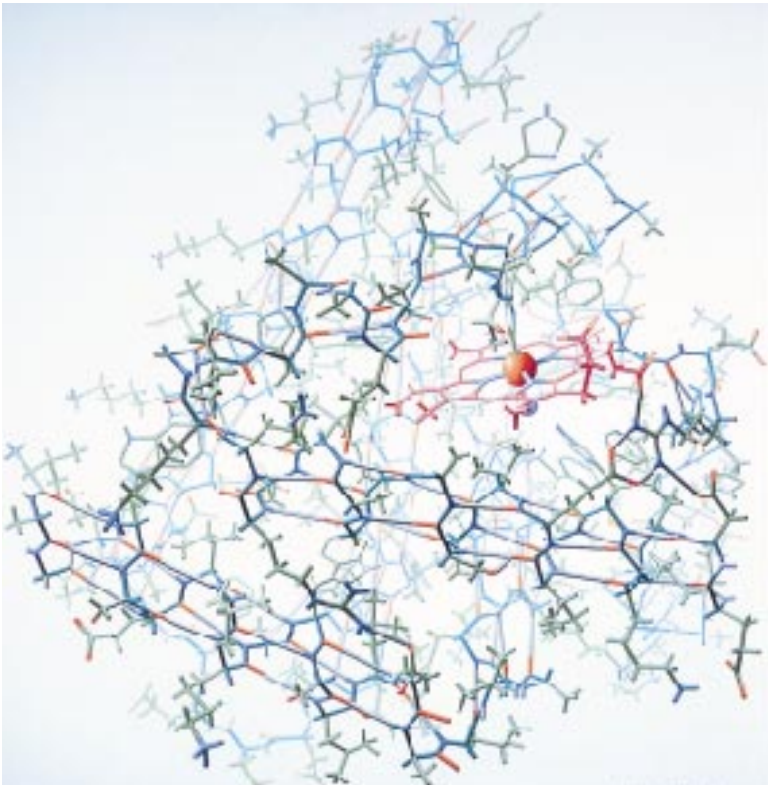
CHAPTER 6



The atomic structure of myoglobin, an oxygen binding protein, is drawn here as a stick model. The overall conformation of a protein such as myoglobin is a function of its amino acid sequence. How do noncovalent forces act on a polypeptide chain to stabilize its unique three-dimensional arrangement of atoms? [Figure copyrighted © by Irving Geis.]

PROTEINS: THREE-DIMENSIONAL STRUCTURE

1. SECONDARY STRUCTURE
   - A. The Peptide Group
   - B. Regular Secondary Structure: The α Helix and the β Sheet
   - C. Fibrous Proteins
   - D. Nonrepetitive Protein Structure
2. TERTIARY STRUCTURE
   - A. Determining Protein Structure
   - B. Motifs (Supersecondary Structures) and Domains
   - C. Protein Families
3. QUATERNARY STRUCTURE AND SYMMETRY
4. PROTEIN FOLDING AND STABILITY
   - A. Forces That Stabilize Protein Structure
   - B. Protein Denaturation and Renaturation
   - C. Protein Folding Pathways
   - D. Protein Dynamics

For many years, it was thought that proteins were colloids of random structure and that the enzymatic activities of certain crystallized proteins were due to unknown entities associated with an inert protein carrier. In 1934, J.D. Bernal and Dorothy Crowfoot Hodgkin showed that a crystal of the protein **pepsin** yielded a discrete diffraction pattern when placed in an X-ray beam. This result provided the first evidence that pepsin was not a random colloid but an ordered array of atoms organized into a large yet uniquely structured molecule.

Even relatively small proteins contain thousands of atoms, almost all of which occupy definite positions in space. The first X-ray structure of a protein, that of sperm whale myoglobin, was reported in 1958 by John Kendrew and co-workers. At the time—only 5 years after James Watson and Francis Crick had elucidated the simple and elegant structure of DNA (Section 3-2B)—protein chemists were chagrined by the complexity and apparent lack of regularity in the structure of myoglobin. In retrospect, such irregularity seems essential for proteins to fulfill their diverse biological roles. However, comparisons of the ~7000 protein structures now known have revealed that proteins actually exhibit a remarkable degree of structural regularity.

As we saw in Section 5-1, the primary structure of a protein is its linear sequence of amino acids. In discussing protein structure, three further levels of structural complexity are customarily invoked:

- **Secondary structure** is the local spatial arrangement of a polypeptide's backbone atoms without regard to the conformations of its side chains.
- **Tertiary structure** refers to the three-dimensional structure of an entire polypeptide.
- Many proteins are composed of two or more polypeptide chains, loosely referred to as subunits. A protein's **quaternary structure** refers to the spatial arrangement of its subunits.

The four levels of protein structure are summarized in Fig. 6-1.

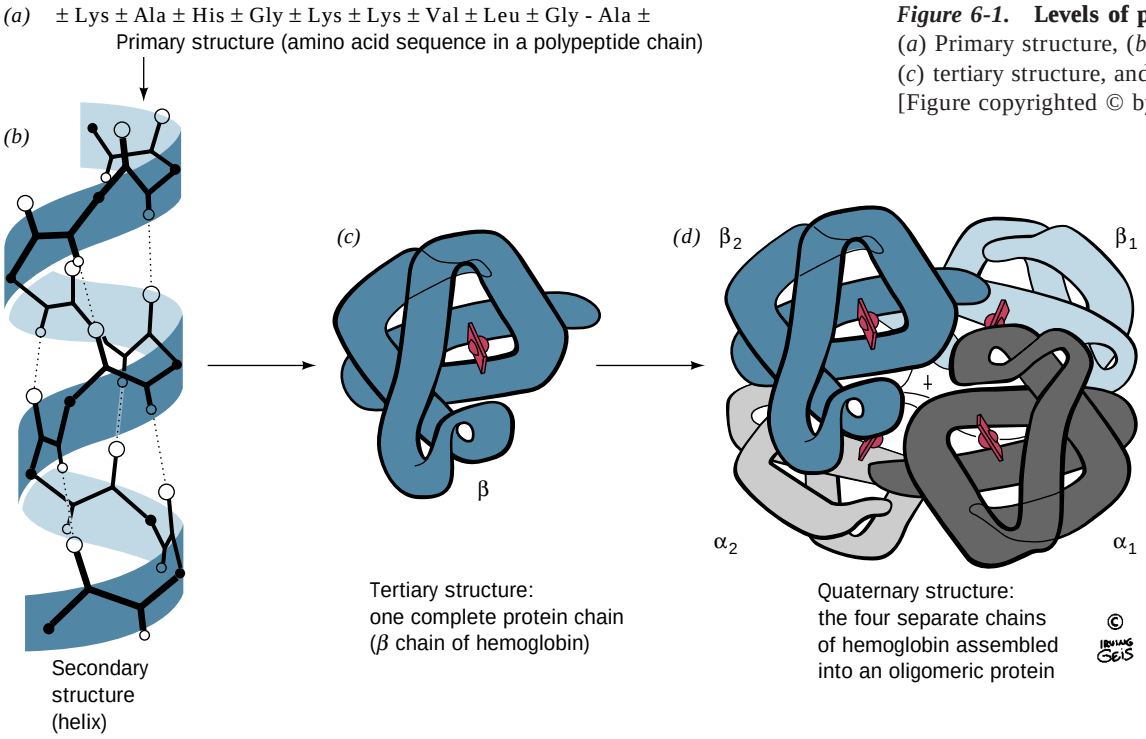


Figure 6-1. **Levels of protein structure.** (*a*) Primary structure, (*b*) secondary structure, (*c*) tertiary structure, and (*d*) quaternary structure. [Figure copyrighted © by Irving Geis.]

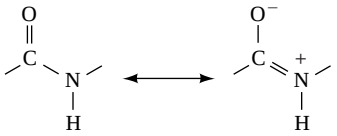
In this chapter, we explore secondary through quaternary structure, including examples of proteins that illustrate each of these levels. We also introduce methods for determining three-dimensional molecular structure and discuss the forces that stabilize folded proteins.

1. SECONDARY STRUCTURE

Protein secondary structure includes the regular polypeptide folding patterns such as helices, sheets, and turns. However, before we discuss these basic structural elements, we must consider the geometric properties of peptide groups, which underlie all higher order structures.

A. The Peptide Group

In the 1930s and 1940s, Linus Pauling and Robert Corey determined the X-ray structures of several amino acids and dipeptides in an effort to elucidate the conformational constraints on a polypeptide chain. These studies indicated that *the peptide group has a rigid, planar structure as a consequence of resonance interactions that give the peptide bond ~40% double-bond character:*



This explanation is supported by the observations that a peptide group's C—N bond is 0.13 Å shorter than its N—C_α single bond and that its C═O bond is 0.02 Å longer than that of aldehydes and ketones. The planar conformation maximizes π-bonding overlap, which accounts for the peptide group's rigidity.

Peptide groups, with few exceptions, assume the **trans conformation,** in which successive C_α atoms are on opposite sides of the peptide bond joining them (Fig. 6-2). The **cis conformation,** in which successive C_α atoms are on the same side of the peptide bond, is ~8 kJ · mol^{-1} less stable than the trans conformation because of steric interference between neighboring side chains. However, this steric interference is reduced in peptide bonds to Pro residues, so *~10% of the Pro residues in proteins follow a cis peptide bond.*

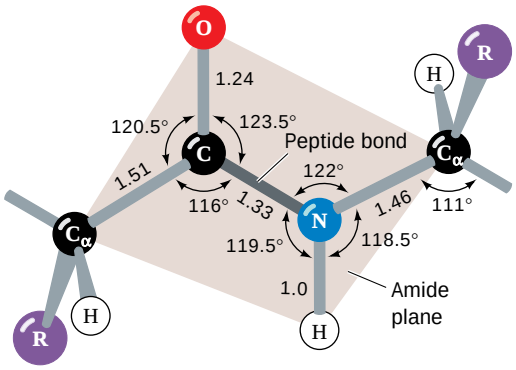


Figure 6-2. **The trans peptide group.** The bond lengths (in angstroms) and angles (in degrees) are derived from X-ray crystal structures. [After Marsh, R.E. and Donohue, J., *Adv. Protein Chem.* 22, 249 (1967).] ● See Kinemage Exercise 3-1.

Torsion Angles between Peptide Groups Describe Polypeptide Chain Conformations

The **backbone** or **main chain** of a protein refers to the atoms that participate in peptide bonds, ignoring the side chains of the amino acid

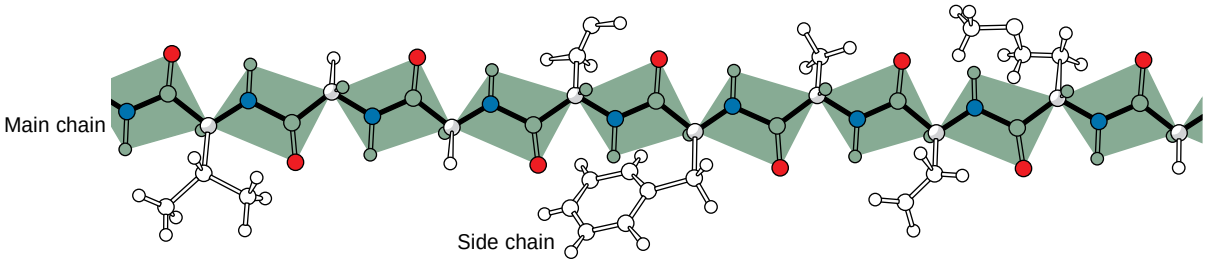


Figure 6-3. **Extended conformation of a polypeptide.** The backbone is shown as a series of planar peptide groups. [Figure copyrighted © by Irving Geis.]

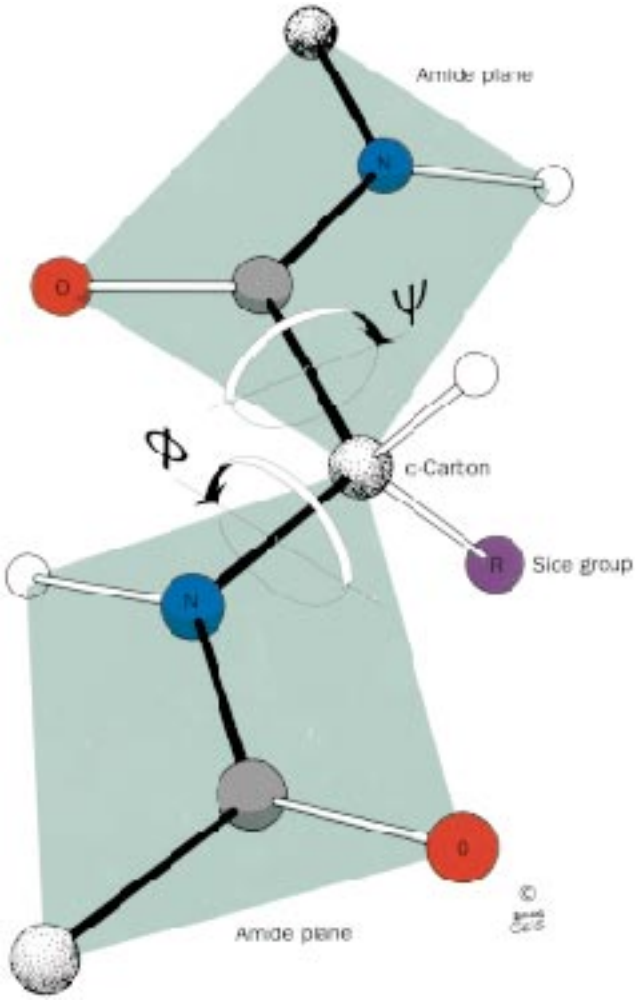


Figure 6-4. **Torsion angles of the polypeptide backbone.** Two planar peptide groups are shown. The only reasonably free movements are rotations around the C_α—N bond (measured as ϕ) and the C_α—C bond (measured as ψ). By convention, both ϕ and ψ are 180° in the conformation shown and increase, as indicated, in the clockwise direction when viewed from C_α. [Figure copyrighted © by Irving Geis.]
See Kinemage Exercise 3-1.

residues. The backbone can be drawn as a linked sequence of rigid planar peptide groups (Fig. 6-3). *The conformation of the backbone can therefore be described by the* ***torsion angles*** (also called **dihedral angles** or rotation angles) *around the C_α—N bond (ϕ) and the C_α—C bond (ψ) of each residue* (Fig. 6-4). These angles, ϕ and ψ, are both defined as 180° when the polypeptide chain is in its fully extended conformation and increase clockwise when viewed from C_α.

The conformational freedom and therefore the torsion angles of a polypeptide backbone are sterically constrained. Rotation around the C_α—N and C_α—C bonds to form certain combinations of ϕ and ψ angles may cause the amide hydrogen, the carbonyl oxygen, or the substituents of C_α of adjacent residues to collide (e.g., Fig. 6-5). Certain conformations of longer polypeptides can similarly produce collisions between residues that are far apart in sequence.

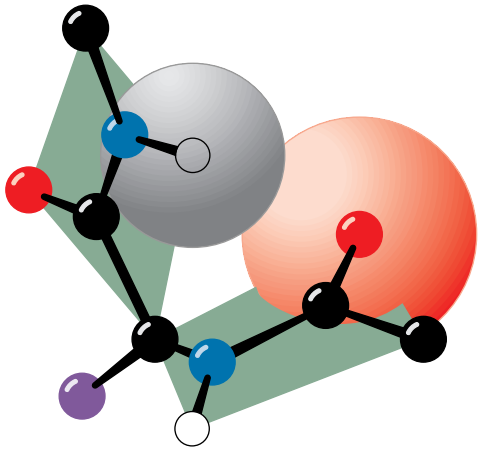


Figure 6-5. **Steric interference between adjacent peptide groups.** Rotation can result in a conformation in which the amide hydrogen of one residue and the carbonyl oxygen of the next are closer than their van der Waals distance. [Figure copyrighted © by Irving Geis.]

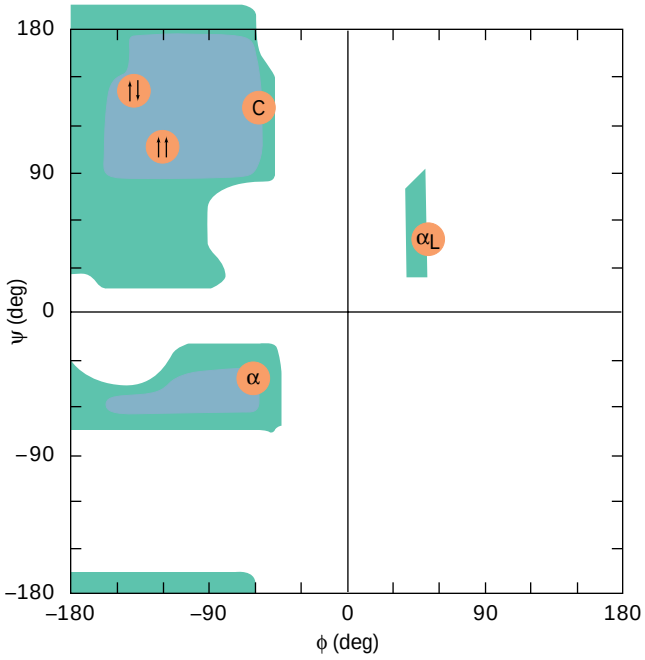


Figure 6-6. **A Ramachandran diagram.** The green-shaded regions indicate the sterically allowed ϕ and ψ angles for all residues except Gly and Pro. The orange circles represent conformational angles of several secondary structures: α, right-handed α helix; ↑↑, parallel β sheet; ↑↓, antiparallel β sheet; C, collagen helix; α_L, left-handed α helix.

The Ramachandran Diagram Indicates Allowed Conformations of Polypeptides

The sterically allowed values of ϕ and ψ can be calculated. Sterically forbidden conformations, such as the one shown in Fig. 6-5, have ϕ and ψ values that would bring atoms closer than the corresponding van der Waals distance (the distance of closest contact between nonbonded atoms). Such information is summarized in a **Ramachandran diagram** (Fig. 6-6), which is named after its inventor, G. N. Ramachandran.

Most areas of the Ramachandran diagram (most combinations of ϕ and ψ) represent forbidden conformations of a polypeptide chain. Only three small regions of the diagram are physically accessible to most residues. The observed ϕ and ψ values of accurately determined structures nearly always fall within these allowed regions of the Ramachandran plot. There are, however, some notable exceptions:

1. The cyclic side chain of Pro limits its range of ϕ values to angles of around −60°, making it, not surprisingly, the most conformationally restricted amino acid residue.
2. Gly, the only residue without a C_β atom, is much less sterically hindered than the other amino acid residues. Hence, its permissible range of ϕ and ψ covers a larger area of the Ramachandran diagram. At Gly residues, polypeptide chains often assume conformations that are forbidden to other residues.

