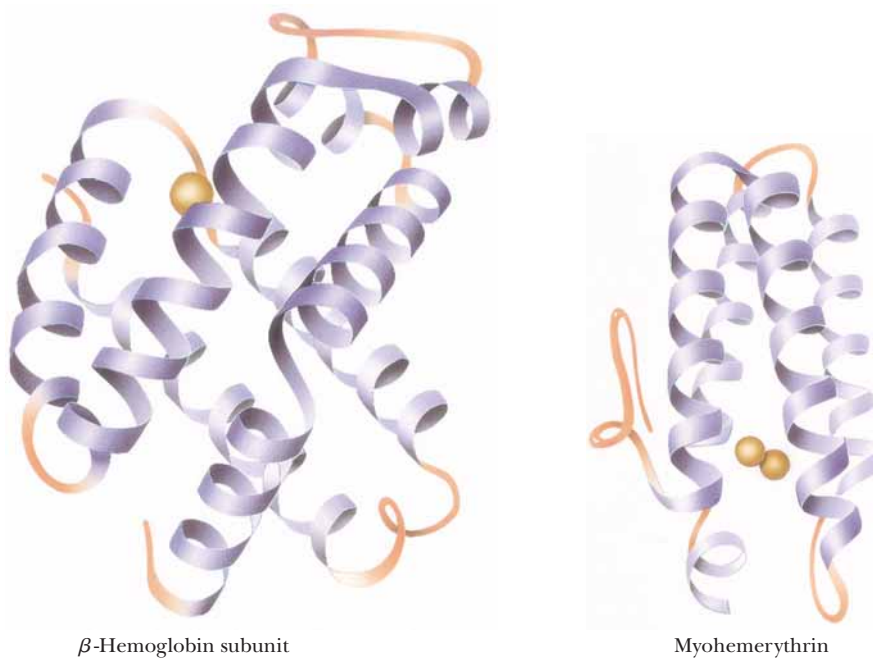how the ribbon indicates the path of the
polypeptide backbone. *(Irving Geis)*

β-Hemoglobin subunit

Myohemerythrin

FIGURE 6.7 • The three-dimensional struc-
tures of two proteins that contain substantial
amounts of α-helix in their structures. The
helices are represented by the regularly coiled
sections of the ribbon drawings. Myohemery-
thrin is the oxygen-carrying protein in certain
invertebrates, including *Sipunculids,* a phylum
of marine worm. *(Jane Richardson)*

(a)

(b)

varies from helix to helix and from protein to protein. On average, there are about 10 residues per helix. Myoglobin, one of the first proteins in which α-helices were observed, has eight stretches of α-helix that form a box to contain the heme prosthetic group. The structures of the α and β subunits of hemoglobin are strikingly similar, with only a few differences at the C- and N-termini and on the surfaces of the structure that contact or interact with the other subunits of this multisubunit protein.

As shown in Figure 6.6, all of the hydrogen bonds point in the same direction along the α-helix axis. Each peptide bond posesses a dipole moment that arises from the polarities of the N—H and C=O groups, and, because these groups are all aligned along the helix axis, the helix itself has a substantial dipole moment, with a partial positive charge at the N-terminus and a partial negative charge at the C-terminus (Figure 6.8). Negatively charged ligands (e.g., phosphates) frequently bind to proteins near the N-terminus of an α-helix. By contrast, positively charged ligands are only rarely found to bind near the C-terminus of an α-helix.

In a typical α-helix of 12 (or n) residues, there are 8 (or $n-4$) hydrogen bonds. As shown in Figure 6.9, the first 4 amide hydrogens and the last 4 carbonyl oxygens cannot participate in helix H-bonds. Also, nonpolar residues sit-

FIGURE 6.9 • Four N—H groups at the N-terminal end of an α-helix and four C=O groups at the C-terminal end cannot participate in hydrogen bonding. The formation of H-bonds with other nearby donor and acceptor groups is referred to as **helix capping.** Capping may also involve appropriate hydrophobic interactions that accomodate nonpolar side chains at the ends of helical segments.

uated near the helix termini can be exposed to solvent. Proteins frequently compensate for these problems by **helix capping**—providing H-bond partners for the otherwise bare N—H and C=O groups and folding other parts of the protein to foster hydrophobic contacts with exposed nonpolar residues at the helix termini.

Careful studies of the **polyamino acids,** polymers in which all the amino acids are identical, have shown that certain amino acids tend to occur in α-helices, whereas others are less likely to be found in them. Polyleucine and polyalanine, for example, readily form α-helical structures. In contrast, polyaspartic acid and polyglutamic acid, which are highly negatively charged at pH 7.0, form only random structures because of strong charge repulsion between the R groups along the peptide chain. At pH 1.5 to 2.5, however, where the side chains are protonated and thus uncharged, these latter species spontaneously form α-helical structures. In similar fashion, polylysine is a random coil at pH values below about 11, where repulsion of positive charges prevents helix formation. At pH 12, where polylysine is a neutral peptide chain, it readily forms an α-helix.

$\mathcal{C}$RITICAL DEVELOPMENTS IN BIOCHEMISTRY

In Bed with a Cold, Pauling Stumbles onto the α-Helix and a Nobel Prize[1]

As high technology continues to transform the modern biochemical laboratory, it is interesting to reflect on Linus Pauling's discovery of the α-helix. It involved only a piece of paper, a pencil, scissors, and a sick Linus Pauling, who had tired of reading detective novels. The story is told in the excellent book *The Eighth Day of Creation* by Horace Freeland Judson:

> From the spring of 1948 through the spring of 1951 . . . rivalry sputtered and blazed between Pauling's lab and (Sir Lawrence) Bragg's — over protein. The prize was to propose and verify in nature a general three-dimensional structure for the polypeptide chain. Pauling was working up from the simpler structures of components. In January 1948, he went to Oxford as a visiting professor for two terms, to lecture on the chemical bond and on molecular structure and biological specificity. "In Oxford, it was April, I believe, I caught cold. I went to bed, and read detective stories for a day, and got bored, and thought why don't I have a crack at that problem of alpha keratin." Confined, and still fingering the polypeptide chain in his mind, Pauling called for paper, pencil, and straightedge and attempted to reduce the problem to an almost Euclidean purity. "I took a sheet of paper — I still have this sheet of paper — and drew, rather roughly, the way that I thought a polypeptide chain would look if it were spread out into a plane." The repetitious herringbone of the chain he could stretch across the paper as simply as this —
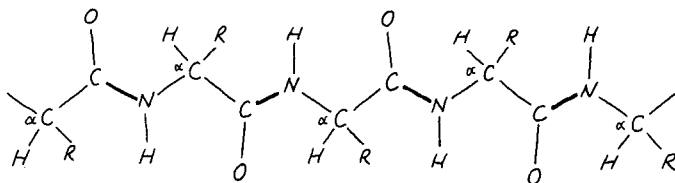
— putting in lengths and bond angles from memory. . . . He knew that the peptide bond, at the carbon-to-nitrogen link, was always rigid:
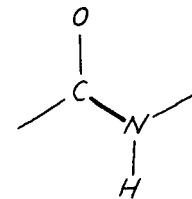
And this meant that the chain could turn corners only at the alpha carbons. . . . "I creased the paper in parallel creases through the alpha carbon atoms, so that I could bend it and make the bonds to the alpha carbons, along the chain, have tetrahedral value. And then I looked to see if I could form hydrogen bonds from one part of the chain to the next." He saw that if he folded the strip like a chain of paper dolls into a helix, and if he got the pitch of the screw right, hydrogen bonds could be shown to form, N—H···O—C, three or four knuckles apart along the backbone, holding the helix in shape. After several tries, changing the angle of the parallel creases in order to adjust the pitch of the helix, he found one where the hydrogen bonds would drop into place, connecting the turns, as straight lines of the right length. He had a model.

---

[1]The discovery of the α-helix structure was only one of many achievements that led to Pauling's Nobel Prize in chemistry in 1954. The official citation for the prize was "for his research into the nature of the chemical bond and its application to the elucidation of the structure of complex substances."

Table 6.1

Helix-Forming and Helix-Breaking Behavior of the Amino Acids

Amino Acid		Helix Behavior*	
A	Ala	H	(I)
C	Cys	Variable	
D	Asp	Variable	
E	Glu	H	
F	Phe	H	
G	Gly	I	(B)
H	His	H	(I)
I	Ile	H	(C)
K	Lys	Variable	
L	Leu	H	
M	Met	H	
N	Asn	C	(I)
P	Pro	B	
Q	Gln	H	(I)
R	Arg	H	(I)
S	Ser	C	(B)
T	Thr	Variable	
V	Val	Variable	
W	Trp	H	(C)
Y	Tyr	H	(C)

*H = helix former; I = indifferent; B = helix breaker; C = random coil; ( ) = secondary tendency.

The tendencies of the amino acids to stabilize or destabilize α-helices are different in typical proteins than in polyamino acids. The occurrence of the common amino acids in helices is summarized in Table 6.1. Notably, proline (and hydroxyproline) act as helix breakers due to their unique structure, which fixes the value of the C_α—N—C bond angle. Helices can be formed from either D- or L-amino acids, but a given helix must be composed entirely of amino acids of one configuration. α-Helices cannot be formed from a mixed copolymer of D- and L-amino acids. An α-helix composed of D-amino acids is left-handed.

Other Helical Structures

There are several other far less common types of helices found in proteins. The most common of these is the 3_{10} helix, which contains 3.0 residues per turn (with 10 atoms in the ring formed by making the hydrogen bond three residues up the chain). It normally extends over shorter stretches of sequence than the α-helix. Other helical structures include the 2_7 ribbon and the π-helix, which has 4.4 residues and 16 atoms per turn and is thus called the 4.4_{16} helix.

The Beta-Pleated Sheet

Another type of structure commonly observed in proteins also forms because of local, cooperative formation of hydrogen bonds. That is the pleated sheet, or β-structure, often called the **β-pleated sheet.** This structure was also first postulated by Pauling and Corey in 1951 and has now been observed in many natural proteins. A β-pleated sheet can be visualized by laying thin, pleated strips of paper side by side to make a "pleated sheet" of paper (Figure 6.10). Each strip of paper can then be pictured as a single peptide strand in which the peptide backbone makes a zigzag pattern along the strip, with the α-carbons lying at the folds of the pleats. The pleated sheet can exist in both par-
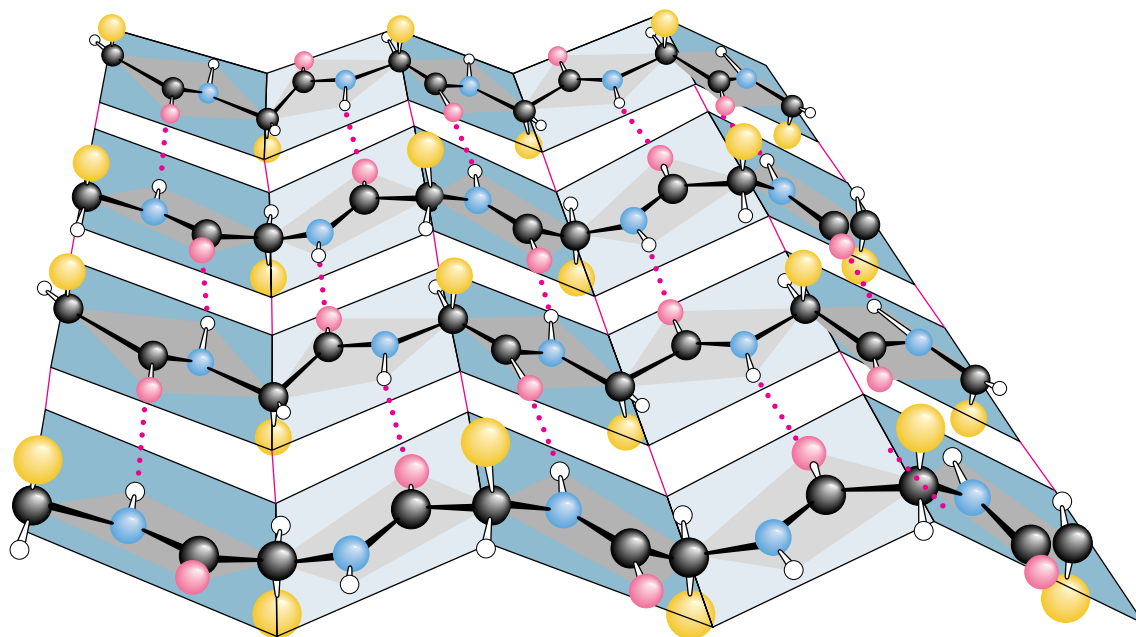

FIGURE 6.10 • A "pleated sheet" of paper with an antiparallel β-sheet drawn on it. *(Irving Geis)*

allel and antiparallel forms. In the **parallel β-pleated sheet,** adjacent chains run in the same direction (N → C or C → N). In the **antiparallel β-pleated sheet,** adjacent strands run in opposite directions.

Each single strand of the β-sheet structure can be pictured as a twofold helix, that is, a helix with two residues per turn. The arrangement of successive amide planes has a pleated appearance due to the tetrahedral nature of the C_α atom. It is important to note that the hydrogen bonds in this structure are essentially *inter*strand rather than *intra*strand. The peptide backbone in the β-sheet is in its most extended conformation (sometimes called the **ε-conformation**). The optimum formation of H bonds in the parallel pleated sheet results in a slightly less extended conformation than in the antiparallel sheet. The H bonds thus formed in the parallel β-sheet are bent significantly. The distance between residues is 0.347 nm for the antiparallel pleated sheet, but only 0.325 nm for the parallel pleated sheet. Figure 6.11 shows examples of both parallel and antiparallel β-pleated sheets. Note that the side chains in the pleated sheet are oriented perpendicular or normal to the plane of the sheet, extending out from the plane on alternating sides.

Parallel β-sheets tend to be more regular than antiparallel β-sheets. The range of ϕ and ψ angles for the peptide bonds in parallel sheets is much smaller than that for antiparallel sheets. Parallel sheets are typically large structures; those composed of less than five strands are rare. Antiparallel sheets, however, may consist of as few as two strands. Parallel sheets characteristically distribute
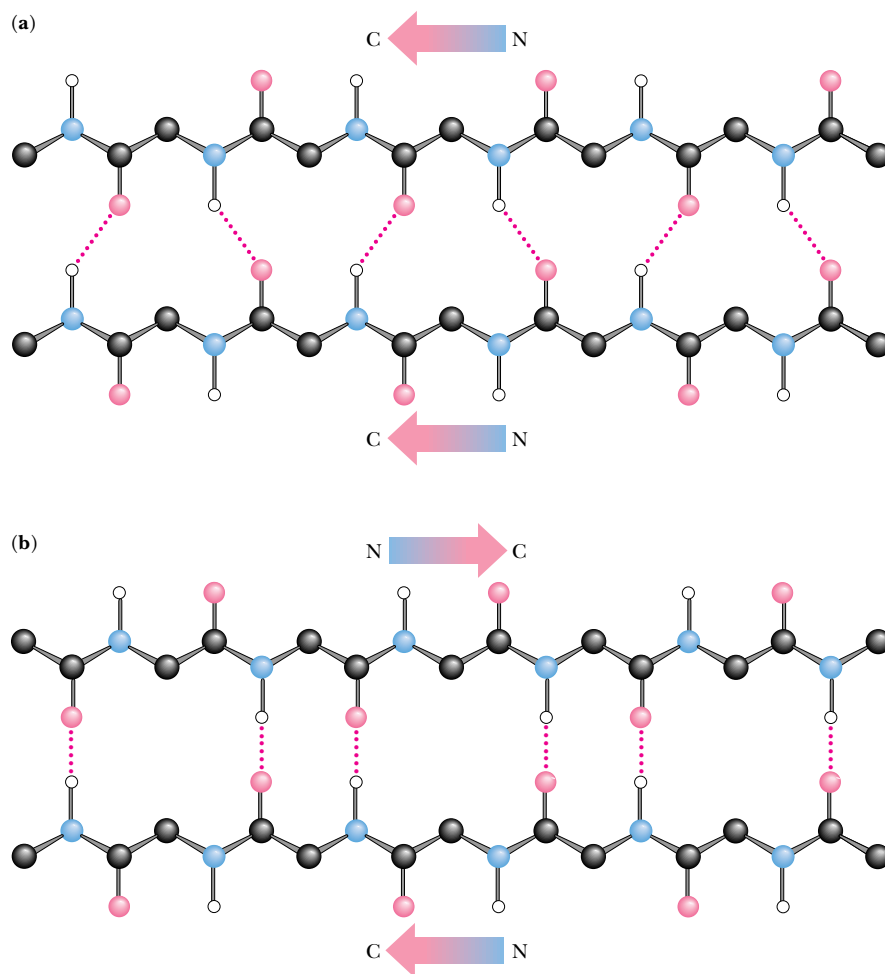
FIGURE 6.11 • The arrangement of hydrogen bonds in (a) parallel and (b) antiparallel β-pleated sheets.

hydrophobic side chains on both sides of the sheet, while antiparallel sheets are usually arranged with all their hydrophobic residues on one side of the sheet. This requires an alternation of hydrophilic and hydrophobic residues in the primary structure of peptides involved in antiparallel β-sheets because alternate side chains project to the same side of the sheet (see Figure 6.10).

Antiparallel pleated sheets are the fundamental structure found in silk, with the polypeptide chains forming the sheets running parallel to the silk fibers. The silk fibers thus formed have properties consistent with those of the β-sheets that form them. They are quite flexible but cannot be stretched or extended to any appreciable degree. Antiparallel structures are also observed in many other proteins, including immunoglobulin G, superoxide dismutase from bovine erythrocytes, and concanavalin A. Many proteins, including carbonic anhydrase, egg lysozyme, and glyceraldehyde phosphate dehydrogenase, possess both α-helices and β-pleated sheet structures within a single polypeptide chain.

The Beta-Turn

Most proteins are globular structures. The polypeptide chain must therefore possess the capacity to bend, turn, and reorient itself to produce the required compact, globular structures. A simple structure observed in many proteins is the **β-turn** (also known as the *tight turn* or *β-bend*), in which the peptide chain forms a tight loop with the carbonyl oxygen of one residue hydrogen-bonded with the amide proton of the residue three positions down the chain. This H bond makes the β-turn a relatively stable structure. As shown in Figure 6.12, the β-turn allows the protein to reverse the direction of its peptide chain. This figure shows the two major types of β-turns, but a number of less common types are also found in protein structures. Certain amino acids, such as proline and glycine, occur frequently in β-turn sequences, and the particular conformation of the β-turn sequence depends to some extent on the amino acids composing it. Due to the absence of a side chain, glycine is sterically the most adaptable of the amino acids, and it accommodates conveniently to other steric constraints in the β-turn. Proline, however, has a cyclic structure and a fixed ϕ angle, so, to some extent, it forces the formation of a β-turn, and in many cases this facilitates the turning of a polypeptide chain upon itself. Such bends promote formation of antiparallel β-pleated sheets.
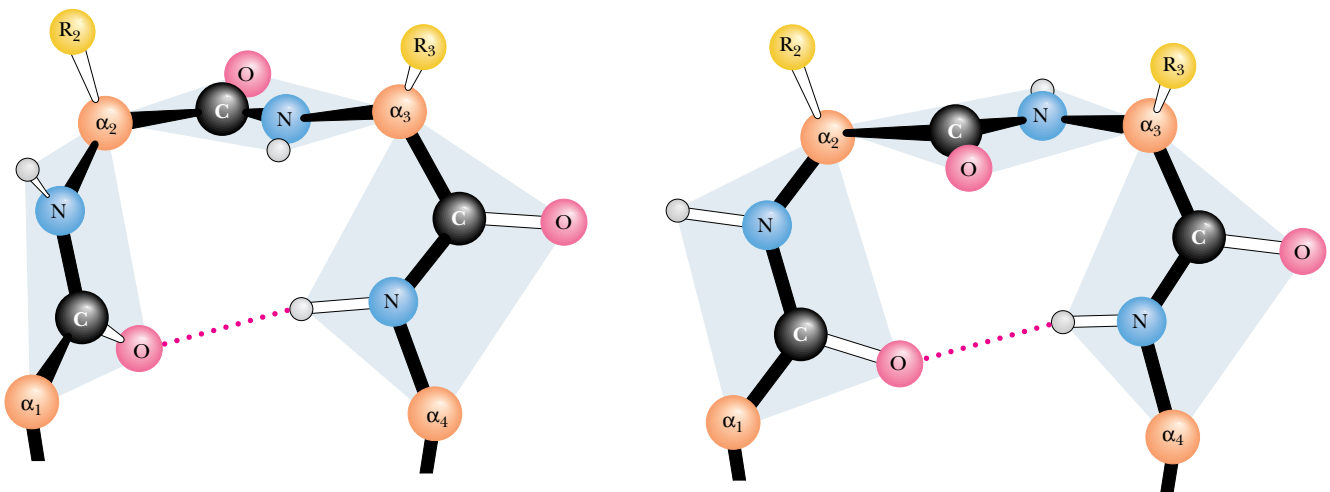

FIGURE 6.12 • The structures of two kinds of β-turns (also called tight turns or β-bends) *(Irving Geis)*
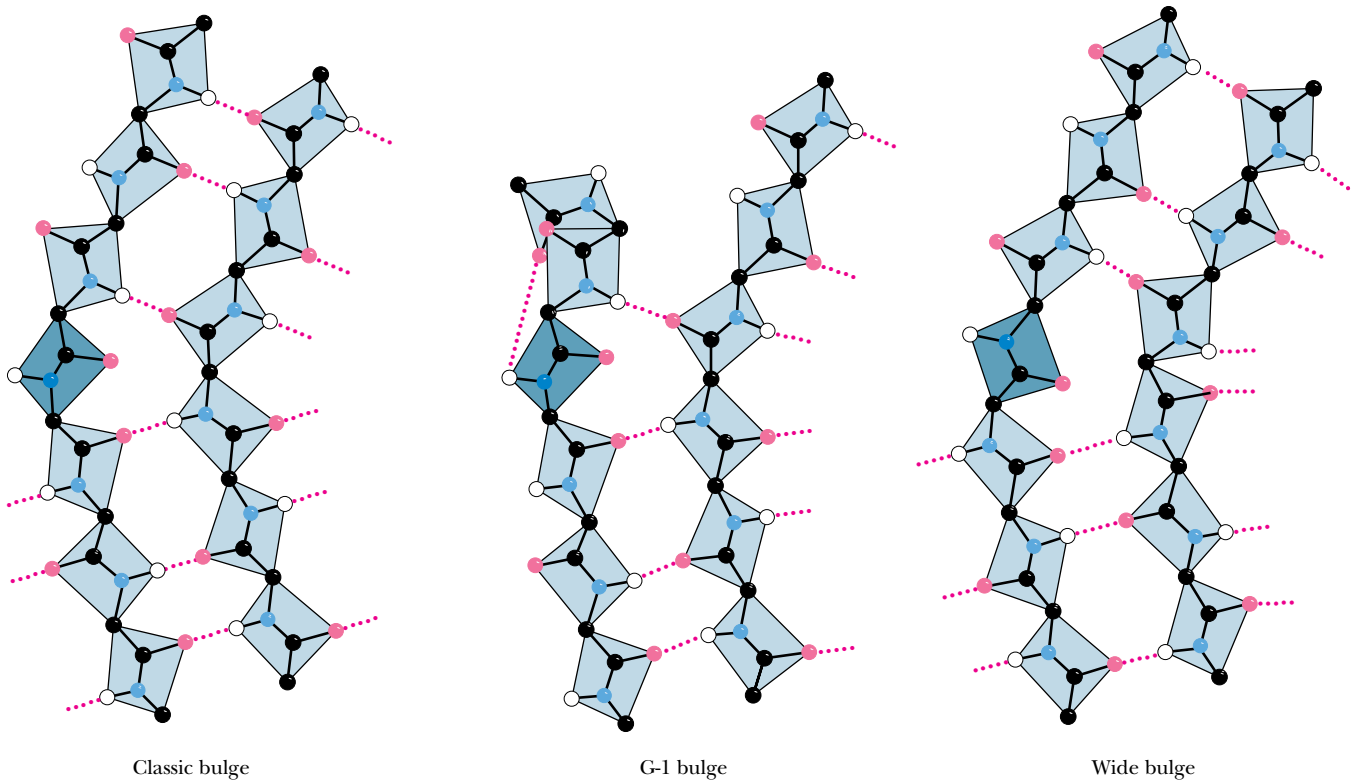
Classic bulge   G-1 bulge   Wide bulge

FIGURE 6.13 • Three different kinds of β-bulge structures involving a pair of adjacent polypeptide chains. *(Adapted from Richardson, J. S., 1981. Advances in Protein Chemistry 34:167–339.)*

The Beta-Bulge

One final secondary structure, the **β-bulge,** is a small piece of nonrepetitive structure that can occur by itself, but most often occurs as an irregularity in antiparallel β-structures. A β-bulge occurs between two normal β-structure hydrogen bonds and comprises two residues on one strand and one residue on the opposite strand. Figure 6.13 illustrates typical β-bulges. The extra residue on the longer side, which causes additional backbone length, is accommodated partially by creating a bulge in the longer strand and partially by forcing a slight bend in the β-sheet. Bulges thus cause changes in the direction of the polypeptide chain, but to a lesser degree than tight turns do. Over 100 examples of β-bulges are known in protein structures.

