DONALD VOET
University of Pennsylvania

JUDITH G. VOET
Swarthmore College

BIOCHEMISTRY

SECOND EDITION

JOHN WILEY & SONS, INC.
New York Chichester Brisbane
Toronto Singapore

Acquisitions Editor: *Nedah Rose*
Marketing Manager: *Catherine Faduska*
Designer: *Madelyn Lesure*
Manufacturing Manager: *Susan Stetzer*
Photo Researcher: *Hilary Newman*
Illustration Coordinator: *Edward Starr*
Illustrators: *J/B Woolsey Associates*
Production Management: *Pamela Kennedy Oborski and Suzanne Ingrao*
Cover and Part Opening Illustrations © *Irving Geis*
This book was set in 9.5/12 Times Roman by Progressive Information Technologies and printed and bound by Von Hoffman Press. The cover was printed by Lehigh Press Lithographers. Color separations by Lehigh Press Colortronics.

Cover Art: Two paintings of horse heart cytochrome *c* by Irving Geis in which the protein is illuminated by its single iron atom. On the front cover the hydrophilic side chains are drawn in green, and on the back cover the hydrophobic side chains are drawn in orange. The paintings are based on an X-ray structure by Richard Dickerson.

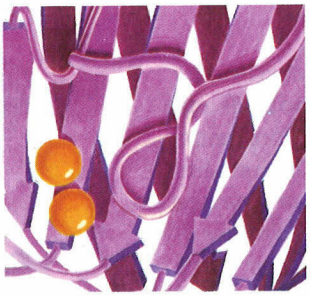
Recognizing the importance of preserving what has been written, it is a policy of John Wiley & Sons, Inc. to have books of enduring value published in the United States printed on acid-free paper, and we exert our best efforts to that end.

ISBN: 0-471-58651-X

Printed in the United States of America

10 9 8 7 6 5 4

CHAPTER



7

Three-Dimensional Structures Of Proteins

1. **Secondary Structure**
   - A. The Peptide Group
   - B. Helical Structures
   - C. Beta Structures
   - D. Nonrepetitive Structures
2. **Fibrous Proteins**
   - A. α Keratin—A Helix of Helices
   - B. Silk Fibroin—A β Pleated Sheet
   - C. Collagen—A Triple Helical Cable
   - D. Elastin—A Nonrepetitive Coil
3. **Globular Proteins**
   - A. Interpretation of Protein X-Ray and NMR Structures
   - B. Tertiary Structure
4. **Protein Stability**
   - A. Electrostatic Forces
   - B. Hydrogen Bonding Forces
   - C. Hydrophobic Forces
   - D. Disulfide Bonds
   - E. Protein Denaturation
5. **Quaternary Structure**
   - A. Subunit Interactions
   - B. Symmetry in Proteins
   - C. Determination of Subunit Composition

Appendix: Viewing Stereo Pictures

The properties of a protein are largely determined by its three-dimensional structure. One might naively suppose that since proteins are composed of the same 20 types of amino acid residues, they would be more or less alike in their properties. Indeed, **denatured** (unfolded) proteins have rather similar characteristics, a kind of homogeneous "average" of their randomly dangling side chains. However, the three-dimensional structure of a **native** (physiologically folded) protein is specified by its primary structure so that it has a unique set of characteristics.

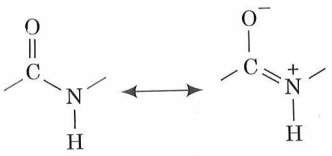
In this chapter, we shall discuss the structural features of proteins, the forces that hold them together, and their hierarchical organization to form complex structures. This will form the basis for understanding the structure–function relationships necessary to comprehend the biochemical roles of proteins. Detailed consideration of the dynamic behavior of proteins and how they fold to their native structures is deferred until Chapter 8.

1. SECONDARY STRUCTURE

A polymer's **secondary structure (2° structure)** is defined as the local conformation of its backbone. For proteins, this has come to mean the specification of regular polypeptide backbone folding patterns: helices, pleated sheets, and turns. However, before we begin our discussion of these basic structural motifs let us consider the geometrical properties of the peptide group because its understanding is prerequisite to that of any structure containing it.

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 structural constraints on the conformations of a polypeptide chain. These studies indicated that *the peptide group has a rigid, planar structure (Fig. 7-1) which, Pauling pointed out, is a consequence of resonance interactions that give the peptide bond an ~40% double-bond character:*



This explanation is supported by the observations that a peptide'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 peptide bond's resonance energy has its maximum value, ~85 kJ · mol^{-1}, when the peptide group is planar because its π-bonding overlap is maximized in this conformation. This overlap, and thus the resonance energy, falls to zero as the peptide bond is twisted to 90° out of planarity, thereby accounting for the planar peptide group's rigidity.

Peptide groups, with few exceptions, assume the trans conformation: that in which successive C_α atoms are on opposite sides of the peptide bond joining them (Fig. 7-1). This is partly a result of steric interference which causes the cis conformation (Fig. 7-2) to be ~8 kJ · mol^{-1} less stable than the trans conformation (this energy difference is somewhat less in peptide bonds followed by a Pro residue; indeed, ~10% of the Pro residues in proteins follow a cis peptide bond, whereas cis peptides are otherwise extremely rare).

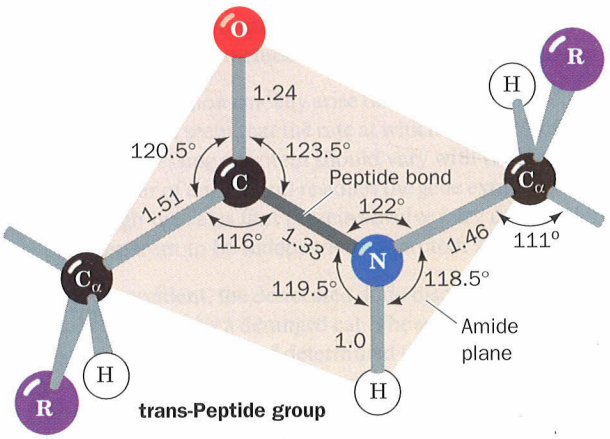


FIGURE 7-1. The standard dimensions (in angstroms, Å, and degrees, °) of the planar trans-peptide group derived by averaging the results of X-ray crystal structure determinations of amino acids and peptides. [After Marsh, R.E. and Donohue, J., *Adv. Protein Chem.* **22,** 249 (1967).]