B. Regular Secondary Structure: The α Helix and the β Sheet

A few elements of protein secondary structure are so widespread that they are immediately recognizable in proteins with widely differing amino acid sequences. Both the **α helix** and the **β sheet** are such elements; they are

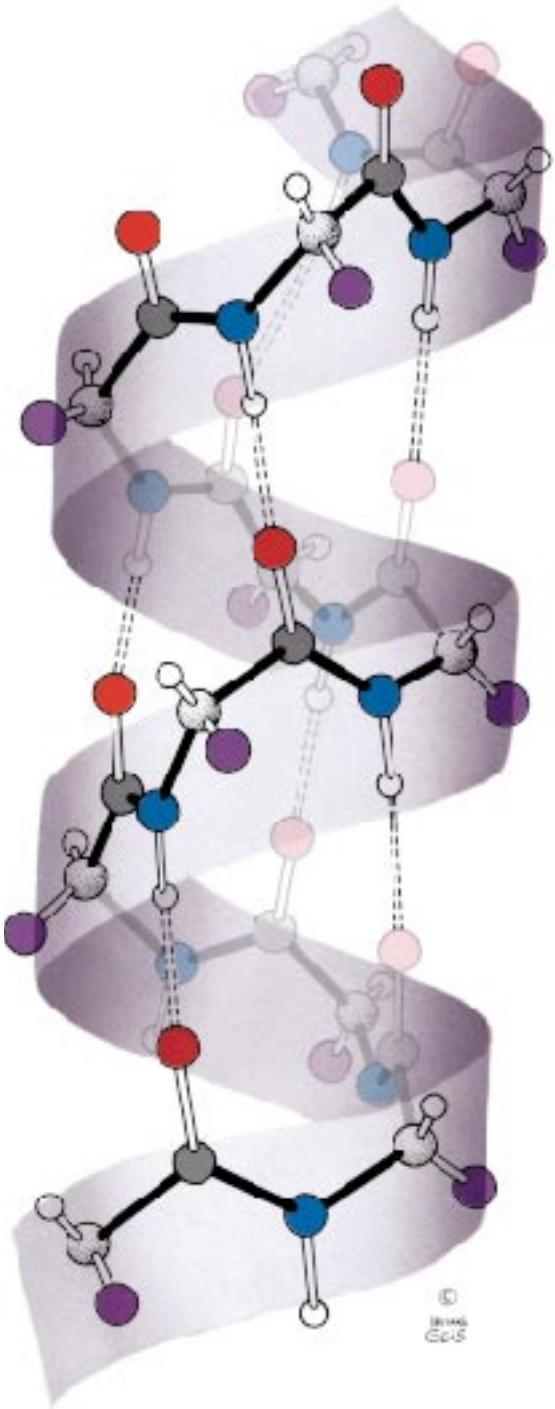


Figure 6-7. *Key to Structure.* **The α helix.** This right-handed helical conformation has 3.6 residues per turn. Dashed lines indicate hydrogen bonds between C═O groups and N—H groups that are four residues farther along the polypeptide chain. [Figure copyrighted © by Irving Geis.] See Kinemage Exercise 3-2.

called **regular secondary structures** because they are composed of sequences of residues with repeating ϕ and ψ values.

The α Helix

Only one polypeptide helix has both a favorable hydrogen bonding pattern and ϕ and ψ values that fall within the fully allowed regions of the Ramachandran diagram: the α helix. Its discovery by Linus Pauling in 1951, through model building, ranks as one of the landmarks of structural biochemistry.

The α helix (Fig. 6-7) is right-handed; that is, it turns in the direction that the fingers of a right hand curl when its thumb points in the direction that the helix rises (see Fig. 3-10). The α helix has 3.6 residues per turn and a **pitch** (the distance the helix rises along its axis per turn) of 5.4 Å. The α helices of proteins have an average length of ~12 residues, which corresponds to over three helical turns, and a length of ~18 Å.

In the α helix, the backbone hydrogen bonds are arranged such that the peptide C═O bond of the nth residue points along the helix axis toward the peptide N—H group of the (n + 4)th residue. This results in a strong hydrogen bond that has the nearly optimum N···O distance of 2.8 Å. Amino acid side chains project outward and downward from the helix (Fig. 6-8), thereby avoiding steric interference with the polypeptide backbone and with each other. The core of the helix is tightly packed; that is, its atoms are in van der Waals contact.

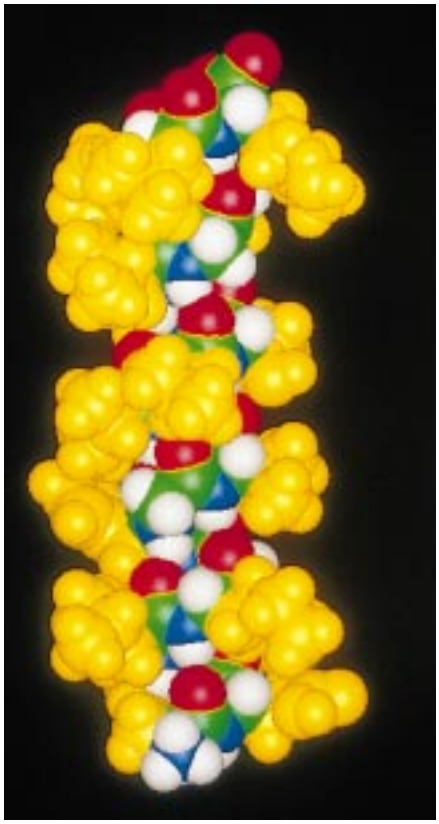


Figure 6-8. **Space-filling model of an α helix.** The backbone atoms are colored with carbon atoms green, nitrogen atoms blue, oxygen atoms red, and hydrogen atoms white. The side chains (*yellow*) project away from the helix. This α helix is a segment of sperm whale myoglobin.

β Sheets

In 1951, the same year Pauling proposed the α helix, Pauling and Corey postulated the existence of a different polypeptide secondary structure, the β sheet. Like the α helix, the β sheet uses the full hydrogen-bonding capacity of the polypeptide backbone. *In β sheets, however, hydrogen bonding occurs between neighboring polypeptide chains rather than within one* as in an α helix.

Sheets come in two varieties:

1. The **antiparallel β sheet,** in which neighboring hydrogen-bonded polypeptide chains run in opposite directions (Fig. 6-9*a*).
2. The **parallel β sheet,** in which the hydrogen-bonded chains extend in the same direction (Fig. 6-9*b*).

The conformations in which these β structures are optimally hydrogen bonded vary somewhat from that of the fully extended polypeptide shown in Fig. 6-3. They therefore have a rippled or pleated edge-on appearance (Fig. 6-10) and for that reason are sometimes called "pleated sheets." Successive side chains of a polypeptide chain in a β sheet extend to opposite sides of the sheet with a two-residue repeat distance of 7.0 Å.

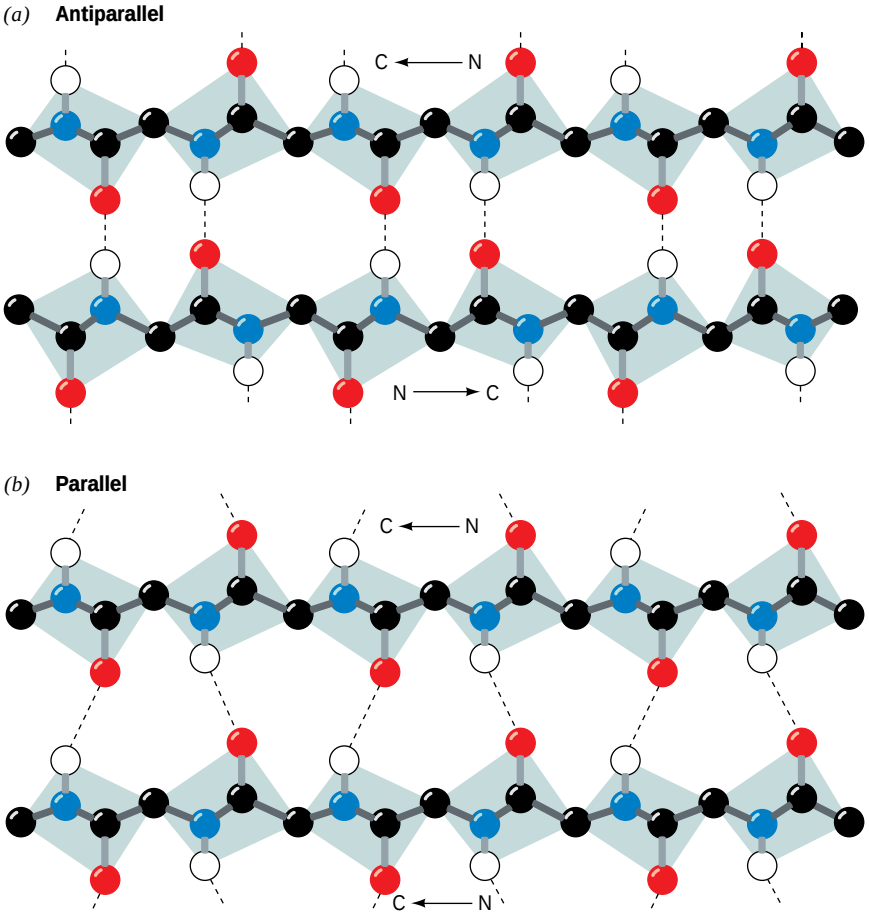


Figure 6-9. *Key to Structure.* **β Sheets.** Dashed lines indicate hydrogen bonds between polypeptide strands. Side chains are omitted for clarity. (*a*) An antiparallel β sheet. (*b*) A parallel β sheet. [Figure copyrighted © by Irving Geis.] See Kinemage Exercise 2-3.

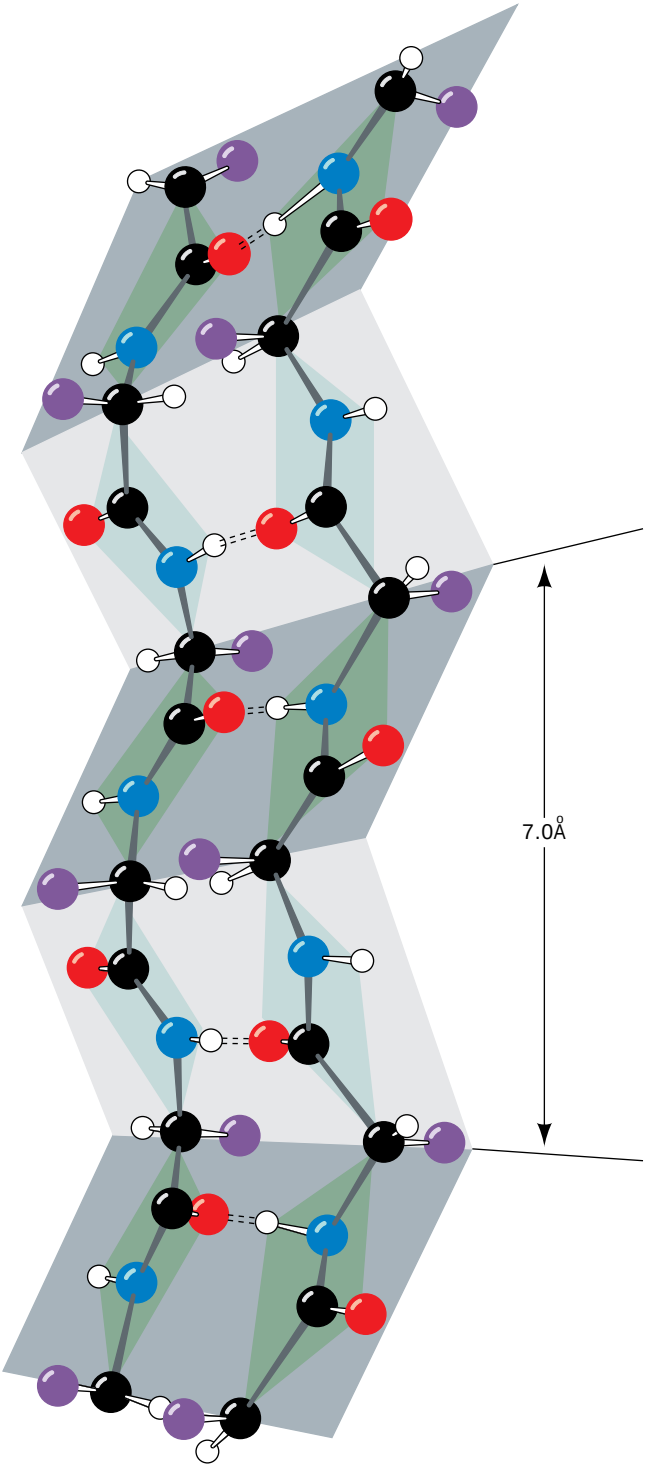


Figure 6-10. **Pleated appearance of a β sheet.** Dashed lines indicate hydrogen bonds. The R groups (*purple*) on each polypeptide chain alternately extend to opposite sides of the sheet and are in register on adjacent chains. [Figure copyrighted © by Irving Geis.]

β Sheets in proteins contain 2 to >12 polypeptide strands, with an average of 6 strands. Each strand may contain up to 15 residues, the average being 6 residues. A six-stranded antiparallel β sheet is shown in Fig. 6-11.

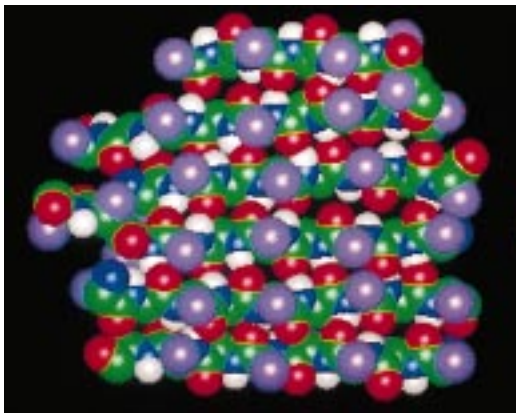


Figure 6-11. **Space-filling model of a β sheet.** The backbone atoms are colored with carbon atoms green, nitrogen atoms blue, oxygen atoms red, and hydrogen atoms white. The R groups are represented by large purple spheres. This six-stranded β sheet is from the jack bean protein concanavalin A.

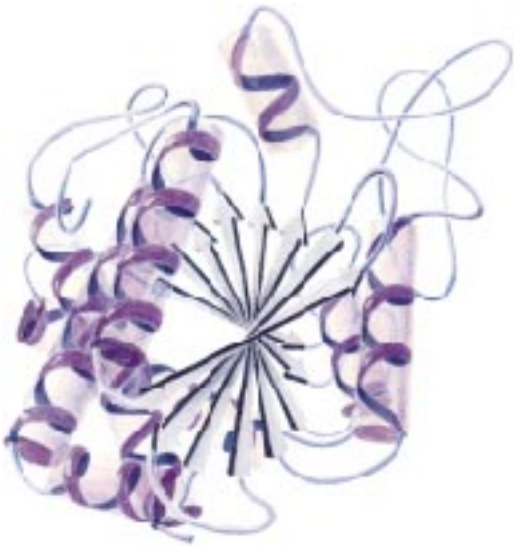


Figure 6-12. **Diagram of a β sheet in bovine carboxypeptidase A.** The polypeptide backbone is represented by a ribbon with α helices drawn as coils and strands of the β sheet drawn as arrows pointing toward the C-terminus. Side chains are not shown. The eight-stranded β sheet forms a saddle-shaped curved surface with a right-handed twist. [After a drawing by Jane Richardson, Duke University.]

Parallel β sheets containing fewer than five strands are rare. This observation suggests that parallel β sheets are less stable than antiparallel β sheets, possibly because the hydrogen bonds of parallel sheets are distorted compared to those of the antiparallel sheets (Fig. 6-9). β Sheets containing mixtures of parallel and antiparallel strands frequently occur.

β Sheets almost invariably exhibit a pronounced right-handed twist when viewed along their polypeptide strands (Fig. 6-12). Conformational energy calculations indicate that the twist is a consequence of interactions between chiral L-amino acid residues in the extended polypeptide chains. The twist actually distorts and weakens the β sheet's interchain hydrogen bonds. The geometry of a particular β sheet is thus a compromise between optimizing the conformational energies of its polypeptide chains and preserving its hydrogen bonding.

The **topology** (connectivity) of the polypeptide strands in a β sheet can be quite complex. The connection between two antiparallel strands may be just a small loop (Fig. 6-13*a*), but the link between tandem parallel strands must be a crossover connection that is out of the plane of the β sheet (Fig. 6-13*b*). The connecting link in either case can be extensive, often containing helices (e.g., Fig. 6-12).

C. Fibrous Proteins

Proteins have historically been classified as either **fibrous** or **globular,** depending on their overall morphology. This dichotomy predates methods for determining protein structure on an atomic scale and does not do justice to proteins that contain both stiff, elongated, fibrous regions as well as more compact, highly folded, globular regions. Nevertheless, the division helps emphasize the properties of fibrous proteins, which often have a protective, connective, or supportive role in living organisms. The three well-characterized fibrous proteins we discuss here—keratin, silk fibroin, and collagen—are highly elongated molecules whose shapes are dominated by a single type of secondary structure. They are therefore useful examples of these structural elements.

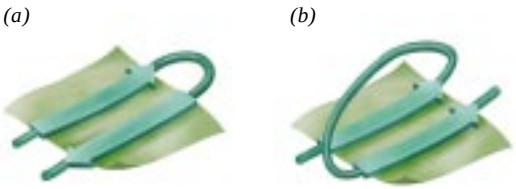


Figure 6-13. **Connections between adjacent strands in β sheets.** (*a*) Antiparallel strands may be connected by a small loop. (*b*) Parallel strands require a more extensive cross-over connection. [After Richardson, J.S., *Adv. Protein Chem.* 34, 196 (1981).]

α Keratin—A Coiled Coil

Keratin is a mechanically durable and chemically unreactive protein that occurs in all higher vertebrates. It is the principal component of their horny outer epidermal layer and its related appendages such as hair, horn, nails, and feathers. Keratins have been classified as either α keratins, which occur in mammals, or β keratins, which occur in birds and reptiles. Mammals each have about 30 keratin variants that are expressed in a tissue-specific manner.

The X-ray diffraction pattern of α keratin resembles that expected for an α helix (hence the name α keratin). However, α keratin exhibits a 5.1-Å spacing rather than the 5.4-Å distance corresponding to the pitch of the

α helix. This discrepancy is the result of *two α keratin polypeptides, each of which forms an α helix, twisting around each other to form a left-handed coil.* The normal 5.4-Å repeat distance of each α helix in the pair is thereby tilted relative to the axis of this assembly, yielding the observed 5.1-Å spacing. The assembly is said to have a **coiled coil** structure because each α helix itself follows a helical path.

The conformation of α keratin's coiled coil is a consequence of its primary structure: The central ~310-residue segment of each polypeptide chain has a 7-residue pseudorepeat, *a-b-c-d-e-f-g*, with nonpolar residues predominating at positions *a* and *d*. Since an α helix has 3.6 residues per turn, α keratin's *a* and *d* residues line up along one side of each α helix (Fig. 6-14*a*). The hydrophobic strip along one helix associates with the hydrophobic strip on another helix. Because the 3.5-residue repeat in α keratin is slightly smaller than the 3.6 residues per turn of a standard α helix, the two keratin helices are inclined about 18° relative to one another, resulting in the coiled coil arrangement. This conformation allows the contacting side chains to interdigitate (Fig. 6-14*b*).

The higher order structure of α keratin is not well understood. The N- and C-terminal domains of each polypeptide facilitate the assembly of coiled coils (dimers) into protofilaments, two of which constitute a protofibril (Fig. 6-15). Four protofibrils constitute a microfibril, which associates with other microfibrils to form a macrofibril. A single mammalian hair consists of layers of dead cells, each of which is packed with parallel macrofibrils.