The secondary structures we have described here are all found commonly in proteins in nature. In fact, it is hard to find proteins that do not contain one or more of these structures. The energetic (mostly H-bond) stabilization afforded by α-helices, β-pleated sheets, and β-turns is important to proteins, and they seize the opportunity to form such structures wherever possible.

6.4 • Protein Folding and Tertiary Structure

The folding of a single polypeptide chain in three-dimensional space is referred to as its **tertiary structure.** As discussed in Section 6.2, all of the information needed to fold the protein into its native tertiary structure is contained within the primary structure of the peptide chain itself. With this in mind, it was disappointing to the biochemists of the 1950s when the early protein structures did not reveal the governing principles in any particular detail. It soon became apparent that the proteins knew how they were supposed to fold into tertiary

shapes, even if the biochemists did not. Vigorous work in many laboratories has slowly brought important principles to light.

First, secondary structures—helices and sheets—form whenever possible as a consequence of the formation of large numbers of hydrogen bonds. Second, α-helices and β-sheets often associate and pack close together in the protein. No protein is stable as a single-layer structure, for reasons that become apparent later. There are a few common methods for such packing to occur. Third, because the peptide segments between secondary structures in the protein tend to be short and direct, the peptide does not execute complicated twists and knots as it moves from one region of a secondary structure to another. A consequence of these three principles is that protein chains are usually folded so that the secondary structures are arranged in one of a few common patterns. For this reason, there are families of proteins that have similar tertiary structure, with little apparent evolutionary or functional relationship among them. Finally, proteins generally fold so as to form the most stable structures possible. The stability of most proteins arises from (1) the formation of large numbers of intramolecular hydrogen bonds and (2) the reduction in the surface area accessible to solvent that occurs upon folding.

Fibrous Proteins

In Chapter 5, we saw that proteins can be grouped into three large classes based on their structure and solubility: *fibrous proteins, globular proteins,* and *membrane proteins.* Fibrous proteins contain polypeptide chains organized approximately parallel along a single axis, producing long fibers or large sheets. Such proteins tend to be mechanically strong and resistant to solubilization in water and dilute salt solutions. Fibrous proteins often play a structural role in nature (see Chapter 5).

α-Keratin

As their name suggests, the structure of the **α-keratins** is dominated by α-helical segments of polypeptide. The amino acid sequence of α-keratin subunits is composed of central α-helix–rich rod domains about 311 to 314 residues in length, flanked by nonhelical N- and C-terminal domains of varying size and composition (Figure 6.14a). The structure of the central rod domain of a typical α-keratin is shown in Figure 6.14b. It consists of four helical strands arranged as twisted pairs of two-stranded **coiled coils.** X-ray diffraction patterns show that these structures resemble α-helices, but with a pitch of 0.51 nm rather than the expected 0.54 nm. This is consistent with a tilt of the helix relative to the long axis of the fiber, as in the two-stranded "rope" in Figure 6.14.

The primary structure of the central rod segments of α-keratin consists of quasi-repeating seven-residue segments of the form $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$. These units are not true repeats, but residues a and d are usually nonpolar amino acids. In α-helices, with 3.6 residues per turn, these nonpolar residues are arranged in an inclined row or stripe that twists around the helix axis. These nonpolar residues would make the helix highly unstable if they were exposed to solvent, but the association of hydrophobic strips on two coiled coils to form the two-stranded rope effectively buries the hydrophobic residues and forms a highly stable structure (Figure 6.14). The helices clearly sacrifice some stability in assuming this twisted conformation, but they gain stabilization energy from the packing of side chains between the helices. In other forms of keratin, covalent disulfide bonds form between cysteine residues of adjacent molecules, making the overall structure rigid, inextensible, and insoluble—important properties for structures such as claws, fingernails, hair, and horns in animals. How and
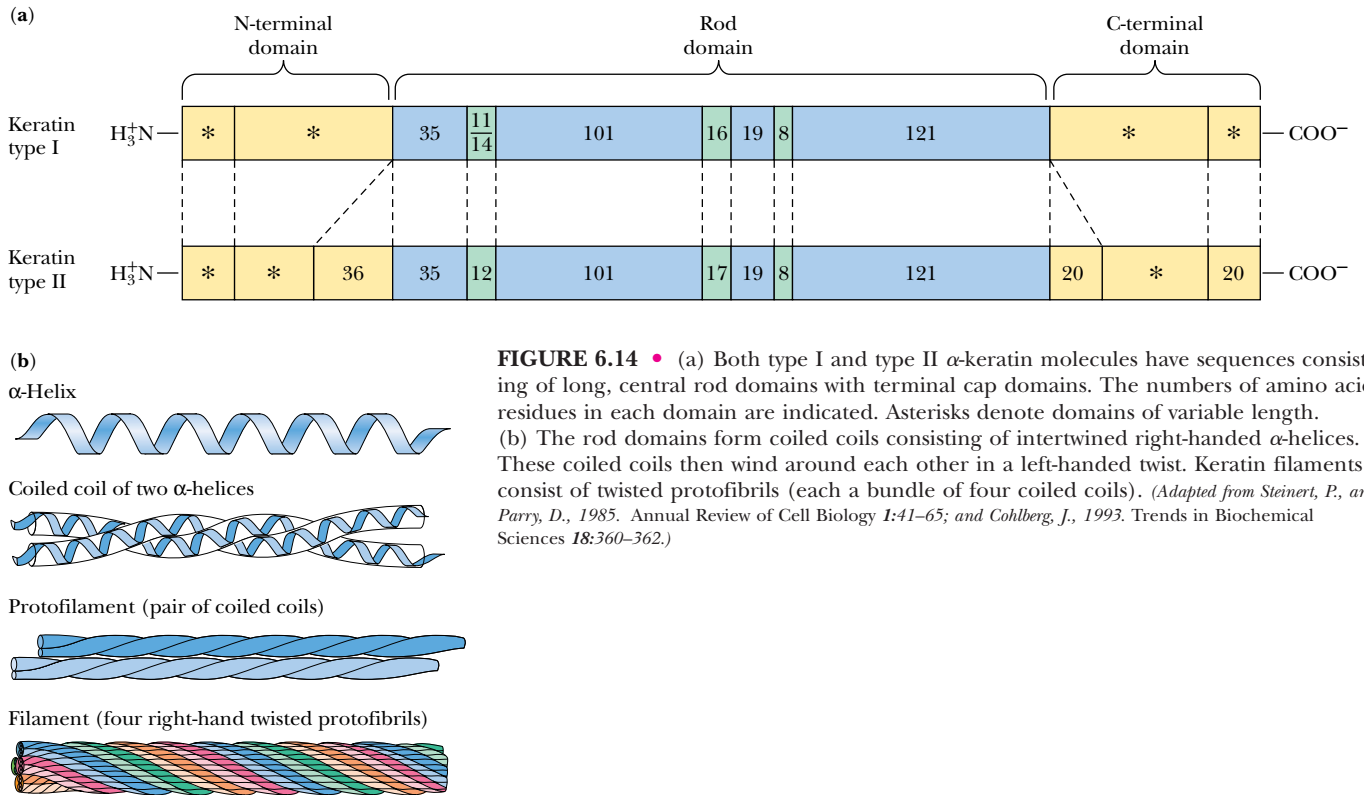
(a)

| | | N-terminal domain | | Rod domain | | | | | | C-terminal domain | | |

(b)

α-Helix

Coiled coil of two α-helices

Protofilament (pair of coiled coils)

Filament (four right-hand twisted protofibrils)

FIGURE 6.14 • (a) Both type I and type II α-keratin molecules have sequences consisting of long, central rod domains with terminal cap domains. The numbers of amino acid residues in each domain are indicated. Asterisks denote domains of variable length. (b) The rod domains form coiled coils consisting of intertwined right-handed α-helices. These coiled coils then wind around each other in a left-handed twist. Keratin filaments consist of twisted protofibrils (each a bundle of four coiled coils). *(Adapted from Steinert, P., and Parry, D., 1985.* Annual Review of Cell Biology **1:**41–65; *and Cohlberg, J., 1993.* Trends in Biochemical Sciences **18:**360–362.)

where these disulfides form determines the amount of curling in hair and wool fibers. When a hairstylist creates a permanent wave (simply called a "permanent") in a hair salon, disulfides in the hair are first reduced and cleaved, then reorganized and reoxidized to change the degree of curl or wave. In contrast, a "set" that is created by wetting the hair, setting it with curlers, and then drying it represents merely a rearrangement of the hydrogen bonds between helices and between fibers. (On humid or rainy days, the hydrogen bonds in curled hair may rearrange, and the hair becomes "frizzy.")

Fibroin and β-Keratin: β-Sheet Proteins

The **fibroin** proteins found in silk fibers represent another type of fibrous protein. These are composed of stacked antiparallel β-sheets, as shown in Figure 6.15. In the polypeptide sequence of silk proteins, there are large stretches in which every other residue is a glycine. As previously mentioned, the residues of a β-sheet extend alternately above and below the plane of the sheet. As a result, the glycines all end up on one side of the sheet and the other residues (mainly alanines and serines) compose the opposite surface of the sheet. Pairs of β-sheets can then pack snugly together (glycine surface to glycine surface or alanine–serine surface to alanine–serine surface). The β-keratins found in bird feathers are also made up of stacked β-sheets.

Collagen: A Triple Helix

Collagen is a rigid, inextensible fibrous protein that is a principal constituent of connective tissue in animals, including tendons, cartilage, bones, teeth, skin, and blood vessels. The high tensile strength of collagen fibers in these struc-
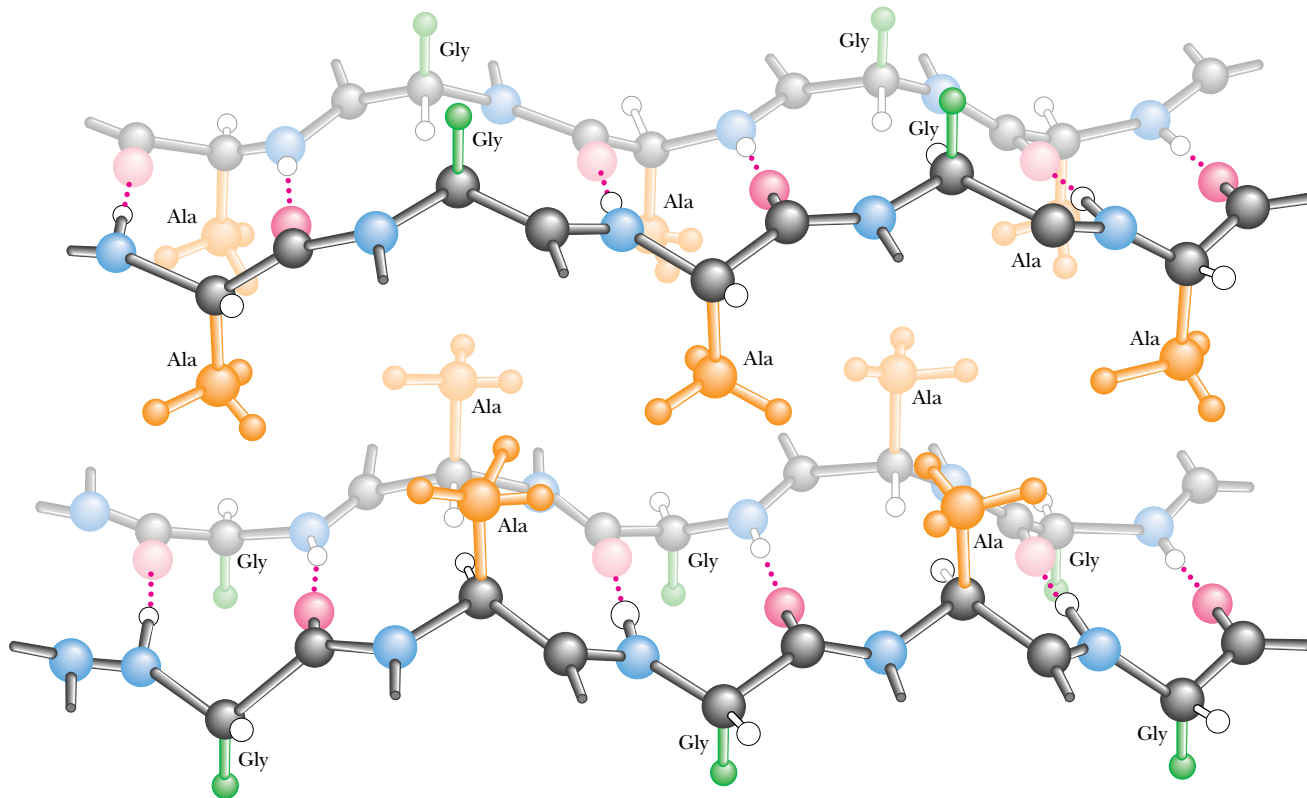
FIGURE 6.15 • Silk fibroin consists of a unique stacked array of β-sheets. The primary structure of fibroin molecules consists of long stretches of alternating glycine and alanine or serine residues. When the sheets stack, the more bulky alanine and serine residues on one side of a sheet interdigitate with similar residues on an adjoining sheet. Glycine hydrogens on the alternating faces interdigitate in a similar manner, but with a smaller intersheet spacing. *(Irving Geis)*

tures makes possible the various animal activities such as running and jumping that put severe stresses on joints and skeleton. Broken bones and tendon and cartilage injuries to knees, elbows, and other joints involve tears or hyper-extensions of the collagen matrix in these tissues.

The basic structural unit of collagen is **tropocollagen,** which has a molecular weight of 285,000 and consists of three intertwined polypeptide chains, each about 1000 amino acids in length. Tropocollagen molecules are about 300 nm long and only about 1.4 nm in diameter. Several kinds of collagen have been identified. *Type I collagen,* which is the most common, consists of two identical peptide chains designated $\alpha 1(I)$ and one different chain designated $\alpha 2(I)$. Type I collagen predominates in bones, tendons, and skin. *Type II collagen,* found in cartilage, and *type III collagen,* found in blood vessels, consist of three identical polypeptide chains.

Collagen has an amino acid composition that is unique and is crucial to its three-dimensional structure and its characteristic physical properties. Nearly one residue out of three is a glycine, and the proline content is also unusually high. Three unusual modified amino acids are also found in collagen: 4-hydroxy-proline (Hyp), 3-hydroxyproline, and 5-hydroxylysine (Hyl) (Figure 6.16). Proline and Hyp together compose up to 30% of the residues of collagen. Interestingly, these three amino acids are formed from normal proline and lysine *after* the collagen polypeptides are synthesized. The modifications are

Charlotte's Web Revisited: Helix–Sheet Composites in Spider Dragline Silk

E. B. White's endearing story *Charlotte's Web* centers around the web-spinning feats of Charlotte the spider. Although the intricate designs of spiderwebs are eye- (and fly-) catching, it might be argued that the composition of web silk itself is even more remarkable. Spider silk is synthesized in special glands in the spider's abdomen. The silk strands produced by these glands are both strong and elastic. *Dragline silk* (that from which the spider hangs) has a tensile strength of 200,000 psi (pounds per square inch)—stronger than steel and similar to Kevlar, the synthetic material used in bulletproof vests! This same silk fiber is also flexible enough to withstand strong winds and other natural stresses.

This combination of strength and flexibility derives from the *composite nature* of spider silk. As keratin protein is extruded from the spider's glands, it endures shearing forces that break the H bonds stabilizing keratin α-helices. These regions then form microcrystalline arrays of β-sheets. These microcrystals are surrounded by the keratin strands, which adopt a highly disordered state composed of α-helices and random coil structures.

The β-sheet microcrystals contribute strength, and the disordered array of helix and coil make the silk strand flexible. The resulting silk strand resembles modern human-engineered composite materials. Certain tennis racquets, for example, consist of fiberglass polymers impregnated with microcrystalline graphite. The fiberglass provides flexibility, and the graphite crystals contribute strength. Modern high technology, for all its sophistication, is merely imitating nature—and Charlotte's web—after all.
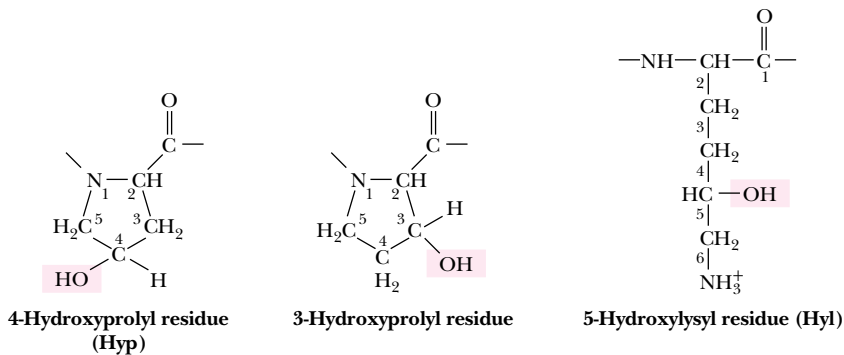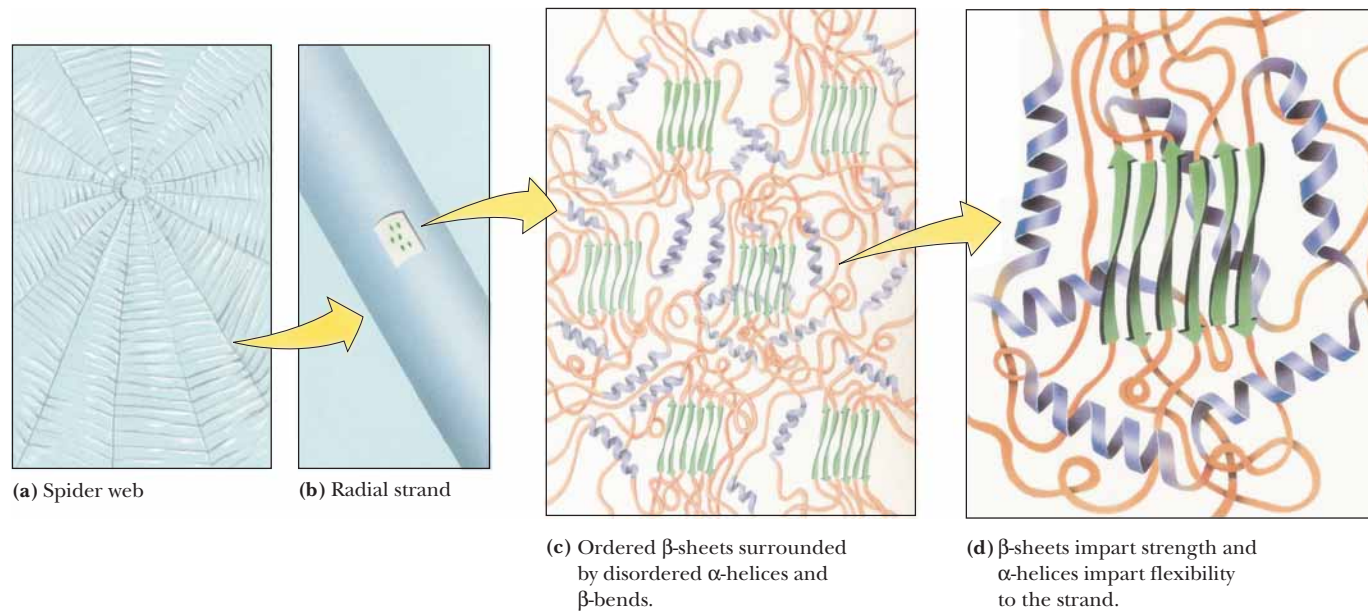
(a) Spider web

(b) Radial strand

(c) Ordered β-sheets surrounded by disordered α-helices and β-bends.

(d) β-sheets impart strength and α-helices impart flexibility to the strand.

4-Hydroxyprolyl residue (Hyp)

3-Hydroxyprolyl residue

5-Hydroxylysyl residue (Hyl)

FIGURE 6.16 • The hydroxylated residues typically found in collagen.
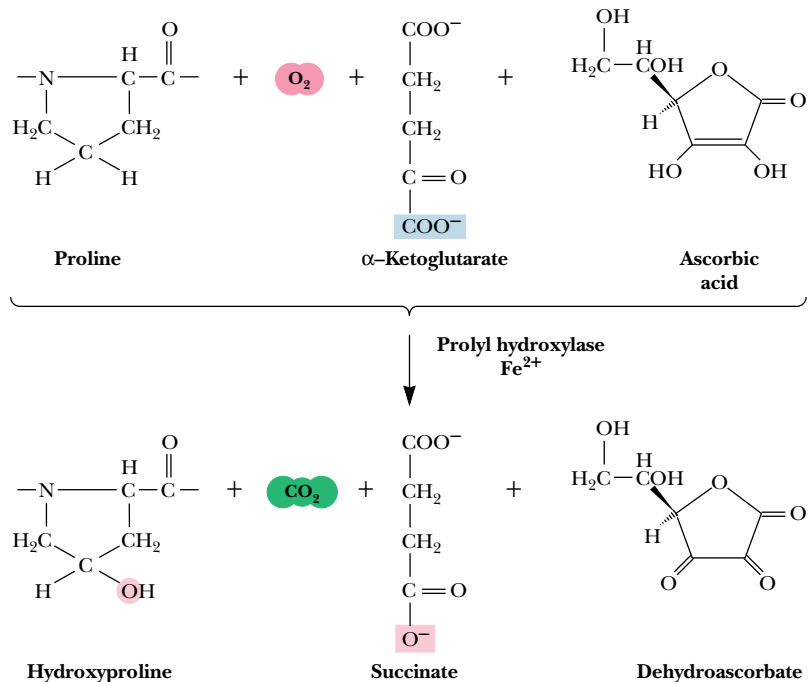
FIGURE 6.17 • Hydroxylation of proline residues is catalyzed by prolyl hydroxylase. The reaction requires α-ketoglutarate and ascorbic acid (vitamin C).
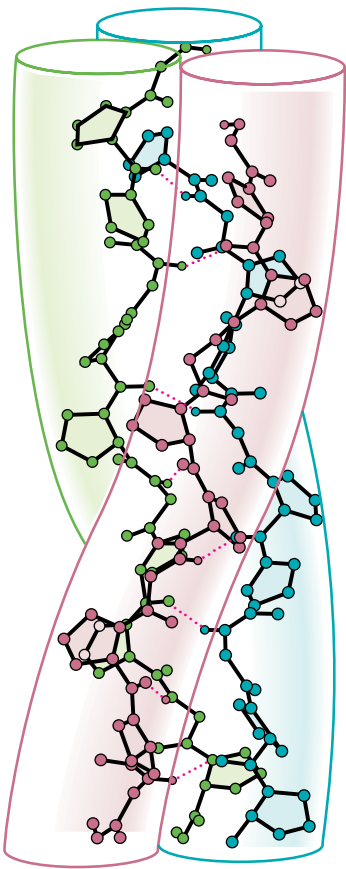

FIGURE 6.18 • Poly(Gly-Pro-Pro), a collagen-like right-handed triple helix composed of three left-handed helical chains. *(Adapted from Miller, M. H., and Scheraga, H. A., 1976, Calculation of the structures of collagen models. Role of interchain interactions in determining the triple-helical coiled-coil conformation. I. Poly(gly-cyl-prolyl-prolyl). Journal of Polymer Science Symposium 54:171–200.)*

effected by two enzymes: *prolyl hydroxylase* and *lysyl hydroxylase*. The prolyl hydroxylase reaction (Figure 6.17) requires molecular oxygen, α-ketoglutarate, and ascorbic acid (vitamin C) and is activated by Fe^{2+}. The hydroxylation of lysine is similar. These processes are referred to as **posttranslational modifications** because they occur after genetic information from DNA has been *translated* into newly formed protein.