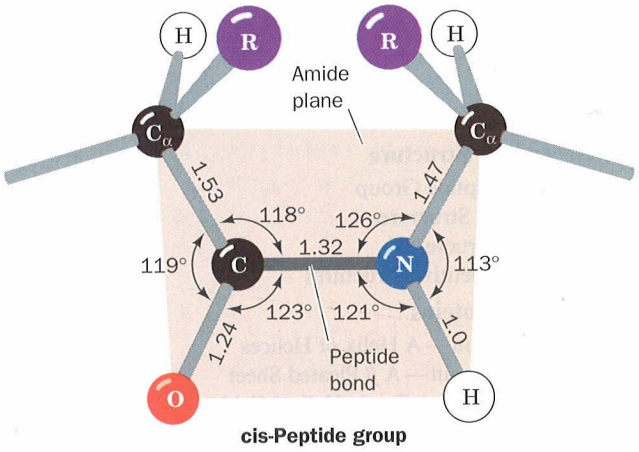


FIGURE 7-2. The cis-peptide group.

Polypeptide Backbone Conformations May Be Described by Their Torsion Angles

The above considerations are important because they indicate that *the backbone of a protein is a linked sequence of rigid planar peptide groups (Fig. 7-3).* We can therefore specify a polypeptide's backbone conformation by the **torsion angles** (rotation angles or **dihedral angles**) about the C_α—N bond (ϕ) and the C_α—C bond (ψ) of each of its amino acid residues. These angles, ϕ and ψ, are both defined as 180° when the polypeptide chain is in its planar, fully extended (all-trans) conformation and increase for a

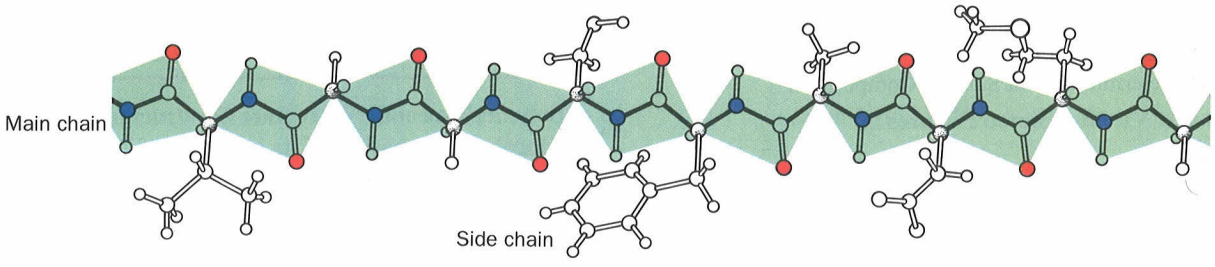


FIGURE 7-3. A polypeptide chain in its fully extended conformation showing the planarity of each of its peptide groups. [Figure copyrighted © by Irving Geis.]

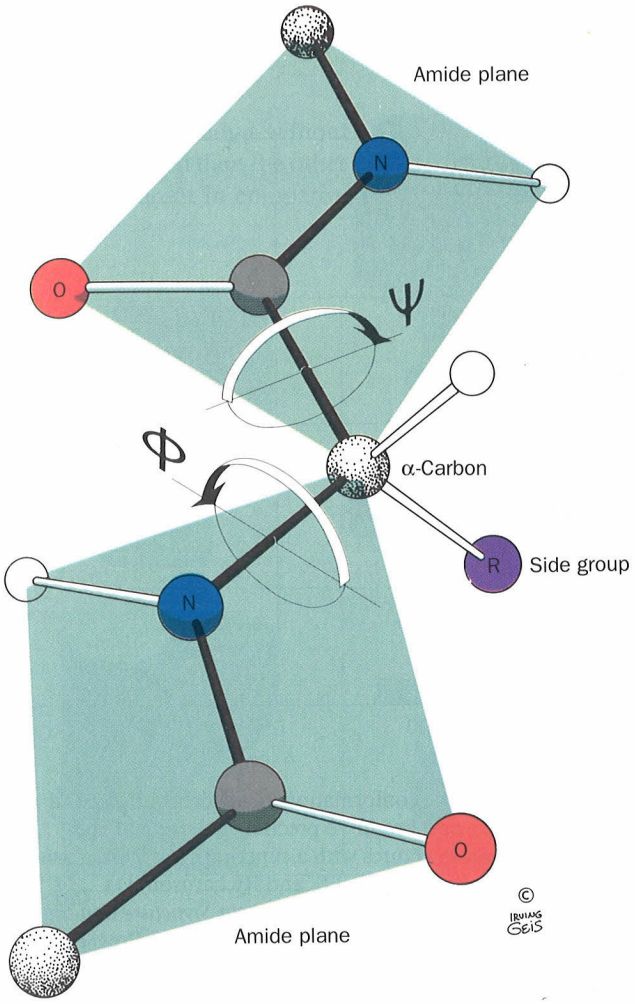
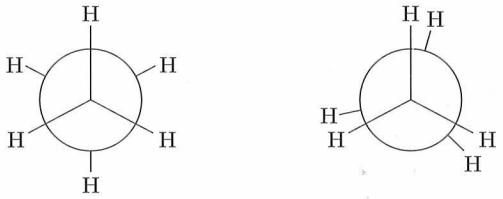


FIGURE 7-4. A portion of a polypeptide chain indicating the torsional degrees of freedom of each peptide unit. The only reasonably free movements are rotations about the C_α—N bond (ϕ) and the C_α—C bond (ψ). The torsion angles are both 180° in the conformation shown and increase, as is indicated, in a clockwise manner when viewed from C_α. [Figure copyrighted © by Irving Geis.]

clockwise rotation when viewed from C_α (Fig. 7-4). (In earlier literature, this fully extended conformation was defined as having $\phi = \psi = 0°$. These values can be made to correspond to the present convention by subtracting 180° from both angles.)

There are several steric constraints on the torsion angles, ϕ and ψ, of a polypeptide backbone that limit its conformational range. The electronic structure of a single (σ) bond, such as a C—C bond, is cylindrically symmetrical about its bond axis so that we might expect such a bond to exhibit free rotation. If this were the case, then in ethane, for example, all torsion angles about the C—C bond would be equally likely. Yet, certain conformations in ethane are favored because of repulsions between electrons in the C—H bonds. The **staggered conformation** (Fig. 7-5*a*), which occurs at a 180° torsion angle, is ethane's most stable arrangement because opposing hydrogen atoms are maximally separated in this position. In contrast, the **eclipsed conformation** (Fig. 7-5*b*), which is characterized by a torsion angle of 0°, is least stable because in this arrangement the hydrogen atoms are closest to each other. The energy difference between the staggered and eclipsed conformations in ethane is ~12 kJ · mol^{-1}, a quantity that represents an **energy barrier** to free rotation about the C—C single bond. Substituents other than hydrogen exhibit greater steric interference; that is, they increase the size of this energy barrier due to their greater bulk. Indeed, with large substituents, some conformations may be sterically forbidden.