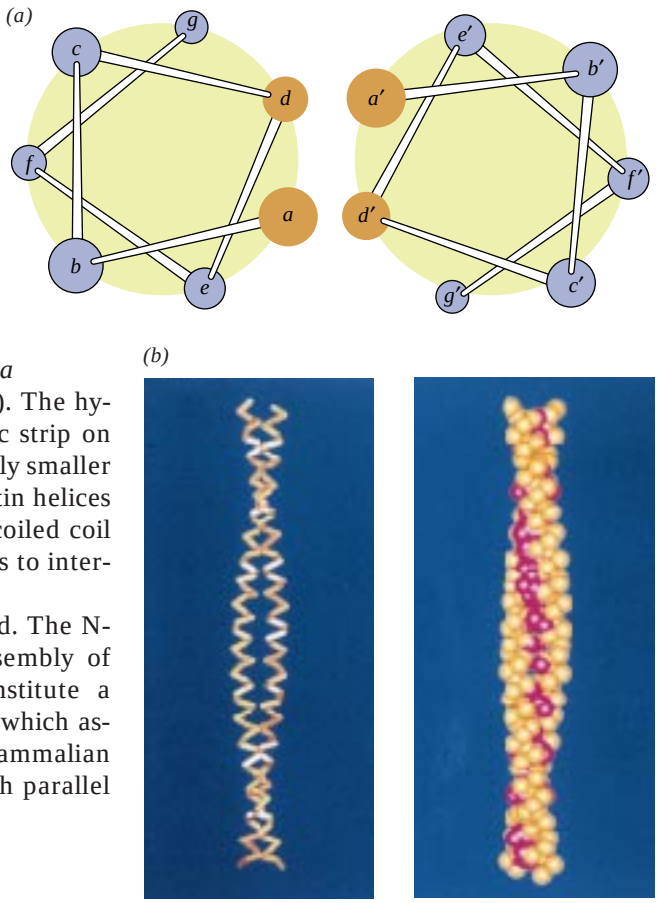


Figure 6-14. **The coiled coil of α keratin.** (*a*) View down the coil axis showing the alignment of nonpolar residues along one side of each α helix. The helices have the pseudorepeating sequence *a-b-c-d-e-f-g* in which residues *a* and *d* are predominately nonpolar. [After McLachlan, A.D. and Stewart, M., *J. Mol. Biol.* 98, 295 (1975).] (*b*) Side view of the polypeptide backbone in skeletal (*left*) and space-filling (*right*) forms. Note that the contacting side chains (red spheres in the space-filling model) interlock. [Courtesy of Carolyn Cohen, Brandeis University.] See Kinemage Exercises 4-1 and 4-2.

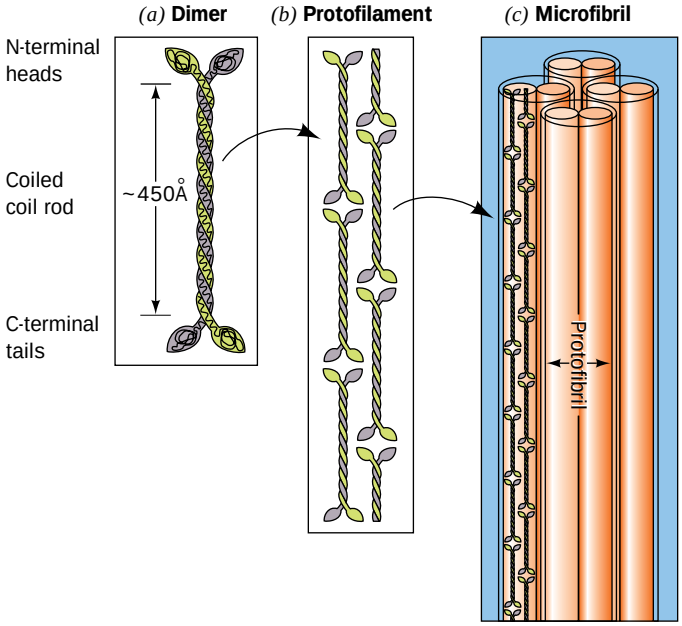


Figure 6-15. **Higher order α keratin structure.** (*a*) Two keratin polypeptides form a dimeric coiled coil. (*b*) Protofilaments are formed from two staggered rows of head-to-tail associated coiled coils. (*c*) Protofilaments dimerize to form a protofibril, four of which form a microfibril. The structures of the latter assemblies are poorly characterized.

α Keratin is rich in Cys residues, which form disulfide bonds that cross-link adjacent polypeptide chains. The α keratins are classified as "hard" or "soft" according to whether they have a high or low sulfur content. Hard keratins, such as those of hair, horn, and nail, are less pliable than soft keratins, such as those of skin and callus, because the disulfide bonds resist deformation. The disulfide bonds can be reductively cleaved with mercaptans (Section 5-3A). Hair so treated can be curled and set in a "permanent wave" by applying an oxidizing agent that reestablishes the disulfide bonds in the new "curled" conformation. Conversely, curly hair can be straightened by the same process.

The springiness of hair and wool fibers is a consequence of the coiled coil's tendency to recover its original conformation after being untwisted by stretching. If some of its disulfide bonds have been cleaved, however, an α keratin fiber can be stretched to over twice its original length. At this point, the polypeptide chains assume a β sheet conformation. β Keratin, such as that in feathers, exhibits a β-like pattern in its native state.

Silk Fibroin—A β Sheet

Insects and arachnids (spiders) produce various silks to fabricate structures such as cocoons, webs, nests, and egg stalks. **Silk fibroin,** the fibrous protein from the cultivated larvae (silkworms) of the moth *Bombyx mori*, consists of antiparallel β sheets whose chains extend parallel to the fiber axis. Sequence studies have shown that long stretches of silk fibroin contain a six-residue repeat:

$$(\text{-Gly-Ser-Gly-Ala-Gly-Ala-})_n$$

Since the side chains from successive residues of a β strand extend to opposite sides of the β sheet (Fig. 6-10), silk's Gly side chains project from one surface of a β sheet and its Ser and Ala side chains project from the opposite surface. *The β sheets stack to form a microcrystalline array in which layers of contacting Gly side chains from neighboring sheets alternate with layers of contacting Ser and Ala side chains* (Fig. 6-16).

The β sheet structure of silk accounts for its mechanical properties. Silk, which is among the strongest of fibers, is only slightly extensible because appreciable stretching would require breaking the covalent bonds of its nearly fully extended polypeptide chains. Yet silk is flexible because neighboring β sheets associate only through relatively weak van der Waals forces.

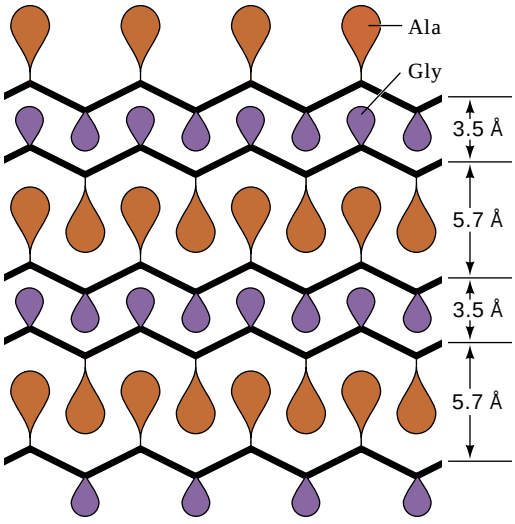
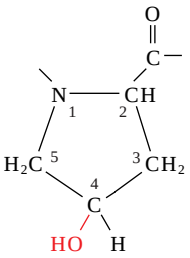


Figure 6-16. **Schematic side view of silk fibroin β sheets.** Alternating Gly and Ala (or Ser) residues extend to opposite sides of each strand so that the Gly side chains (*purple*) from one sheet nestle efficiently between those of the neighboring sheet and likewise for the Ser and Ala side chains (*brown*). The intersheet spacings consequently have the alternating values of 3.5 and 5.7 Å. [Figure copyrighted © by Irving Geis.]

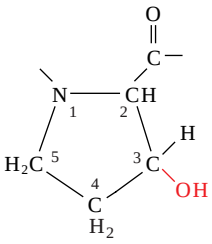
Collagen—A Triple Helix

Collagen, which occurs in all multicellular animals, is the most abundant vertebrate protein. Its strong, insoluble fibers are the major stress-bearing components of connective tissues such as bone, teeth, cartilage, tendon, and the fibrous matrices of skin and blood vessels. A single collagen molecule consists of three polypeptide chains. Mammals have about 30 genetically distinct chains that are assembled into at least 19 collagen varieties found in different tissues in the same individual. One of the most common collagens, called Type I, consists of two α_1(I) chains and one α_2(I) chain. It has a molecular mass of ~285 kD, a width of ~14 Å, and a length of ~3000 Å.

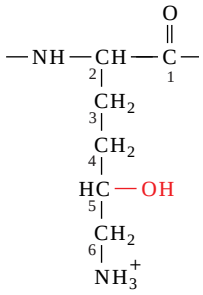
Collagen has a distinctive amino acid composition: Nearly one-third of its residues are Gly; another 15 to 30% of its residues are Pro and **4-hydroxyprolyl** (**Hyp**). **3-Hydroxyprolyl** and **5-hydroxylysyl** (**Hyl**) residues also occur in collagen, but in smaller amounts.



4-Hydroxyprolyl residue (Hyp)

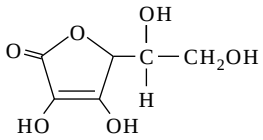


3-Hydroxyprolyl residue



5-Hydroxylysyl residue (Hyl)

These nonstandard residues are formed after the collagen polypeptides are synthesized. For example, Pro residues are converted to Hyp in a reaction catalyzed by **prolyl hydroxylase.** This enzyme requires **ascorbic acid** (**vitamin C**) to maintain its activity.



Ascorbic acid (vitamin C)

The disease **scurvy** results from the dietary deficiency of vitamin C (see Box 6-1).

Box 6-1

BIOCHEMISTRY IN HEALTH AND DISEASE

Collagen Diseases

Some collagen diseases have dietary causes. In scurvy (caused by vitamin C deficiency), Hyp production decreases because prolyl hydroxylase requires vitamin C. Thus, in the absence of vitamin C, newly synthesized collagen cannot form fibers properly, resulting in skin lesions, fragile blood vessels, poor wound healing, and, ultimately, death. Scurvy was common in sailors on long voyages whose diets were devoid of fresh foods. The introduction of limes to the diet of the British navy by the renowned explorer Capt. James Cook alleviated scurvy and led to the nickname "limey" for the British sailor.

The disease lathyrism is caused by regular ingestion of the seeds from the sweet pea *Lathyrus odoratus,* which contain a compound that specifically inactivates lysyl oxidase. The resulting reduced cross-linking of collagen fibers produces serious abnormalities of the bones, joints, and large blood vessels.

Several rare heritable disorders of collagen are known. Mutations of Type I collagen, which constitutes the major structural protein in most human tissues, usually result in osteogenesis imperfecta (brittle bone disease). The severity of this disease varies with the nature and position of the mutation: Even a single amino acid change can have lethal consequences. For example, the central Gly → Ala substitution in the model polypeptide shown in Fig. 6-18*b* locally distorts the already internally crowded collagen helix. This ruptures the hydrogen bond from the backbone N—H of each Ala (normally Gly) to the carbonyl group of the adjacent Pro in a neighboring chain, thereby reducing the stability of the collagen structure.

Mutations may affect the structure of the collagen molecule or how it forms fibrils. These mutations tend to be dominant because they affect either the folding of the triple helix or fibril formation even when normal chains are also involved.

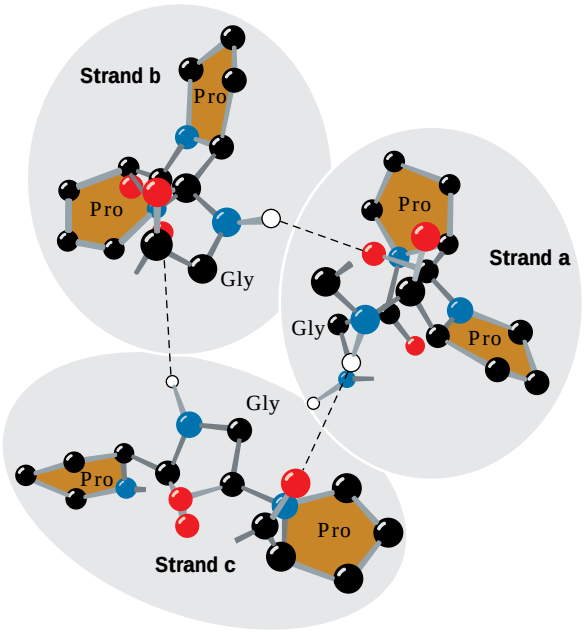
Many collagen disorders are characterized by deficiencies in the amount of a particular collagen type synthesized, or by abnormal activities of collagen-processing enzymes such as lysyl hydroxylase and lysyl oxidase. One group of at least 10 different collagen deficiency diseases, the Ehlers–Danlos syndromes, are all characterized by the hyperextensibility of the joints and skin. The "India-rubber man" of circus fame had an Ehlers–Danlos syndrome.

Figure 6-17. **The collagen triple helix.** Left-handed polypeptide helices are twisted together to form a right-handed superhelical structure. [Figure copyrighted © by Irving Geis.]

The amino acid sequence of a typical collagen polypeptide consists of monotonously repeating triplets of sequence Gly-X-Y over a segment of ~1000 residues, where X is often Pro, and Y is often Hyp. Hyl sometimes appears at the Y position. Collagen's Pro residues prevent it from forming an α helix (Pro residues cannot assume the α-helical backbone conformation and lack the backbone N—H groups that form the intrahelical hydrogen bonds shown in Fig. 6-7). Instead, *the collagen polypeptide assumes a left-handed helical conformation with about three residues per turn. Three parallel chains wind around each other with a gentle, right-handed, ropelike twist to form the triple-helical structure of a collagen molecule* (Fig. 6-17).

Every third residue of each polypeptide chain passes through the center of the triple helix, which is so crowded that only a Gly side chain can fit there. This crowding explains the absolute requirement for a Gly at every third position of a collagen polypeptide chain. The three polypeptide chains are staggered so that Gly, X, and Y residues from each of the three chains occur at the same level along the helix axis. The peptide groups are oriented such that the N—H of each Gly makes a strong hydrogen bond with the carbonyl oxygen of an X residue on a neighboring chain (Fig. 6-18*a*). The bulky and relatively inflexible Pro and Hyp residues confer rigidity on the entire assembly.

This model of the collagen structure has been confirmed by Barbara Brodsky and Helen Berman, who determined the X-ray crystal structure of the collagenlike polypeptide (Pro-Hyp-Gly)$_4$-(Pro-Hyp-Ala)-(Pro-Hyp-Gly)$_5$. Three of these polypeptides associate to form a triple-helical structure that closely resembles the above model (Fig. 6-18*b*). The X-ray structure further reveals that the 87-Å-long cylindrical molecule is surrounded by a sheath of ordered water molecules that apparently stabilizes the collagen structure. These water mol-



(a)

(b)

Figure 6-18. **Molecular interactions in collagen.** (*a*) Hydrogen bonding in the collagen triple helix. This view down the helix axis shows one Gly and two Pro residues (X and Y) in each chain. The residues are staggered so that one Gly, X, and Y occur at every level along the axis. The dashed lines represent hydrogen bonds between each Gly N—H group and the oxygen of the succeeding Pro residue on a neighboring chain. Every third residue on each chain must be Gly because no other residue can fit near the helix axis. The bulky Pro side chains are on the periphery of the helix, where they are sterically unhindered. [After Yonath, A. and Traub, W., *J. Mol. Biol.* 43**,** 461 (1969).] (*b*) Space-filling model of a collagenlike peptide. The three parallel polypeptide chains (*blue, purple,* and *green*) are staggered by one residue. Ala residues (*yellow*), which replace the normally occurring Gly residue in each chain cause a significant distortion of the normal collagen structure. [Courtesy of Helen Berman, Rutgers University.] See Kinemage Exercises 4-3 and 4-4**.**

ecules form a hydrogen-bonded network that is anchored to the polypeptides in large part through hydrogen bonds to the 4-OH group of Hyp.

Collagen's well-packed, rigid, triple-helical structure is responsible for its characteristic tensile strength. The twist in the helix cannot be pulled out under tension because its component polypeptide chains are twisted in the opposite direction (Fig. 6-17). Successive levels of fiber bundles in high-quality ropes and cables, as well as in other proteins such as keratin (Fig. 6-14), are likewise oppositely twisted.

Several types of collagen molecules assemble to form loose networks or thick fibrils arranged in bundles or sheets, depending on the tissue. The collagen molecules in fibrils are organized in staggered arrays that are stabilized by hydrophobic interactions resulting from the close packing of triple-helical units. Collagen is also covalently cross-linked, which accounts for its poor solubility. The cross-links cannot be disulfide bonds, as in keratin, because collagen is almost devoid of Cys residues. Instead, the cross-links are derived from Lys and His side chains in reactions such as those shown in Fig. 6-19. **Lysyl oxidase,** the enzyme that converts Lys residues to those of the aldehyde **allysine,** is the only enzyme implicated in this cross-linking process. Up to four side chains can be covalently bonded to each other. The cross-links do not form at random but tend to occur near the N- and C-termini of the collagen molecules. The degree of cross-linking in a particular tissue increases with age. This is why meat from older animals is tougher than meat from younger animals.

D. Nonrepetitive Protein Structure

The majority of proteins are globular proteins that, unlike the fibrous proteins discussed in the preceding section, may contain several types of regular secondary structure, including α helices, β sheets, and other recognizable elements. A significant portion of a protein's structure may also be irregular or unique.

Irregular Structures

Segments of polypeptide chains whose successive residues do not have similar ϕ and ψ values are sometimes called coils. However, you should not confuse this term with the appellation **random coil,** which refers to the totally disordered and rapidly fluctuating conformations assumed by **denatured** (fully unfolded) proteins in solution. In **native** (folded) proteins, *nonrepetitive structures are no less ordered than are helices or β sheets; they are simply irregular and hence more difficult to describe.*

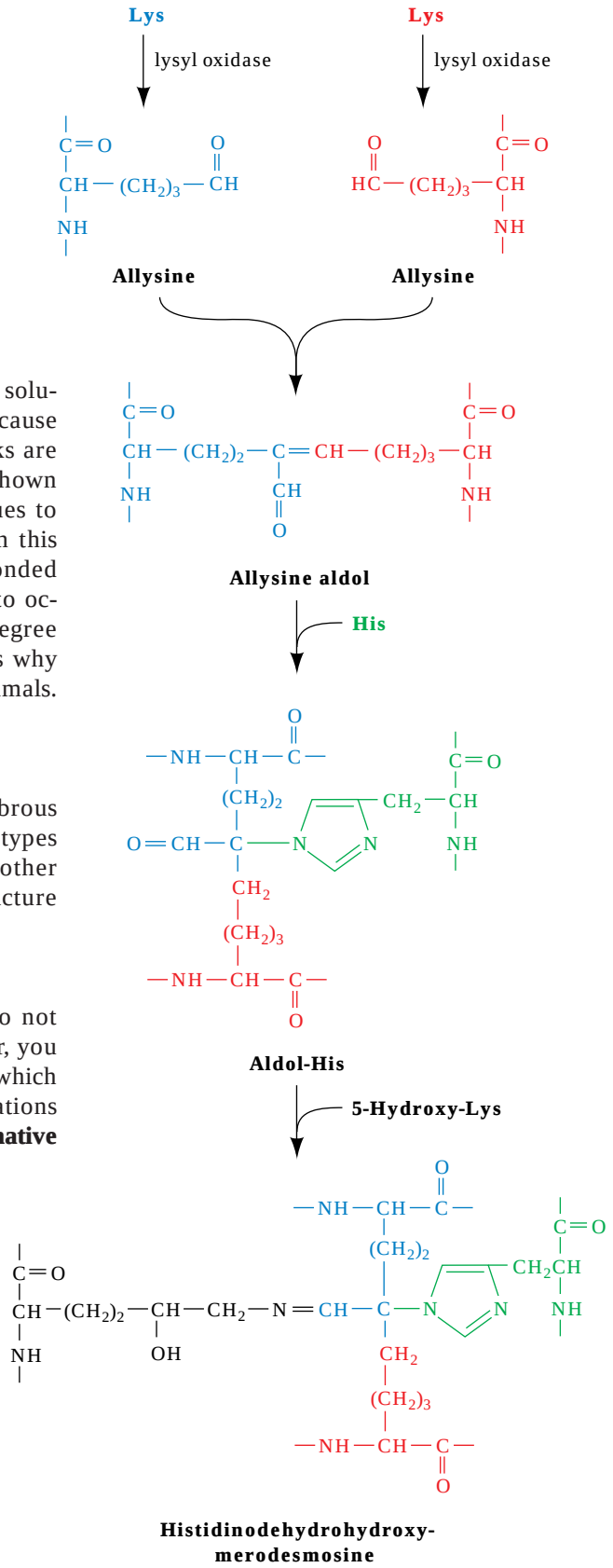


Figure 6-19. **A reaction pathway for cross-linking side chains in collagen.** The first step is the lysyl oxidase–catalyzed oxidative deamination of Lys to form the aldehyde allysine. Two allysines then undergo an aldol condensation to form allysine aldol. This product can react with His to form aldol histidine, which can in turn react with 5-hydroxylysine to form a Schiff base (an imine bond), thereby cross-linking four side chains.