Because of their high content of glycine, proline, and hydroxyproline, collagen fibers are incapable of forming traditional structures such as α-helices and β-sheets. Instead, collagen polypeptides intertwine to form a unique **triple helix,** with each of the three strands arranged in a helical fashion (Figure 6.18). Compared to the α-helix, the collagen helix is much more extended, with a rise per residue along the triple helix axis of 2.9 Å, compared to 1.5 Å for the α-helix. There are about 3.3 residues per turn of each of these helices. *The triple helix is a structure that forms to accommodate the unique composition and sequence of collagen.* Long stretches of the polypeptide sequence are repeats of a Gly-x-y motif, where x is frequently Pro and y is frequently Pro or Hyp. In the triple helix, every third residue faces or contacts the crowded center of the structure. This area is so crowded that only Gly can fit, and thus every third residue must be a Gly (as observed). Moreover, the triple helix is a *staggered* structure, such that Gly residues from the three strands stack along the center of the triple helix and the Gly from one strand lies adjacent to an x residue from the second strand and to a y from the third. This allows the N—H of each Gly residue to hydrogen bond with the C═O of the adjacent x residue. The triple helix structure is further stabilized and strengthened by the formation of interchain H bonds involving hydroxyproline.

Collagen types I, II, and III form strong, organized **fibrils,** consisting of staggered arrays of tropocollagen molecules (Figure 6.19). The periodic
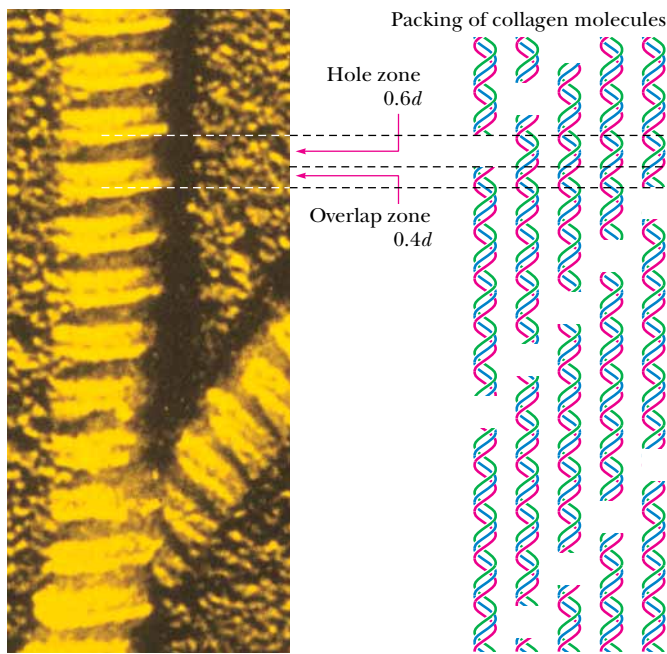
FIGURE 6.19 • In the electron microscope, collagen fibers exhibit alternating light and dark bands. The dark bands correspond to the 40-nm gaps or "holes" between pairs of aligned collagen triple helices. The repeat distance, *d,* for the light- and dark-banded pattern is 68 nm. The collagen molecule is 300 nm long, which corresponds to 4.41*d.* The molecular repeat pattern of five staggered collagen molecules corresponds to 5*d.* *(J. Gross, Biozentrum/Science Photo Library)*

arrangement of triple helices in a head-to-tail fashion results in banded patterns in electron micrographs. The banding pattern typically has a periodicity (repeat distance) of 68 nm. Because collagen triple helices are 300 nm long, 40-nm gaps occur between adjacent collagen molecules in a row along the long axis of the fibrils and the pattern repeats every five rows (5×68 nm $= 340$ nm). The 40-nm gaps are referred to as *hole regions,* and they are important in at least two ways. First, sugars are found covalently attached to 5-hydroxylysine residues in the hole regions of collagen (Figure 6.20). The occurrence of carbohydrate in the hole region has led to the proposal that it plays a role in organizing fibril assembly. Second, the hole regions may play a role in bone formation. Bone consists of microcrystals of **hydroxyapatite,** $Ca_5(PO_4)_3OH$, embedded in a matrix of collagen fibrils. When new bone tissue forms, the formation of new hydroxyapatite crystals occurs at intervals of 68 nm. The hole regions of collagen fibrils may be the sites of nucleation for the mineralization of bone.

The collagen fibrils are further strengthened and stabilized by the formation of both *intramolecular* (within a tropocollagen molecule) and *intermolecular* (between tropocollagen molecules in the fibril) cross-links. Intramolecular cross-links are formed between lysine residues in the (nonhelical) N-terminal region of tropocollagen in a unique pair of reactions shown in Figure 6.21. The enzyme *lysyl oxidase* catalyzes the formation of aldehyde groups at the lysine side chains in a copper-dependent reaction. The aldehyde groups of two such side chains then link covalently in a spontaneous nonenzymatic *aldol condensation.* The intermolecular cross-linking of tropocollagens involves the formation of a unique **hydroxypyridinium** structure from one lysine and two hydroxy-
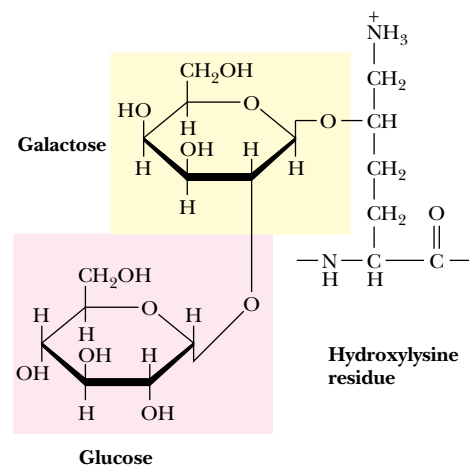

FIGURE 6.20 • A disaccharide of galactose and glucose is covalently linked to the 5-hydroxyl group of hydroxylysines in collagen by the combined action of the enzymes galactosyl transferase and glucosyl transferase.

$\mathcal{H}$UMAN BIOCHEMISTRY

Collagen-Related Diseases

Collagen provides an ideal case study of the molecular basis of physiology and disease. For example, the nature and extent of collagen cross-linking depends on the age and function of the tissue. Collagen from young animals is predominantly un-cross-linked and can be extracted in soluble form, whereas collagen from older animals is highly cross-linked and thus insoluble. The loss of flexibility of joints with aging is probably due in part to increased cross-linking of collagen.

Several serious and debilitating diseases involving collagen abnormalities are known. **Lathyrism** occurs in animals due to the regular consumption of seeds of *Lathyrus odoratus*, the sweet pea, and involves weakening and abnormalities in blood vessels, joints, and bones. These conditions are caused by **β-amino-propionitrile** (see figure), which covalently inactivates lysyl oxidase and leads to greatly reduced intramolecular cross-linking of collagen in affected animals (or humans).

$$N \equiv C - CH_2 - CH_2 - \overset{+}{N}H_3$$
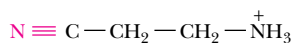
β-Aminopropionitrile (present in sweet peas) covalently inactivates lysyl oxidase, preventing intramolecular cross-linking of collagen and causing abnormalities in joints, bones, and blood vessels.

Scurvy results from a dietary vitamin C deficiency and involves the inability to form collagen fibrils properly. This is the result of reduced activity of prolyl hydroxylase, which is vitamin C–dependent, as previously noted. Scurvy leads to lesions in the skin and blood vessels, and, in its advanced stages, it can lead to grotesque disfiguration and eventual death. Although rare in the modern world, it was a disease well known to sea-faring explorers in earlier times who did not appreciate the importance of fresh fruits and vegetables in the diet.

A number of rare genetic diseases involve collagen abnormalities, including *Marfan's syndrome* and the *Ehlers–Danlos syndromes*, which result in hyperextensible joints and skin. The formation of *atherosclerotic plaques*, which cause arterial blockages in advanced stages, is due in part to the abnormal formation of collagenous structures in blood vessels.
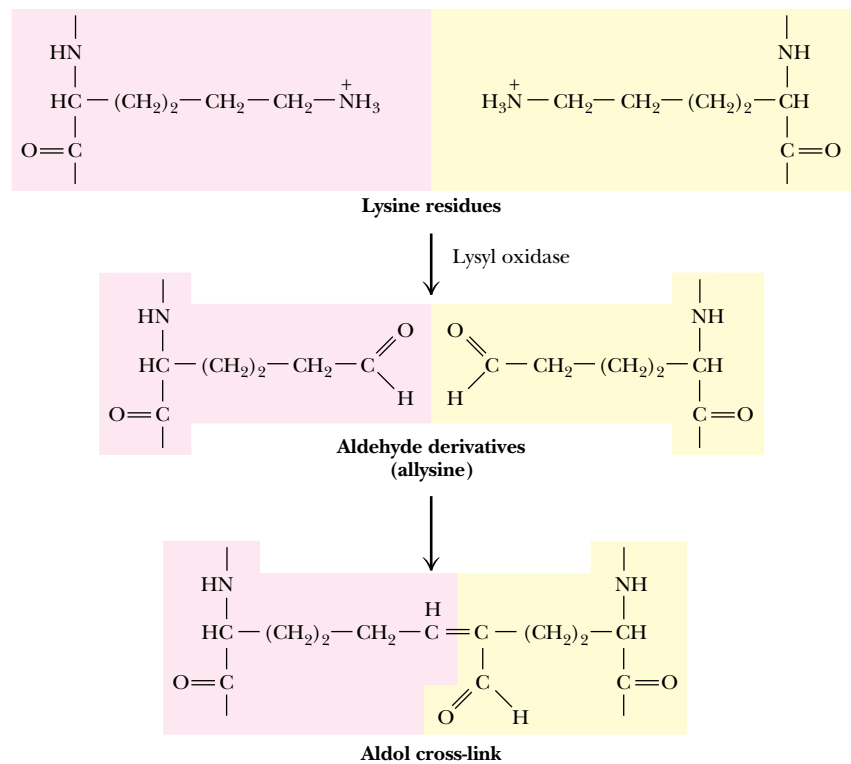

FIGURE 6.21 • Collagen fibers are stabilized and strengthened by Lys–Lys cross-links. Aldehyde moieties formed by lysyl oxidase react in a spontaneous nonenzymatic aldol reaction.

lysine residues (Figure 6.22). These cross-links form between the N-terminal region of one tropocollagen and the C-terminal region of an adjacent tropocollagen in the fibril.

Globular Proteins

Fibrous proteins, although interesting for their structural properties, represent only a small percentage of the proteins found in nature. **Globular proteins,** so named for their approximately spherical shape, are far more numerous.

Helices and Sheets in Globular Proteins

Globular proteins exist in an enormous variety of three-dimensional structures, but nearly all contain substantial amounts of the α-helices and β-sheets that form the basic structures of the simple fibrous proteins. For example, myoglobin, a small, globular, oxygen-carrying protein of muscle (17 kD, 153 amino acid residues), contains eight α-helical segments, each containing 7 to 26 amino acid residues. These are arranged in an apparently irregular (but invariant) fashion (see Figure 5.7). The space between the helices is filled efficiently and tightly with (mostly hydrophobic) amino acid side chains. Most of the polar side chains in myoglobin (and in most other globular proteins) face the outside of the protein structure and interact with solvent water. Myoglobin's structure is unusual because most globular proteins contain a relatively small amount of α-helix. A more typical globular protein (Figure 6.23) is *bovine ribonuclease A*, a small protein (14.6 kD, 129 residues) that contains a few short helices, a broad section of antiparallel β-sheet, a few β-turns, and several peptide segments without defined secondary structure.
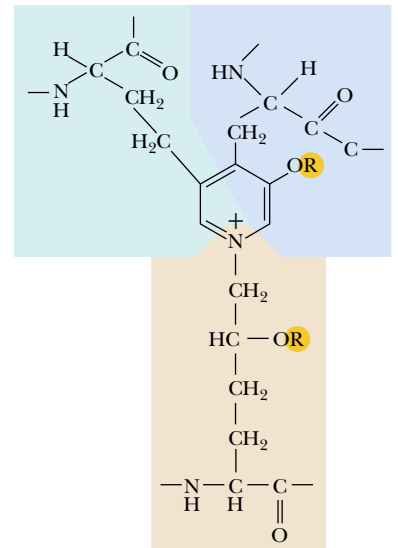
FIGURE 6.22 • The hydroxypyridinium structure formed by the cross-linking of a Lys and two hydroxy Lys residues.
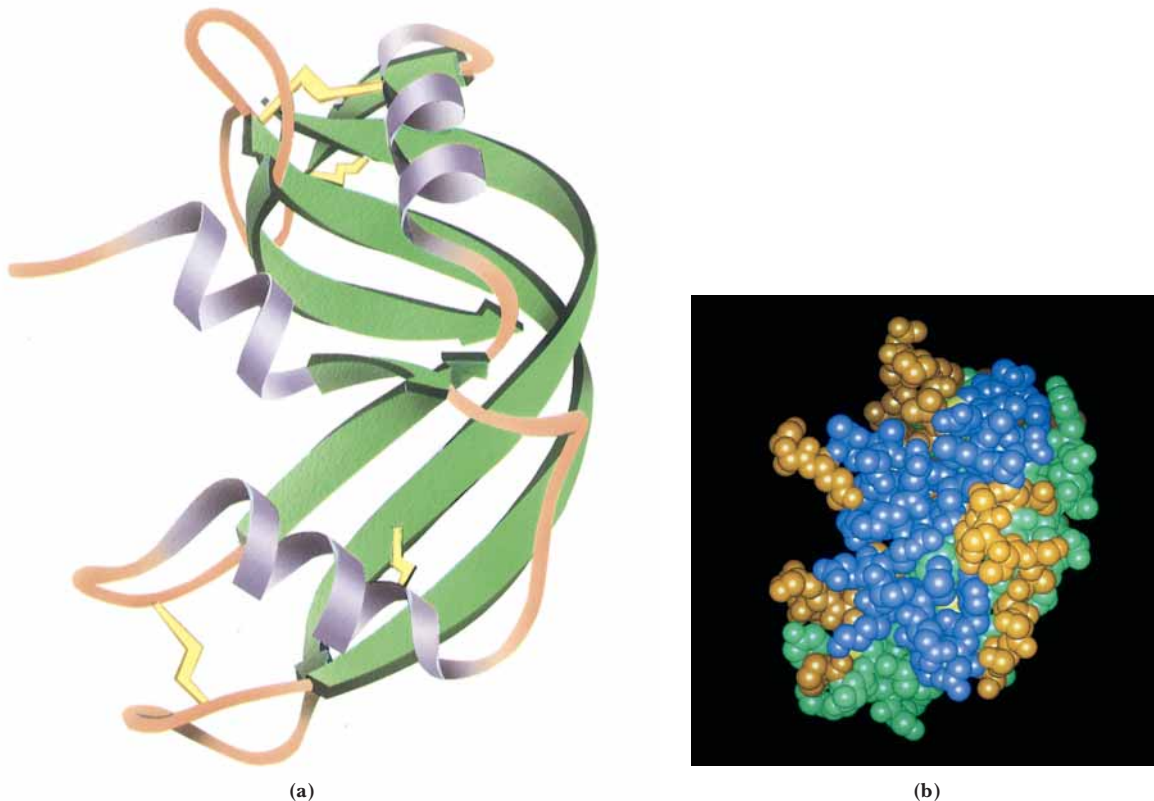

(a)

(b)

FIGURE 6.23 • The three-dimensional structure of bovine ribonuclease A, showing the α-helices as ribbons. *(Jane Richardson)*
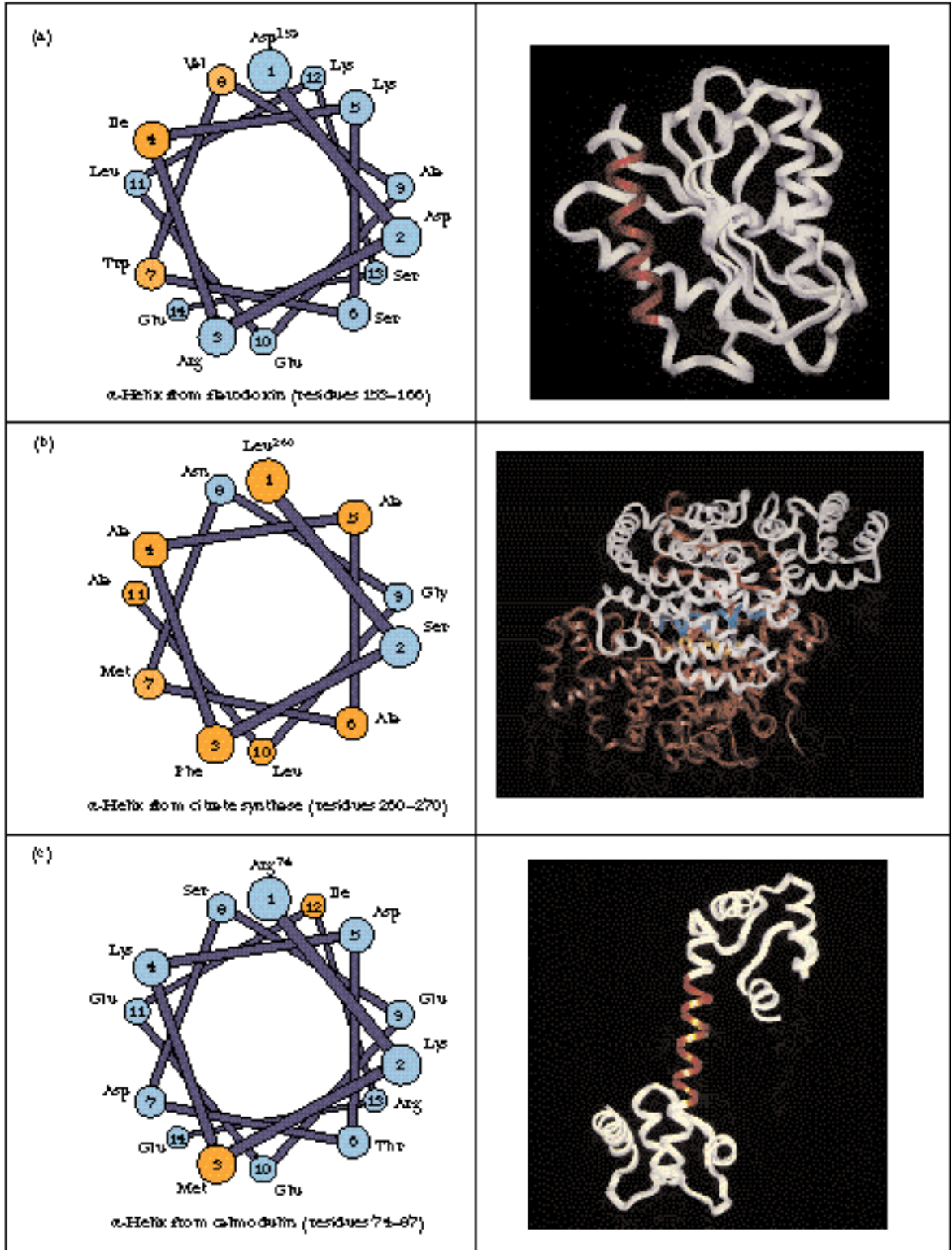
FIGURE 6.24 • (a) The alpha helix consisting of residues 153–166 (red) in flavodoxin from *Anabaena* is a surface helix and is amphipathic. (b) The two helices (yellow and blue) in the interior of the citrate synthase dimer (residues 260–270 in each monomer) are mostly hydrophobic. (c) The exposed helix (residues 74–87—red) of calmodulin is entirely accessible to solvent and consists mainly of polar and charged residues.

Why should the cores of most globular and membrane proteins consist almost entirely of α-helices and β-sheets? The reason is that the highly polar N—H and C=O moieties of the peptide backbone must be neutralized in the hydrophobic core of the protein. The extensively H-bonded nature of α-helices and β-sheets is ideal for this purpose, and these structures effectively stabilize the polar groups of the peptide backbone in the protein core.

In globular protein structures, it is common for one face of an α-helix to be exposed to the water solvent, with the other face toward the hydrophobic interior of the protein. The outward face of such an **amphiphilic helix** consists mainly of polar and charged residues, whereas the inward face contains mostly nonpolar, hydrophobic residues. A good example of such a surface helix is that of residues 153 to 166 of **flavodoxin** from *Anabaena* (Figure 6.24). Note that the **helical wheel presentation** of this helix readily shows that one face contains four hydrophobic residues and that the other is almost entirely polar and charged.

Less commonly, an α-helix can be completely buried in the protein interior or completely exposed to solvent. **Citrate synthase** is a dimeric protein in which α-helical segments form part of the subunit–subunit interface. As shown in Figure 6.24, one of these helices (residues 260 to 270) is highly hydrophobic and contains only two polar residues, as would befit a helix in the protein core. On the other hand, Figure 6.24 also shows the solvent-exposed helix (residues 74 to 87) of **calmodulin,** which consists of 10 charged residues, 2 polar residues, and only 2 nonpolar residues.

Packing Considerations

The secondary and tertiary structures of myoglobin and ribonuclease A illustrate the importance of packing in tertiary structures. Secondary structures pack closely to one another and also intercalate with (insert between) extended polypeptide chains. If the sum of the van der Waals volumes of a protein's constituent amino acids is divided by the volume occupied by the protein, packing densities of 0.72 to 0.77 are typically obtained. This means that, even with close packing, approximately 25% of the total volume of a protein is not occupied by protein atoms. Nearly all of this space is in the form of very small cavities. Cavities the size of water molecules or larger do occasionally occur, but they make up only a small fraction of the total protein volume. It is likely that such cavities provide flexibility for proteins and facilitate conformation changes and a wide range of protein dynamics (discussed later).

Ordered, Nonrepetitive Structures

In any protein structure, the segments of the polypeptide chain that cannot be classified as defined secondary structures, such as helices or sheets, have been traditionally referred to as *coil* or *random coil.* Both these terms are misleading. Most of these segments are neither coiled nor random, in any sense of the words. These structures are every bit as highly organized and stable as the defined secondary structures. They are just more variable and difficult to describe. These so-called coil structures are strongly influenced by side-chain interactions. Few of these interactions are well understood, but a number of interesting cases have been described. In his early studies of myoglobin structure, John Kendrew found that the —OH group of threonine or serine often forms a hydrogen bond with a backbone NH at the beginning of an α-helix. The same stabilization of an α-helix by a serine is observed in the three-dimensional structure of pancreatic trypsin inhibitor (Figure 6.25). Also in this same structure, an asparagine residue adjacent to a β-strand is found to form H bonds that stabilize the β-structure.