(*a*) **Staggered** (*b*) **Eclipsed**

FIGURE 7-5. Newman projections indicating the (*a*) staggered conformation and (*b*) eclipsed conformation of ethane.

Allowed Conformations of Polypeptides Are Indicated by the Ramachandran Diagram

The sterically allowed values of ϕ and ψ can be determined by calculating the distances between the atoms of a tripeptide at all values of ϕ and ψ for the central peptide unit. Sterically forbidden conformations, such as the one shown in Fig. 7-6, are those in which any nonbonding interatomic distance is less than its corresponding van der Waals

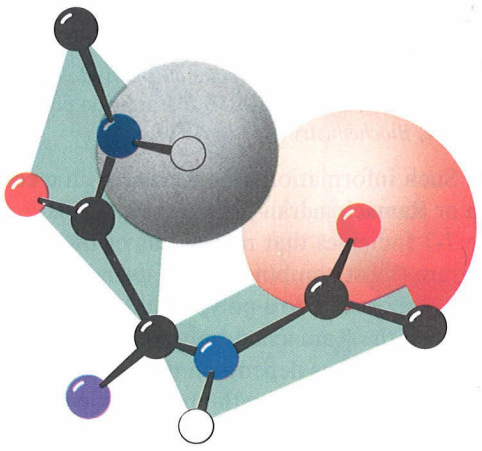


FIGURE 7-6. Steric interference between the carbonyl oxygen and the amide hydrogen on adjacent residues prevents the occurrence of the conformation $\phi = -60°$, $\psi = 30°$. [Figure copyrighted © by Irving Geis.]

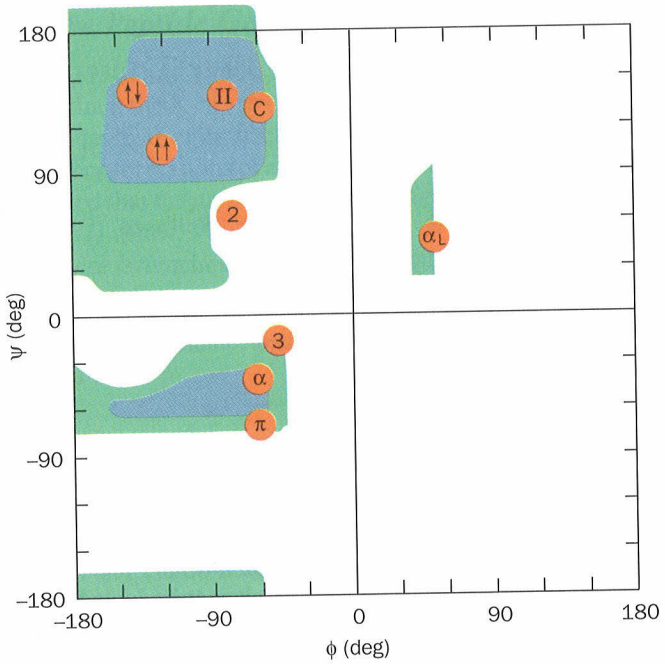


FIGURE 7-7. A Ramachandran diagram (named in honor of its inventor, G.N. Ramachandran) shows the sterically allowed ϕ and ψ angles for poly-L-alanine. The diagram was calculated using the van der Waals distances in Table 7-1. Regions of "normally allowed" ϕ and ψ angles are shaded in blue, whereas green-shaded regions correspond to conformations having "outer limit" van der Waals distances. The conformation angles, ϕ and ψ, of several secondary structures are indicated below:

Secondary Structure	ϕ (deg)	ψ (deg)
Right-handed α helix (α)	−57	−47
Parallel β pleated sheet ($\uparrow\uparrow$)	−119	113
Antiparallel β pleated sheet ($\uparrow\downarrow$)	−139	135
Right-handed 3_{10} helix (3)	−49	−26
Right-handed π helix (π)	−57	−70
2.2_7 ribbon (2)	−78	59
Left-handed polyglycine II and poly-L-proline II helices (II)	−79	150
Collagen (C)	−51	153
Left-handed α helix (α_L)	57	47

[After Flory, P.J., *Statistical Mechanics of Chain Molecules, p.* 253, Interscience (1969); *and* IUPAC–IUB Commission on Biochemical Nomenclature, *Biochemistry* **9**, 3475 (1970).]

distance. Such information is summarized in a **conformation map** or **Ramachandran diagram** (Fig. 7-7).

Figure 7-7 indicates that most areas of the Ramachandran diagram (most combinations of ϕ and ψ) are conformationally inaccessible to a polypeptide chain. The particular regions of the Ramachandran diagram that represent allowed conformations depend on the van der Waals radii chosen to calculate it. But with any realistic set of values, such as that in Table 7-1, *only three small regions of the conformational map are physically accessible to a polypeptide chain.* Nevertheless, as we shall see, all of the common types of regular secondary structures found in proteins fall within allowed regions of the Ramachandran diagram. In-

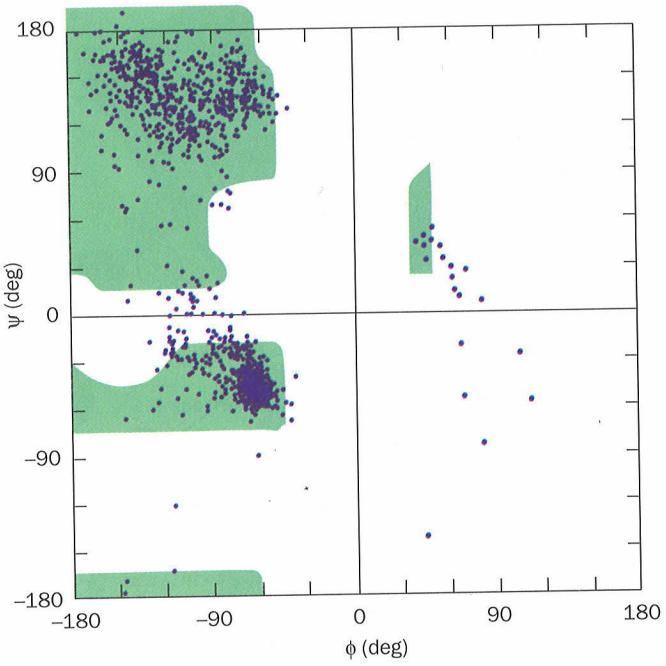


FIGURE 7-8. The conformation angle distribution of all residues but Gly and Pro in 12 precisely determined high-resolution X-ray structures with a superimposed Ramachandran diagram. [After Richardson, J.S. and Richardson, D.C., *in* Fasman, G.D. (Ed.), *Prediction of Protein Structure and the Principles of Protein Conformation, p.* 6, Plenum Press (1989).]