Box 6-2
BIOCHEMISTRY IN CONTEXT

Protein Structure Prediction and Protein Design

Hundreds of thousands of protein sequences are known either through direct protein sequencing (Section 5-3) or, more commonly, through nucleic acid sequencing (Section 3-4). Yet the structures of only ~7000 of these proteins have as yet been determined by X-ray crystallography or NMR techniques. Determining the function of a newly discovered protein often requires knowledge of its three-dimensional structure.

There are currently several major approaches to protein structure prediction. The simplest and most reliable approach, homology modeling, aligns the sequence of interest with the sequence of a homologous protein of known structure—compensating for amino acid substitutions, insertions, and deletions—through modeling and energy minimization calculations. This method yields reliable models for proteins that have as little as 25% sequence identity with a protein of known structure, although, of course, the accuracy of the model increases with the degree of sequence identity.

Distantly related proteins may be structurally similar even though they have diverged to such an extent that their sequences show no obvious resemblance. Threading is a computational technique that attempts to determine the unknown structure of a protein by ascertaining whether it is consistent with a known protein structure. It does so by placing (threading) the residues of the unknown protein along the backbone of a known protein structure and then determining whether the amino acid side chains of the unknown protein are stable in that arrangement. This method is not yet reliable, although it has yielded encouraging results.

Empirical methods based on experimentally determined statistical information such as the α helix and β sheet propensities deduced by Chou and Fasman (Table 6-1) have been moderately successful in predicting the secondary structures of proteins. Their advantage is their simplicity (they don't require a computer).

Since the native structure of a protein depends only on its amino acid sequence, it should be possible, in principle, to predict the structure of a protein based only on its chemical and physical properties (e.g., the hydrophobicity, size, hydrogen-bonding propensity, and charge of each of its amino acid residues). Such *ab initio* (from the beginning) methods are still in their infancy. They are moderately successful in predicting simple structures such as a single α helix but have failed miserably when tested with larger polypeptides whose structures have been experimentally determined. Nevertheless, a recently developed algorithm that simulates the hierarchical protein folding pathway has yielded structural models that are surprisingly similar to those of the corresponding observed protein structures.

Protein design, the experimental inverse of protein structure prediction, has provided insights into protein folding and stability. Protein design begins with a target structure such as a simple sandwich of β sheets or a bundle of four α helices. It attempts to construct an amino acid sequence that will form that structure. The designed polypeptide is then chemically or biologically synthesized, and its structure is determined. Fortunately, protein folding seems to be governed more by extended sequences of amino acids than by individual residues, which allows some room for error in designing polypeptides. Experimental results suggest that the greatest challenge of protein design may lie not in getting the polypeptide to fold to the desired conformation but in preventing it from folding into other, unwanted conformations. In this respect, science lags far behind nature. However, the recent successful design, using computationally based techniques, of a 28-residue polypeptide that stably folds to the desired structure (the smallest known polypeptide that is capable of folding into a unique structure without the aid of disulfide bonds, metal ions, or other subunits) indicates that significant progress in our understanding of protein folding has been made.

Variations in Standard Secondary Structure

Variations in amino acid sequence as well as the overall structure of the folded protein can distort the regular conformations of secondary structural elements. For example, the α helix frequently deviates from its ideal conformation in its initial and final turns of the helix. Similarly, a strand of polypeptide in a β sheet may contain an "extra" residue that is not hydrogen bonded to a neighboring strand, producing a distortion known as a **β bulge.**

Many of the limits on amino acid composition and sequence (Section 5-1) may be due in part to conformational constraints in the three-dimensional structure of proteins. For example, a Pro residue produces a kink in an α helix or β sheet. Similarly, steric clashes between several

sequential amino acid residues with large branched side chains (e.g., Ile and Tyr) can destabilize α helices.

Analysis of known protein structures by Peter Chou and Gerald Fasman revealed the propensity *P* of a residue to occur in an α helix or a β sheet (Table 6-1). Chou and Fasman also discovered that certain residues not only have a high propensity for a particular secondary structure but they tend to disrupt or break other secondary structures. Such data are useful for predicting the secondary structures of proteins with known amino acid sequences (see Box 6-2).

The presence of certain residues outside of α helices or β sheets may also be nonrandom. For example, α helices are often flanked by residues such as Asn and Gln, whose side chains can fold back to form hydrogen bonds with one of the four terminal residues of the helix, a phenomenon termed **helix capping.** Recall that the four residues at each end of an α helix are not fully hydrogen bonded to neighboring backbone segments (Fig. 6-7).

Turns and Loops

Segments with regular secondary structure such as α helices or strands of β sheets are typically joined by stretches of polypeptide that abruptly change direction. Such **reverse turns** or **β bends** (so named because they often connect successive strands of antiparallel β sheets) almost always occur at protein surfaces. Most reverse turns involve four successive amino acid residues more or less arranged in one of two ways, Type I and Type II, that differ by a 180° flip of the peptide unit linking residues 2 and 3 (Fig. 6-20). Both types of turns are stabilized by a hydrogen bond, although deviations from these ideal conformations often disrupt this hydrogen bond. In Type II turns, the oxygen atom of residue 2 crowds the C_β atom of residue 3, which is therefore usually Gly. Residue 2 of either type of turn is often Pro since it can assume the required conformation.

Almost all proteins with more than 60 residues contain one or more loops of 6 to 16 residues, called **Ω loops.** These loops, which have the

Table 6-1. **Propensities of Amino Acid Residues for α Helical and β Sheet Conformations**

Residue	P_α	P_β
Ala	1.42	0.83
Arg	0.98	0.93
Asn	0.67	0.89
Asp	1.01	0.54
Cys	0.70	1.19
Gln	1.11	1.10
Glu	1.51	0.37
Gly	0.57	0.75
His	1.00	0.87
Ile	1.08	1.60
Leu	1.21	1.30
Lys	1.16	0.74
Met	1.45	1.05
Phe	1.13	1.38
Pro	0.57	0.55
Ser	0.77	0.75
Thr	0.83	1.19
Trp	1.08	1.37
Tyr	0.69	1.47
Val	1.06	1.70

Source: Chou, P.Y. and Fasman, G.D., *Annu. Rev. Biochem.* 47, 258 (1978).

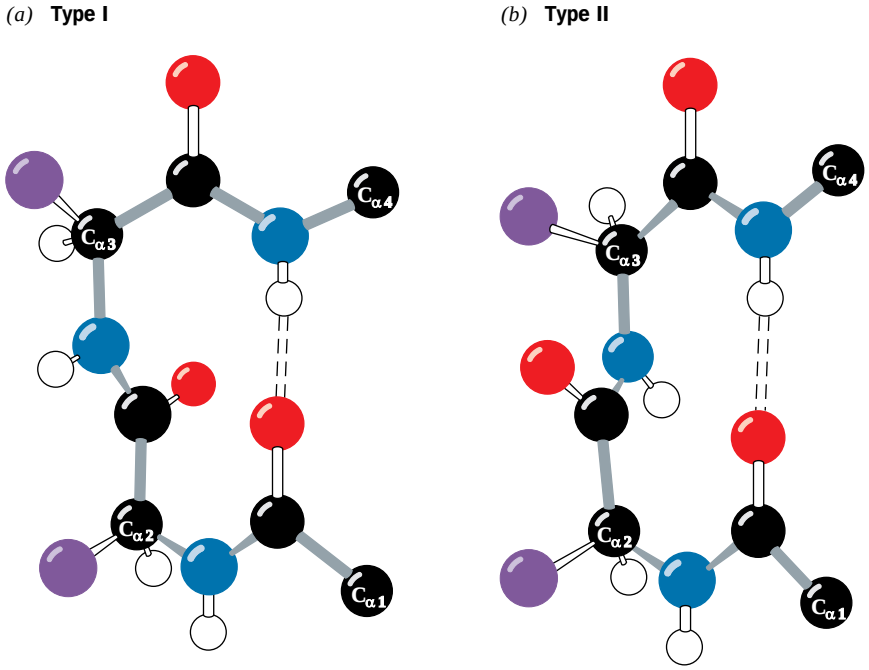


Figure 6-20. **Reverse turns in polypeptide chains.** Dashed lines represent hydrogen bonds. (*a*) Type I. (*b*) Type II. [Figure copyrighted © by Irving Geis.] See Kinemage Exercise 3-4.

Figure 6-21. **Space-filling model of an Ω loop.** Only backbone atoms are shown; the side chains would fill the loop. This structure is residues 40 to 54 from cytochrome *c*. [Courtesy of George Rose, The Johns Hopkins University.]

necked-in shape of the Greek uppercase letter omega (Fig. 6-21), are compact globular entities because their side chains tend to fill in their internal cavities. Since Ω loops are almost invariably located on the protein surface, they may have important roles in biological recognition processes.

2. TERTIARY STRUCTURE

The tertiary structure of a protein describes the folding of its secondary structural elements and specifies the positions of each atom in the protein, including those of its side chains. The known protein structures have come to light through **X-ray crystallographic** or **nuclear magnetic resonance** (**NMR**) studies. The atomic coordinates of most of these structures are deposited in a database known as the Protein Data Bank (PDB). These data are readily available via the Internet (http://www.pbd.bnl.gov), which allows the tertiary structures of a variety of proteins to be analyzed and compared. The common features of protein tertiary structure reveal much about the biological functions of the proteins and their evolutionary origins.

A. Determining Protein Structure

X-Ray crystallography is one of the most powerful methods for studying macromolecular structure. According to optical principles, the uncertainty in locating an object is approximately equal to the wavelength of the radiation used to observe it. X-Rays can directly image a molecule because X-ray wavelengths are comparable to covalent bond distances (~1.5 Å; individual molecules cannot be seen in a light microscope because visible light has a minimum wavelength of 4000 Å).

When a crystal of the molecule to be visualized is exposed to a collimated (parallel) beam of X-rays, the atoms in the molecule scatter the X-rays, with the scattered rays canceling or reinforcing each other in a process known as diffraction. The resulting **diffraction pattern** is recorded on photographic film (Fig. 6-22) or by a radiation counter. The intensities of the

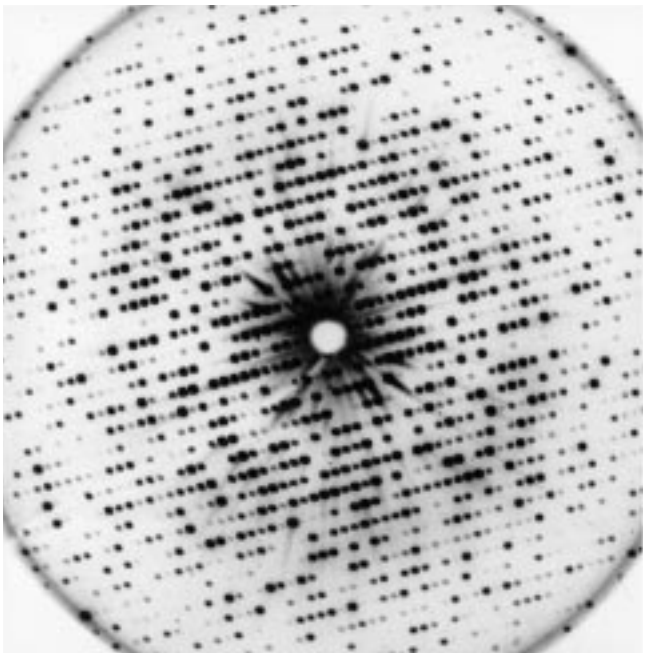


Figure 6-22. **An X-ray diffraction photograph of a crystal of sperm whale myoglobin.** The intensity of each diffraction maximum (the darkness of each spot) is a function of the crystal's electron density. [Courtesy of John Kendrew, Cambridge University, U.K.]

diffraction maxima (darkness of the spots on the film) are then used to mathematically construct the three-dimensional image of the crystal structure. The photograph in Fig. 6-22 represents only a small portion of the total diffraction information available from a crystal of myoglobin, a small globular protein. In contrast, fibrous proteins do not crystallize but, instead, can be drawn into fibers whose X-ray diffraction patterns contain only a few spots and thus contain comparatively little structural information. Likewise, the diffraction pattern of a DNA fiber (Fig. 3-8) is relatively simple.

X-Rays interact almost exclusively with the electrons in matter, not the atomic nuclei. An X-ray structure is therefore an image of the electron density of the object under study. This information can be shown as a three-dimensional **contour map** (Fig. 6-23). Hydrogen atoms, which have only one electron, are not visible in macromolecular X-ray structures.

The X-ray structures of small organic molecules can be determined with a resolution on the order of ~1 Å. Few protein crystals have this degree of organization. Furthermore, not all proteins can be coaxed to crystallize, that is, to precipitate in ordered three-dimensional arrays. The protein crystals that do form (Fig. 6-24) differ from those of most small organic molecules in being highly hydrated; protein crystals are typically 40 to 60% water by volume. The large solvent content gives protein crystals a soft, jellylike consistency so that the molecules are typically disordered by a few angstroms. This limits their resolution to about 2 to 3.5 Å, although a few protein crystals are better ordered (have higher resolution).

A resolution of a few angstroms is too coarse to clearly reveal the positions of individual atoms, but the distinctive shape of the polypeptide backbone can usually be traced. The positions and orientations of its side chains can therefore be deduced. However, since many side chains have similar sizes and shapes, *knowledge of the protein's primary structure is required to fit the sequence of amino acids to its electron density map.* Mathematical techniques can then refine the atomic positions to within ~0.1 Å in high-resolution structures.

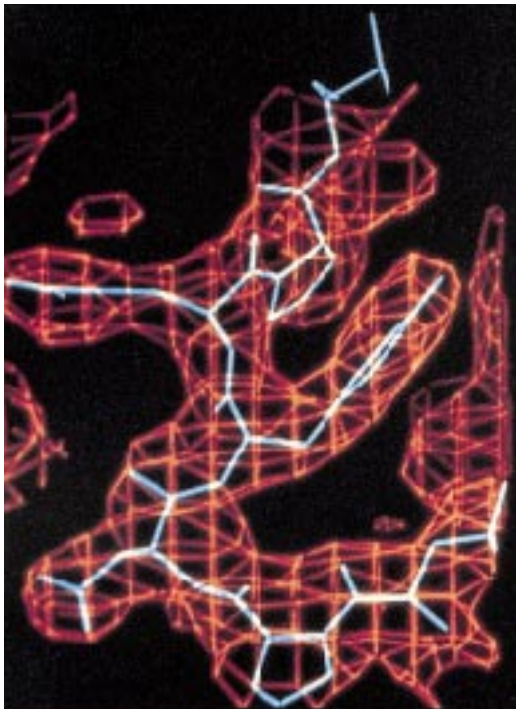
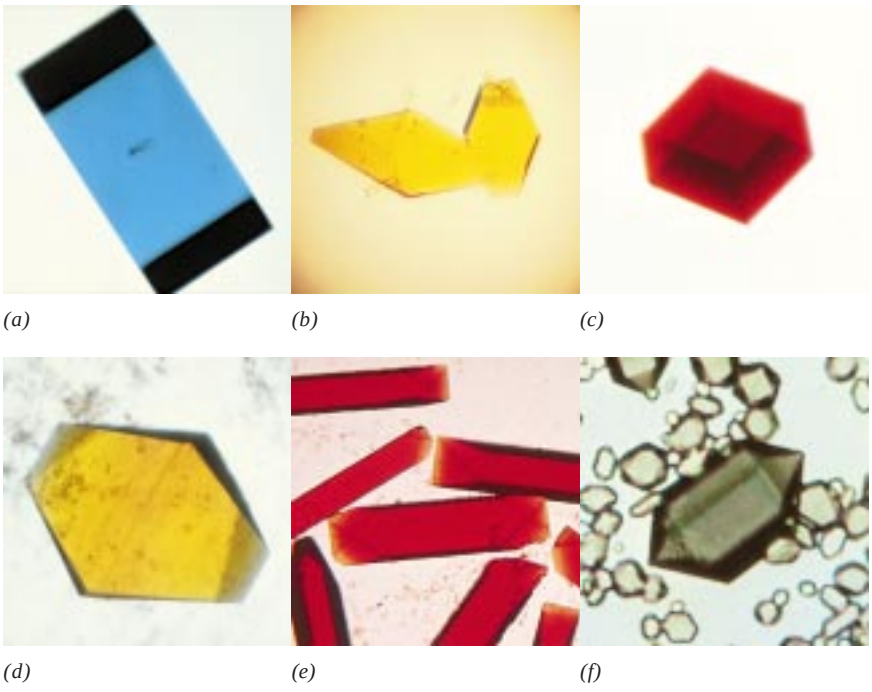


Figure 6-23. **An electron density map.** The three-dimensional outline of the electron density (*orange*) is shown with a superimposed atomic model of the corresponding polypeptide segment (*white*). This structure is a portion of human rhinovirus (the cause of the common cold). [Courtesy of Michael Rossmann, Purdue University.]



(a) *(b)* *(c)*

(d) *(e)* *(f)*

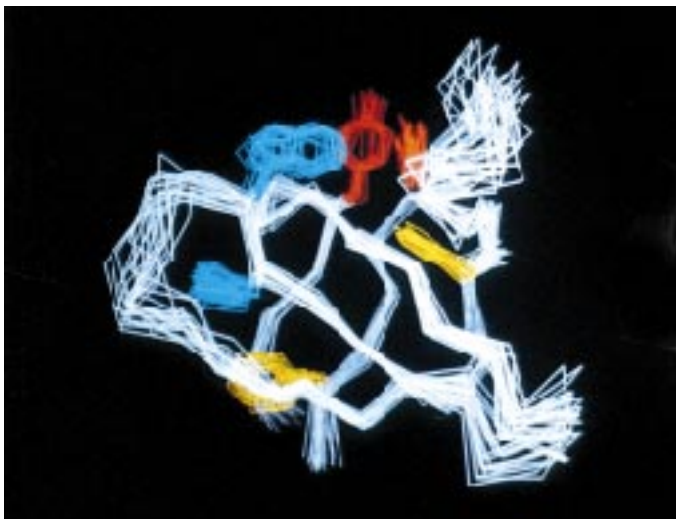
Figure 6-24. **Protein crystals.** (*a*) Azurin from *Pseudomonas aeruginosa,* (*b*) flavodoxin from *Desulfovibrio vulgaris,* (*c*) rubredoxin from *Clostridium pasteurianum,* (*d*) azidomet myohemerythrin from the marine worm *Siphonosoma funafuti,* (*e*) lamprey hemoglobin, and (*f*) bacteriochlorophyll *a* protein from *Prosthecochloris aestuarii.* These crystals are colored because the proteins contain light-absorbing groups; proteins are colorless in the absence of such groups. [Parts *a–c* courtesy of Larry Siecker, University of Washington; parts *d* and *e* courtesy of Wayne Hendrikson, Columbia University; and part *f* courtesy of John Olsen, Brookhaven National Laboratories, and Brian Matthews, University of Oregon.]