Ser^{47}

Asn^{43}
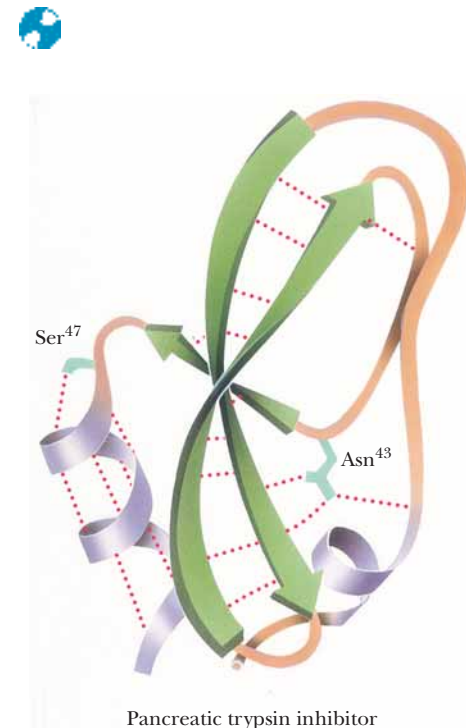
Pancreatic trypsin inhibitor

FIGURE 6.25 • The three-dimensional structure of bovine pancreatic trypsin inhibitor. Note the stabilization of the α-helix by a hydrogen bond to Ser^{47} and the stabilization of the β-sheet by Asn^{43}.
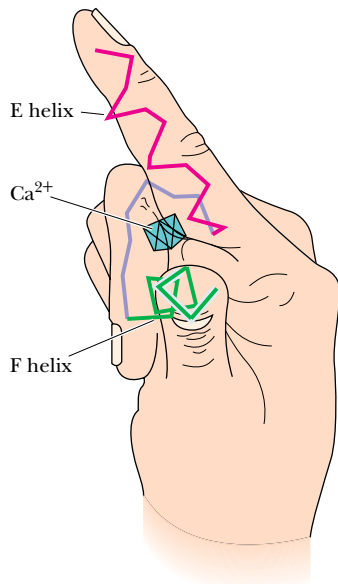
FIGURE 6.26 • A representation of the so-called E–F hand structure, which forms calcium-binding sites in a variety of proteins. The stick drawing shows the peptide backbone of the E–F hand motif. The "E" helix extends along the index finger, a loop traces the approximate arrangement of the curled middle finger, and the "F" helix extends outward along the thumb. A calcium ion (Ca^{2+}) snuggles into the pocket created by the two helices and the loop. Kretsinger and coworkers originally assigned letters alphabetically to the helices in parvalbumin, a protein from carp. The E–F hand derives its name from the letters assigned to the helices at one of the Ca^{2+}-binding sites.

Nonrepetitive but well-defined structures of this type form many important features of enzyme active sites. In some cases, a particular arrangement of "coil" structure providing a specific type of functional site recurs in several functionally related proteins. The peptide loop that binds iron–sulfur clusters in both ferredoxin and high potential iron protein is one example. Another is the central loop portion of the *E–F hand structure* that binds a calcium ion in several calcium-binding proteins, including calmodulin, carp parvalbumin, troponin C, and the intestinal calcium-binding protein. This loop, shown in Figure 6.26, connects two short α-helices. The calcium ion nestles into the pocket formed by this structure.

Flexible, Disordered Segments

In addition to nonrepetitive but well-defined structures, which exist in all proteins, genuinely disordered segments of polypeptide sequence also occur. These sequences either do not show up in electron density maps from X-ray crystallographic studies or give diffuse or ill-defined electron densities. These segments either undergo actual motion in the protein crystals themselves or take on many alternate conformations in different molecules within the protein crystal. Such behavior is quite common for long, charged side chains on the surface of many proteins. For example, 16 of the 19 lysine side chains in myoglobin have uncertain orientations beyond the δ-carbon, and five of these are disordered beyond the β-carbon. Similarly, a majority of the lysine residues are disordered in trypsin, rubredoxin, ribonuclease, and several other proteins. Arginine residues, however, are usually well ordered in protein structures. For the four proteins just mentioned, 70% of the arginine residues are highly ordered, compared to only 26% of the lysines.

Motion in Globular Proteins

Although we have distinguished between well-ordered and disordered segments of the polypeptide chain, it is important to realize that even well-ordered side chains in a protein undergo motion, sometimes quite rapid. These motions should be viewed as momentary oscillations about a single, highly stable conformation. *Proteins are thus best viewed as dynamic structures.* The allowed motions may be motions of individual atoms, groups of atoms, or even whole sections of the protein. Furthermore, they may arise from either thermal energy or specific, triggered conformational changes in the protein. **Atomic fluctuations** such as vibrations typically are random, very fast, and usually occur over small distances (less than 0.5 Å), as shown in Table 6.2. These motions arise from the kinetic energy within the protein and are a function of temperature. These very fast motions can be modeled by molecular dynamics calculations and studied by X-ray diffraction.

A class of slower motions, which may extend over larger distances, is **collective motions.** These are movements of groups of atoms covalently linked in such a way that the group moves as a unit. Such groups range in size from a few atoms to hundreds of atoms. Whole structural domains within a protein may be involved, as in the case of the flexible antigen-binding domains of immunoglobulins, which move as relatively rigid units to selectively bind separate antigen molecules. Such motions are of two types—(1) those that occur quickly but infrequently, such as tyrosine ring flips, and (2) those that occur slowly, such as *cis-trans* isomerizations of prolines. These collective motions also arise from thermal energies in the protein and operate on a time scale of 10^{-12} to 10^{-3} sec. These motions can be studied by nuclear magnetic resonance (NMR) and fluorescence spectroscopy.

Table 6.2

Motion and Fluctuations in Proteins			
Type of Motion	Spatial Displacement (Å)	Characteristic Time (sec)	Source of Energy
Atomic vibrations	0.01–1	10^{-15}–10^{-11}	Kinetic energy
Collective motions	0.01–5 or more	10^{-12}–10^{-3}	Kinetic energy
1. Fast: Tyr ring flips; methyl group rotations 2. Slow: hinge bending between domains			
Triggered conformation changes	0.5–10 or more	10^{-9}–10^{3}	Interactions with triggering agent

Adapted from Petsko and Ringe (1984).

Conformational changes involve motions of groups of atoms (individual side chains, for example) or even whole sections of proteins. These motions occur on a time scale of 10^{-9} to 10^{3} sec, and the distances covered can be as large as 1 nm. These motions may occur in response to specific stimuli or arise from specific interactions within the protein, such as hydrogen bonding, electrostatic interactions, and ligand binding. More will be said about conformational changes when enzyme catalysis and regulation are discussed (see Chapters 14 and 15).

Forces Driving the Folding of Globular Proteins

As already pointed out, the driving force for protein folding and the resulting formation of a tertiary structure is the formation of the most stable structure possible. Two forces are at work here. The peptide chain must both (1) satisfy the constraints inherent in its own structure and (2) fold so as to "bury" the hydrophobic side chains, minimizing their contact with solvent. The polypeptide itself does not usually form simple straight chains. Even in chain segments where helices and sheets are not formed, an extended peptide chain, being composed of L-amino acids, has a tendency to twist slightly in a right-handed direction. As shown in Figure 6.27, this tendency is apparently the basis for the formation of a variety of tertiary structures having a right-handed sense. Principal among these are the right-handed twists in arrays of β-sheets and right-handed cross-overs in parallel β-sheet arrays. Right-handed twisted β-sheets are found at the center of a number of proteins and provide an extended, highly stable structural core. Phosphoglycerate mutase, adenylate kinase, and carbonic anhydrase, among others, exist as smoothly twisted planes or saddle-shaped structures. Triose phosphate isomerase, soybean trypsin inhibitor, and domain 1 of pyruvate kinase contain right-handed twisted cylinders or barrel structures at their cores.

Connections between β-strands are of two types—hairpins and cross-overs. **Hairpins,** as shown in Figure 6.27, connect adjacent antiparallel β-strands. **Cross-overs** are necessary to connect adjacent (or nearly adjacent) parallel β-strands. Nearly all cross-over structures are right-handed. Only in subtilisin and phosphoglucoisomerase have isolated left-handed cross-overs been identi-
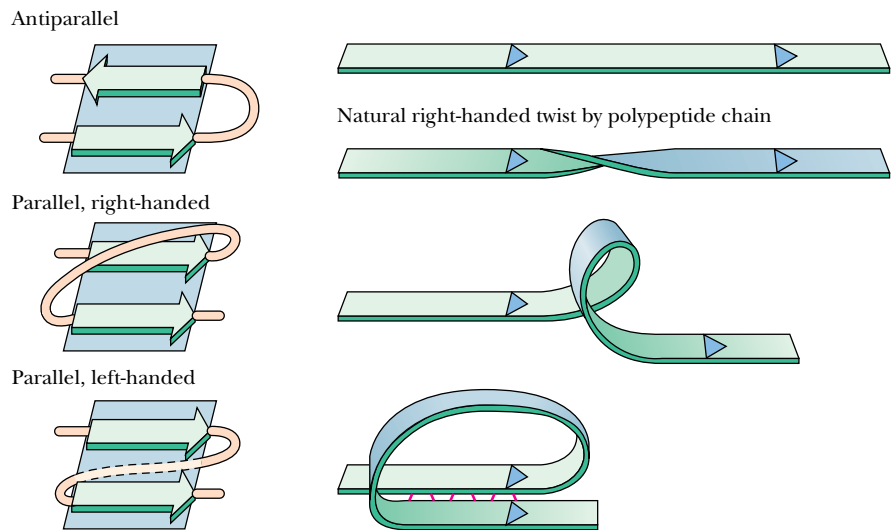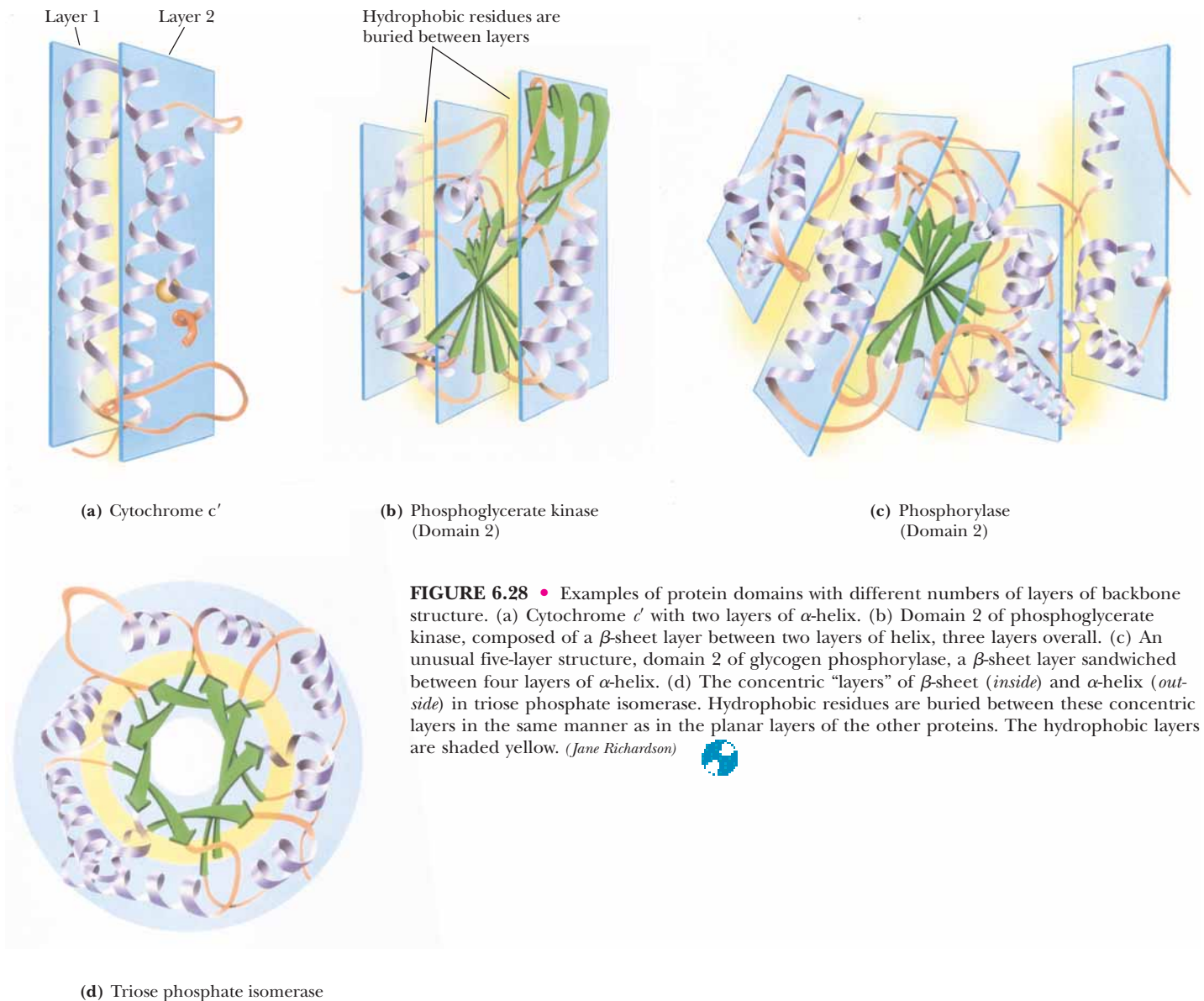
Antiparallel

Natural right-handed twist by polypeptide chain

Parallel, right-handed

Parallel, left-handed

FIGURE 6.27 • The natural right-handed twist exhibited by polypeptide chains, and the variety of structures that arise from this twist.

fied. In many cross-over structures, the cross-over connection itself contains an α-helical segment. This is referred to as a **βαβ-loop.** As shown in Figure 6.27, the strong tendency in nature to form right-handed cross-overs, the wide occurrence of α-helices in the cross-over connection, and the right-handed twists of β-sheets can all be understood as arising from the tendency of an extended polypeptide chain of L-amino acids to adopt a right-handed twist structure. This is a chiral effect. Proteins composed of D-amino acids would tend to adopt left-handed twist structures.

The second driving force that affects the folding of polypeptide chains is the need to bury the hydrophobic residues of the chain, protecting them from solvent water. From a topological viewpoint, then, all globular proteins must have an "inside" where the hydrophobic core can be arranged and an "outside" toward which the hydrophilic groups must be directed. The sequestration of hydrophobic residues away from water is the dominant force in the arrangement of secondary structures and nonrepetitive peptide segments to form a given tertiary structure. Globular proteins can be classified mainly on the basis of the particular kind of core or backbone structure they use to accomplish this goal. The term *hydrophobic core,* as used here, refers to a region in which hydrophobic side chains cluster together, away from the solvent. *Backbone* refers to the polypeptide backbone itself, excluding the particular side chains. Globular proteins can be pictured as consisting of "layers" of backbone, with hydrophobic core regions between them. Over half the known globular protein structures have two layers of backbone (separated by one hydrophobic core). Roughly one-third of the known structures are composed of three backbone layers and two hydrophobic cores. There are also a few known four-layer structures and one known five-layer structure. A few structures are not easily classified in this way, but it is remarkable that most proteins fit into one of these classes. Examples of each are presented in Figure 6.28.

Classification of Globular Proteins

In addition to classification based on layer structure, proteins can be grouped according to the type and arrangement of secondary structure. There are four such broad groups: antiparallel α-helix, parallel or mixed β-sheet, antiparallel β-sheet, and the small metal- and disulfide-rich proteins.

Layer 1   Layer 2   Hydrophobic residues are
                                      buried between layers

(a) Cytochrome c′   **(b)** Phosphoglycerate kinase   **(c)** Phosphorylase
                                      (Domain 2)   (Domain 2)

FIGURE 6.28 • Examples of protein domains with different numbers of layers of backbone structure. (a) Cytochrome c′ with two layers of α-helix. (b) Domain 2 of phosphoglycerate kinase, composed of a β-sheet layer between two layers of helix, three layers overall. (c) An unusual five-layer structure, domain 2 of glycogen phosphorylase, a β-sheet layer sandwiched between four layers of α-helix. (d) The concentric "layers" of β-sheet (*inside*) and α-helix (*outside*) in triose phosphate isomerase. Hydrophobic residues are buried between these concentric layers in the same manner as in the planar layers of the other proteins. The hydrophobic layers are shaded yellow. (*Jane Richardson*)

(d) Triose phosphate isomerase

It is important to note that the similarities of tertiary structure within these groups do not necessarily reflect similar or even related functions. Instead, **functional homology** usually depends on structural similarities on a smaller and more intimate scale.

Antiparallel α-Helix Proteins

Antiparallel α-helix proteins are structures heavily dominated by α-helices. The simplest way to pack helices is in an antiparallel manner, and most of the proteins in this class consist of bundles of antiparallel helices. Many of these exhibit a slight (15°) left-handed twist of the helix bundle. Figure 6.29 shows a representative sample of antiparallel α-helix proteins. Many of these are regular, uniform structures, but in a few cases (uteroglobin, for example) one of the helices is tilted away from the bundle. Tobacco mosaic virus protein has small, highly
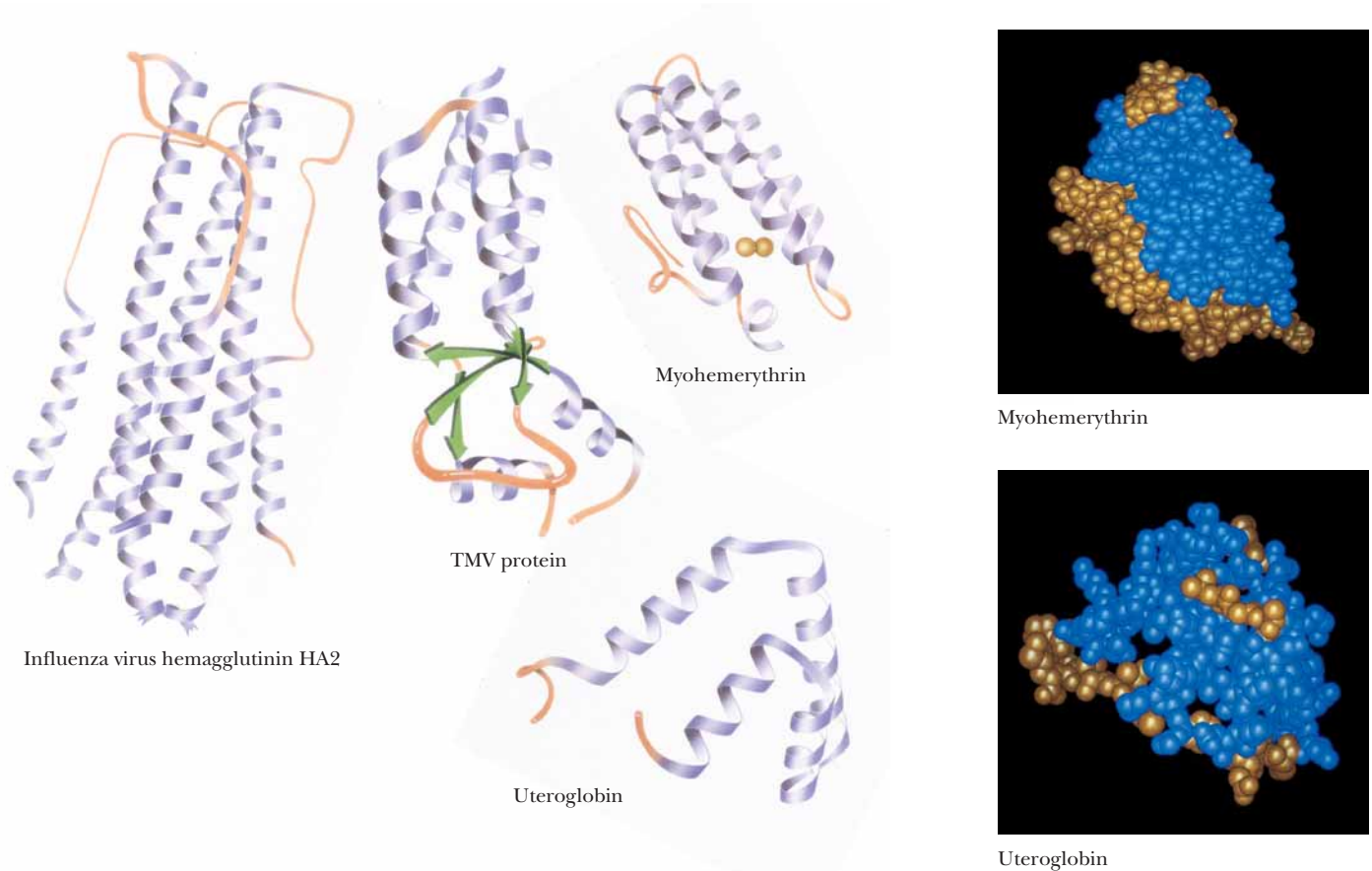
Influenza virus hemagglutinin HA2

TMV protein

Myohemerythrin

Uteroglobin

Myohemerythrin

Uteroglobin

FIGURE 6.29 • Several examples of antiparallel α-proteins. (*Jane Richardson*)

twisted antiparallel β-sheets on one end of the helix bundle with two additional helices on the other side of the sheet. Notice in Figure 6.29 that most of the antiparallel α-helix proteins are made up of four-helix bundles.

The so-called globin proteins are an important group of α-helical proteins. These include hemoglobins and myoglobins from many species. The globin structure can be viewed as two layers of helices, with one of these layers perpendicular to the other and the polypeptide chain moving back and forth between the layers.

Parallel or Mixed β-Sheet Proteins

The second major class of protein structures contains structures based around **parallel** or **mixed β-sheets.** Parallel β-sheet arrays, as previously discussed, distribute hydrophobic side chains on both sides of the sheet. This means that neither side of parallel β-sheets can be exposed to solvent. Parallel β-sheets are thus typically found as core structures in proteins, with little access to solvent.

Another important parallel β-array is the eight-stranded **parallel β-barrel,** exemplified in the structures of triose phosphate isomerase and pyruvate kinase (Figure 6.30). Each β-strand in the barrel is flanked by an antiparallel α-helix. The α-helices thus form a larger cylinder of parallel helices concentric with the β-barrel. Both cylinders thus formed have a right-handed twist. Another parallel β-structure consists of an internal twisted wall of parallel or mixed β-sheet protected on both sides by helices or other substructures. This structure is called the **doubly wound parallel β-sheet** because the structure can be
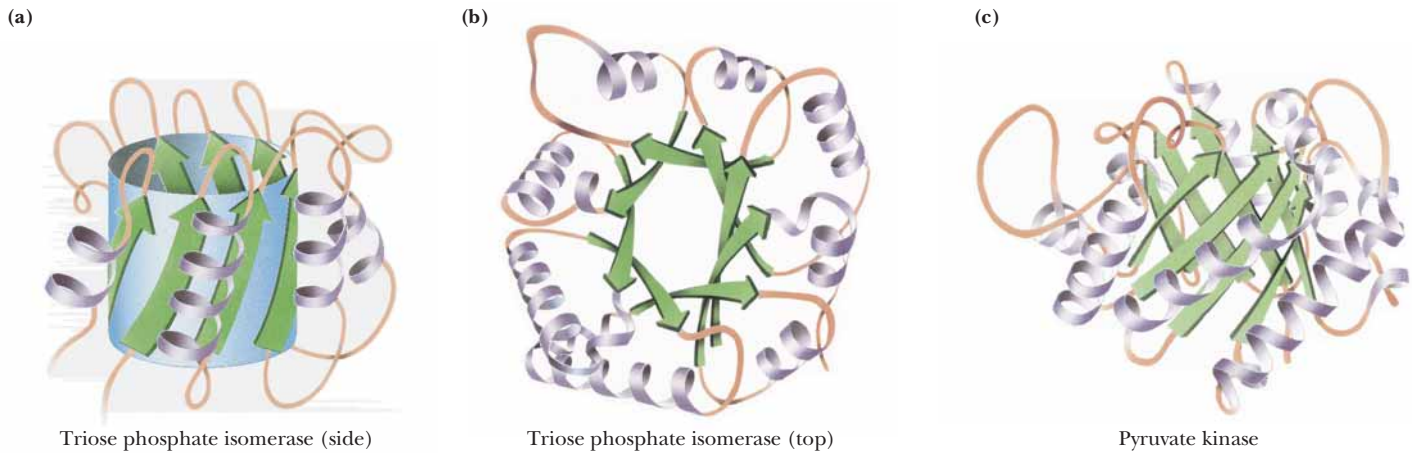
(a)

Triose phosphate isomerase (side)

(b)

Triose phosphate isomerase (top)

(c)

Pyruvate kinase

FIGURE 6.30 • Parallel β-array proteins—the eight-stranded β-barrels of triose phosphate isomerase (a, *side view,* and b, *top view*) and (c) pyruvate kinase. *(Jane Richardson)*

imagined to have been wound by strands beginning in the middle and going outward in opposite directions. The essence of this structure is shown in Figure 6.31. Whereas the barrel structures have four layers of backbone structure, the doubly wound sheet proteins have three major layers and thus two hydrophobic core regions.