TABLE 7-1. VAN DER WAALS DISTANCES FOR INTERATOMIC CONTACTS

Contact Type	Normally Allowed (Å)	Outer Limit (Å)
H · · · H	2.0	1.9
H · · · O	2.4	2.2
H · · · N	2.4	2.2
H · · · C	2.4	2.2
O · · · O	2.7	2.6
O · · · N	2.7	2.6
O · · · C	2.8	2.7
N · · · N	2.7	2.6
N · · · C	2.9	2.8
C · · · C	3.0	2.9
C · · · CH_2	3.2	3.0
CH_2 · · · CH_2	3.2	3.0

Source: Ramachandra, G.N. and Sasisekharan, V., *Adv. Protein Chem.***23,** 326 (1968).

deed, the observed conformational angles of most non-Gly residues in proteins whose X-ray structures have been determined lie in these allowed regions (Fig. 7-8).

Most points that fall in forbidden regions of Fig. 7-8 lie between its two fully allowed areas near $\psi = 0$. However, these "forbidden" conformations, which arise from the collision of successive amide groups, are allowed if twists of only a few degrees about the peptide bond are permitted. This is not unreasonable since the peptide bond offers little resistance to small deformations from planarity.

Gly, the only residue without a C_β atom, is much less sterically hindered than the other amino acid residues. This is clearly apparent in comparing the Ramachandran diagram for Gly in a polypeptide chain (Fig. 7-9) with that of other residues (Fig. 7-7). In fact, Gly often occupies positions where a polypeptide backbone makes a sharp turn which, with any other residue, would be subject to steric interference.

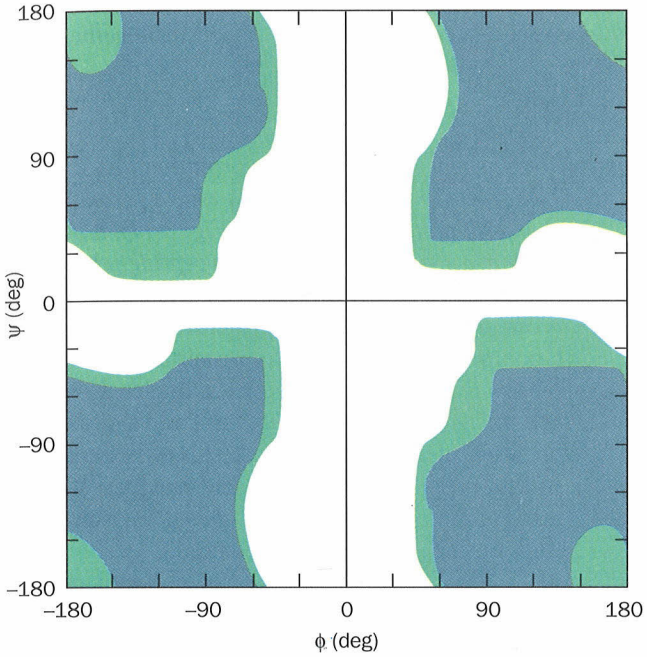


FIGURE 7-9. The Ramachandran diagram of Gly residues in a polypeptide chain. "Normally allowed" regions are shaded in blue, whereas green-shaded regions correspond to "outer limit" atomic distances. Gly residues have far greater conformational freedom than do other (bulkier) amino acid residues as the comparison of this figure with Fig. 7-7 indicates. [After Ramachandran, G.N. and Sasisekharan, V., *Adv. Protein Chem.* **23,** 332 (1968).]

Figure 7-7 was calculated for three consecutive Ala residues. Similar plots for larger residues that are unbranched at C_β, such as Phe, are nearly identical. In Ramachandran diagrams of residues that are branched at C_β, such as Thr, the allowed regions are somewhat smaller than for Ala. The cyclic side chain of Pro limits its ϕ to the range $-60° \pm 25°$, making it, not surprisingly, the most conformationally restricted amino acid residue. The conformations of residues in chains longer than tripeptides are even more restricted than the Ramachandran diagram indicates because, for instance, a polypeptide chain cannot assume a conformation in which it passes through itself. We shall see, however, that despite the great restrictions that peptide bond planarity and side chain bulk place on the conformations of a polypeptide chain, every unique primary structure has a correspondingly unique three-dimensional structure.

B. Helical Structures

Helices are the most striking elements of protein 2° structure. If a polypeptide chain is twisted by the same amount about each of its C_α atoms, it assumes a helical conformation. As an alternative to specifying its ϕ and ψ angles, a helix may be characterized by the number, n, of peptide units per helical turn, and its **pitch,** p, the distance the helix rises along its axis per turn. Several examples of helices are diagrammed in Fig. 7-10. Note that a helix has chirality;