Box 6-3
BIOCHEMISTRY IN FOCUS

Protein Structure Determination by NMR

The determination of the three-dimensional structures of small proteins (<250 residues) in aqueous solution has become possible since the mid-1980s, through the development of two-dimensional (2D) NMR spectroscopy (and, more recently, of 3D and 4D techniques), in large part by Kurt Wüthrich. A sample is placed in a magnetic field so that the spins of its protons are aligned. When radiofrequency pulses are applied, the protons are excited and then emit signals whose frequency depends on the molecular environment of the protons. Proteins contain so many atoms that their standard (one-dimensional) NMR spectra consist almost entirely of overlapping signals that are impossible to interpret.

In 2D NMR, the characteristics of the applied signal are varied in order to obtain additional information from interactions between protons that are <5 Å apart in space or are covalently connected by only one or two other atoms. The resulting set of distances, together with known geometric constraints such as covalent bond distances and angles, group planarity, chirality, and van der Waals radii, are used to compute the protein's three-dimensional structure. However, since interproton distance measurements are imprecise, they are insufficient to imply a unique structure. For this reason, the NMR structure of a protein (or any other macromolecule with a well-defined structure) is often presented as an ensemble of closely related structures. The above NMR structure of a 64-residue polypeptide is presented as 20 superimposed C_α traces (*white*), each of which is consistent with the NMR data and geometric constraints. A few side chains (*red*, *yellow*, and *blue*) are also shown.



[Figure courtesy of Stuart Schreiber, Harvard University.]

Another consequence of the large solvent content of protein crystals is that *crystalline proteins maintain their native conformations and therefore their functions.* Indeed, the degree of hydration of proteins in crystals is similar to that in cells. Thus, the X-ray crystal structures of proteins often provide a basis for understanding their biological activities. Crystal structures have been used as starting points for designing drugs that can interact specifically with target proteins under physiological conditions.

Recent advances in NMR spectroscopy have permitted the determination of the structures of proteins and nucleic acids in solution (but limited to a size of <30 kD). Thus, NMR techniques can be used to elucidate the structures of proteins and other macromolecules that fail to crystallize (see Box 6-3). In the several instances in which both the X-ray and NMR structures of a particular protein were determined, the two structures had few, if any, significant differences.

Visualizing Proteins

The huge number of atoms in proteins makes it difficult to visualize them using the same sorts of models employed for small organic molecules. Ball-and-stick representations showing all or most atoms in a protein (as in Figs. 6-7 and 6-10) are exceedingly cluttered, and space-filling models (as in Figs. 6-8 and 6-11) obscure the internal details of the protein. Accordingly, computer-generated or artistic renditions (e.g., Fig. 6-12) are often more useful for representing protein structures. The course of the polypeptide chain can be followed by tracing the positions of its C_α atoms or by representing helices as helical ribbons or cylinders, and β sheets as sets of flat arrows pointing from the N- to the C-termini.

B. Motifs (Supersecondary Structures) and Domains

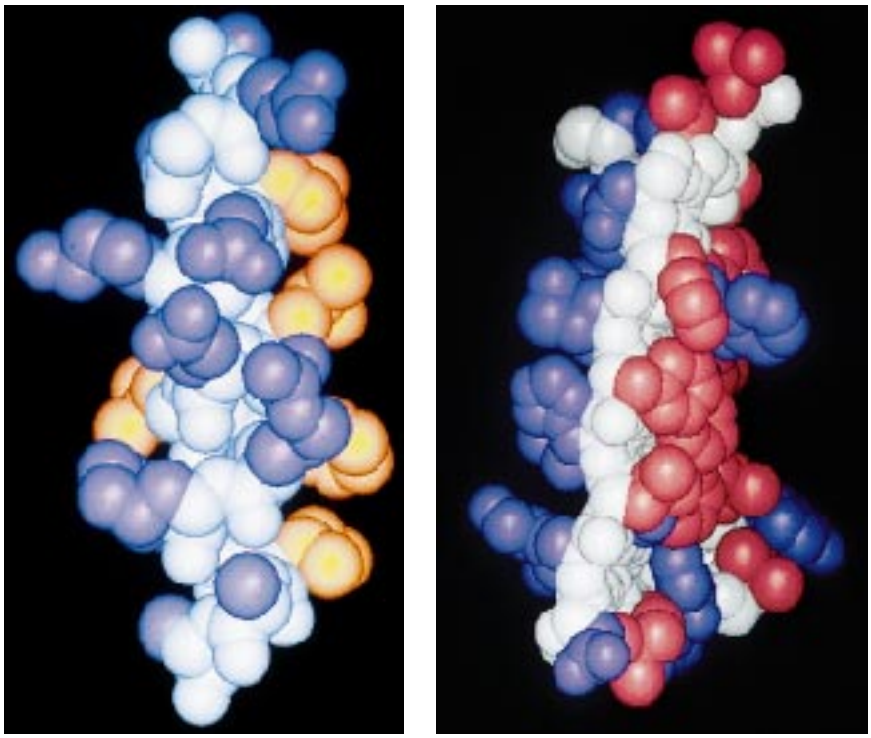
In the years since Kendrew solved the structure of myoglobin, nearly 7000 protein structures have been reported. No two are exactly alike, but they exhibit remarkable consistencies.

Side Chain Location Varies with Polarity

The primary structures of globular proteins generally lack the repeating sequences that support the regular conformations seen in fibrous proteins. However, *the amino acid side chains in globular proteins are spatially distributed according to their polarities:*

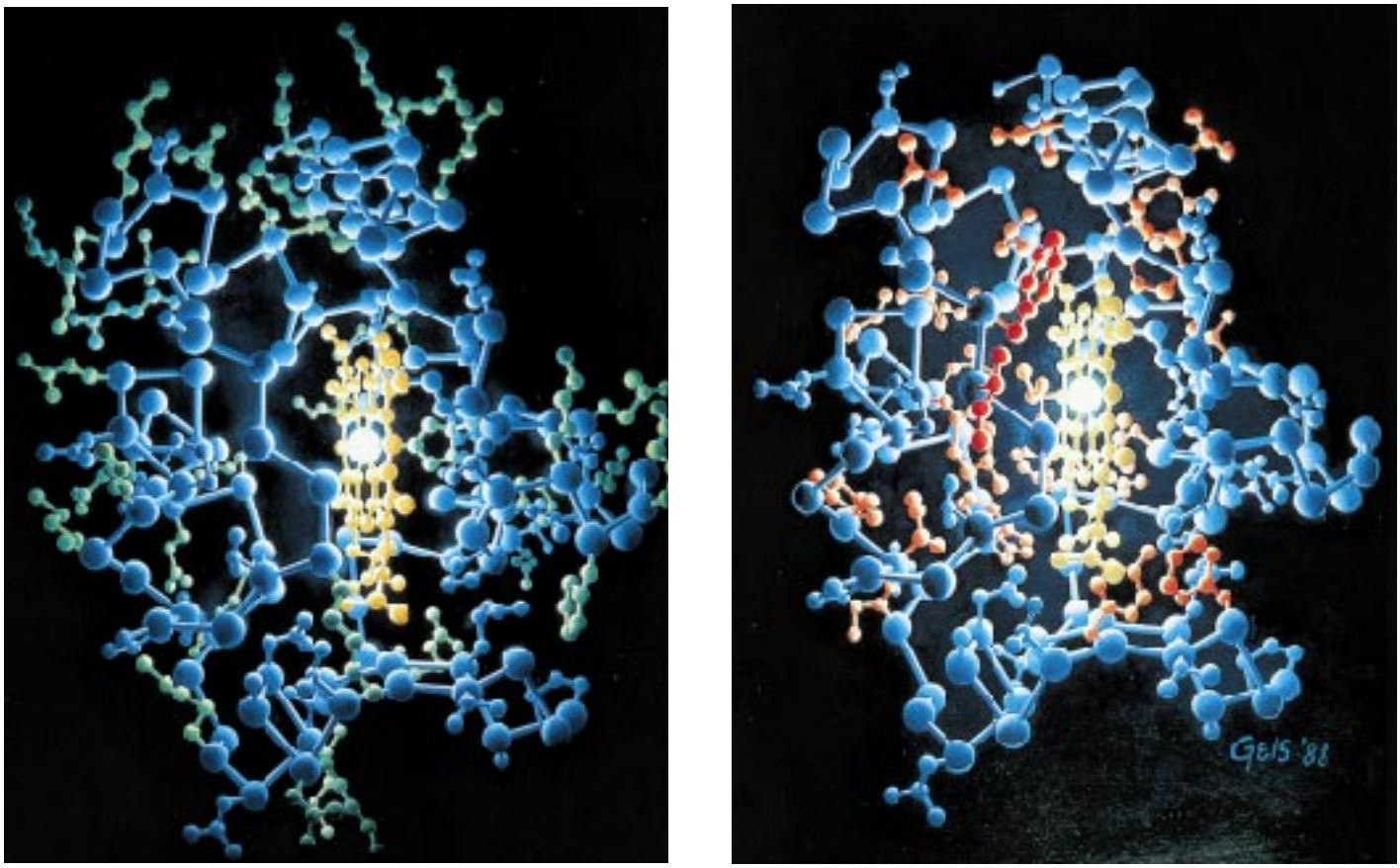
1. The nonpolar residues Val, Leu, Ile, Met, and Phe occur mostly in the interior of a protein, out of contact with the aqueous solvent. The hydrophobic effects that promote this distribution are largely responsible for the three-dimensional structure of native proteins.
2. The charged polar residues Arg, His, Lys, Asp, and Glu are usually located on the surface of a protein in contact with the aqueous solvent. This is because immersing an ion in the virtually anhydrous interior of a protein is energetically unfavorable.
3. The uncharged polar groups Ser, Thr, Asn, Gln, and Tyr are usually on the protein surface but also occur in the interior of the molecule. When buried in the protein, these residues are almost always hydrogen bonded to other groups; in a sense, the formation of a hydrogen bond "neutralizes" their polarity. This is also the case with the polypeptide backbone.

These general principles of side chain distribution are evident in individual elements of secondary structure (Fig. 6-25) as well as in whole proteins



(a) *(b)*

Figure 6-25. **Side chain locations in an α helix and a β sheet.** In these space-filling models, the main chain is white, nonpolar side chains are yellow or brown, and polar side chains are purple. (*a*) An α helix from sperm whale myoglobin. Note that the nonpolar residues are primarily on one side of the helix. (*b*) An antiparallel β sheet from concanavalin A (*side view*). The protein interior is to the right and the exterior is to the left.



(a) *(b)*

Figure 6-26. **Side chain distribution in horse heart cytochrome *c*.** In these paintings, based on the X-ray structure determined by Richard Dickerson, the protein is illuminated by its single iron atom centered in a heme group. Hydrogen atoms are not shown. In (*a*) the hydrophilic side chains are green, and in (*b*) the hydrophobic side chains are orange. [Figures copyrighted © by Irving Geis.] See Kinemage Exercise 5.

(Fig. 6-26). Polar side chains tend to extend toward—and thereby help form—the protein's surface, whereas nonpolar side chains largely extend toward—and thereby occupy—its interior.

Most proteins are quite compact, with their interior atoms packed together even more efficiently than the atoms in a crystal of small organic molecules. Nevertheless, the atoms of protein side chains almost invariably have low-energy arrangements. Evidently, interior side chains adopt relaxed conformations despite the profusion of intramolecular interactions. Closely packed protein interiors generally exclude water. When water molecules are present, they often occupy specific positions where they can form hydrogen bonds, sometimes acting as a bridge between two hydrogen-bonding protein groups.

Helices and Sheets Can Be Combined in Various Ways

The major types of secondary structural elements, α helices and β sheets, occur in globular proteins in varying proportions and combinations. Some proteins, such as hemoglobin subunits, consist only of α helices spanned by short connecting links (Fig. 6-27*a*). Others, such as **concanavalin A,** have a large proportion of β sheets and are devoid of α helices (Fig. 6-27*b*). Most proteins, such as **triose phosphate isomerase** (Fig. 6-27*c*) and carboxypeptidase A (Fig. 6-12), have significant amounts of both types of secondary structure (on average, ~31% α helix and ~28% β sheet).

Figure 6-27. **Examples of globular proteins.** Different proteins contain different proportions and arrangements of secondary structural elements. In these models, α helices are drawn as helical ribbons, and strands of β sheets are drawn as flat arrows pointing toward the C-terminus. (*a*) A hemoglobin subunit, with its heme group shown as a skeletal model. ● See Kinemage Exercise 6-2. (*b*) Jack bean concanavalin A. The spheres represent metal ions. (*c*) Triose phosphate isomerase from chicken muscle. [After drawings by Jane Richardson, Duke University.] ● See Kinemage Exercise 12-1.

Certain groupings of secondary structural elements, called **supersecondary structures** or **motifs,** occur in many unrelated globular proteins:

1. The most common form of supersecondary structure is the **βαβ motif,** in which an α helix connects two parallel strands of a β sheet (Fig. 6-28*a*).
2. Another common supersecondary structure, the **β hairpin** motif, consists of antiparallel strands connected by relatively tight reverse turns (Fig. 6-28*b*).
3. In an **αα motif,** two successive antiparallel α helices pack against each other with their axes inclined. This permits energetically favorable

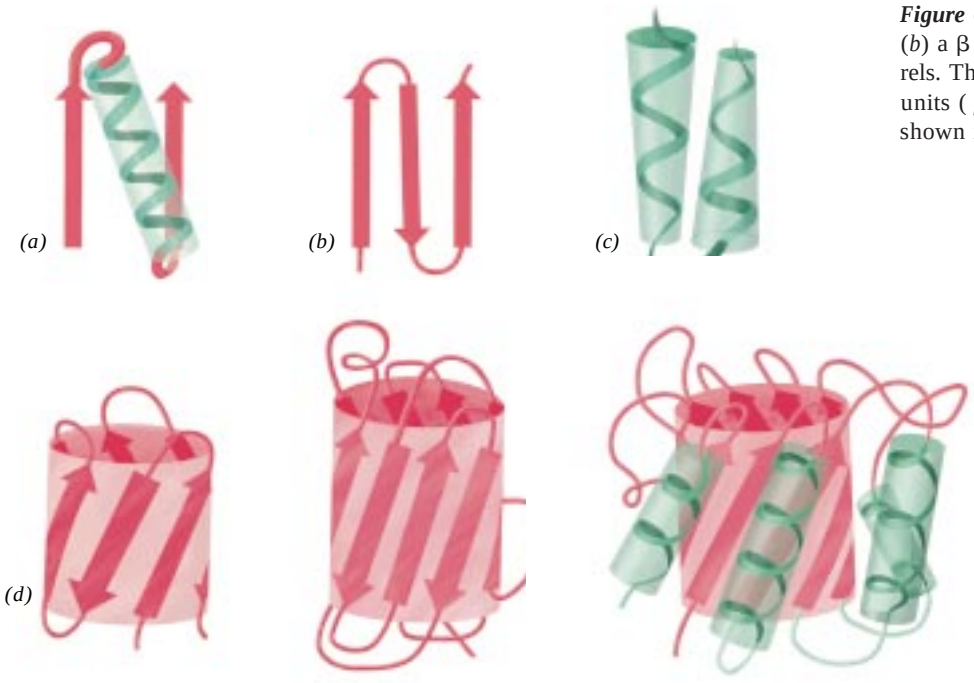


Figure 6-28. **Protein motifs.** (*a*) A βαβ motif, (*b*) a β hairpin, (*c*) an αα motif, and (*d*) β barrels. The β barrel composed of overlapping βαβ units (*far right*) is known as an α/β barrel. It is shown in top view in Fig. 6-27*c*.

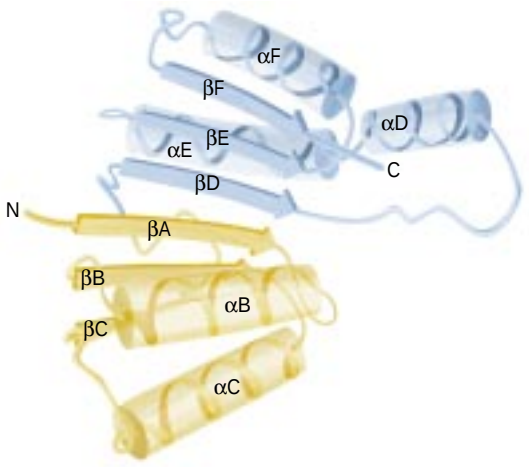


Figure 6-29. **An idealized dinucleotide-binding (Rossmann) fold.** The two structurally similar βαβαβ units (*yellow* and *blue*) can each bind a nucleotide portion of the dinucleotide NAD^+ (not shown). [After Rossmann, M.G., Liljas, A., Bränden, C.-I., and Banaszak, L.J., *in* Boyer, P.D. (Ed.), *The Enzymes*, Vol. 11 (3rd ed.), p. 68, Academic Press (1975).] See Kinemage Exercise 21-1.

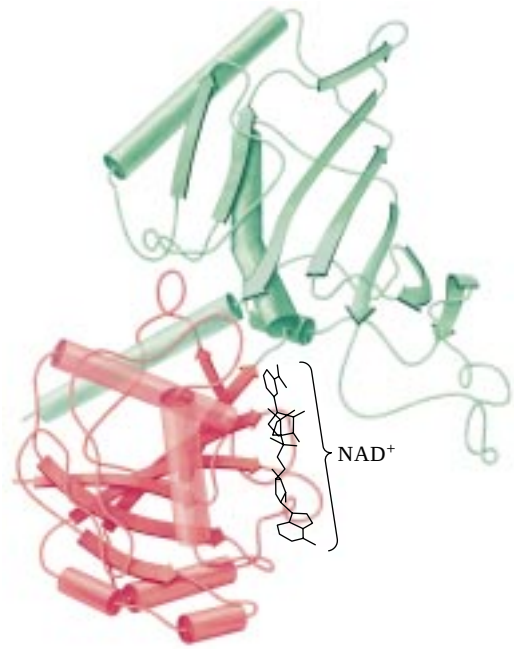


Figure 6-30. **The two-domain protein glyceraldehyde-3-phosphate dehydrogenase.** The first domain (*red*) binds NAD^+ (*black*), and the second domain (*green*) binds glyceraldehyde-3-phosphate (not shown). [After Biesecker, G., Harris, J.I., Thierry, J.C., Walker, J.E., and Wonacott, A., *Nature* 266, 331 (1977).]

intermeshing of their contacting side chains (Fig. 6-28*c*). Such associations stabilize the coiled coil conformation of α keratin (Section 6-1C).