Hexokinase domain 1

Flavodoxin

Flavodoxin

Phosphoglycerate mutase

Phosphoglycerate mutase
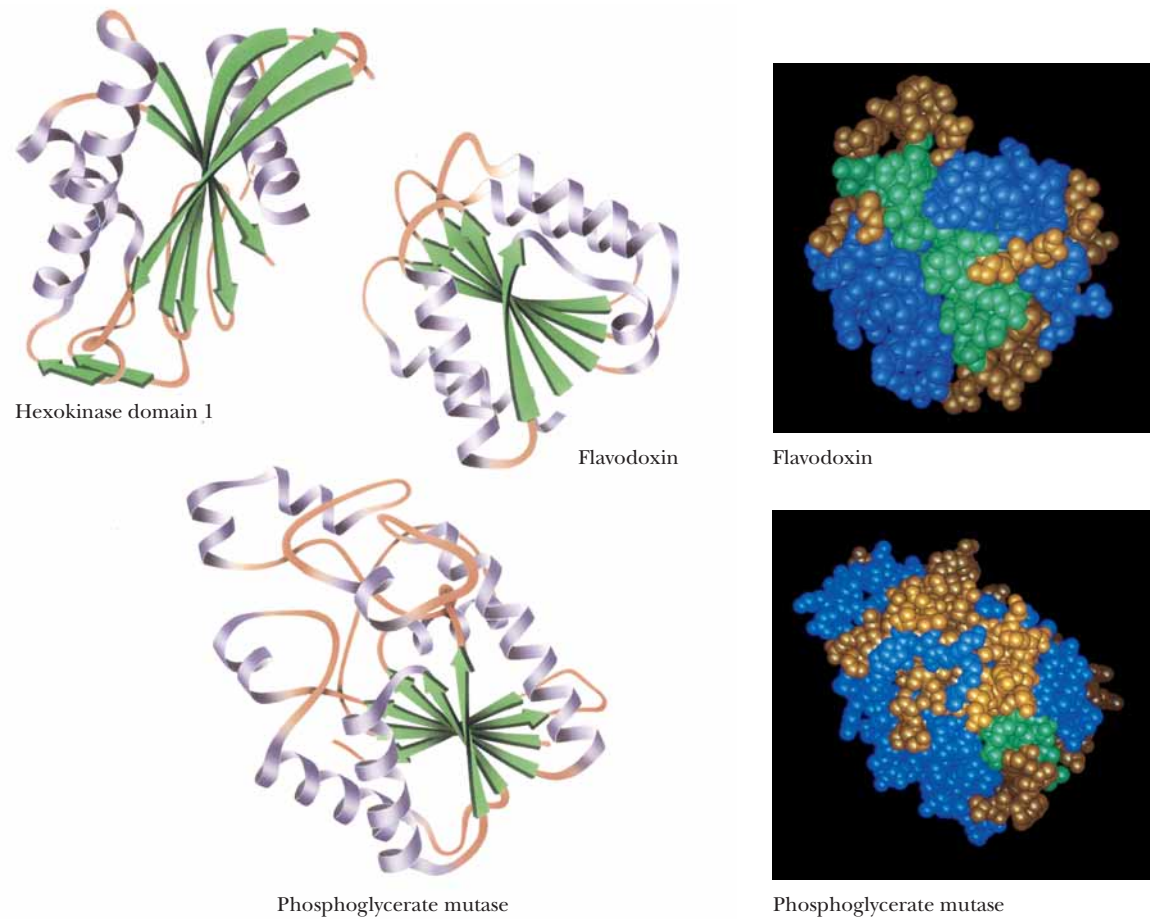
FIGURE 6.31 • Several typical doubly wound parallel β-sheet proteins. *(Jane Richardson)*

The Coiled Coil Motif in Proteins

The **coiled coil** motif was first identified in 1953 by Linus Pauling, Robert Corey, and Francis Crick as the main structural element of fibrous proteins such as keratin and myosin. Since that time, many proteins have been found to contain one or more coiled coil segments or domains. A coiled coil is a bundle of α-helices that are wound into a superhelix. Two, three, or four helical segments may be found in the bundle, and they may be arranged parallel or antiparallel to one another. Coiled coils are characterized by a distinctive and regular packing of side chains in the core of the bundle. This regular meshing of side chains requires that they occupy equivalent positions turn after turn. This is not possible for undistorted α-helices, which have 3.6 residues per turn. The positions of side chains on their surface shift continuously along the helix surface (see figure). However, giving the right-handed α-helix a left-handed twist reduces the number of residues per turn to 3.5, and, because 3.5 times 2 equals 7.0, the positions of the side chains repeat after two turns (seven residues). Thus, a **heptad repeat** pattern in the peptide sequence is diagnostic of a coiled coil structure. The figure shows a sampling of coiled coil structures (highlighted in color) in various proteins.
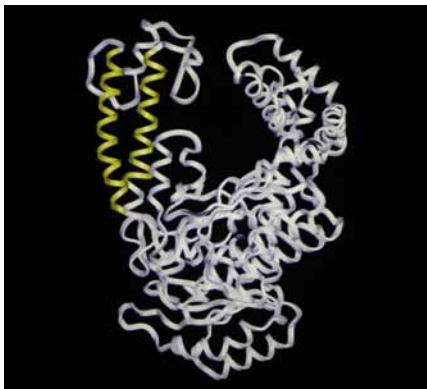

Pitch
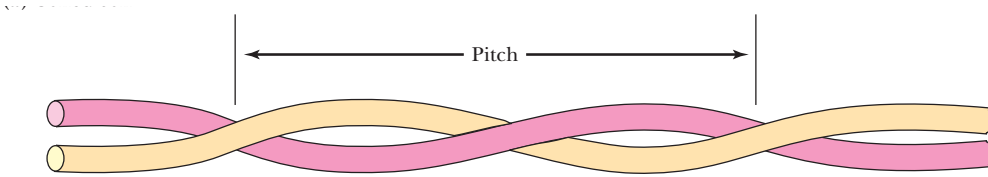
(b) Periodicity of hydrophobic residues

Undistorted   Supercoiled   Left-handed coiled coil

Helices with a heptad repeat of hydrophobic residues
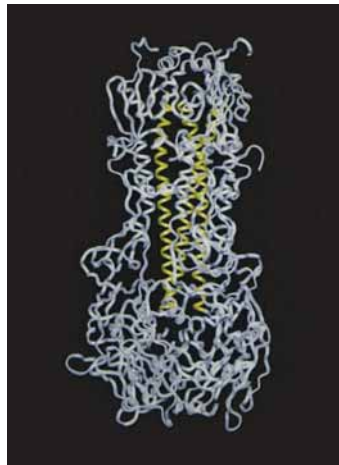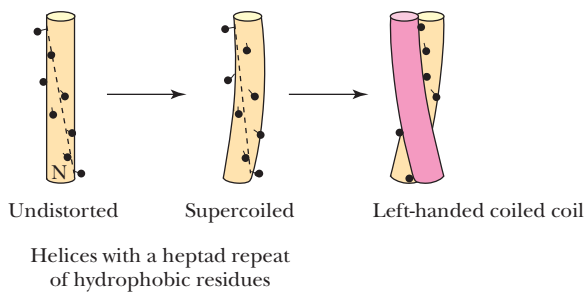
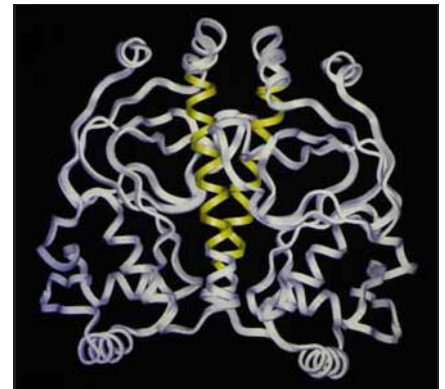

Influenza hemagglutinin
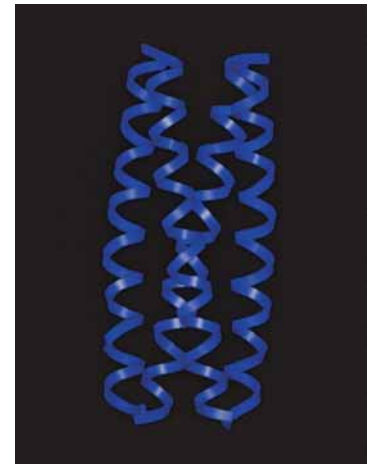
GCN4 leucine/isoleucine mutant

DNA polymerase

Seryl tRNA synthetase

Catabolite activator protein

Antiparallel β-Sheet Proteins

Another important class of tertiary protein conformations is the **antiparallel β-sheet** structures. Antiparallel β-sheets, which usually arrange hydrophobic residues on just one side of the sheet, can exist with one side exposed to solvent. The minimal structure for an antiparallel β-sheet protein is thus a two-layered structure, with hydrophobic faces of the two sheets juxtaposed and the opposite faces exposed to solvent. Such domains consist of β-sheets arranged in a cylinder or barrel shape. These structures are usually less symmetric than the singly wound parallel barrels and are not as efficiently hydrogen bonded, but they occur much more frequently in nature. Barrel structures tend to be either all parallel or all antiparallel and usually consist of even numbers of β-strands. Good examples of antiparallel structures include soybean trypsin inhibitor, rubredoxin, and domain 2 of papain (Figure 6.32). Topology diagrams of antiparallel β-sheet barrels reveal that many of them arrange the polypeptide sequence in an interlocking pattern reminiscent of patterns found on ancient Greek vases (Figure 6.33) and are thus referred to as a *Greek key topology*. Several of these, including concanavalin A and γ-crystallin, contain an extra swirl in the Greek key pattern (see Figure 6.33). Antiparallel arrangements of β-strands can also form sheets as well as barrels. Glyceraldehyde-3-phosphate dehydrogenase, *Streptomyces* subtilisin inhibitor, and glutathione reductase are examples of single-sheet, double-layered topology (Figure 6.34).

Metal- and Disulfide-Rich Proteins

Other than the structural classes just described and a few miscellaneous structures that do not fit nicely into these categories, there is only one other major class of protein tertiary structures—the small metal-rich and disulfide-rich structures. These proteins or fragments of proteins are usually small (<100 residues), and their conformations are heavily influenced by their high content of either liganded metals or disulfide bonds. The structures of disulfide-rich proteins are unstable if their disulfide bonds are broken. Figure 6.35 shows several representative disulfide-rich proteins, including insulin, phospholipase A_2, and crambin (from the seeds of *Crambe abyssinica*), as well as several metal-rich proteins, including ferredoxin and high potential iron protein (HiPIP). The structures of some of these proteins bear a striking resemblance to structural classes that have already been discussed. For example, phospholipase A_2
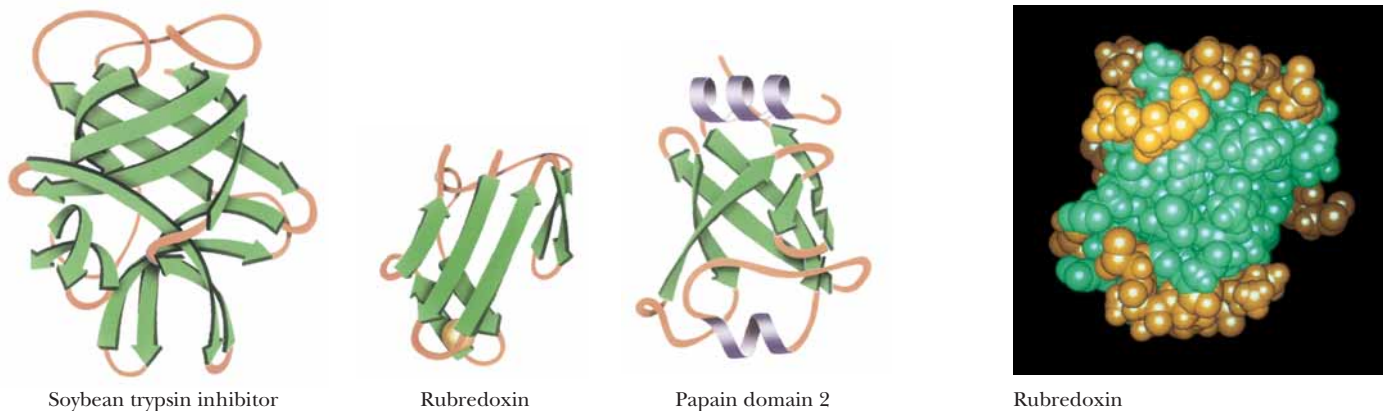

| Soybean trypsin inhibitor | Rubredoxin | Papain domain 2 | Rubredoxin |

FIGURE 6.32 • Examples of antiparallel β-sheet structures in proteins. *(Jane Richardson)*

FIGURE 6.33 • Examples of the so-called Greek key antiparallel β-barrel structure in proteins.
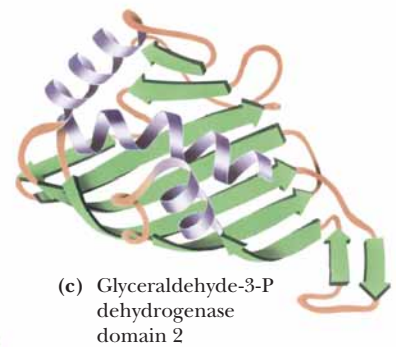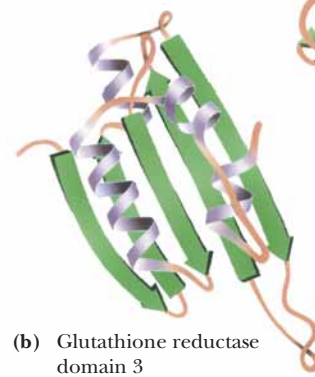

"Greek key" topology

Concanavalin A

Concanavalin A

γ-Crystallin

γ-Crystallin

Streptomyces subtilisin inhibitor

(a) *Streptomyces* subtilisin inhibitor

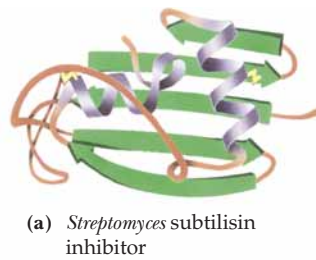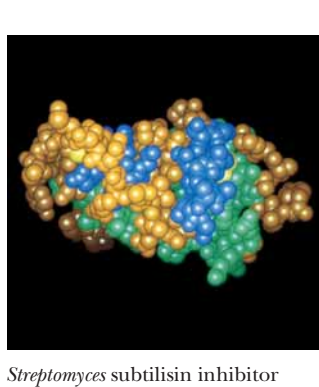**(c)** Glyceraldehyde-3-P dehydrogenase domain 2
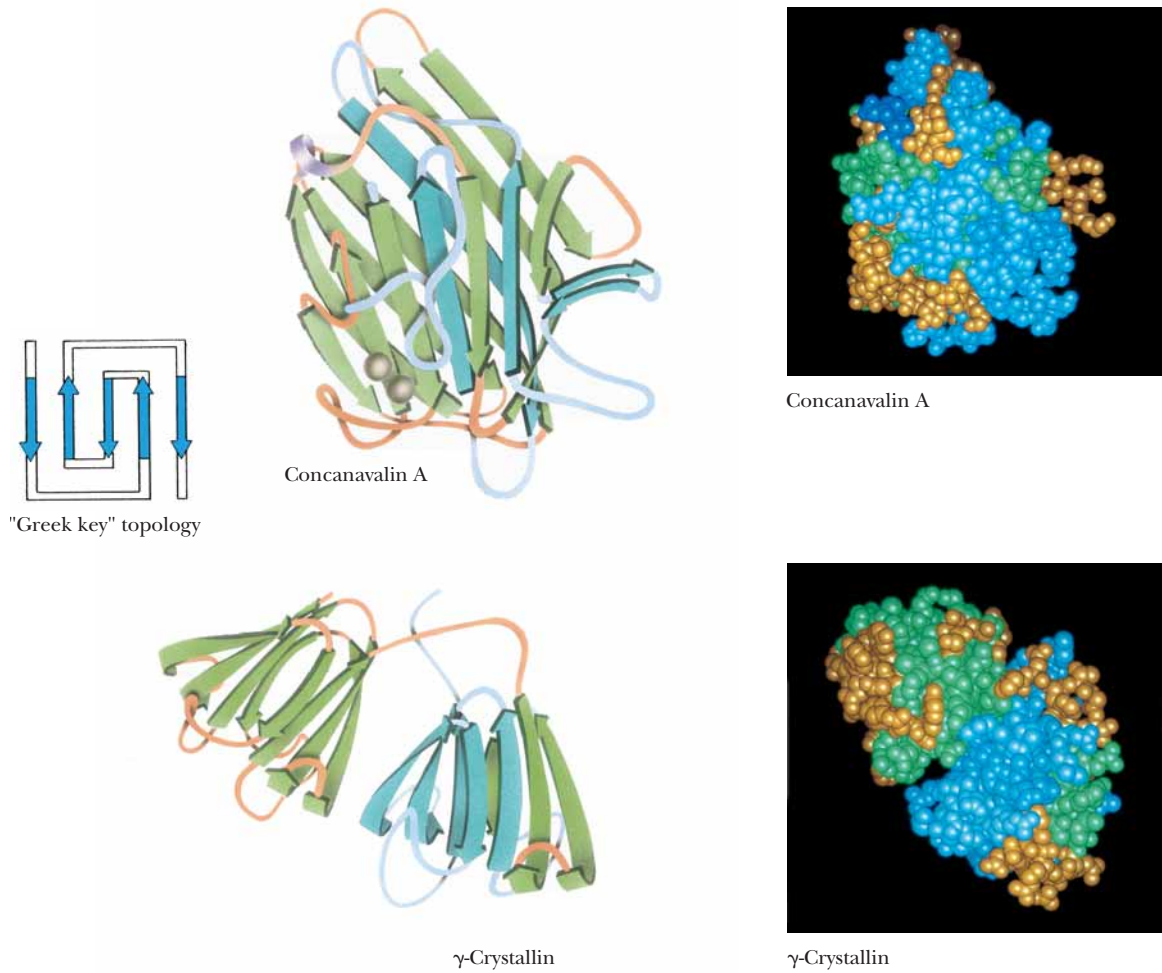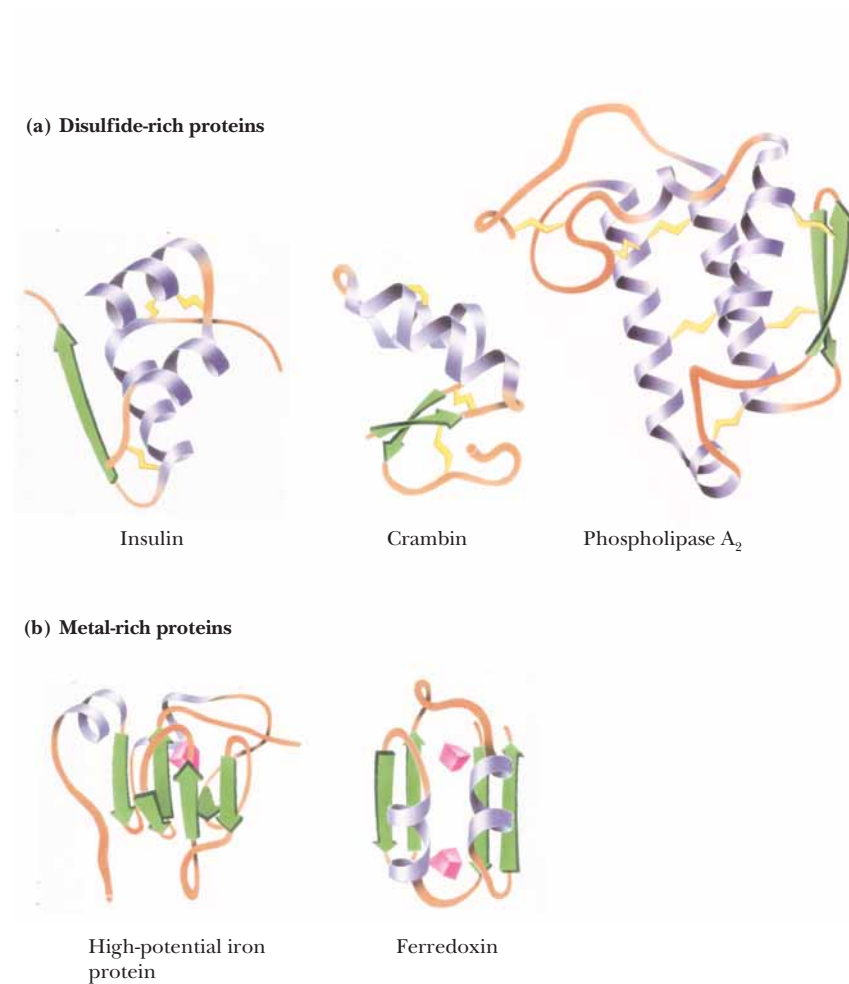

(b) Glutathione reductase domain 3

FIGURE 6.34 • Sheet structures formed from antiparallel arrangements of β-strands. (a) *Streptomyces* subtilisin inhibitor, (b) glutathione reductase domain 3, and (c) the second domain of glyceraldehyde-3-phosphate dehydrogenase represent minimal antiparallel β-sheet domain structures. In each of these cases, an antiparallel β-sheet is largely exposed to solvent on one face and covered by helices and random coils on the other face. *(Jane Richardson)*

FIGURE 6.35 • Examples of the (a) disulfide-rich and (b) metal-rich proteins. (*Jane Richardson*)

(a) Disulfide-rich proteins

Insulin   Crambin   Phospholipase A$_2$

(b) Metal-rich proteins

High-potential iron   Ferredoxin
protein

Insulin

Crambin

Phospholipase A$_2$

is a distorted α-helix cluster, whereas HiPIP is a distorted β-barrel structure. Others among this class (such as insulin and crambin), however, are not easily likened to any of the standard structure classes.

## Molecular Chaperones: Proteins That Help Fold Globular Proteins

The landmark experiments by Christian Anfinsen on the refolding of ribonuclease clearly show that the refolding of a denatured protein *in vitro* can be a spontaneous process. As noted above, this refolding is driven by the small Gibbs free energy difference between the unfolded and folded states. It has also been generally assumed that all the information necessary for the correct folding of a polypeptide chain is contained in the primary structure and requires no additional molecular factors. However, the folding of proteins in the cell is a different matter. The highly concentrated protein matrix in the cell may adversely affect the folding process by causing aggregation of some unfolded or partially folded proteins. Also, it may be necessary to accelerate slow steps in the folding process or to suppress or reverse incorrect or premature folding. Recent studies have uncovered a family of proteins, known as **molecular chaperones,**
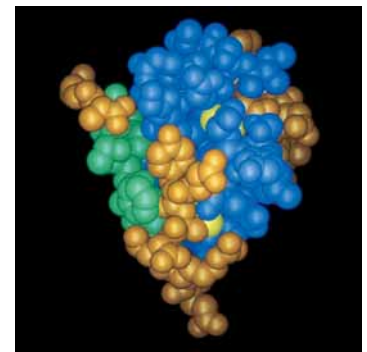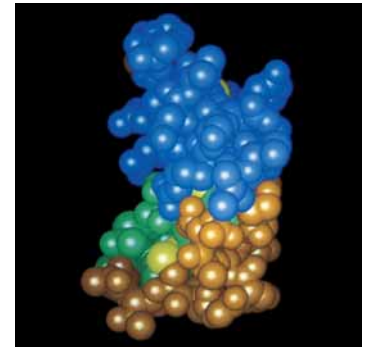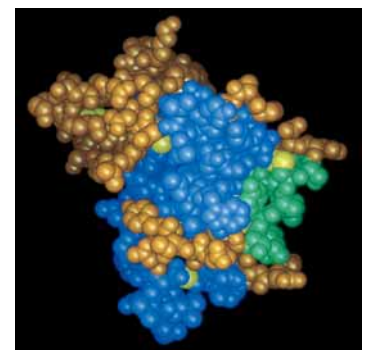
CRITICAL DEVELOPMENTS IN BIOCHEMISTRY

Thermodynamics of the Folding Process in Globular Proteins

Section 6.1 considered the noncovalent binding energies that stabilize a protein structure. However, the folding of a protein depends ultimately on the difference in Gibbs free energy (ΔG) between the folded (F) and unfolded (U) states at some temperature *T*:

$$\Delta G = G_F - G_U = \Delta H - T\Delta S$$
$$= (H_F - H_U) - T(S_F - S_U)$$

In the unfolded state, the peptide chain and its R groups interact with solvent water, and any measurement of the free energy change upon folding must consider contributions to the enthalpy change (ΔH) and the entropy change (ΔS) both for the polypeptide chain and for the solvent:

$$\Delta G_{total} = \Delta H_{chain} + \Delta H_{solvent}$$
$$- T\Delta S_{chain} - T\Delta S_{solvent}$$

If each of the four terms on the right side of this equation is understood, the thermodynamic basis for protein folding should be clear. A summary of the signs and magnitudes of these quantities for a typical protein is shown in the accompanying figure. The folded protein is a highly ordered structure compared to the unfolded state, so ΔS_{chain} is a negative number and thus $-T\Delta S_{chain}$ is a positive quantity in the equation. The other terms depend on the nature of the particular ensemble of R groups. The nature of ΔH_{chain} depends on both residue–residue interactions and residue–solvent interactions. Nonpolar groups in the folded protein interact mainly with one another via weak van der Waals forces. Interactions between nonpolar groups and water in the unfolded state are stronger because the polar water molecules induce dipoles in the nonpolar groups, producing a significant electrostatic interaction. As a result, ΔH_{chain} is positive for nonpolar groups and favors the unfolded state. $\Delta H_{solvent}$ for nonpolar groups, however, is negative and favors the folded state. This is because folding allows many water molecules to interact (favorably) with one another rather than (less favorably) with the nonpolar side chains. The magnitude of ΔH_{chain} is smaller than that of $\Delta H_{solvent}$, but both these terms are small and usually do not dominate the folding process. However, $\Delta S_{solvent}$ for nonpolar

groups is large and positive and strongly favors the folded state. This is because nonpolar groups force order upon the water solvent in the unfolded state.