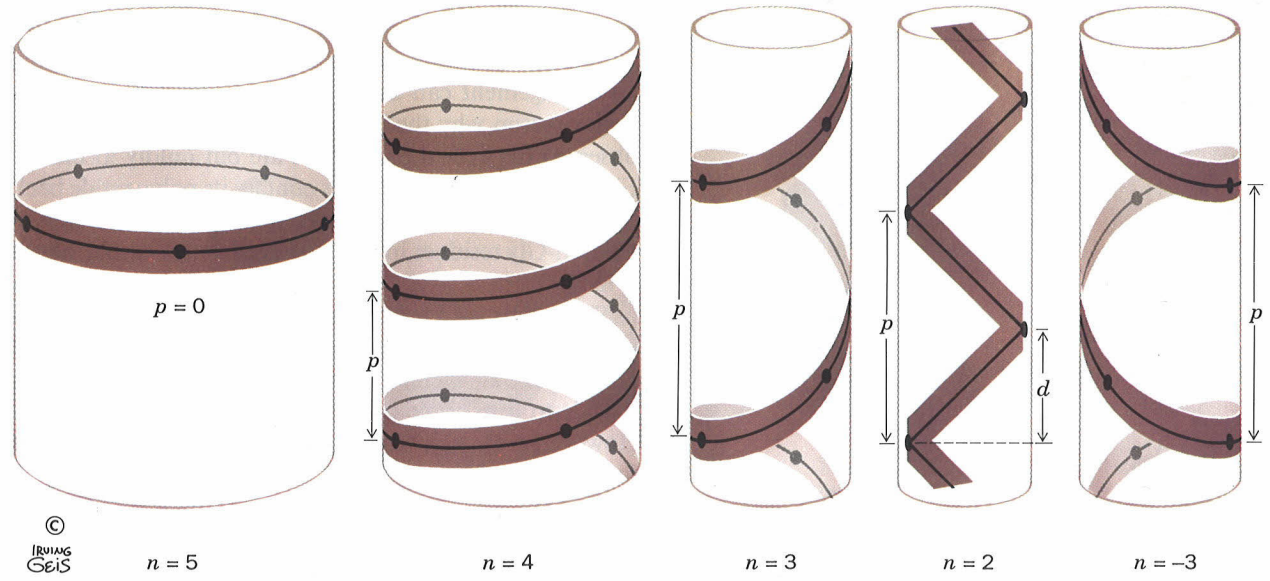


FIGURE 7-10. Examples of helices indicating the definitions of the helical pitch, p, the number of repeating units per turn, n, and the helical rise per repeating unit, $d = p/n$. Right- and left-handed helices are defined, respectively, as having positive and negative values of n. For $n = 2$, the helix degenerates to a nonchiral ribbon. For $p = 0$, the helix degenerates to a closed ring. [Figure copyrighted © by Irving Geis.]

that is, it may be either right handed or left handed (a right-handed helix turns in the direction that the fingers of a right hand curl when its thumb points along the helix axis in the direction that the helix rises). In proteins, moreover, n need not be an integer and, in fact, rarely is.

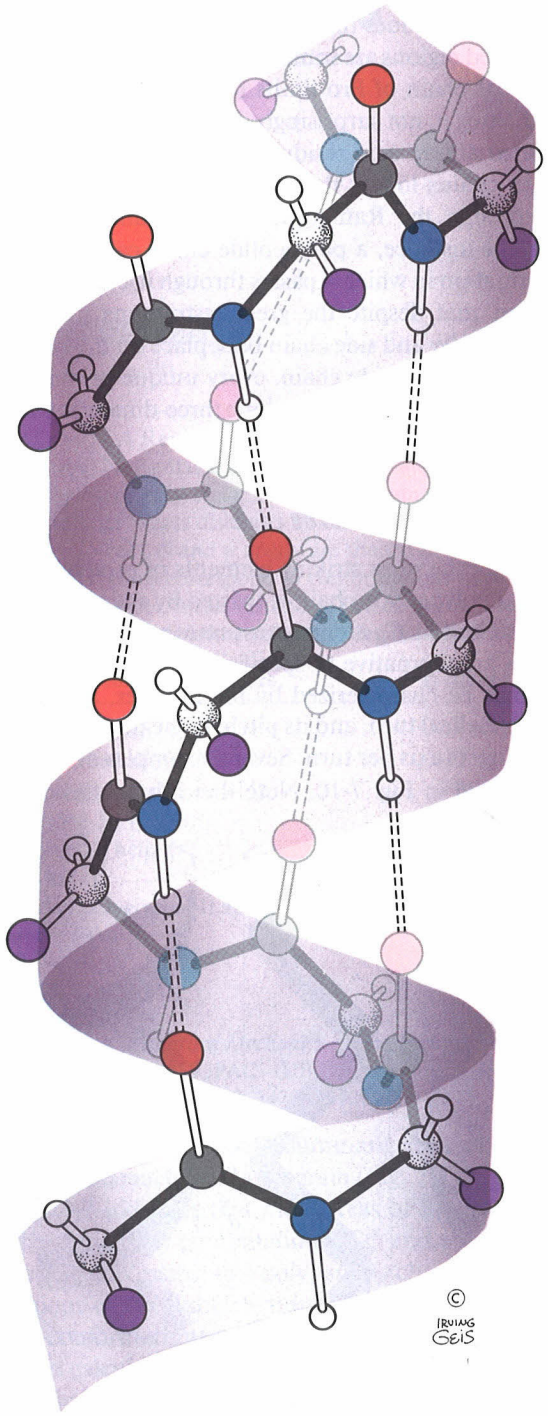


FIGURE 7-11. The right-handed α helix. Hydrogen bonds between the N—H groups and the C=O groups that are four residues back along the polypeptide chain are indicated by dashed lines. [Figure copyrighted © by Irving Geis.]

A polypeptide helix must, of course, have conformation angles that fall within the allowed regions of the Ramachandran diagram. As we have seen, this greatly limits the possibilities. Furthermore, if a particular conformation is to have more than a transient existence, it must be more than just allowed—it must be stabilized. The "glue" that holds polypeptide helices and other 2° structures together is, in part, hydrogen bonds.

The α Helix

Only one helical polypeptide conformation has simultaneously allowed conformation angles and a favorable hydrogen bonding pattern: the ***α helix*** *(Fig. 7-11), a particularly rigid arrangement of the polypeptide chain.* Its discovery through model building, by Pauling in 1951, ranks as one of the landmarks of structural biochemistry.

For a polypeptide made from L-α-amino acid residues, the α helix is right handed with torsion angles $\phi = -57°$ and $\psi = -47°$, $n = 3.6$ residues per turn, and a pitch of 5.4 Å. (An α helix of D-α-amino acid residues is the mirror image of that made from L-amino acid residues: It is left handed with conformation angles $\phi = +57°$ and $\psi = +47°$ but with the same values of n and p.)

Figure 7-11 indicates that the hydrogen bonds of an α helix are arranged such that the peptide C=O bond of the nth residue points along the helix towards 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. In addition, the core of the α helix is tightly packed; that is, its atoms are in van der Waals contact across the helix, thereby maximizing their association energies (Section 7-4A). The R groups, whose positions, as we saw, are not fully dealt with by the Ramachandran diagram, all project backward (downward in Fig. 7-11) and outward from the helix so as to avoid steric interference with the polypeptide backbone and with each other. Such an arrangement can also be seen in Fig. 7-12. Indeed, a major reason why the left-handed α helix has never been observed (its helical parameters are but mildly forbidden; Fig. 7-7) is that its side chains contact its polypeptide backbone too closely. Note, however, that 1 to 2% of the individual non-Gly residues in proteins assume this conformation (Fig. 7-8).