4. Extended β sheets often roll up to form **β barrels.** Three different types of β barrels are shown in Fig. 6-28*d*.

Motifs may have functional as well as structural significance. For example, Michael Rossmann showed that a βαβαβ unit, in which the β strands form a parallel sheet with α helical connections, often acts as a nucleotide-binding site. In most proteins that bind dinucleotides (such as nicotinamide adenine dinucleotide, NAD^+; Section 3-1), two such βαβαβ units combine to form a motif known as a **dinucleotide-binding fold,** or **Rossmann fold** (Fig. 6-29).

Large Polypeptides Form Domains

Polypeptide chains containing more than ~200 residues usually fold into two or more globular clusters known as **domains,** which give these proteins a bi- or multilobal appearance. Most domains consist of 100 to 200 amino acid residues and have an average diameter of ~25 Å. Each subunit of the enzyme **glyceraldehyde-3-phosphate dehydrogenase,** for example, has two distinct domains (Fig. 6-30). A polypeptide chain wanders back and forth within a domain, but neighboring domains are usually connected by only one or two polypeptide segments. *Consequently, many domains are structurally independent units that have the characteristics of small globular proteins.* Nevertheless, the domain structure of a protein is not necessarily obvious since its domains may make such extensive contacts with each other that the protein appears to be a single globular entity.

An inspection of the various protein structures diagrammed in this chapter reveals that domains consist of two or more layers of secondary structural elements. The reason for this is clear: At least two such layers are required to seal off a domain's hydrophobic core from its aqueous environment.

Domains often have a specific function such as the binding of a small molecule. In Fig. 6-30, for example, NAD^+ binds to the first domain of glyceraldehyde-3-phosphate dehydrogenase (note its dinucleotide-binding fold). In multidomain proteins, binding sites often occupy the clefts between domains; that is, the small molecules are bound by groups from two domains. In such cases, the relatively pliant covalent connection between the domains allows flexible interactions between the protein and the small molecule.

C. Protein Families

The thousands of known protein structures, comprising an even greater number of separate domains, can be grouped into families by examining the overall paths followed by their polypeptide chains. When folding patterns are compared without regard to the amino acid sequence or the presence of surface loops, the number of unique structural domains drops to only a few hundred. (Although not all protein structures are known, estimates place an upper limit of about 1000 on the total number of unique protein domains in nature.) Surprisingly, a few dozen folding patterns account for about half of all known protein structures.

There are several possible reasons for the limited number of known domain structures. The numbers may reflect database bias; that is, the collection of known protein structures may not be a representative sample of all protein structures. However, the rapidly increasing number of proteins whose structures have been determined makes this possibility less and less

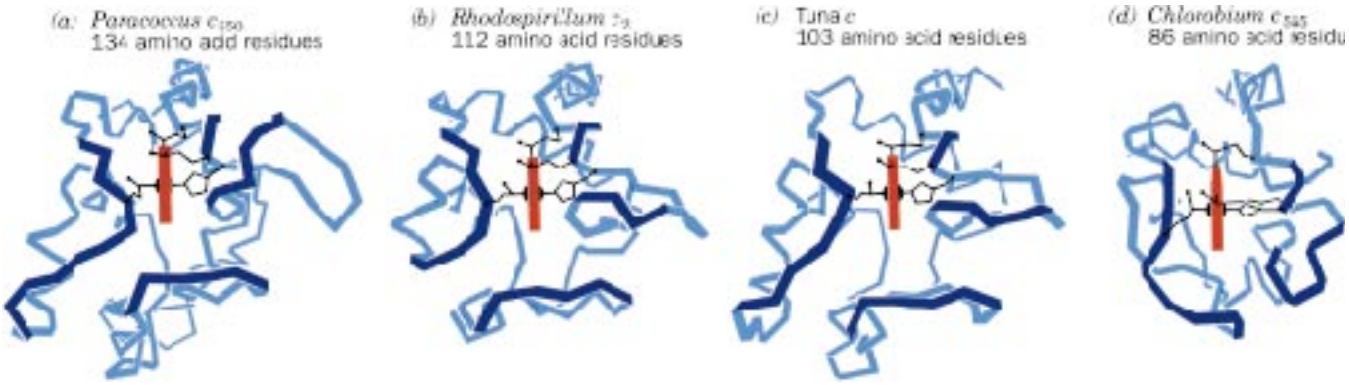


Figure 6-31. **Three-dimensional structures of *c*-type cytochromes.** The polypeptide backbones (*blue*) are shown in analogous orientations such that their heme groups (*red*) are viewed edge-on. The Cys, Met, and His side chains that covalently link the heme to the protein are also shown. (*a*) Cytochrome c_{550} from *Paracoccus denitrificans* (134 residues), (*b*) cytochrome c_2 from *Rhodospirillum rubrum* (112 residues), (*c*) cytochrome *c* from tuna (103 residues), and (*d*) cytochrome c_{555} from *Chlorobium thiosulfatophilum* (86 residues). [Figures copyrighted © by Irving Geis.] ● See Kinemage Exercise 5.

plausible. More likely, the common protein structures may be evolutionary sinks—domains that arose and persisted because of their ability (1) to form stable folding patterns; (2) to tolerate amino acid deletions, substitutions, and insertions, thereby making them more likely to survive evolutionary changes; and (3) to support essential biological functions.

Polypeptides with similar sequences tend to adopt similar backbone conformations. This is certainly true for evolutionarily related proteins that carry out similar functions. For example, the cytochromes *c* of different species are highly conserved proteins with closely similar sequences (see Table 5-6) and three-dimensional structures.

Cytochrome *c* occurs only in eukaryotes, but prokaryotes contain proteins known as ***c*-type cytochromes,** which perform the same general function (that of an electron carrier). The *c*-type cytochromes from different species exhibit only low degrees of sequence similarity to each other and to eukaryotic cytochromes *c*. Yet their X-ray structures are clearly similar, particularly in polypeptide chain folding and side chain packing in the protein interior (Fig. 6-31). The major structural differences among *c*-type cytochromes lie in the various polypeptide loops on their surfaces. The sequences of the *c*-type cytochromes have diverged so far from one another that, in the absence of their X-ray structures, they can be properly aligned only through the use of recently developed and mathematically sophisticated computer programs. Thus, *it appears that the essential structural and functional elements of proteins, rather than their amino acid residues, are conserved during evolution.*

Structural similarities in proteins with only distantly related functions are commonly observed. For example, many NAD^+-binding enzymes that participate in widely different metabolic pathways contain similar dinucleotide-binding folds (see Fig. 6-30) coupled to diverse domains that carry out specific enzymatic reactions.

3. QUATERNARY STRUCTURE AND SYMMETRY

Most proteins, particularly those with molecular masses >100 kD, consist of more than one polypeptide chain. These polypeptide subunits associate with a specific geometry. The spatial arrangement of these subunits is known as a protein's quaternary structure.

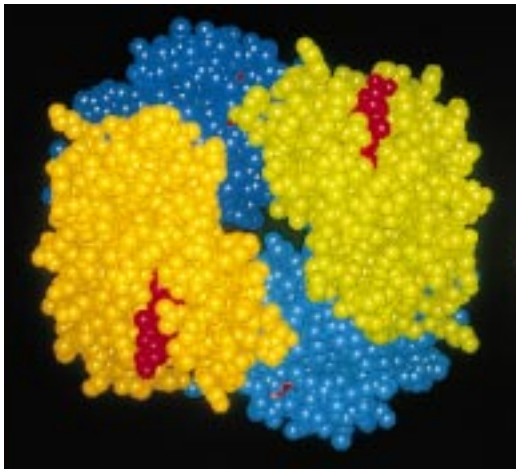


Figure 6-32. **Quaternary structure of hemoglobin.** In this space-filling model, the α_1, α_2, β_1, and β_2 subunits are colored yellow, green, cyan, and blue, respectively. Heme groups are red.

There are several reasons why multisubunit proteins are so common. In large assemblies of proteins, such as collagen fibrils, the advantages of subunit construction over the synthesis of one huge polypeptide chain are analogous to those of using prefabricated components in constructing a building: Defects can be repaired by simply replacing the flawed subunit; the site of subunit manufacture can be different from the site of assembly into the final product; and the only genetic information necessary to specify the entire edifice is the information specifying its few different self-assembling subunits. In the case of enzymes, increasing a protein's size tends to better fix the three-dimensional positions of its reacting groups. *Increasing the size of an enzyme through the association of identical subunits is more efficient than increasing the length of its polypeptide chain since each subunit has an active site. More importantly, the subunit construction of many enzymes provides the structural basis for the regulation of their activities* (Sections 7-2E and 12-3).

Subunits Usually Associate Noncovalently

A multisubunit protein may consist of identical or nonidentical polypeptide chains. Hemoglobin, for example, has the subunit composition $\alpha_2\beta_2$ (Fig. 6-32). Proteins with more than one subunit are called **oligomers,** and their identical units are called **protomers.** A protomer may therefore consist of one polypeptide chain or several unlike polypeptide chains. In this sense, hemoglobin is a dimer of $\alpha\beta$ protomers.

The contact regions between subunits closely resemble the interior of a single-subunit protein. They contain closely packed nonpolar side chains, hydrogen bonds involving the polypeptide backbones and their side chains, and, in some cases, interchain disulfide bonds.

Subunits Are Symmetrically Arranged

In the vast majority of oligomeric proteins, the protomers are symmetrically arranged; that is, each protomer occupies a geometrically equivalent position in the oligomer. Proteins cannot have inversion or mirror symmetry, however, because bringing the protomers into coincidence would require converting chiral L residues to D residues. Thus, *proteins can have only* ***rotational symmetry.***

In the simplest type of rotational symmetry, **cyclic symmetry,** protomers are related by a single axis of rotation (Fig. 6-33*a*). Objects with two-, three-, or *n*-fold rotational axes are said to have C_2, C_3, or C_n symmetry, respectively. C_2 symmetry is the most common; higher cyclic symmetries are relatively rare.

Dihedral symmetry (D_n), a more complicated type of rotational symmetry, is generated when an *n*-fold rotation axis intersects a two-fold rotation axis at right angles (Fig. 6-33*b*). An oligomer with D_n symmetry consists of $2n$ protomers. D_2 symmetry is the most common type of dihedral symmetry in proteins.

Other possible types of rotational symmetry are those of a tetrahedron, cube, and icosahedron (Fig. 6-33*c*). Some multienzyme complexes and spherical viruses are built on these geometric plans.

4. PROTEIN FOLDING AND STABILITY

Incredible as it may seem, thermodynamic measurements indicate that *native proteins are only marginally stable under physiological conditions.* The free energy required to denature them is ~0.4 $kJ \cdot mol^{-1}$ per amino acid residue, so a fully folded 100-residue protein is only about 40 $kJ \cdot mol^{-1}$

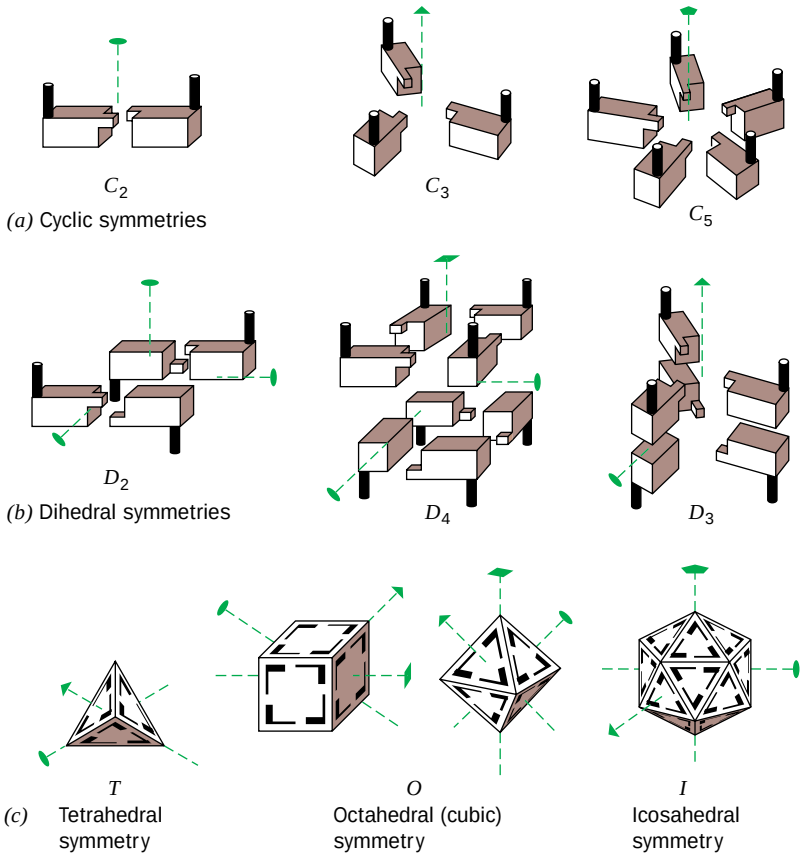


Figure 6-33. **Some symmetries for oligomeric proteins.** The oval, the triangle, the square, and the pentagon at the ends of the dashed green lines indicate, respectively, the unique two-fold, three-fold, four-fold, and five-fold rotational axes of the objects shown. (*a*) Assemblies with cyclic (*C*) symmetry. (*b*) Assemblies with dihedral (*D*) symmetry. In these objects, a 2-fold axis is perpendicular to another rotational axis. (*c*) Assemblies with the rotational symmetries of a tetrahedron (*T*), a cube or octahedron (*O*), and an icosahedron (*I*). [Figure copyrighted © by Irving Geis.]

more stable than its unfolded form (for comparison, the energy required to break a typical hydrogen bond is ~20 $kJ \cdot mol^{-1}$). The various noncovalent influences on proteins—hydrophobic effects, electrostatic interactions, and hydrogen bonding—each have energies that may total thousands of kilojoules per mole over an entire protein molecule. Consequently, a protein structure is the result of a delicate balance among powerful countervailing forces. In this section, we discuss the forces that stabilize proteins and the processes by which proteins achieve their most stable folded state.

A. Forces That Stabilize Protein Structure

Protein structures are governed primarily by hydrophobic effects and, to a lesser extent, by interactions between polar residues and other types of bonds.

The Hydrophobic Effect

The hydrophobic effect, which causes nonpolar substances to minimize their contacts with water (Section 2-1C)*, is the major determinant of native protein structure.* The aggregation of nonpolar side chains in the interior of

Table 6-2. **Hydropathy Scale for Amino Acid Side Chains**

Side Chain	*Hydropathy*
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	−0.4
Thr	−0.7
Ser	−0.8
Trp	−0.9
Tyr	−1.3
Pro	−1.6
His	−3.2
Glu	−3.5
Gln	−3.5
Asp	−3.5
Asn	−3.5
Lys	−3.9
Arg	−4.5

Source: Kyte, J. and Doolittle, R.F., *J. Mol. Biol.* 157, 110 (1982).

a protein is favored by the increase in entropy of the water molecules that would otherwise form ordered "cages" around the hydrophobic groups. The combined hydrophobic and hydrophilic tendencies of individual amino acid residues in proteins can be expressed as **hydropathies** (Table 6-2). The greater a side chain's hydropathy, the more likely it is to occupy the interior of a protein and vice versa. Hydropathies are good predictors of which portions of a polypeptide chain are inside a protein, out of contact with the aqueous solvent, and which portions are outside (Fig. 6-34).

Site-directed mutagenesis experiments in which individual interior residues have been replaced by a number of others suggest that the factors that affect stability are, in order, the hydrophobicity of the substituted residue, its steric compatibility, and, last, the volume of its side chain.

Electrostatic Interactions

In the closely packed interiors of native proteins, van der Waals forces, which are relatively weak (Section 2-1A), are nevertheless an important stabilizing influence. This is because these forces act over only short distances and hence are lost when the protein is unfolded.

Perhaps surprisingly, *hydrogen bonds, which are central features of protein structures, make only minor contributions to protein stability.* This is because hydrogen-bonding groups in an unfolded protein form energetically equivalent hydrogen bonds with water molecules. Nevertheless, hydrogen bonds are important determinants of native protein structures, because if a protein folded in a way that prevented a hydrogen bond from forming, the stabilizing energy of that hydrogen bond would be lost. Hydrogen bonding therefore fine-tunes tertiary structure by "selecting" the unique native structure of a protein from among a relatively small number of hydrophobically stabilized conformations.

The association of two ionic protein groups of opposite charge (e.g., Lys and Asp) is known as an **ion pair** or **salt bridge.** About 75% of the charged residues in proteins are members of ion pairs that are located mostly on the protein surface. Despite the strong electrostatic attraction between the

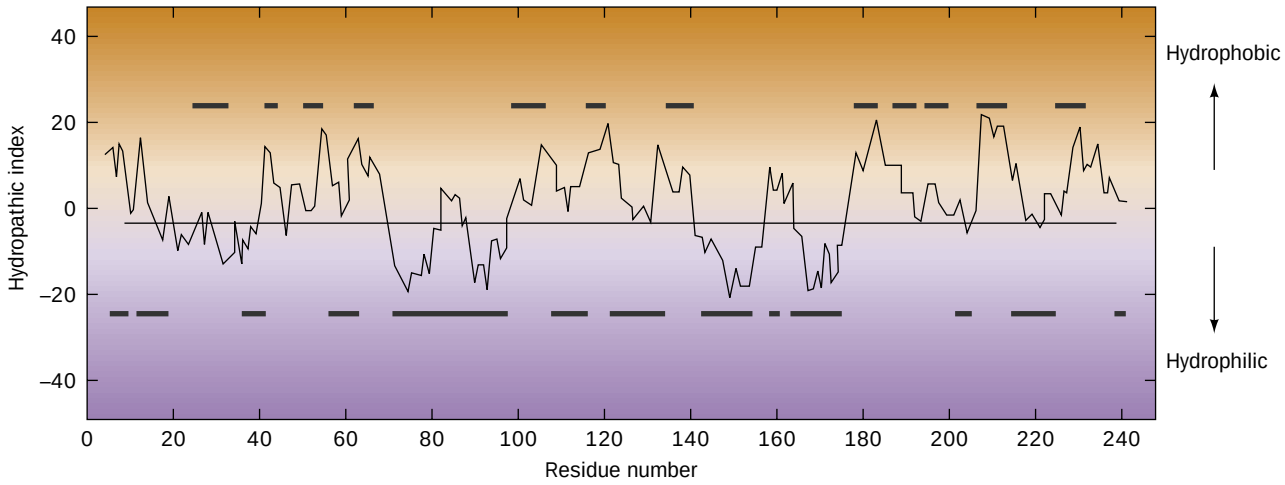


Figure 6-34. **A hydropathic index plot for bovine chymotrypsinogen.** The sum of the hydropathies of nine consecutive residues is plotted versus residue sequence number. A large positive hydropathic index indicates a hydrophobic region of the polypeptide, whereas a large negative value indicates a hydrophilic region. The upper bars denote the protein's interior regions, as determined by X-ray crystallography, and the lower bars denote the protein's exterior regions. [After Kyte, J. and Doolittle, R.F., *J. Mol. Biol.* 157, 111 (1982).]

oppositely charged members of an ion pair, these interactions contribute little to the stability of a native protein. This is because the free energy of an ion pair's charge–charge interactions usually fails to compensate for the loss of entropy of the side chains and the loss of solvation free energy when the charged groups form an ion pair. This accounts for the observation that ion pairs are poorly conserved among homologous proteins.