For polar side chains, ΔH_{chain} is positive and $\Delta H_{solvent}$ is negative. Because solvent molecules are ordered to some extent around polar groups, $\Delta S_{solvent}$ is small and positive. As shown in the figure, ΔG_{total} for the polar groups of a protein is near zero. Comparison of all the terms considered here makes it clear that *the single largest contribution to the stability of a folded protein is $\Delta S_{solvent}$ for the nonpolar residues.*
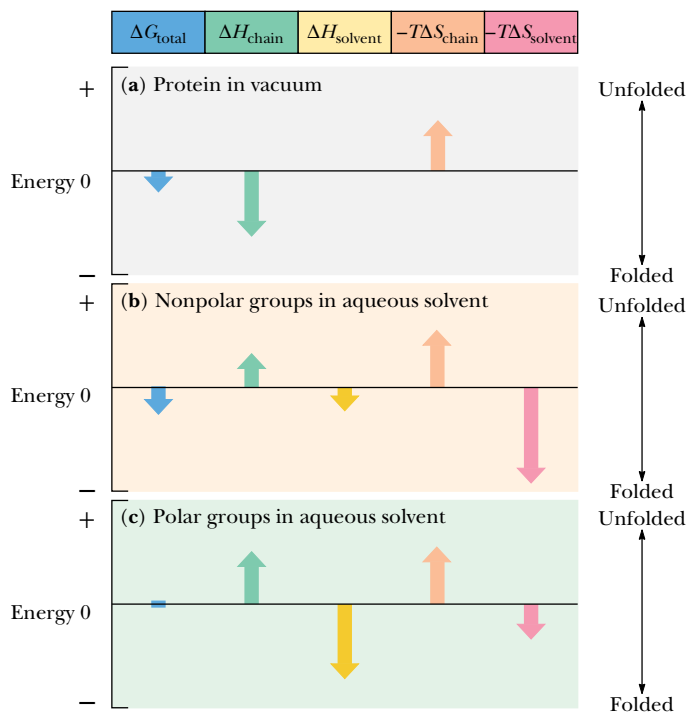
that appear to be essential for the correct folding of certain polypeptide chains *in vivo,* for their assembly into oligomers, and for preventing inappropriate liaisons with other proteins during their synthesis, folding, and transport. Many of these proteins were first identified as **heat shock proteins,** which are induced in cells by elevated temperature or other stress. The most thoroughly studied proteins are **hsp70,** a 70-kD heat shock protein, and the so-called **chaperonins,** also known as **cpn60s** or **hsp60s,** a class of 60-kD heat shock proteins. A well-characterized **hsp60** chaperonin is **GroEL,** an *E. coli* protein that has been shown to affect the folding of several proteins.

The way in which molecular chaperones interact with polypeptides during the folding process is not completely understood. What *is* clear is that chaperones bind effectively to the exposed hydrophobic regions of partially folded structures. These folding intermediates are less compact than the native folded proteins. They contain large amounts of secondary and even some tertiary

structure, but they undergo relatively large conformational fluctuations. It is possible that chaperone proteins recognize exposed helices or other secondary structure elements on their target proteins. This initial interaction may then allow the chaperone to guide or regulate the subsequent events of folding (Figure 6.36).

(1) The rapid and reversible formation of local secondary structures

(2) Formation of domains through the cooperative aggregation of folding nuclei

(3) "Molten globule" formation of the assembled domains

(4) An adjustment in the conformation of the domains
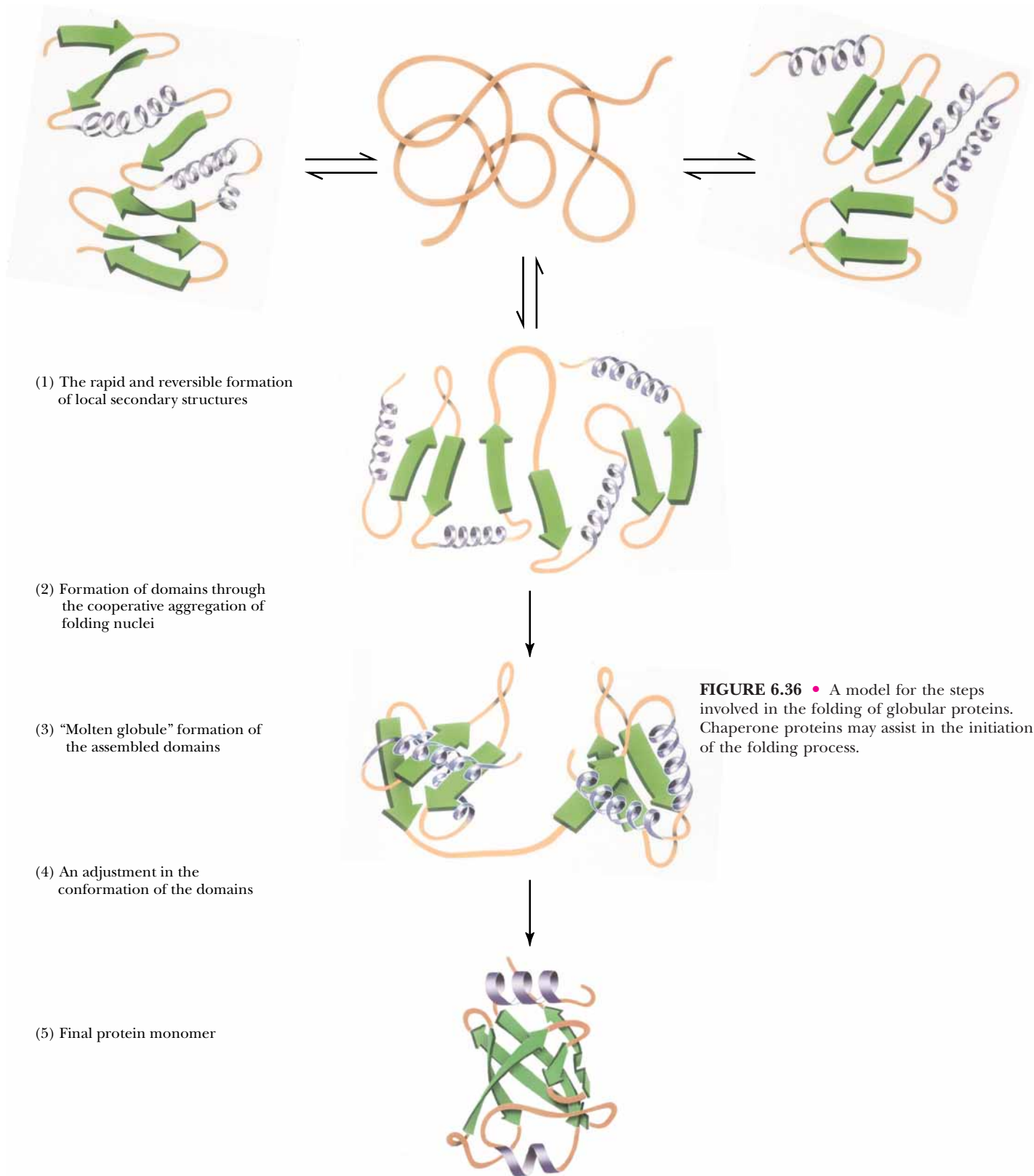
(5) Final protein monomer

FIGURE 6.36 • A model for the steps involved in the folding of globular proteins. Chaperone proteins may assist in the initiation of the folding process.

$\mathcal{H}$UMAN BIOCHEMISTRY

A Mutant Protein That Folds Slowly Can Cause Emphysema and Liver Damage

Lungs enable animals to acquire oxygen from the air and to give off CO_2 produced in respiration. Exchange of oxygen and CO_2 occurs in the alveoli—air sacs surrounded by capillaries that connect the pulmonary veins and arteries. The walls of alveoli consist of the elastic protein **elastin.** Inhalation expands the alveoli and exhalation compresses them. A pair of human lungs contains 300 million alveoli, and the total area of the alveolar walls in contact with capillaries is about 70 m^2—an area about the size of a tennis court! White blood cells naturally secrete **elastase**—a serine protease—which can attack and break down the elastin of the alveolar walls. However, **α_1-antitrypsin**—a 52-kD protein belonging to the **serpin** (*ser*ine *p*rotease *in*hibitor) family—normally binds to elastase, preventing alveolar damage. The structural gene for α_1-antitrypsin is extremely polymorphic (i.e., occurs as many different sequence variants), and several versions of this gene encode a protein that is poorly secreted into the circulation. Deficiency of α_1-antitrypsin in the blood can lead to destruction of the alveolar walls by white cell elastase, resulting in **emphysema** —a condition in which the alveoli are destroyed, leaving large air sacs that cannot be compressed during exhalation.

α_1-Antitrypsin normally adopts a highly ordered tertiary structure composed of three β-sheets and eight α-helices (see figure). Elastase and other serine proteases interact with a reactive, inhibitory site involving two amino acids—Met^{358} and Ser^{359}— on the so-called **reactive-center loop.** Formation of a tight complex between elastase and α_1-antitrypsin renders the elastase inactive. The most common α_1-antitrypsin deficiency involves the so-called **Z-variant** of the protein, in which lysine is substituted for glutamate at position 342 ($Glu^{342}\rightarrow Lys$). Residue 342 lies at the amino-terminal base of the **reactive-center loop,** and glutamate at this position normally forms a crucial salt bridge with Lys^{290} on an adjacent strand of sheet A (see figure). In normal α_1-antitrypsin, the reactive-center loop is fully exposed and can interact readily with elastase. However, in the Z-variant, the $Glu^{342}\rightarrow Lys$ substitution destabilizes sheet A, separating the strands slightly and allowing the reactive-center loop of one molecule to insert into the β-sheet of another. Repetition of this anomalous association of α_1-antitrypsin molecules results in "loop-sheet" polymerization and the formation of large protein aggregates. α_1-Antitrypsin is synthesized in hepatocytes of the liver and is normally secreted into the circulation. Accumulation of Z-variant protein aggregates in the endoplasmic reticulum of liver cells leads to deficiency of circulating α_1-antitrypsin. In some cases, the accumulation of these protein aggregates can also cause liver damage.

Myeong-Hee Yu and coworkers at the Korea Institute of Science and Technology have studied the folding kinetics of normal and Z-variant α_1-antitrypsin and have found that the Z-variant of α_1-antitrypsin folds identically to—but much more slowly than—normal α_1-antitrypsin. Newly synthesized Z-variant protein, incubated for 5 hours at 30°C, eventually adopts a native and active conformation and can associate tightly with elastase. However, incubation of the Z-variant at 37°C results in loop-sheet polymerization and self-aggregation of the protein. These results imply that emphysema arising in individuals carrying the Z-variant of α_1-antitrypsin is due to the slow folding kinetics of the protein rather than the adoption of an altered three-dimensional structure.
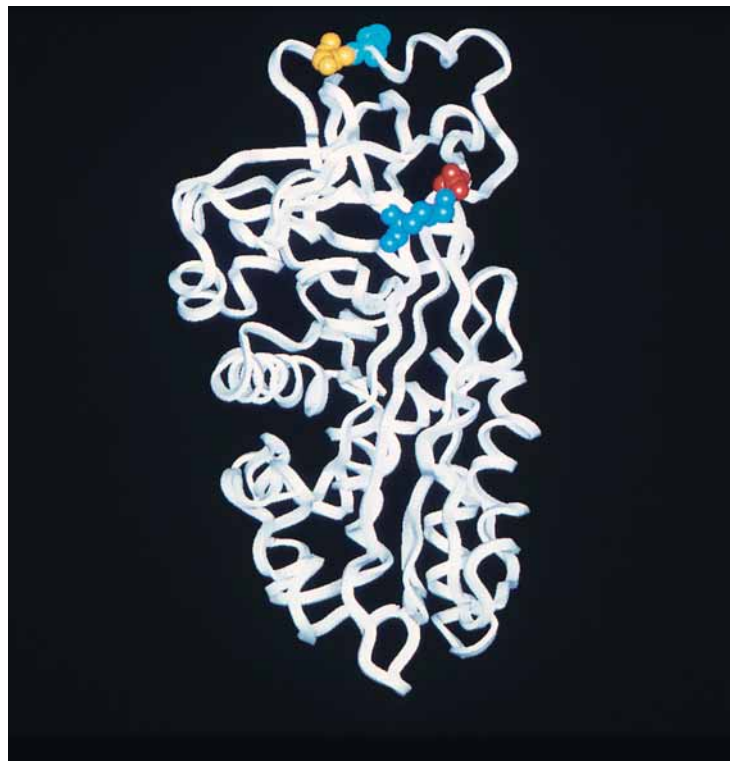
α_1-Antitrypsin. Note Met^{358} (blue) and Ser^{354} (yellow) at top, as well as Glu^{342} (red) and Lys^{290} (blue—upper right).

Protein Domains: Nature's Modular Strategy for Protein Design

Now that many thousands of proteins have been sequenced (more than 100,000 sequences are known), it has become obvious that certain protein sequences that give rise to distinct structural domains are used over and over again in modular fashion. These **protein modules** may occur in a wide variety of pro-
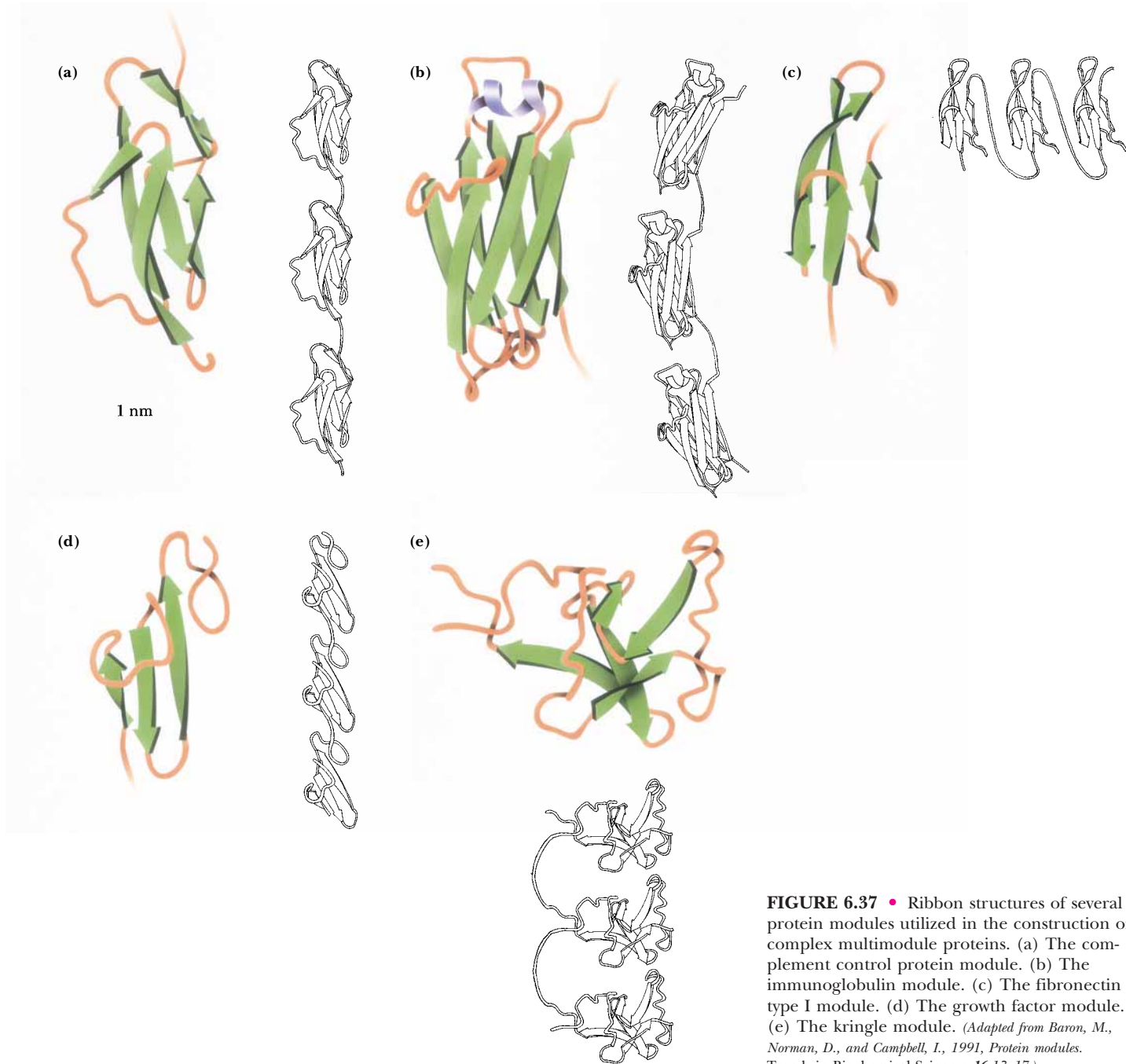
FIGURE 6.37 • Ribbon structures of several protein modules utilized in the construction of complex multimodule proteins. (a) The complement control protein module. (b) The immunoglobulin module. (c) The fibronectin type I module. (d) The growth factor module. (e) The kringle module. *(Adapted from Baron, M., Norman, D., and Campbell, I., 1991, Protein modules.* Trends in Biochemical Sciences *16:13–17.)*

teins, often being used for different purposes, or they may be used repeatedly in the same protein. Figure 6.37 shows the tertiary structures of five protein modules, and Figure 6.38 presents several proteins that contain versions of these modules. These modules typically contain about 40 to 100 amino acids and often adopt a stable tertiary structure when isolated from their parent protein. One of the best-known examples of a protein module is the **immunoglobulin module,** which has been found not only in immunoglobulins but also in a wide variety of cell surface proteins, including cell adhesion molecules and growth factor receptors, and even in *twitchin,* an intracellular protein found in muscle. It is likely that more protein modules will be identified. (The role of protein modules in signal transduction is discussed in Chapter 34.)

FIGURE 6.38 • A sampling of proteins that consist of mosaics of individual protein modules. The modules shown include γCG, a module containing γ-carboxyglutamate residues; G, an epidermal growth-factor–like module; K, the "kringle" domain, named for a Danish pastry; C, which is found in complement proteins; F1, F2, and F3, first found in fibronectin; I, the immunoglobulin superfamily domain; N, found in some growth factor receptors; E, a module homologous to the calcium-binding E–F hand domain; and LB, a lectin module found in some cell surface proteins. *(Adapted from Baron, M., Norman, D., and Campbell, I., 1991, Protein modules.* Trends in Biochemical Sciences **16**:13–17.)
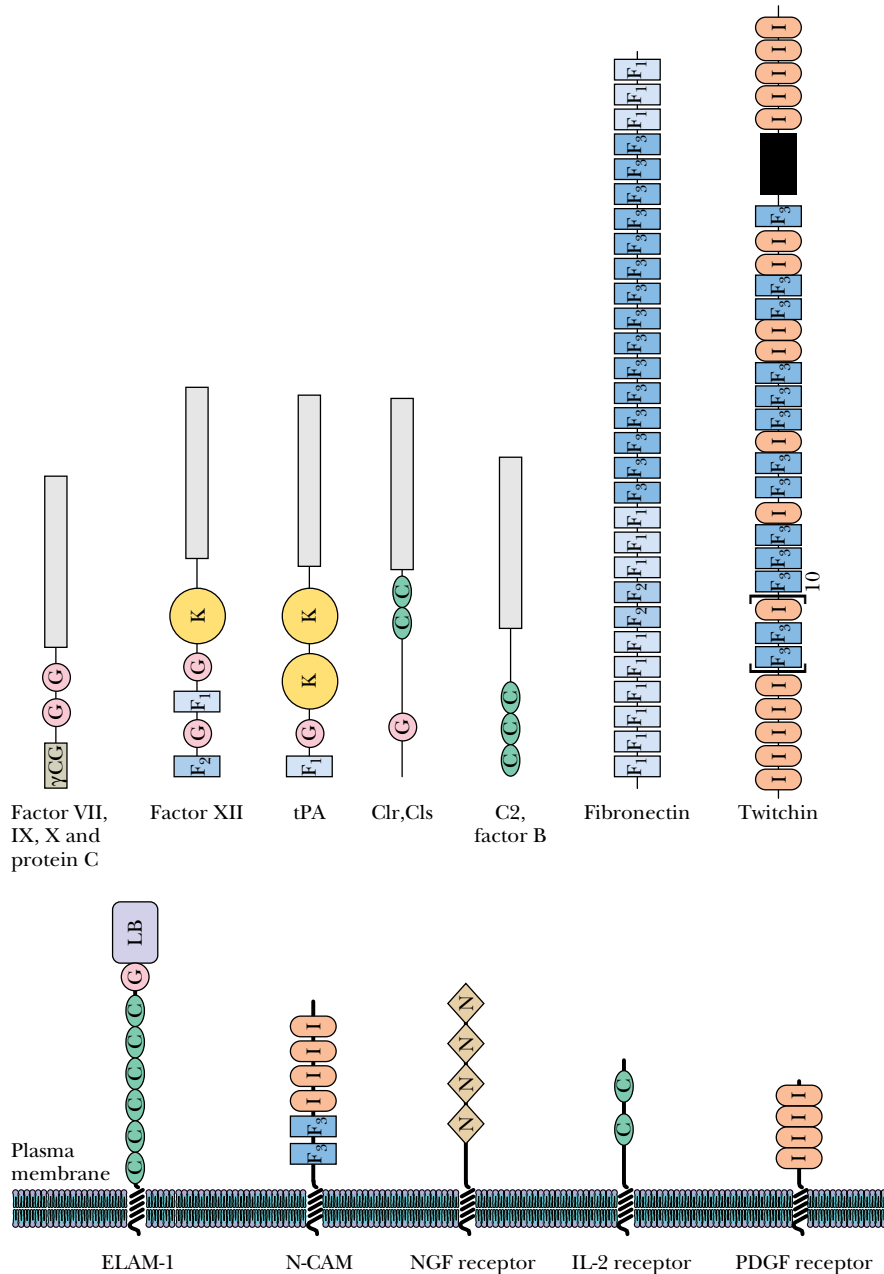
How Do Proteins Know How to Fold?

Christian Anfinsen's experiments demonstrated that proteins can fold reversibly. A corollary result of Anfinsen's work is that the native structures of at least some globular proteins are thermodynamically stable states. But the matter of how a given protein achieves such a stable state is a complex one. Cyrus Levinthal pointed out in 1968 that so many conformations are possible for a typical protein that the protein does not have sufficient time to reach its most stable conformational state by sampling all the possible conformations. This argument, termed "Levinthal's paradox," goes as follows: consider a protein of 100 amino acids. Assume that there are only two conformational possibilities per amino acid, or $2^{100} = 1.27 \times 10^{30}$ possibilities. Allow 10^{-13} sec for

the protein to test each conformational possibility in search of the overall energy minimum:

$$(10^{-13} \text{ sec})(1.27 \times 10^{30}) = 1.27 \times 10^{17} \text{ sec} = 4 \times 10^9 \text{ years}$$

Levinthal's paradox led protein chemists to hypothesize that proteins must fold by specific "folding pathways," and many research efforts have been devoted to the search for these pathways.