The α helix is a common secondary structural element of both fibrous and globular proteins. In globular proteins, α helices have an average span of ~12 residues, which corresponds to over three helical turns and a length of 18 Å. However, α helices with as many as 53 residues have been found.

Other Polypeptide Helices

Figure 7-13 indicates how hydrogen bonded polypeptide helices may be constructed. The first two, the **2.2_7 ribbon** and the **3_{10} helix,** are described by the notation n_m where n, as before, is the number of residues per helical turn, and m is the number of atoms, including H, in the ring that is

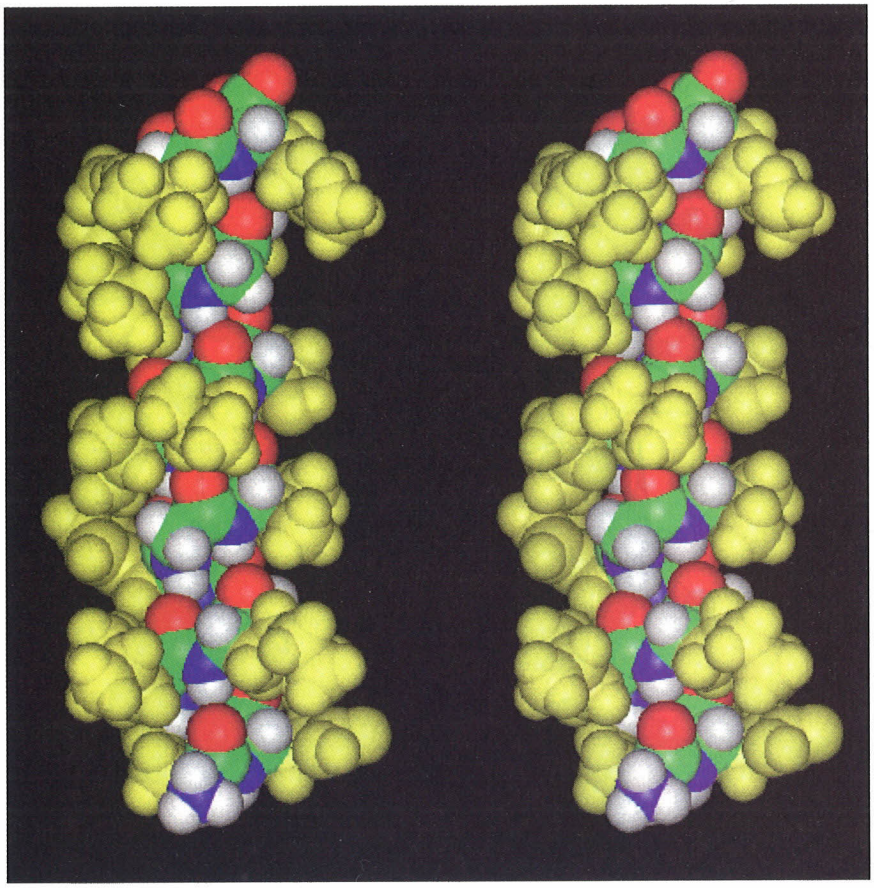


FIGURE 7-12. A stereo, space-filling representation of an α helical segment of sperm whale myoglobin (its E helix) as determined by X-ray crystal structure analysis. In the main chain, carbon atoms are green, nitrogen atoms are blue, oxygen atoms are red, and hydrogen atoms are white. The side chains are yellow. Instructions for viewing stereo diagrams are given in the appendix to this chapter.

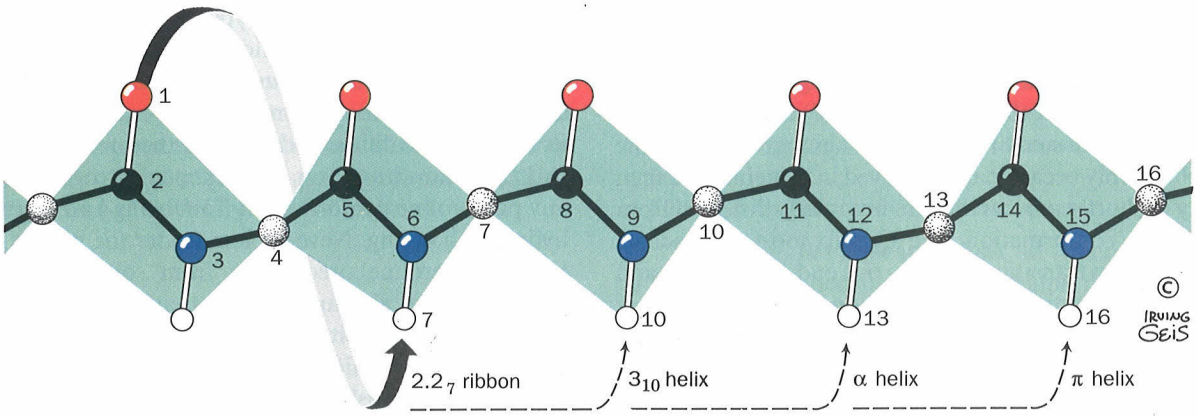


FIGURE 7-13. The hydrogen bonding pattern of several polypeptide helices. In the cases shown, the polypeptide chain curls around such that the C=O group on residue n forms a hydrogen bond with the N—H groups on residues $(n+2)$, $(n+3)$, $(n+4)$, or $(n+5)$.
[Figure copyrighted © by Irving Geis.]