Chemical Cross-links

Disulfide bonds (Fig. 4-6) within and between polypeptide chains form as a protein folds to its native conformation. Some polypeptides whose Cys residues have been derivatized to prevent disulfide bond formation can still assume their fully active conformations, suggesting that disulfide bonds are not essential stabilizing forces. They may, however, be important for "locking in" a particular backbone folding pattern as the protein proceeds from its fully extended state to its mature form.

Disulfide bonds are rare in intracellular proteins because the cytoplasm is a reducing environment. Most disulfide bonds occur in proteins that are secreted from the cell into the more oxidizing extracellular environment. The relatively hostile extracellular world (e.g., uncontrolled temperature and pH) apparently requires the additional structural constraints conferred by disulfide bonds.

Metal ions may also function to internally cross-link proteins. For example, at least ten motifs collectively known as **zinc fingers** have been described in nucleic acid–binding proteins. These structures contain about 25–60 residues arranged around one or two Zn^{2+} ions that are tetrahedrally coordinated by the side chains of Cys, His, and occasionally Asp or Glu (Fig. 6-35). The Zn^{2+} allows relatively short stretches of polypeptide chain to fold into stable units that can interact with nucleic acids. Zinc fingers are too small to be stable in the absence of Zn^{2+}. Zinc is ideally suited to its structural role in intracellular proteins: Its filled *d* electron shell permits it to interact strongly with a variety of ligands (e.g., sulfur, nitrogen, or oxygen) from different amino acid residues. In addition, zinc has only one stable oxidation state (unlike, for example, copper and iron), so it does not undergo oxidation–reduction reactions in the cell.

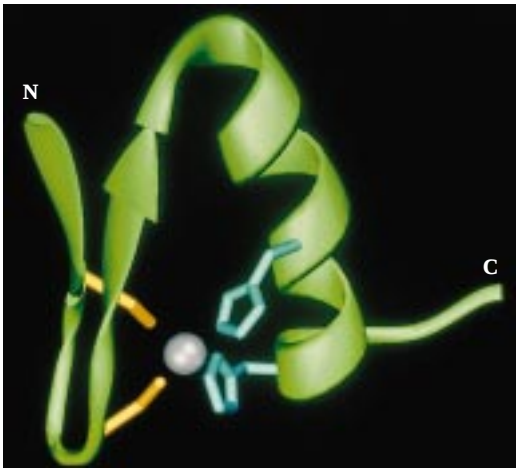


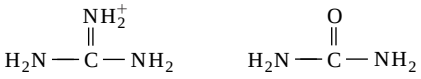
Figure 6-35. **A zinc finger motif.** This structure, from the DNA-binding protein Zif268, is known as a Cys_2–His_2 zinc finger because the zinc atom (*silver*) is coordinated by two Cys residues (*yellow*) and two His residues (*cyan*). [Based on an X-ray structure by Carl Pabo, MIT.]

B. Protein Denaturation and Renaturation

The low conformational stabilities of native proteins make them easily susceptible to denaturation by altering the balance of the weak nonbonding forces that maintain the native conformation. Proteins can be denatured by a variety of conditions and substances:

1. Heating causes a protein's conformationally sensitive properties, such as optical rotation (Section 4-2), viscosity, and UV absorption, to change abruptly over a narrow temperature range. Such a sharp transition indicates that the entire polypeptide unfolds or "melts" **cooperatively,** that is, nearly simultaneously. Most proteins have melting points well below 100°C. Among the exceptions are the proteins of thermophilic bacteria, organisms that inhabit hot springs or submarine volcanic vents with temperatures near 100°C. Amazingly, the X-ray structures of these heat-stable proteins are only subtly different from those of their low-temperature homologs.
2. pH variations alter the ionization states of amino acid side chains, thereby changing protein charge distributions and hydrogen bonding requirements.

3. Detergents associate with the nonpolar residues of a protein, thereby interfering with the hydrophobic interactions responsible for the protein's native structure.
4. The **chaotropic agents** guanidinium ion and urea,



Guanidinium ion **Urea**

in concentrations in the range 5 to 10 M, are the most commonly used protein denaturants. Chaotropic agents are ions or small organic molecules that increase the solubility of nonpolar substances in water. Their effectiveness as denaturants stems from their ability to disrupt hydrophobic interactions, although their mechanism of action is not well understood.

Denatured Proteins Can Be Renatured

In 1957, the elegant experiments of Christian Anfinsen on **ribonuclease A** (**RNase A**) showed that proteins can be denatured reversibly. RNase A, a 124-residue single-chain protein, is completely unfolded and its four disulfide bonds reductively cleaved in an 8 M urea solution containing 2-mercaptoethanol (Fig. 6-36). Dialyzing away the urea and reductant and exposing the resulting solution to O_2 at pH 8 (which oxidizes the SH groups to form disulfides) yields a protein that is virtually 100% enzymatically active and physically indistinguishable from native RNase A. The protein must therefore **renature** spontaneously.

The renaturation of RNase A demands that its four disulfide bonds reform. The probability of one of the eight Cys residues randomly forming a disulfide bond with its proper mate among the other seven Cys residues is 1/7; that one of the remaining six Cys residues then randomly forming its proper disulfide bond is 1/5; etc. The overall probability of RNase A reforming its four native disulfide links at random is

$$\frac{1}{7} \times \frac{1}{5} \times \frac{1}{3} \times \frac{1}{1} = \frac{1}{105}$$

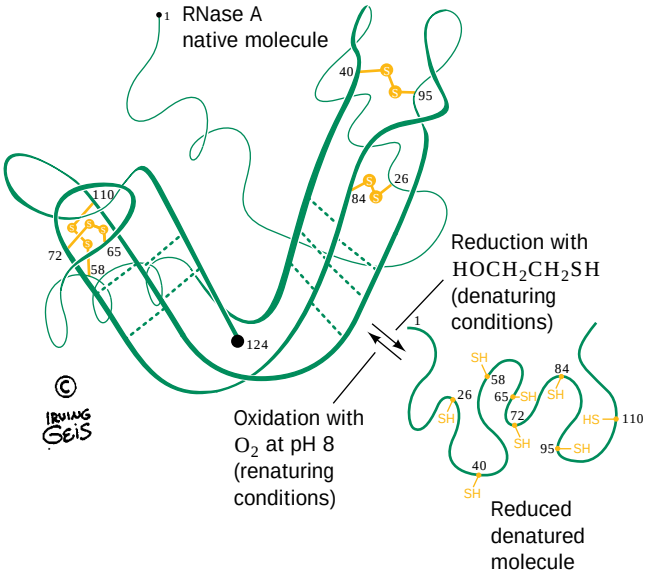


Figure 6-36. **The reductive denaturation and oxidative renaturation of RNase A.** [Figure copyrighted © by Irving Geis.]

Clearly, the disulfide bonds do not randomly re-form under renaturing conditions, since, if they did, only 1% of the refolded protein would be catalytically active. Indeed, if the RNase A is reoxidized in 8 M urea so that its disulfide bonds re-form while the polypeptide chain is a random coil, then after removal of the urea, the RNase A is, as expected, only ~1% active. This "scrambled" protein can be made fully active by exposing it to a trace of 2-mercaptoethanol, which breaks the improper disulfide bonds and allows the proper bonds to form. *Anfinsen's work demonstrated that proteins can fold spontaneously into their native conformations under physiological conditions. This implies that a protein's primary structure dictates its three-dimensional structure.*

C. Protein Folding Pathways

Studies of protein stability and renaturation suggest that protein folding is directed largely by the residues that occupy the interior of the folded protein. But *how* does a protein fold to its native conformation? One might guess that this process occurs through the protein's random exploration of all the conformations available to it until it eventually stumbles onto the correct one. A simple calculation first made by Cyrus Levinthal, however, convincingly demonstrates that this cannot possibly be the case: Assume that an n-residue protein's 2^n torsion angles, ϕ and ψ, each have three stable conformations. This yields $3^{2n} \approx 10^n$ possible conformations for the protein (a gross underestimate because we have completely neglected its side chains). Then, if the protein could explore a new conformation every 10^{-13} s (the rate at which single bonds reorient), the time t, in seconds, required for the protein to explore all the conformations available to it is

$$t = \frac{10^n}{10^{13}}$$

For a small protein of 100 residues, $t = 10^{87}$ s, which is immensely greater than the apparent age of the universe (20 billion years, or 6×10^{17} s).

In fact, many proteins fold to their native conformations in less than a few seconds. This is because *proteins fold to their native conformations via directed pathways rather than stumbling on them through random conformational searches.* Thus, as a protein folds, its conformational stability increases sharply (i.e., its free energy decreases sharply), which makes folding a one-way process. A hypothetical folding pathway is diagrammed in Fig. 6-37.

Experimental observations indicate that protein folding begins with the formation of local segments of secondary structure (α helices and β sheets). This early stage of protein folding is extremely rapid, with much of the native secondary structure in small proteins appearing within 5 ms of the initiation of folding. Since native proteins contain compact hydrophobic cores, it is likely that the driving force in protein folding is what has been termed a **hydrophobic collapse.** The collapsed state is known as a **molten globule,** a species that has much of the secondary structure of the native protein but little of its tertiary structure. Over the next 5 to 1000 ms, the secondary structure becomes stabilized and tertiary structure begins to form. During this intermediate stage, the nativelike elements are thought to take the form of subdomains that are not yet properly docked to form domains. In the final stage of folding, which for small single-domain proteins occurs over the next few seconds, the protein undergoes a series of complex motions in which it attains its relatively rigid internal side chain packing and hydrogen bonding while it expels the remaining water molecules from its hydrophobic core.

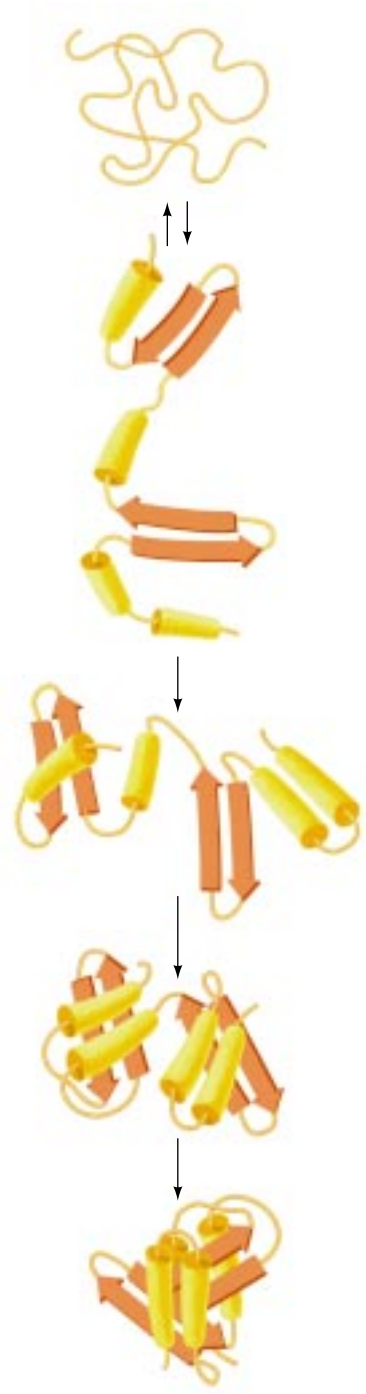


Figure 6-37. **Hypothetical protein folding pathway.** This example shows a linear pathway for folding a two-domain protein. [After Goldberg, M.E., *Trends Biochem. Sci.* 10, 389 (1985).]

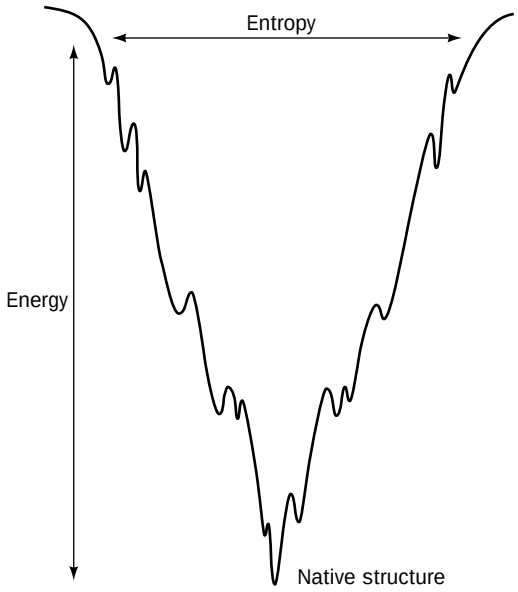


Figure 6-38. **Energy–entropy diagram for protein folding.** The width of the diagram represents entropy, and the depth, the energy. The unfolded polypeptide proceeds from a high-entropy, disordered state (*wide*) to a single low-entropy (*narrow*), low-energy native conformation. [After Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z., and Socci, N.D., *Proc. Natl. Acad. Sci.* 92, 3626 (1995).]

In multidomain and multisubunit proteins, the respective units then assemble in a similar manner, with a few slight conformational adjustments required to produce the protein's native tertiary or quaternary structure. Thus, *proteins appear to fold in a hierarchical manner, with small local elements of structure forming and then coalescing to yield larger elements, which coalesce with other such elements to form yet larger elements, etc.*

Folding, like denaturation, appears to be a cooperative process, with small elements of structure accelerating the formation of additional structures. A folding protein must proceed from a high-energy, high-entropy state to a low-energy, low-entropy state. This energy–entropy relationship is diagrammed in Fig. 6-38. An unfolded polypeptide has many possible conformations (high entropy). As it folds into an ever-decreasing number of possible conformations, its entropy and free energy decrease. The energy–entropy diagram is not a smooth valley but a jagged landscape. Minor clefts and gullies represent conformations that are temporarily trapped until, through random thermal activation, they overcome a slight "uphill" free energy barrier and can then proceed to a lower energy conformation. Evidently, *proteins have evolved to have efficient folding pathways as well as stable native conformations.* Nevertheless, misfolded proteins do occur in nature, and their accumulation is believed to be the cause of a variety of neurological diseases (see Box 6-4).

Protein Disulfide Isomerase

Even under optimal experimental conditions, proteins often fold more slowly *in vitro* than they fold *in vivo*. One reason is that folding proteins often form disulfide bonds not present in the native proteins, which then slowly form native disulfide bonds through the process of disulfide interchange. **Protein disulfide isomerase** (**PDI**) catalyzes this process. Indeed, the observation that RNase A folds so much faster *in vivo* than *in vitro* led Anfinsen to discover this enzyme.

PDI binds to a wide variety of unfolded polypeptides via a hydrophobic patch on its surface. A Cys —SH group on PDI reacts with a disulfide group on the polypeptide to form a mixed disulfide and a Cys —SH group on the polypeptide (Fig. 6-39*a*). Another disulfide group on the polypeptide, brought into proximity by the spontaneous folding of the polypeptide, is attacked by this Cys —SH group. The newly liberated Cys —SH group then repeats this process with another disulfide bond, and so on, ultimately yielding the polypeptide containing only native disulfide bonds, along with regenerated PDI.

Oxidized (disulfide-containing) PDI also catalyzes the initial formation of a polypeptide's disulfide bonds by a similar mechanism (Fig. 6-39*b*). In this case, the reduced PDI reaction product must be reoxidized by cellular oxidizing agents in order to repeat the process.

Molecular Chaperones

Proteins begin to fold as they are being synthesized, so the renaturation of a denatured protein *in vitro* may not mimic the folding of a protein *in vivo*. In addition, proteins fold *in vivo* in the presence of extremely high concentrations of other proteins with which they can potentially interact. ***Molecular chaperones*** *are essential proteins that bind to unfolded and partially folded polypeptide chains to prevent the improper association of exposed hydrophobic segments that might lead to non-native folding as well as polypeptide aggregation and precipitation.* This is especially important for multidomain and multisubunit proteins, whose components must fold fully before they can properly associate with each other. Molecular chap-

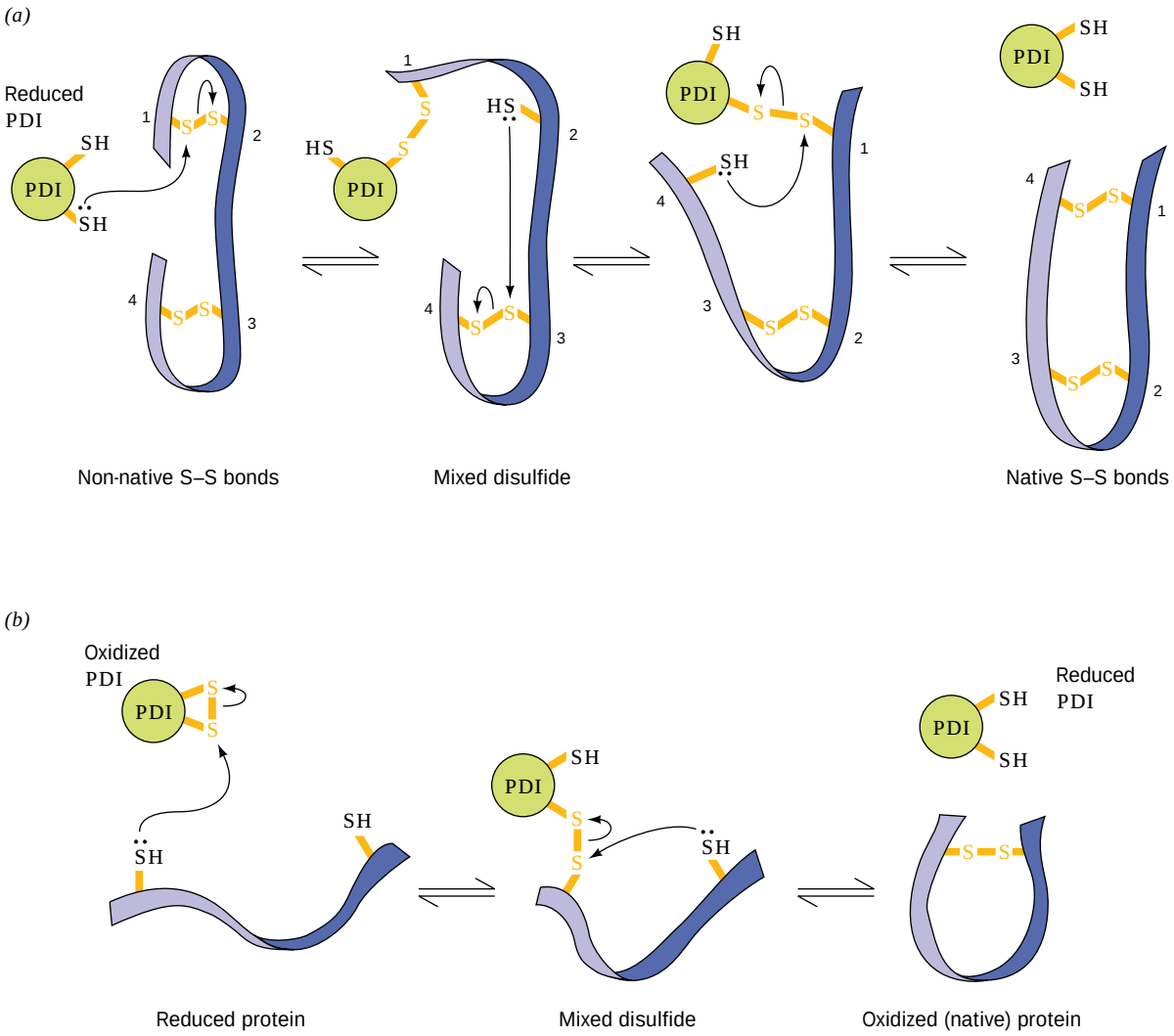


Figure 6-39. **Mechanism of protein disulfide isomerase.** (*a*) Reduced (SH-containing) PDI catalyzes the rearrangement of a polypeptide's non-native disulfide bonds via disulfide interchange reactions to yield native disulfide bonds. (*b*) Oxidized (disulfide-containing) PDI catalyzes the initial formation of a polypeptide's disulfide bonds through the formation of a mixed disulfide. Reduced PDI can then react with a cellular oxidizing agent to regenerate oxidized PDI.

erones also allow misfolded proteins to refold into their native conformations.