Implicit in the presumption of folding pathways is the existence of intermediate, partially folded conformational states. The notion of intermediate states on the pathway to a tertiary structure raises the possibility that segments of a protein might independently adopt local and well-defined secondary structures (α-helices and β-sheets). The tendency of a peptide segment to prefer a particular secondary structure depends in turn on its amino acid composition and sequence.

Surveys of the frequency with which various residues appear in helices and sheets show that some residues, such as alanine, glutamate, and methionine, occur much more frequently in α-helices than do others. In contrast, glycine and proline are the least likely residues to be found in an α-helix. Likewise, certain residues, including valine, isoleucine, and the aromatic amino acids, are more likely to be found in β-sheets than other residues, and aspartate, glutamate, and proline are much less likely to be found in β-sheets.

Such observations have led to many efforts to predict the occurrence of secondary structure in proteins from knowledge of the peptide sequence. Such **predictive algorithms** consider the composition of short segments of a polypeptide. If these segments are rich in residues that are found frequently in helices or sheets, then that segment is judged likely to adopt the corresponding secondary structure. The predictive algorithm designed by Peter Chou and Gerald Fasman in 1974 used data like that in Figure 6.39 to classify the 20 amino acids for their α-helix–forming and β-sheet–forming **propensities, P_α and P_β** (Table 6.3). These residues are classified as strong helix formers (H_α), helix formers (h_α), weak helix formers (I_α), indifferent helix formers (i_α), helix breakers (b_α), and strong helix breakers (B_α). Similar classes were established by Chou
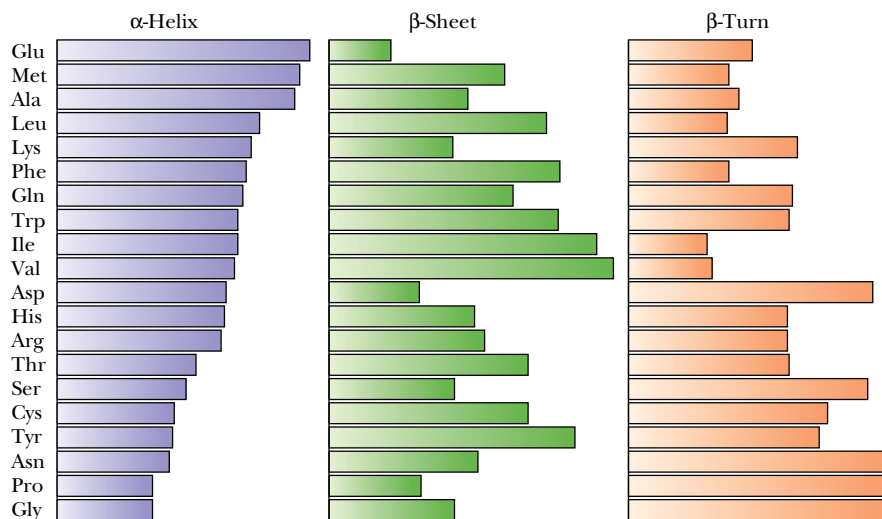
FIGURE 6.39 • Relative frequencies of occurrence of amino acid residues in α-helices, β-sheets, and β-turns in proteins of known structure. *(Adapted from Bell, J. E., and Bell, E. T., 1988, Proteins and Enzymes, Englewood Cliffs, NJ: Prentice-Hall.)*

Table 6.3

Chou–Fasman Helix and Sheet Propensities (P_α and P_β) of the Amino Acids

Amino Acid	P_α	Helix Classification	P_β	Sheet Classification
A Ala	1.42	H_α	0.83	i_β
C Cys	0.70	i_α	1.19	h_β
D Asp	1.01	I_α	0.54	B_β
E Glu	1.51	H_α	0.37	B_β
F Phe	1.13	h_α	1.38	h_β
G Gly	0.57	B_α	0.75	b_β
H His	1.00	I_α	0.87	h_β
I Ile	1.08	h_α	1.60	H_β
K Lys	1.16	h_α	0.74	b_β
L Leu	1.21	H_α	1.30	h_β
M Met	1.45	H_α	1.05	h_β
N Asn	0.67	b_α	0.89	i_β
P Pro	0.57	B_α	0.55	B_β
Q Gln	1.11	h_α	1.10	h_β
R Arg	0.98	i_α	0.93	i_β
S Ser	0.77	i_α	0.75	b_β
T Thr	0.83	i_α	1.19	h_β
V Val	1.06	h_α	1.70	H_β
W Trp	1.08	h_α	1.37	h_β
Y Tyr	0.69	b_α	1.47	H_β

Source: Chou, P. Y., and Fasman, G. D., 1978. *Annual Review of Biochemistry* **47:**258.

and Fasman for β-sheet–forming ability. Such algorithms are only modestly successful in predicting the occurrence of helices and sheets in proteins.

George Rose and Rajgopal Srinivasan at Johns Hopkins University have taken a different, and very successful, approach to the prediction of protein structures. They begin by assuming that protein folding is both *local* and *hierarchical*. "Local" in this context means that each amino acid's folding behavior is influenced by other residues nearby in the sequence. "Hierarchical" means that folded structures develop from the smallest structural units and work up to more and more complex entities. These and other assumptions are the basis for a computer program called **LINUS**—for Local Independently Nucleated Units of Structure—which Rose and Srinivasan have used to generate remarkably accurate predicted structures for a number of small proteins. LINUS considers groups of three amino acids in a sequence—for example, residues 2, 3, and 4 in a sequence of 50. The initial assumption is that this group of amino acids will (randomly) adopt one of four possible structures—helix, sheet, turn, or "loop" (any tertiary structure that is not helix, sheet, or turn). The program then asks whether this assumed "ministructure" is energetically suited to the six amino acids on either side. The program then moves on to the next set of three residues—amino acids 3, 4, and 5 in the present case—and randomly selects one of the four secondary structures for this unit, evaluates the energetic consequences in terms of the six residues on either side, and then continues in like manner to groups 4, 5, and 6, then 5, 6, and 7, and so on to the end of the protein. The program carries out this random selection and testing

routine a total of 5000 times for the entire protein. Once this is done, the program analyzes all the trials, looking for local groups of amino acids that seem to prefer one of the four conformations 70% of the time. Such groups are then held in those conformations while the program repeats the entire process, this time comparing energetic preferences with respect to 12 amino acids on either side of selected groups of three residues, then 18 amino acids on either side, 24, 32, and so on up to 48 residues. The results of calculations with LINUS on several proteins whose structures are known are shown in Figure 6.40.
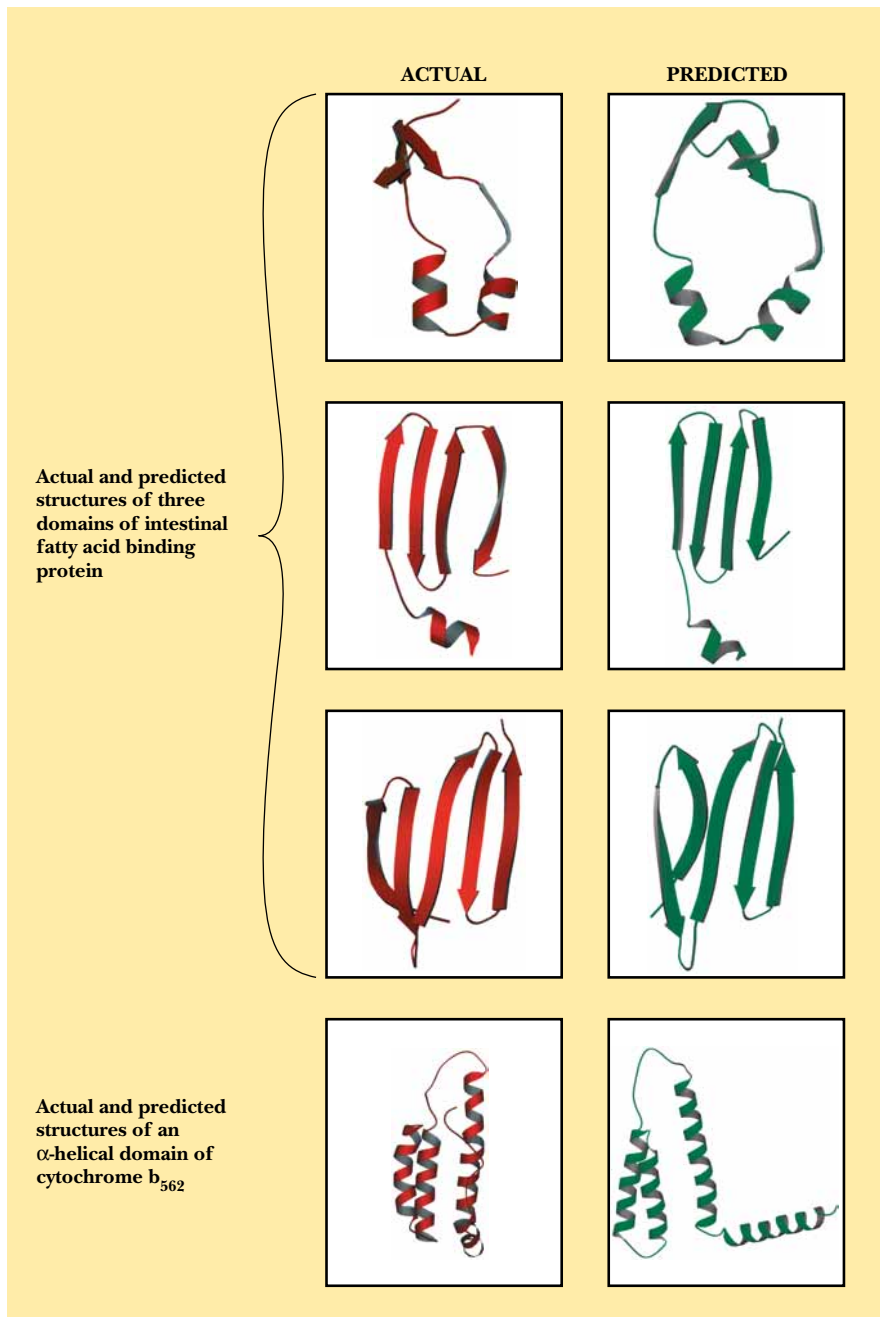

FIGURE 6.40 • A comparison of the structures of four protein domains and predictions of these structures by the program LINUS by Rose and Srinivasan. *(Professor George Rose/Johns Hopkins University)*

ACTUAL   PREDICTED

Actual and predicted structures of three domains of intestinal fatty acid binding protein

Actual and predicted structures of an α-helical domain of cytochrome b$_{562}$

6.5 • Subunit Interactions and Quaternary Structure

Many proteins exist in nature as **oligomers,** complexes composed of (often symmetric) noncovalent assemblies of two or more monomer subunits. In fact, subunit association is a common feature of macromolecular organization in biology. Most intracellular enzymes are oligomeric and may be composed either of a single type of monomer subunit (*homomultimers*) or of several different kinds of subunits (*heteromultimers*). The simplest case is a protein composed of identical subunits. Liver alcohol dehydrogenase, shown in Figure 6.41, is such a protein. More complicated proteins may have several different subunits in one, two, or more copies. Hemoglobin, for example, contains two each of two different subunits and is referred to as an $\alpha_2\beta_2$-complex. An interesting counterpoint to these relatively simple cases is made by the proteins that form polymeric structures. Tubulin is an $\alpha\beta$-dimeric protein that polymerizes to form microtubules of the formula $(\alpha\beta)_n$. The way in which separate folded monomeric protein subunits associate to form the oligomeric protein constitutes the **quaternary structure** of that protein. Table 6.4 lists several proteins and their subunit compositions (see also Table 5.1). Clearly, proteins with two to four subunits dominate the list, but many cases of higher numbers exist.
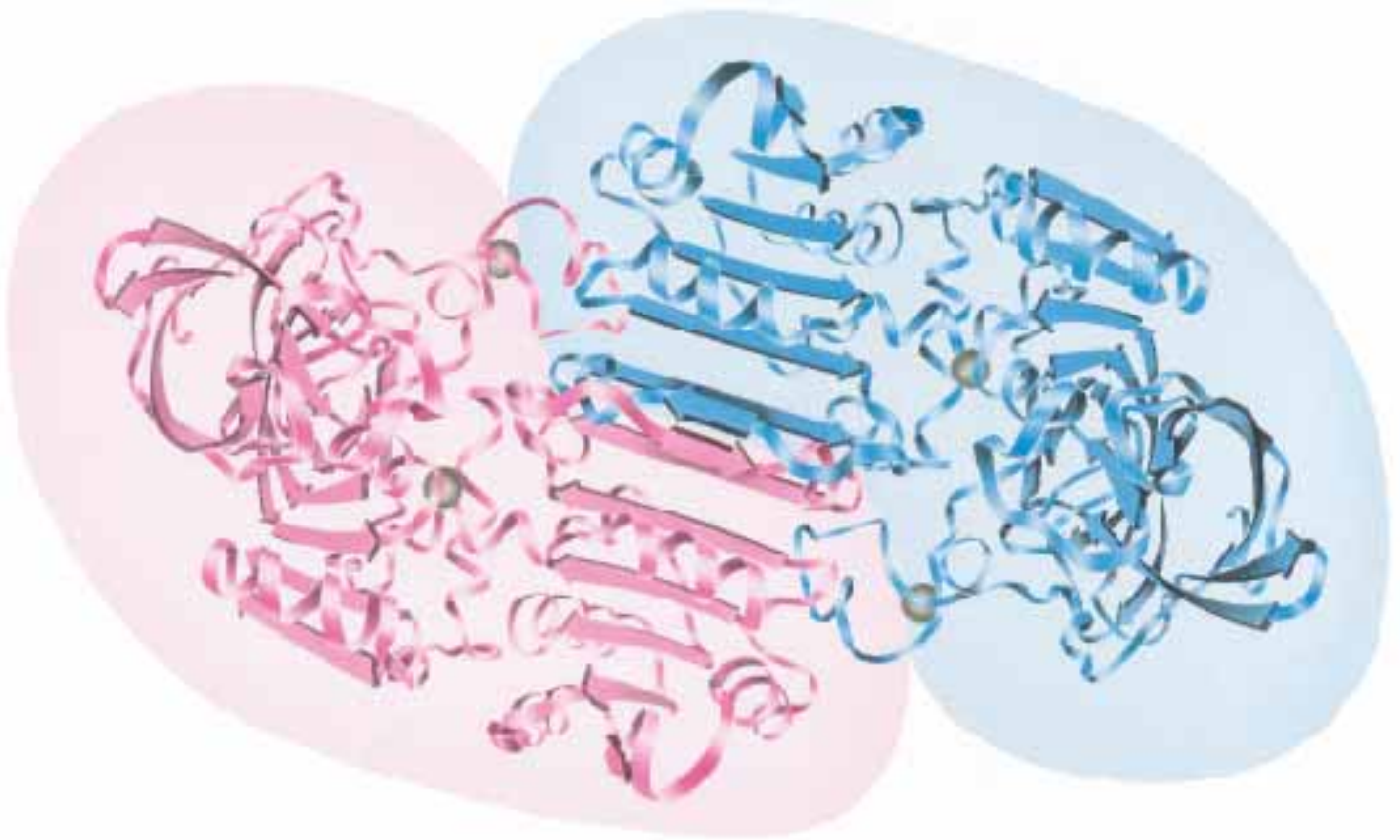

FIGURE 6.41 • The quaternary structure of liver alcohol dehydrogenase. Within each subunit is a six-stranded parallel sheet. Between the two subunits is a two-stranded antiparallel sheet. The point in the center is a C_2 symmetry axis. *(Jane Richardson)*

The subunits of an oligomeric protein typically fold into apparently independent globular conformations and then interact with other subunits. The particular surfaces at which protein subunits interact are similar in nature to the interiors of the individual subunits. These interfaces are closely packed and involve both polar and hydrophobic interactions. Interacting surfaces must therefore possess complementary arrangements of polar and hydrophobic groups.

Oligomeric associations of protein subunits can be divided into those between **identical** subunits and those between **nonidentical** subunits. Interactions among identical subunits can be further distinguished as either **isologous** or **heterologous.** In isologous interactions, the interacting surfaces are identical, and the resulting structure is necessarily dimeric and closed, with a twofold axis of symmetry (Figure 6.42). If any additional interactions occur to form a trimer or tetramer, these must use different interfaces on the protein's surface. Many proteins, including concanavalin and prealbumin, form tetramers by means of two sets of isologous interactions, one of which is shown in Figure 6.43. Such structures possess three different twofold axes of symmetry. In contrast, heterologous associations among subunits involve nonidentical interfaces. These surfaces must be complementary, but they are generally not symmetric. As shown in Figure 6.43, heterologous interactions are necessarily open-ended. This can give rise either to a closed cyclic structure, if geometric constraints exist, or to large polymeric assemblies. The closed cyclic structures are far more common and include the trimers of aspartate transcarbamoylase catalytic subunits and the tetramers of neuraminidase and hemerythrin.
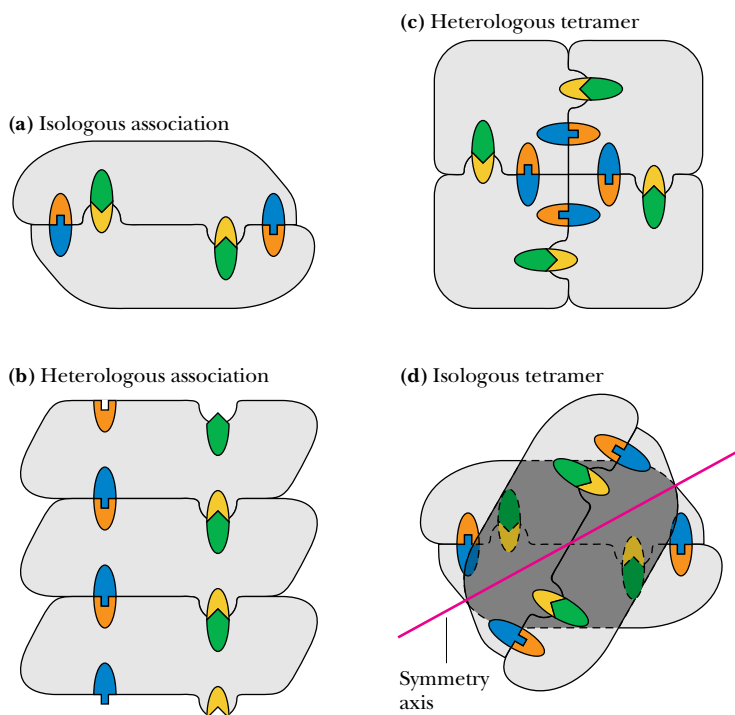
Table 6.4

Aggregation Symmetries of Globular Proteins

Protein	Number of Subunits
Alcohol dehydrogenase	2
Immunoglobulin	4
Malate dehydrogenase	2
Superoxide dismutase	2
Triose phosphate isomerase	2
Glycogen phosphorylase	2
Alkaline phosphatase	2
6-Phosphogluconate dehydrogenase	2
Wheat germ agglutinin	2
Phosphoglucoisomerase	2
Tyrosyl-tRNA synthetase	2
Glutathione reductase	2
Aldolase	3
Bacteriochlorophyll protein	3
TMV protein disc	17
Concanavalin A	4
Glyceraldehyde-3-phosphate dehydrogenase	4
Lactate dehydrogenase	4
Prealbumin	4
Pyruvate kinase	4
Phosphoglycerate mutase	4
Hemoglobin	2 + 2
Insulin	6
Aspartate transcarbamoylase	6 + 6
Glutamine synthetase	12
Apoferritin	24
Coat of tomato bushy stunt virus	180

FIGURE 6.42 • Isologous and heterologous associations between protein subunits. (a) An isologous interaction between two subunits with a twofold axis of symmetry perpendicular to the plane of the page. (b) A heterologous interaction that could lead to the formation of a long polymer. (c) A heterologous interaction leading to a closed structure—a tetramer. (d) A tetramer formed by two sets of isologous interactions.
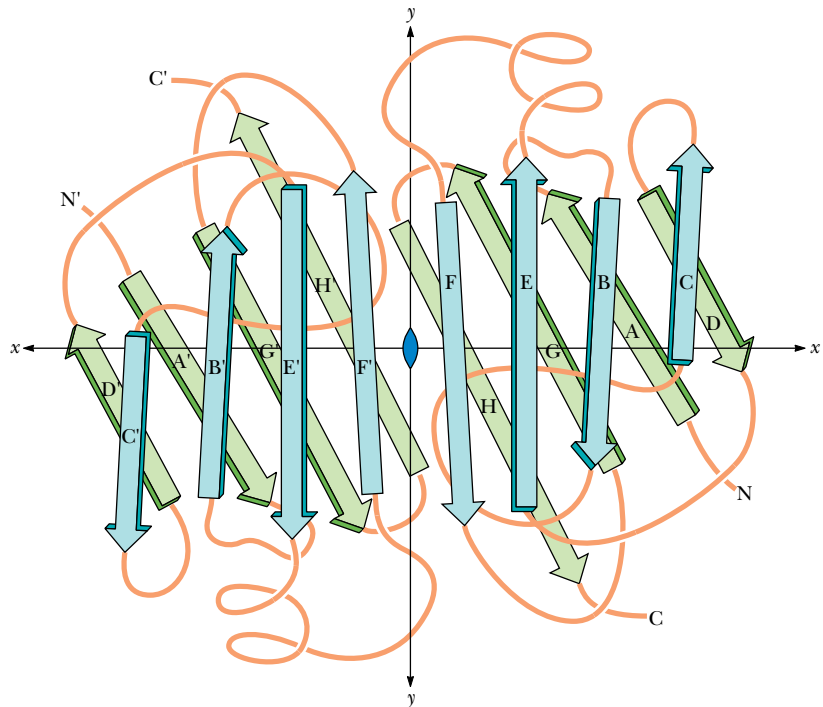
FIGURE 6.43 • The polypeptide backbone of the prealbumin dimer. The monomers associate in a manner that continues the β-sheets. A tetramer is formed by isologous inter- actions between the side chains extending outward from sheet D′A′G′H′HGAD in both dimers, which pack together nearly at right angles to one another. *(Jane Richardson)*