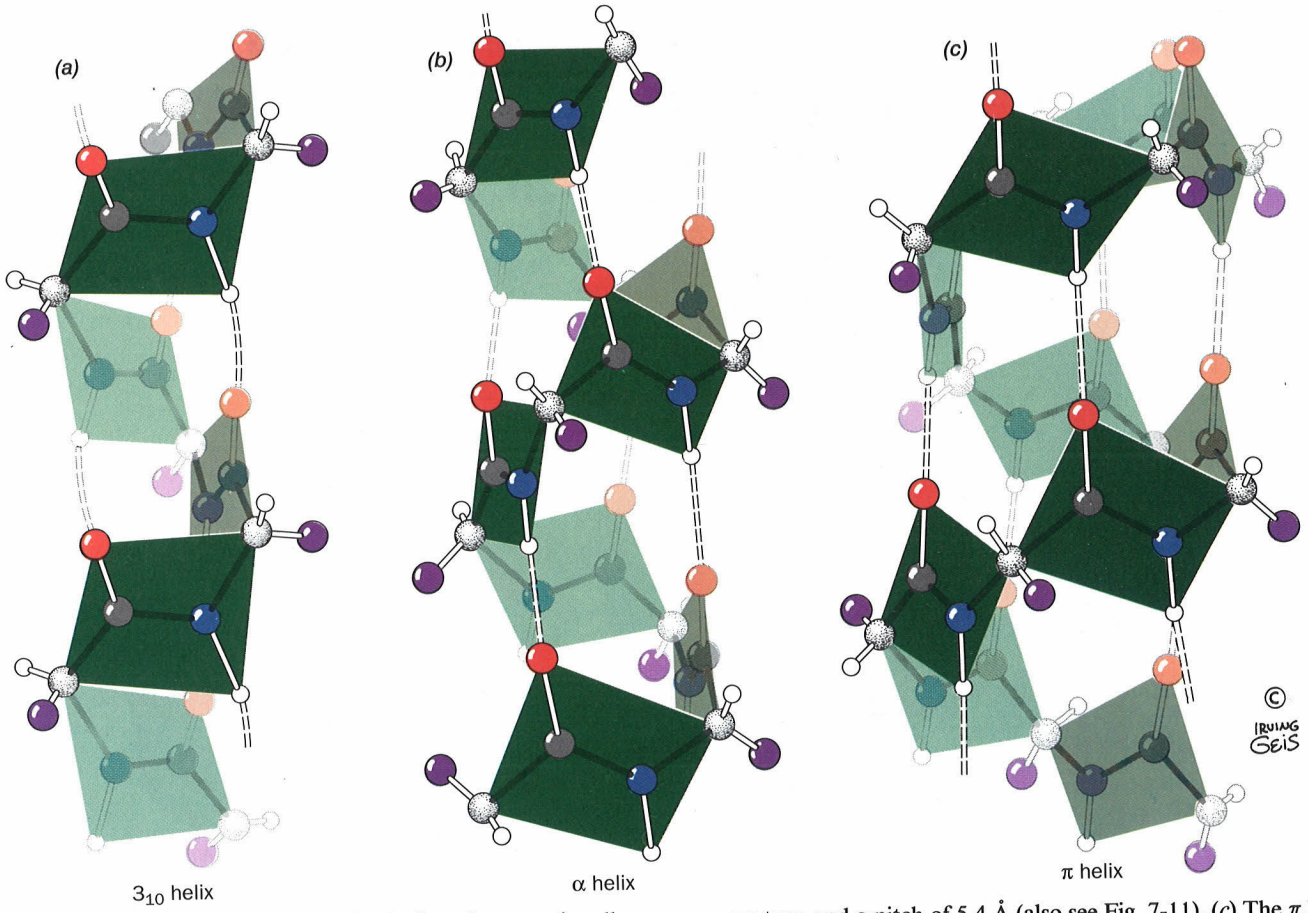


FIGURE 7-14. Two polypeptide helices that occasionally occur in proteins compared with the commonly occurring α helix. (*a*) The 3_{10} helix is characterized by 3.0 peptide units per turn and a pitch of 6.0 Å, which makes it thinner and more elongated than the α helix. (*b*) The α helix has 3.6 peptide units per turn and a pitch of 5.4 Å (also see Fig. 7-11). (*c*) The π helix, with 4.4 peptide units per turn and a pitch of 5.2 Å, is wider and shorter than the α helix. The peptide planes are indicated. [Figure copyrighted © by Irving Geis.]

closed by the hydrogen bond. With this notation, an α helix is a 3.6_{13} helix.

The right-handed 3_{10} helix (Fig. 7-14*a*), which has a pitch of 6.0 Å, is thinner and rises more steeply than does the α helix (Fig. 7-14*b*). Its torsion angles place it in a mildly forbidden zone of the Ramachandran diagram that is rather near the position of the α helix (Fig. 7-7) and its R groups experience some steric interference. This explains why the 3_{10} helix is only occasionally observed in proteins, and then mostly in short segments that are frequently distorted from the ideal 3_{10} conformation. The 3_{10} helix most often occurs as a single-turn transition between one end of an α helix and the adjoining portion of a polypeptide chain.

The **π helix** (4.4_{16} helix), which also has a mildly forbidden conformation (Fig. 7-7), has only rarely been observed and then only as segments of longer helices. This is probably because its comparatively wide and flat conformation (Fig. 7-14*c*) results in an axial hole that is too small to admit water molecules but yet too wide to allow van der Waals associations across the helix axis; this greatly reduces its stability relative to more closely packed conformations. The 2.2_7 ribbon, which as Fig. 7-7 indicates, has strongly forbidden conformation angles, has never been observed.

Certain synthetic homopolypeptides assume conformations that are models for helices in particular proteins. **Polyproline** is unable to assume any common secondary structure due to the conformational constraints imposed by its cyclic pyrrolidine side chains. Furthermore, the lack of a hydrogen substituent on its backbone nitrogen precludes any polyproline conformation from being knit together by hydrogen bonding. Nevertheless, under the proper conditions, polyproline precipitates from solution as a left-handed helix of all-trans peptides that has 3.0 residues per helical turn and a pitch of 9.4 Å (Fig. 7-15). This rather extended conformation permits the Pro side chains to avoid each other. Curiously, **polyglycine,** the least conformationally constrained polypeptide, precipitates from solution as a helix whose parameters are essentially identical to those of polyproline, the most conformationally constrained polypeptide (although the polyglycine helix may be either right

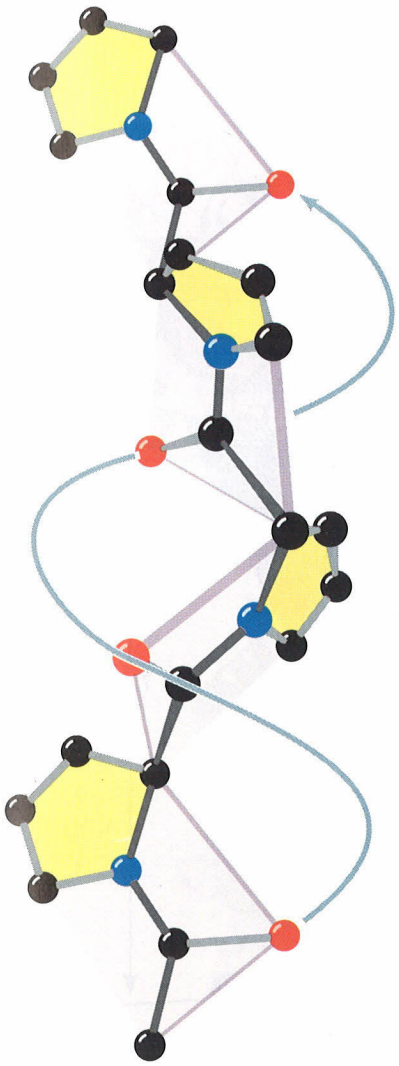


FIGURE 7-15. The **polyproline II** helix. Polyglycine forms a nearly identical helix **(polyglycine II).** [Figure copyrighted © by Irving Geis.]