Many molecular chaperones were first described as **heat shock proteins** (**Hsp**) because their rate of synthesis is increased at elevated temperatures. Presumably, the additional chaperones are required to recover heat-denatured proteins or to prevent misfolding under conditions of environmental stress. There are two major classes of molecular chaperones in both prokaryotes and eukaryotes: the **Hsp70** family of 70-kD proteins and the **chaperonins,** which are large multisubunit proteins.

An Hsp70 protein binds to a newly synthesized polypeptide, possibly as soon as the first 30 amino acids have been polymerized at the ribosome. The Hsp70 chaperone probably helps prevent premature folding.

Chaperonins consist of two types of proteins:

1. The **Hsp60 proteins** (**GroEL** in *E. coli*), which are composed of 14 identical ~60-kD subunits arranged in two stacked rings of 7 subunits

Box 6-4

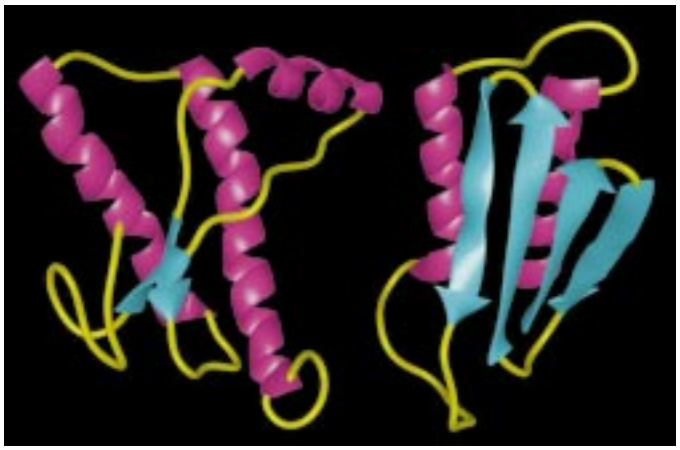
BIOCHEMISTRY IN HEALTH AND DISEASE

Diseases Related to Protein Folding

A number of neurological diseases are characterized by insoluble protein aggregates, called **amyloid deposits**, in brain and other tissues. These diseases, including the most common, **Alzheimer's disease**, result from the precipitation of a single kind of protein. Protein chemists are familiar with partially folded polypeptides that precipitate indiscriminately *in vitro*, probably due to exposure of hydrophobic patches. It is somewhat surprising that *in vivo*, in the presence of high concentrations of many other proteins, partially folded protein molecules aggregate only with other molecules of the same protein. What is even more surprising is that the molecules that precipitate are apparently misfolded versions of molecules that are normally present in the same tissues.

The **β amyloid protein** that forms the fibrous deposits, or **plaques**, in the brains of Alzheimer's patients is a 40-residue segment that is cleaved from a larger precursor protein. Its normal function is not known. Some mutations in the precursor increase β amyloid production. The increased concentration of the peptide apparently increases the likelihood that a misfolded peptide will initiate the aggregation that eventually produces an amyloid plaque. Plaques appear to be the primary cause, not a side effect, of the neurological deterioration in the disease.

Abnormal protein folding is almost certainly a factor in a group of infectious disorders that include **bovine spongiform encephalopathy** ("mad cow disease"), **scrapie** in sheep, and several rare fatal human diseases including **Creutzfeldt–Jakob disease**. According to a hypothesis advanced by Stanley Pruisner, these diseases are caused by a protein known as a **prion**. The prion protein is present in normal brain tissue, where it apparently has a mostly α helical conformation (the structural model on the left, which was derived using structure-prediction techniques).



[Figure adapted from Fred Cohen, University of California at San Francisco.]

In diseased brains, the same protein—now apparently a mixture of β sheet and α helices (model on the right)—forms insoluble fibrous aggregates that damage brain cells. The infectious nature of prions is believed to result from the ability of an abnormally folded prion protein to catalyze the misfolding of normal prion proteins that then aggregate.

each, thereby forming a hollow cylinder with D_7 symmetry (Section 6-3).

2. The **Hsp10 proteins** (**GroES** in *E. coli*), which consist of 7 identical ~10-kD subunits arranged with 7-fold rotational (C_7) symmetry to form a dome-shaped complex.

The X-ray structure of the GroEL–GroES complex determined by Paul Sigler shows, in agreement with previous electron microscopy studies, that one open end of the GroEL cylinder is capped by a GroES complex (Fig. 6-40). The interior of the cylinder provides a protected environment in which a protein can fold without aggregating with other partially folded proteins.

The interior of the GroEL cylinder contains hydrophobic patches that bind the exposed groups of its enclosed and improperly folded protein. The GroEL subunits bind ATP and catalyze its hydrolysis to ADP and inorganic phosphate (P_i), a process that motivates a conformational change that masks the hydrophobic patches. The bound protein is thereby released and stimulated to continue folding. The binding and release, in effect, frees the partially folded protein from its entrapment in a local free energy minimum, such as those in Fig. 6-38, which permits the folding protein to con-

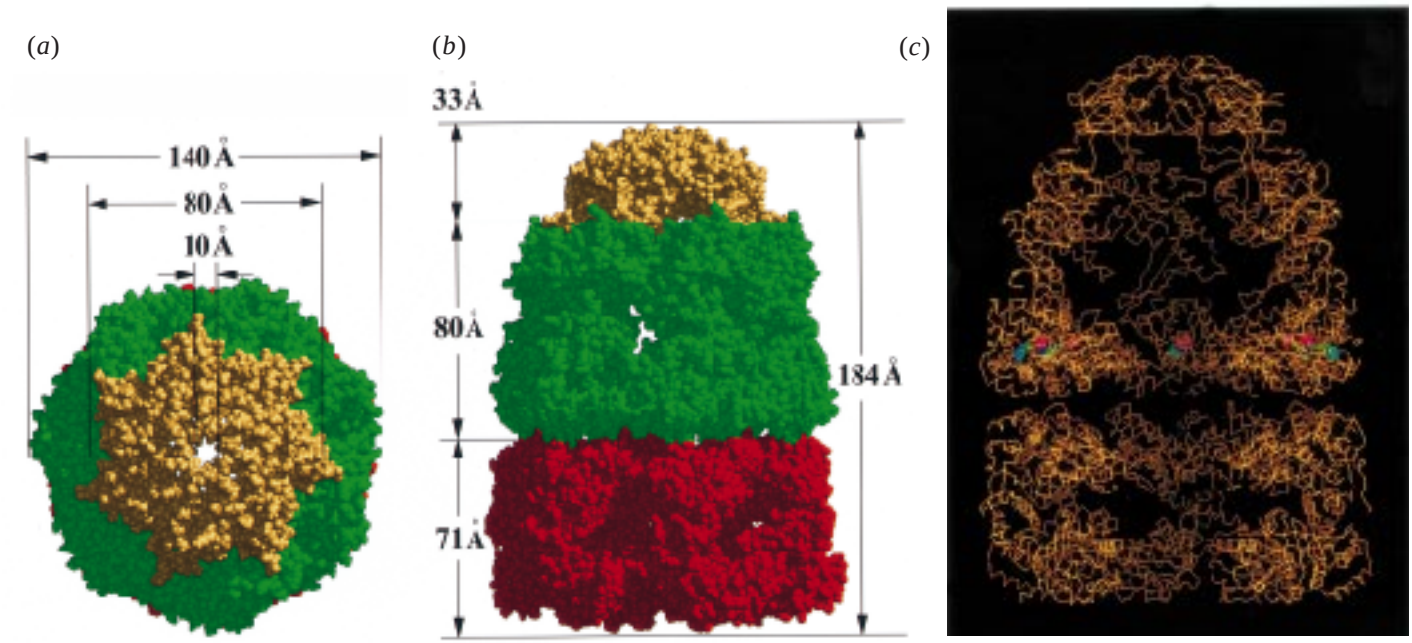


Figure 6-40. **The X-ray structure of the GroEL–GroES–$(ADP)_7$ chaperonin complex.** Two seven-membered rings of GroEL subunits stack to form a hollow cylinder that is capped at one end by a seven-membered ring of GroES subunits. (*a*) A space-filling representation of the complex as viewed down its 7-fold axis of rotation. The GroES ring is gold, its adjacent GroEL ring is green, and the adjoining GroEL ring is red. (*b*) As in *a* but viewed perpendicularly to the 7-fold axis. Note the different conformations of the two GroEL rings. (*c*) The C_α backbones of the complex as viewed perpendicularly to the 7-fold axis and cut away in the plane containing this axis. The ADPs, which are bound to the lower portion of each subunit in the upper GroEL ring, are shown in space-filling form. Note the large cavity formed by the GroES cap and the upper GroEL ring in which a polypeptide can fold in isolation. [Courtesy of Paul Sigler, Yale University.]

tinue its descent down its funnel toward the native state. In the protected environment inside the hollow GroEL–GroES (Hsp60–Hsp10) barrel, a >70-kD protein can fold out of contact with other proteins with which it otherwise might aggregate. *This cycle of ATP-driven binding, release, and refolding is repeated until the protein achieves its native conformation.* The Hsp70 proteins are thought to follow a similar pathway of binding and ATP-driven release of a folding protein.

D. Protein Dynamics

The precision with which protein structures are determined may leave the false impression that proteins have fixed and rigid structures. In fact, *proteins are flexible and rapidly fluctuating molecules whose structural mobilities are functionally significant.* Groups ranging in size from individual side chains to entire domains or subunits may be displaced by up to several angstroms through random intramolecular movements or in response to a trigger such as the binding of a small molecule. Extended side chains, such as Lys, and the N- and C-termini of polypeptide chains are especially prone to wave around in solution because there are few forces holding them in place.

Theoretical calculations by Martin Karplus indicate that a protein's native structure probably consists of a large collection of rapidly interconverting conformations that have essentially equal stabilities (Fig. 6-41). Conformational flexibility, or **breathing,** with structural displacement of up to ~2 Å, allows small molecules to diffuse in and out of the interior of certain proteins.

Figure 6-41. **Molecular dynamics of myoglobin.** Several "snapshots" of the protein calculated at intervals of 5×10^{-12} s are superimposed. The backbone is blue, the heme group is yellow, and the His side chain linking the heme to the protein is orange. [Courtesy of Martin Karplus, Harvard University.]

SUMMARY

1. Four levels of structural complexity are used to describe the three-dimensional shapes of proteins.
2. The conformational flexibility of the peptide group is described by its ϕ and ψ torsion angles.
3. The α helix is a regular secondary structure in which hydrogen bonds form between backbone groups four residues apart. In the β sheet, hydrogen bonds form between the backbones of separate polypeptide strands.
4. Fibrous proteins are characterized by a single type of secondary structure: α keratin is a left-handed coil of two α helices; silk fibroin is an array of stacked β sheets; and collagen is a left-handed triple helix with three residues per turn.
5. Nonrepetitive structures include variations in regular secondary structures, turns, and loops.
6. The tertiary structures of proteins, which can be determined by X-ray crystallography or NMR techniques, may contain motifs (supersecondary structures) and domains.
7. The nonpolar side chains of a globular protein tend to occupy the protein's interior, whereas the polar side chains tend to define its surface.
8. Protein structures can be grouped into families according to their folding patterns. Structural elements are more likely to be evolutionarily conserved than are amino acid sequences.
9. The individual units of multisubunit proteins are usually symmetrically arranged.
10. Native protein structures are only slightly more stable than their denatured forms. The hydrophobic effect is the primary determinant of protein stability. Hydrogen bonding and ion pairing contribute relatively little to a protein's stability.
11. Studies of protein denaturation and renaturation indicate that the primary structure of a protein determines its three-dimensional structure.
12. Proteins fold to their native conformations via directed pathways in which small elements of structure coalesce into larger structures.
13. Protein disulfide isomerase and molecular chaperones facilitate protein folding *in vivo*.
14. Proteins have some conformational flexibility that results in small molecular motions.

REFERENCES

General

Branden, C. and Tooze, J., *Introduction to Protein Structure,* Garland Publishing (1991).

Chothia, C. and Finkelstein, A.V., The classification and origins of protein folding patterns, *Annu. Rev. Biochem.* **59,** 1007–1039 (1990).

Creighton, T.E., *Proteins* (2nd ed.), Chapters 4–6, Freeman (1993).

Darby, N.J. and Creighton, T.E., *Protein Structure,* IRL Press (1993).

Holm, L. and Sander, C., Mapping the protein universe, *Science* **273,** 595–602 (1996). [Discusses the limited numbers of structural domains.]

Kyte, J., *Structure in Protein Chemistry,* Garland Publishing (1995).

Specific Proteins

Baldwin, M.A., Cohen, F.E., and Pruisner, S.B., Prion protein isoforms, a convergence of biological and structural investigations, *J. Biol. Chem.* **270,** 19197–19200 (1995).

Bella, J., Eaton, M., Brodsky, B., and Berman, H.M., Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution, *Science* **266,** 75–81 (1994).

Kaplan, D., Adams, W.W., Farmer, B., and Viney, C., *Silk Polymers,* American Chemical Society (1994).

Prokop, D.J. and Kivirikko, K.I., Collagens: biology, disease, and potentials for therapy, *Annu. Rev. Biochem.* **64,** 403–434 (1995).

Techniques

Karplus, M. and Petsko, G.A., Molecular dynamics simulations in biology, *Nature* **347,** 631–639 (1990).

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

Wagner, G., Hyberts, S.G., and Havel, T.F., NMR structure determination in solution. A critique and comparison with X-ray crystallography, *Annu. Rev. Biophys. Biomol. Struct.* **21,** 167–198 (1992).

Folding and Stabilization

Aurora, R. and Rose, G.D., Helix capping, *Protein Sci.* **7,** 21–38 (1998).

Bardwell, J.C.A. and Beckwith, J., The bonds that tie: catalyzed disulfide bond formation, *Cell* **74,** 769–771 (1993).

Bukau, B. and Horwich, A.L., The Hsp70 and Hsp60 chaperone machines, *Cell* **92,** 351–356 (1998).

Cordes, M.H.J., Davidson, A.R., and Sauer, R.T., Sequence space, folding and protein design, *Curr. Opin. Struct. Biol.* **6,** 3–10 (1996).

Dahiyat, B. and Mayo, S.L., De novo protein design: Fully automated sequence selection, *Science* **278,** 82–87 (1997).

Dill, K.A. and Chan, H.S., From Levinthal pathways to funnels, *Nature Struct. Biol.* **4,** 10–19 (1997). [A highly readable review of modern theories of protein folding.]

Frydman, J. and Hartl, F.U., Principles of chaperone-assisted protein folding: differences between in vitro and in vivo mechanisms, *Science* **272,** 1497–1502 (1996).

Honig, B. and Yang, A.-S., Free energy balance in protein folding, *Adv. Prot. Chem.* **46,** 27–58 (1995).

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

Rost, B. and Sander, C., Bridging the protein-sequence–structure gap by structure predictions, *Annu. Rev. Biophys. Biomol. Struct.* **25,** 113–136 (1996).

Schwabe, J.W.R. and Klug, A., Zinc mining for protein domains, *Nature Struct. Biol.* **1,** 345–349 (1994).

Stickle, D.F., Presta, L.G., Dill, K.A., and Rose, G.D., Hydrogen bonding in globular proteins, *J. Mol. Biol.* **226,** 1143–1159 (1992).

KEY TERMS

secondary structure
tertiary structure
quaternary structure
trans conformation
cis conformation
backbone
torsion (dihedral) angle
ϕ
ψ
Ramachandran diagram
α helix
pitch
parallel β sheet
antiparallel β sheet
topology
fibrous protein
globular protein
coiled coil
denatured
native
β bulge
helix capping
reverse turn (β bend)
Ω loop
X-ray crystallography
NMR
diffraction pattern
contour map
supersecondary structure (motif)
$\beta\alpha\beta$ motif
β hairpin
$\alpha\alpha$ motif
β barrel
dinucleotide-binding (Rossmann) fold
domain
oligomer
protomer
rotational symmetry
cyclic symmetry
dihedral symmetry
hydropathy
ion pair (salt bridge)
zinc finger
renaturation
cooperativity
chaotropic agent
hydrophobic collapse
molten globule
molecular chaperone
heat shock protein
breathing

STUDY EXERCISES

1. Explain why the conformational freedom of peptide bonds is limited.
2. What distinguishes regular and irregular secondary structures?
3. Describe the hydrogen bonding pattern of an α helix.
4. Why are β sheets pleated?
5. What properties do fibrous proteins confer on substances such as hair, horns, bones, and tendons?
6. Why do turns and loops most often occur on the protein surface?
7. Which side chains usually occur in a protein's interior? On its surface?
8. Give some reasons why the number of possible protein structures is much less than the number of amino acid sequences.
9. List the advantages of multiple subunits in proteins.
10. Why can't proteins have mirror symmetry?
11. Describe the forces that stabilize proteins.
12. Describe the energy and entropy changes that occur during protein folding.
13. How does protein renaturation *in vitro* differ from protein folding *in vivo?*

PROBLEMS

1. Draw a cis peptide bond and identify the groups that experience steric interference.
2. Helices can be described by the notation n_m where n is the number of residues per helical turn and m is the number of atoms, including H, in the ring that is closed by the hydrogen bond. (a) What is this notation for the α helix? (b) Is the 3_{10} helix steeper or shallower than the α helix?
3. Calculate the length in angstroms of a 100-residue segment of the α keratin coiled coil.
4. Is it possible for a native protein to be entirely irregular, that is, without α helices, β sheets, or other repetitive secondary structure?
5. (a) Is Trp or Gln more likely to be on a protein's surface? (b) Is Ser or Val less likely to be in the protein's interior? (c) Is Leu or Ile less likely to be found in the middle of an α helix? (d) Is Cys or Ser more likely to be in a β sheet?
6. Describe the structure of glyceraldehyde-3-phosphate dehydrogenase (Fig. 6-30) as a linear sequence of α helix ("α") and strands of β sheet ("β") starting from the N-terminus. The N-terminal domain is red.
7. What types of rotational symmetry are possible for a protein with (a) four or (b) six identical subunits?
8. Given enough time, can all denatured proteins spontaneously renature?

9. Describe the intra- and intermolecular bonds/interactions that are broken or retained when collagen is heated to produce gelatin.
10. Under physiological conditions, polylysine assumes a random coil conformation. Under what conditions might it form an α helix?
11. It is often stated that proteins are quite large compared to the molecules they bind. However, what constitutes a large number depends on your point of view. Calculate the ratio of the volume of a hemoglobin molecule (65 kD) to that of the four O_2 molecules that it binds and the ratio of the volume of a typical office ($4 \times 4 \times 3$ m) to that of the typical (70 kg) office worker that occupies it. Assume that the molecular volumes of hemoglobin and O_2 are in equal proportions to their molecular masses and that the office worker has a density of 1.0 g/cm^3. Compare these ratios. Is this the result you expected?
12. Which of the following polypeptides is most likely to form an α helix? Which is least likely to form a β strand?

    a. CRAGNRKIVLETY

    b. SEDNFGAPKSILW

    c. QKASVEMAVRNSG

    [Problem by Bruce Wightman, Muhlenberg College.]