Symmetry of Quaternary Structures

One useful way to consider quaternary interactions in proteins involves the symmetry of these interactions. Globular protein subunits are always asym- metric objects. All of the polypeptide's α-carbons are asymmetric, and the polypeptide nearly always folds to form a low-symmetry structure. (The long helical arrays formed by some synthetic polypeptides are an exception.) Thus, protein subunits do not have mirror reflection planes, points, or axes of inver- sion. The only symmetry operation possible for protein subunits is a rotation. The most common symmetries observed for multisubunit proteins are cyclic symmetry and dihedral symmetry. In **cyclic symmetry,** the subunits are arranged around a single rotation axis, as shown in Figure 6.44. If there are two sub- units, the axis is referred to as a *twofold rotation axis*. Rotating the quaternary structure 180° about this axis gives a structure identical to the original one. With three subunits arranged about a threefold rotation axis, a rotation of 120° about that axis gives an identical structure. **Dihedral symmetry** occurs when a structure possesses at least one twofold rotation axis perpendicular to another n-fold rotation axis. This type of subunit arrangement (Figure 6.44) occurs in concanavalin A (where $n = 2$) and in insulin (where $n = 3$). Higher symmetry groups, including the tetrahedral, octahedral, and icosahedral symmetries, are much less common among multisubunit proteins, partly because of the large number of asymmetric subunits required to assemble truly symmetric tetrahe- dra and other high symmetry groups. For example, a truly symmetric tetrahe- dral protein structure would require 12 identical monomers arranged in tri- angles, as shown in Figure 6.45. Simple four-subunit tetrahedra of protein monomers, which actually possess dihedral symmetry, are more common in biological systems.
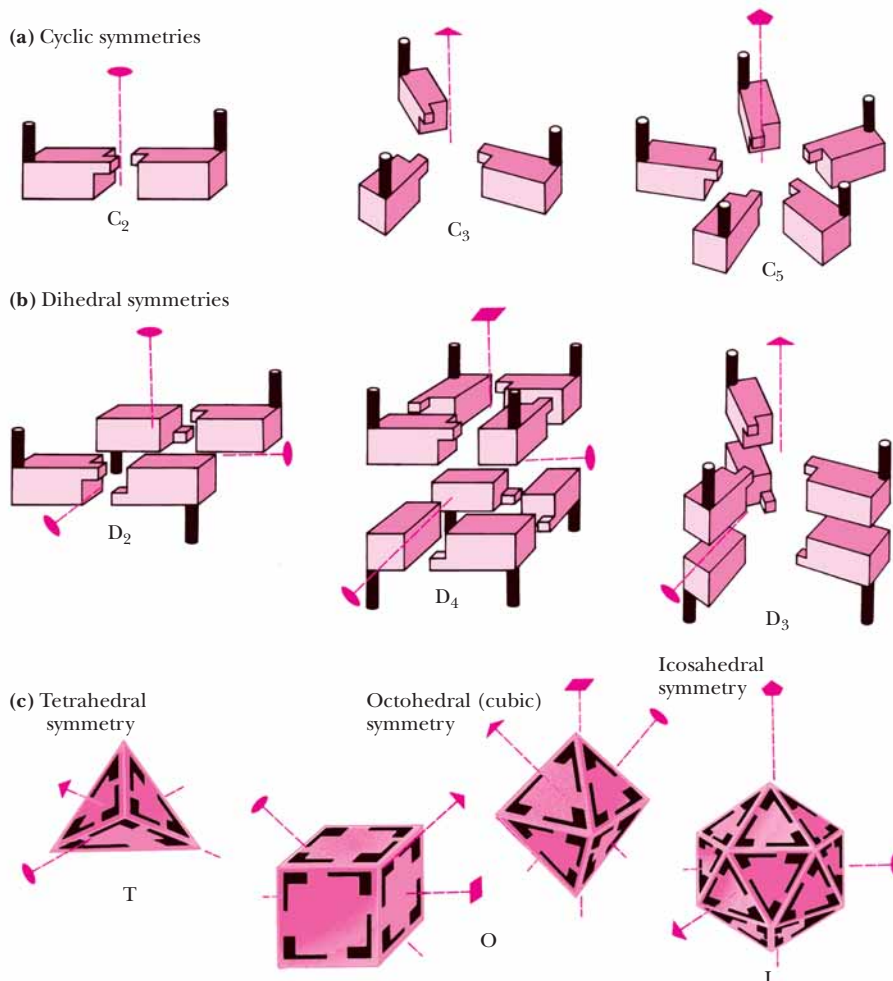
(a) Cyclic symmetries

C_2

C_3

C_5

(b) Dihedral symmetries

D_2

D_4

D_3

(c) Tetrahedral symmetry

Octohedral (cubic) symmetry

Icosahedral symmetry

T

O

I

FIGURE 6.44 • Several possible symmetric arrays of identical protein subunits, including (a) cyclic symmetry, (b) dihedral symmetry, and (c) cubic symmetry, including examples of tetrahedral (T), octahedral (O), and icosahedral (I) symmetry. *(Irving Geis)*

Forces Driving Quaternary Association

The forces that stabilize quaternary structure have been evaluated for a few proteins. Typical dissociation constants for simple two-subunit associations range from 10^{-8} to 10^{-16} M. These values correspond to free energies of association of about 50 to 100 kJ/mol at 37°C. Dimerization of subunits is accompanied by both favorable and unfavorable energy changes. The favorable interactions include van der Waals interactions, hydrogen bonds, ionic bonds, and hydrophobic interactions. However, considerable entropy loss occurs when subunits interact. When two subunits move as one, three translational degrees of freedom are lost for one subunit because it is constrained to move with the other one. In addition, many peptide residues at the subunit interface, which were previously free to move on the protein surface, now have their movements restricted by the subunit association. This unfavorable energy of association is in the range of 80 to 120 kJ/mol for temperatures of 25 to 37°C. Thus, to achieve stability, the dimerization of two subunits must involve approximately 130 to 220 kJ/mol of favorable interactions.[1] Van der Waals interactions at protein interfaces are numerous, often running to several hundred for a typical monomer–monomer association. This would account for about 150 to 200

[1]For example, 130 kJ/mol of favorable interaction minus 80 kJ/mol of unfavorable interaction equals a net free energy of association of 50 kJ/mol.

Intramolecular disulfide bridges
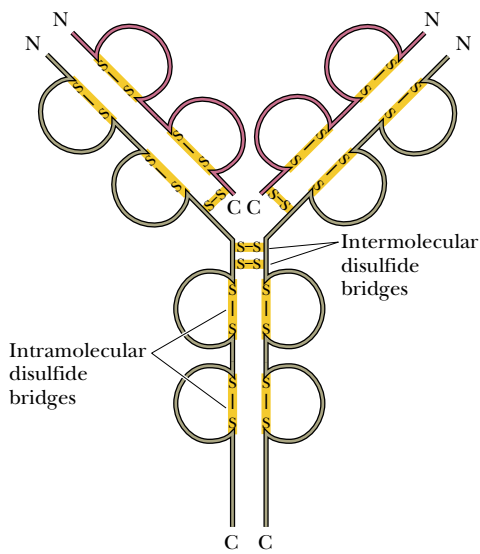
Intermolecular disulfide bridges

FIGURE 6.45 • Schematic drawing of an immunoglobulin molecule showing the intramolecular and intermolecular disulfide bridges. (A space-filling model of the same molecule is shown in Figure 1.11.)

kJ/mol of favorable free energy of association. However, when solvent is removed from the protein surface to form the subunit–subunit contacts, nearly as many van der Waals associations are lost as are made. One subunit is simply trading water molecules for peptide residues in the other subunit. As a result, the energy of subunit association due to van der Waals interactions actually contributes little to the stability of the dimer. Hydrophobic interactions, however, are generally very favorable. For many proteins, the subunit association process effectively buries as much as 20 nm^2 of surface area previously exposed to solvent, resulting in as much as 100 to 200 kJ/mol of favorable hydrophobic interactions. Together with whatever polar interactions occur at the protein–protein interface, this is sufficient to account for the observed stabilization that occurs when two protein subunits associate.

An additional and important factor contributing to the stability of subunit associations for some proteins is the formation of disulfide bonds between different subunits. All antibodies are $\alpha_2\beta_2$-tetramers composed of two heavy chains (53 to 75 kD) and two relatively light chains (23 kD). In addition to *intrasubunit* disulfide bonds (four per heavy chain, two per light chain), two *intersubunit* disulfide bridges hold the two heavy chains together and a disulfide bridge links each of the two light chains to a heavy chain (Figure 6.45).

Modes and Models for Quaternary Structures

When a protein is composed of only one kind of polypeptide chain, the manner in which the subunits interact and the arrangement of the subunits to produce the quaternary structure are usually simple matters. Sometimes, however, the same protein derived from several different species can exhibit different modes of quaternary interactions. Hemerythrin, the oxygen-carrying protein in certain species of marine invertebrates, is composed of a compact arrangement of four antiparallel α-helices. It is capable of forming dimers, trimers, tetramers, octamers, and even higher aggregates (Figure 6.46).

When two or more distinct peptide chains are involved, the nature of their interactions can be quite complicated. Multimeric proteins with more than one kind of subunit often display different affinities between different pairs of sub-
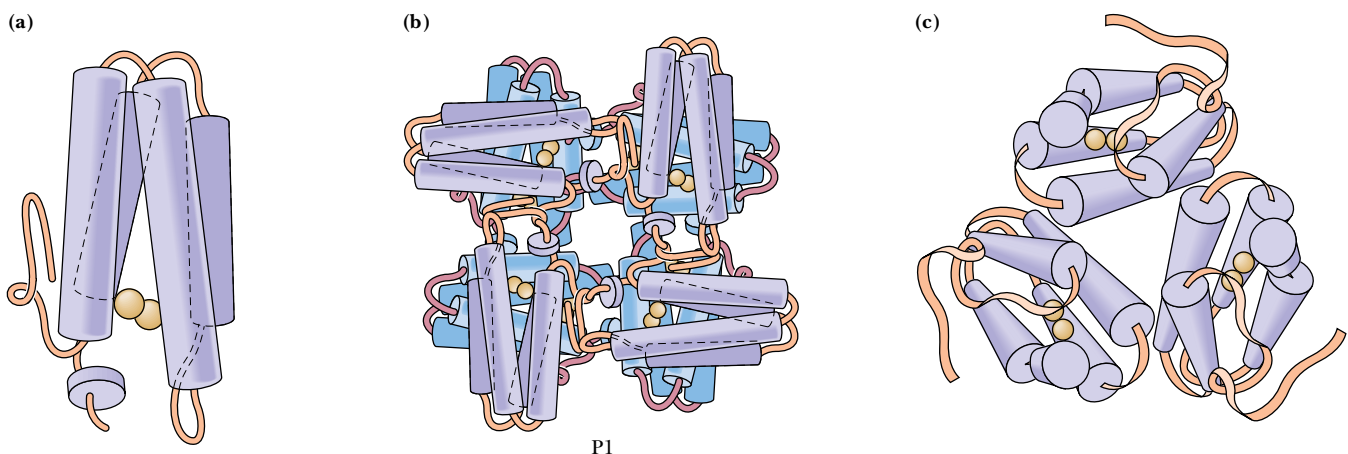
(a)

(b)

P1

(c)

FIGURE 6.46 • The oligomeric states of hemerythrin in various marine worms. (a) The hemerythrin in *Thermiste zostericola* crystallized as a monomer; (b) the octameric hemerythrin crystallized from *Phascolopsis gouldii;* (c) the trimeric hemerythrin crystallized from *Siphonosoma* collected in mangrove swamps in Fiji.

$\mathcal{A}$ **D E E P E R   L O O K**

Immunoglobulins—All the Features of Protein Structure Brought Together

The immunoglobulin structure in Figure 6.45 represents the confluence of all the details of protein structure that have been thus far discussed. As for all proteins, the primary structure determines other aspects of structure. There are numerous elements of secondary structure, including β-sheets and tight turns. The tertiary structure consists of 12 distinct domains, and the protein adopts a heterotetrameric quaternary structure. To make matters more interesting, both intrasubunit and intersubunit disulfide linkages act to stabilize the discrete domains and to stabilize the tetramer itself.

One more level of sophistication awaits. As discussed in Chapter 29, the amino acid sequences of both light and heavy immunoglobulin chains are not constant! Instead, the primary structure of these chains is highly variable in the N-terminal regions (first 108 residues). Heterogeneity of amino acid sequence leads to variations in the conformation of these variable regions. This variation accounts for antibody diversity and the ability of antibodies to recognize and bind a virtually limitless range of antigens. This full potential of antibody:antigen recognition enables organisms to mount immunologic responses to almost any antigen that might challenge the organism.

units. Whereas strongly denaturing solvents may dissociate the protein entirely into monomers, more subtle denaturing conditions may dissociate the oligomeric structure in a carefully controlled stepwise manner. Hemoglobin is a good example. Strong denaturants dissociate hemoglobin into α- and β-monomers. Using mild denaturing conditions, however, it is possible to dissociate hemoglobin almost completely into αβ-dimers, with few or no free monomers occurring. In this sense, hemoglobin behaves functionally like a two-subunit protein, with each "subunit" composed of an αβ-dimer.

Open Quaternary Structures and Polymerization

All of the quaternary structures we have considered to this point have been **closed** structures, with a limited capacity to associate. Many proteins in nature associate to form **open** heterologous structures, which can polymerize more or less indefinitely, creating structures that are both esthetically attractive and functionally important to the cells or tissue in which they exist. One such protein is **tubulin,** the αβ-dimeric protein that polymerizes into long, tubular structures, which are the structural basis of cilia flagella and the cytoskeletal matrix. The *microtubule* thus formed (Figure 6.47) may be viewed as consisting of 13 parallel filaments arising from end-to-end aggregation of the tubulin dimers. Human immunodeficiency virus, HIV, the causative agent of AIDS (also discussed in Chapter 16), is enveloped by a spherical shell composed of hundreds of coat protein subunits, a large-scale quaternary association.

Structural and Functional Advantages of Quaternary Association

There are several important reasons for protein subunits to associate in oligomeric structures.

Stability

One general benefit of subunit association is a favorable reduction of the protein's surface-to-volume ratio. The surface-to-volume ratio becomes smaller as the radius of any particle or object becomes larger. (This is because surface area is a function of the radius squared and volume is a function of the radius cubed.) Because interactions within the protein usually tend to stabilize the protein energetically and because the interaction of the protein surface with
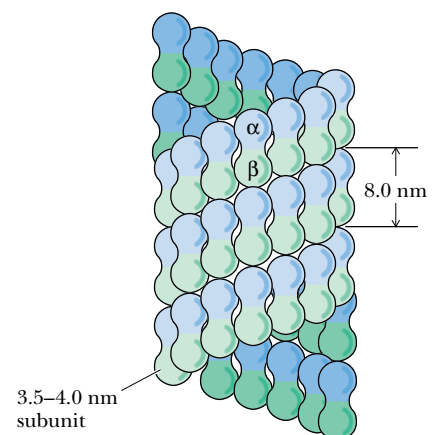
α
β
8.0 nm

3.5–4.0 nm
subunit

FIGURE 6.47 • The structure of a typical microtubule, showing the arrangement of the α- and β-monomers of the tubulin dimer.

solvent water is often energetically unfavorable, decreased surface-to-volume ratios usually result in more stable proteins. Subunit association may also serve to shield hydrophobic residues from solvent water. Subunits that recognize either themselves or other subunits avoid any errors arising in genetic translation by binding mutant forms of the subunits less tightly.

Genetic Economy and Efficiency

Oligomeric association of protein monomers is genetically economical for an organism. Less DNA is required to code for a monomer that assembles into a homomultimer than for a large polypeptide of the same molecular mass. Another way to look at this is to realize that virtually all of the information that determines oligomer assembly and subunit–subunit interaction is contained in the genetic material needed to code for the monomer. For example, HIV protease, an enzyme that is a dimer of identical subunits, performs a catalytic function similar to homologous cellular enzymes that are single polypeptide chains of twice the molecular mass (Chapter 16).
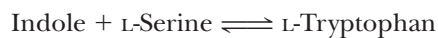
Bringing Catalytic Sites Together

Many enzymes (see Chapters 14 to 16) derive at least some of their catalytic power from oligomeric associations of monomer subunits. This can happen in several ways. The monomer may not constitute a complete enzyme active site. Formation of the oligomer may bring all the necessary catalytic groups together to form an active enzyme. For example, the active sites of bacterial glutamine synthetase are formed from pairs of adjacent subunits. The dissociated monomers are inactive.

Oligomeric enzymes may also carry out different but related reactions on different subunits. Thus, tryptophan synthase is a tetramer consisting of pairs of different subunits, $\alpha_2\beta_2$. Purified α-subunits catalyze the following reaction:

$$\text{Indoleglycerol phosphate} \rightleftharpoons \text{Indole} + \text{Glyceraldehyde-3-phosphate}$$

and the β-subunits catalyze this reaction:

$$\text{Indole} + \text{L-Serine} \rightleftharpoons \text{L-Tryptophan}$$

Indole, the product of the α-reaction and the reactant for the β-reaction, is passed directly from the α-subunit to the β-subunit and cannot be detected as a free intermediate.

Cooperativity

There is another, more important reason for monomer subunits to associate into oligomeric complexes. Most oligomeric enzymes regulate catalytic activity by means of subunit interactions, which may give rise to cooperative phenomena. Multisubunit proteins typically possess multiple binding sites for a given ligand. If the binding of ligand at one site changes the affinity of the protein for ligand at the other binding sites, the binding is said to be **cooperative.** Increases in affinity at subsequent sites represent positive cooperativity, whereas decreases in affinity correspond to negative cooperativity. The points of contact between protein subunits provide a mechanism for communication between the subunits. This in turn provides a way in which the binding of ligand to one subunit can influence the binding behavior at the other subunits. Such cooperative behavior, discussed in greater depth in Chapter 15, is the underlying mechanism for regulation of many biological processes.

Faster-Acting Insulin: Genetic Engineering Solves a Quaternary Structure Problem

Insulin is a peptide hormone, secreted by the pancreas, that regulates glucose metabolism in the body. Insufficient production of insulin or failure of insulin to stimulate target sites in liver, muscle, and adipose tissue leads to the serious metabolic disorder known as *diabetes mellitus*. Diabetes afflicts millions of people worldwide. Diabetic individuals typically exhibit high levels of glucose in the blood, but insulin injection therapy allows diabetic individuals to maintain normal levels of blood glucose.

Insulin is composed of two peptide chains covalently linked by disulfide bonds (Figures 5.17 and 6.35). This "monomer" of insulin is the active form that binds to receptors in target cells. However, in solution, insulin spontaneously forms dimers, which themselves aggregate to form hexamers. The surface of the insulin molecule that self-associates to form hexamers is also the surface that binds to insulin receptors in target cells. Thus, hexamers of insulin are inactive.

Insulin released from the pancreas is monomeric and acts rapidly at target tissues. However, when insulin is administered (by injection) to a diabetic patient, the insulin hexamers dissociate slowly, and the patient's blood glucose levels typically drop slowly (over a period of several hours).

In 1988, G. Dodson showed that insulin could be genetically engineered to prefer the monomeric (active) state. Dodson and his colleagues used recombinant DNA technology (discussed in Chapter 13) to produce insulin with an aspartate residue replacing a proline at the contact interface between adjacent subunits. The negative charge on the Asp side chain creates electrostatic repulsion between subunits and increases the dissociation constant for the hexamer $\rightleftarrows$ monomer equilibrium. Injection of this mutant insulin into test animals produced more rapid decreases in blood glucose than did ordinary insulin. The Danish pharmaceutical company Novo is conducting clinical trials of the mutant insulin, which may eventually replace ordinary insulin in treatment of diabetes.

PROBLEMS

1. The central rod domain of a keratin protein is approximately 312 residues in length. What is the length (in Å) of the keratin rod domain? If this same peptide segment were a true α-helix, how long would it be? If the same segment were a β-sheet, what would its length be?

2. A teenager can grow 4 inches in a year during a "growth spurt." Assuming that the increase in height is due to vertical growth of collagen fibers (in bone), calculate the number of collagen helix turns synthesized per minute.

3. Discuss the potential contributions to hydrophobic and van der Waals interactions and ionic and hydrogen bonds for the side chains of Asp, Leu, Tyr, and His in a protein.

4. Figure 6.40 shows that Pro is the amino acid least commonly found in α-helices but most commonly found in β-turns. Discuss the reasons for this behavior.

5. For flavodoxin in Figure 6.32, identify the right-handed cross-overs and the left-handed cross-overs in the parallel β-sheet.

6. Choose any three regions in the Ramachandran plot and discuss the likelihood of observing that combination of ϕ and ψ in a peptide or protein. Defend your answer using suitable molecular models of a peptide.

7. A new protein of unknown structure has been purified. Gel filtration chromatography reveals that the native protein has a molecular weight of 240,000. Chromatography in the presence of 6 M guanidine hydrochloride yields only a peak for a protein of M_r 60,000. Chromatography in the presence of 6 M guanidine hydrochloride and 10 mM β-mercaptoethanol yields peaks for proteins of M_r 34,000 and 26,000. Explain what can be determined about the structure of this protein from these data.

8. Two polypeptides, A and B, have similar tertiary structures, but A normally exists as a monomer, whereas B exists as a tetramer, B_4. What differences might be expected in the amino acid composition of A versus B?

9. The hemagglutinin protein in influenza virus contains a remarkably long α-helix, with 53 residues.
a. How long is this α-helix (in nm)?
b. How many turns does this helix have?
c. Each residue in an α-helix is involved in two H bonds. How many H bonds are present in this helix?

FURTHER READING

Abeles, R., Frey, P., and Jencks, W., 1992. *Biochemistry*. Boston: Jones and Bartlett.

Aurora, R., Creamer, T., Srinivasan, R., and Rose, G. D., 1997. Local interactions in protein folding: Lessons from the α-helix. *The Journal of Biological Chemistry* **272:**1413–1416.

Branden, C., and Tooze, J., 1991. *Introduction to Protein Structure*. New York: Garland Publishing, Inc.

Cantor, C. R., and Schimmel, P. R., 1980. *Biophysical Chemistry,* Part I: The Conformation of Biological Macromolecules. New York: W. H. Freeman and Co.

Chothia, C., 1984. Principles that determine the structure of proteins. *Annual Review of Biochemistry* **53:**537–572.

Creighton, T. E., 1983. *Proteins: Structure and Molecular Properties*. New York: W. H. Freeman and Co.

Creighton, T. E., 1997. How important is the molten globule for correct protein folding? *Trends in Biochemical Sciences* **22:**6–11.

Dickerson, R. E., and Geis, I., 1969. *The Structure and Action of Proteins.* New York: Harper and Row.

Dill, K. A., and Chan, H. S., 1997. From Levinthal to pathways to funnels. *Nature Structural Biology* **4:**10–19.

Englander, S. W., and Mayne, L., 1992. Protein folding studied using hydrogen exchange labeling and two-dimensional NMR. *Annual Review of Biophysics and Biomolecular Structure* **21:**243–265.

Hardie, D. G., and Coggins, J. R., eds., 1986. *Multidomain Proteins: Structure and Evolution.* New York: Elsevier.

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

Judson, H. F., 1979. *The Eighth Day of Creation.* New York: Simon and Schuster.

Klotz, I. M., 1996. Equilibrium constants and free energies in unfolding of proteins in urea solutions. *Proceedings of the National Academy of Sciences* **93:**14411–14415.

Klotz, I. M., Langerman, N. R., and Darnell, D. W., 1970. Quaternary structure of proteins. *Annual Review of Biochemistry* **39:**25–62.

Lupas, A., 1996. Coiled coils: New structures and new functions. *Trends in Biochemical Sciences* **21:**375–382.

Petsko, G. A., and Ringe, D., 1984. Fluctuations in protein structure from X-ray diffraction. *Annual Review of Biophysics and Bioengineering* **13:**331–371.

Privalov, P. L., and Makhatadze, G. I., 1993. Contributions of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *Journal of Molecular Biology* **232:**660–679.

Raschke, T. M., and Marqusee, S., 1997. The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature Structural Biology* **4:**298–304.

Richardson, J. S., 1981. The anatomy and taxonomy of protein structure. *Advances in Protein Chemistry* **34:**167–339.

Richardson, J. S., and Richardson, D. C., 1988. Amino acid preferences for specific locations at the ends of α-helices. *Science* **240:**1648–1652.

Rossman, M. G., and Argos, P., 1981. Protein folding. *Annual Review of Biochemistry* **50:**497–532.

Salemme, F. R., 1983. Stuctural properties of protein β-sheets. *Progress in Biophysics and Molecular Biology* **42:**95–133.

Schulze, A. J., Huber, R., Bode, W., and Engh, R. A., 1994. Structural aspects of serpin inhibition. *FEBS Letters* **344:**117–124.

Sifers, R. M., 1995. Defective protein folding as a cause of disease. *Nature Structural Biology* **2:**355–367.

Srinivasan, R., and Rose, G. D., 1995. LINUS: A hierarchic procedure to predict the fold of a protein. *Proteins: Structure, Function and Genetics* **22:**81–99.

Stein, P. E., and Carrell, R. W., 1995. What do dysfunctional serpins tell us about molecular mobility and disease? *Nature Structural Biology* **2:**96–113.

Thomas, P. J., Qu, B.-H., and Pedersen, P. L., 1995. Defective protein folding as a basis of human disease. *Trends in Biochemical Sciences* **20:**456–459.

Torchia, D. A., 1984. Solid state NMR studies of protein internal dynamics. *Annual Review of Biophysics and Bioengineering* **13:**125–144.

Wagner, G., Hyberts, S., and Havel, T., 1992. NMR structure determination in solution: A critique and comparison with X-ray crystallography. *Annual Review of Biophysics and Biomolecular Structure* **21:**167–242.

Xiong, H., Buckwalter, B., Shieh, H.-M., and Hecht, M. H., 1995. Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides. *Proceedings of the National Academy of Sciences* **92:**6349–6353.