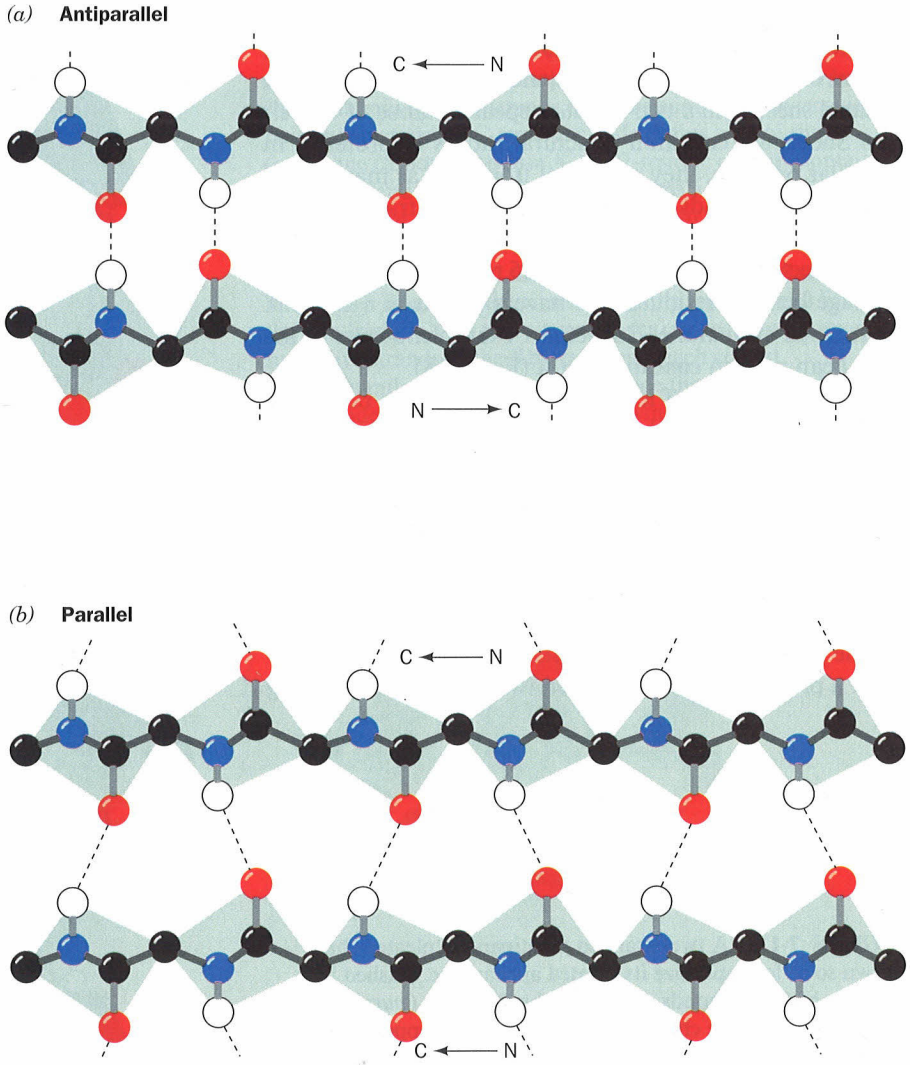


FIGURE 7-16. The hydrogen bonding associations in β pleated sheets. Side chains are omitted for clarity. (*a*) The antiparallel β pleated sheet. (*b*) The parallel β pleated sheet. [Figure copyrighted © by Irving Geis.]

or left handed because Gly is nonchiral). The structures of the polyglycine and polyproline helices are of biological significance because they form the basic structural motif of collagen, a structural protein that contains a remarkably high proportion of both Gly and Pro (Section 7-2C).

C. *Beta Structures*

In 1951, the year that they proposed the α helix, Pauling and Corey also postulated the existence of a different polypeptide secondary structure, the **β pleated sheet.** As with the α helix, the β pleated sheet's conformation has repeating ϕ and ψ angles that fall in the allowed region of the Ramachandran diagram (Fig. 7-7) and utilizes the full hydrogen bonding capacity of the polypeptide backbone. *In β pleated sheets, however, hydrogen bonding occurs between neighboring polypeptide chains* rather than within one as in α helices.

β Pleated sheets come in two varieties:

1. The **antiparallel β pleated sheet,** in which neighboring hydrogen bonded polypeptide chains run in opposite directions (Fig. 7-16*a*).
2. The **parallel β pleated sheet,** in which the hydrogen bonded chains extend in the same direction (Fig. 7-16*b*).

The conformations in which these **β structures** are optimally hydrogen bonded vary somewhat from that of a fully extended polypeptide ($\phi = \psi = \pm 180°$) as indicated in Fig. 7-7. They therefore have a rippled or pleated edge-on

appearance (Fig. 7-17), which accounts for the appellation "pleated sheet." In this conformation, successive side chains of a polypeptide chain extend to opposite sides of the pleated sheet with a two-residue repeat distance of 7.0 Å.

β Sheets are common structural motifs in proteins. In globular proteins, they consist of from 2 to as many as 15 polypeptide strands, the average being 6 strands, which have an aggregate width of ~25 Å. The polypeptide chains in a β sheet are known to be up to 15 residues long, with the average being 6 residues that have a length of ~21 Å. A 6-stranded antiparallel β sheet, for example, occurs in the jack bean protein **concanavalin A** (Fig. 7-18).

Parallel β sheets of less 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 in comparison to those of the antiparallel sheets (Fig. 7-16). Mixed parallel–antiparallel β sheets are common but occur with only ~40% of the frequency that would be expected for the random mixing of strand directions.

The β pleated sheets in globular proteins invariably exhibit a pronounced right-handed twist when viewed along

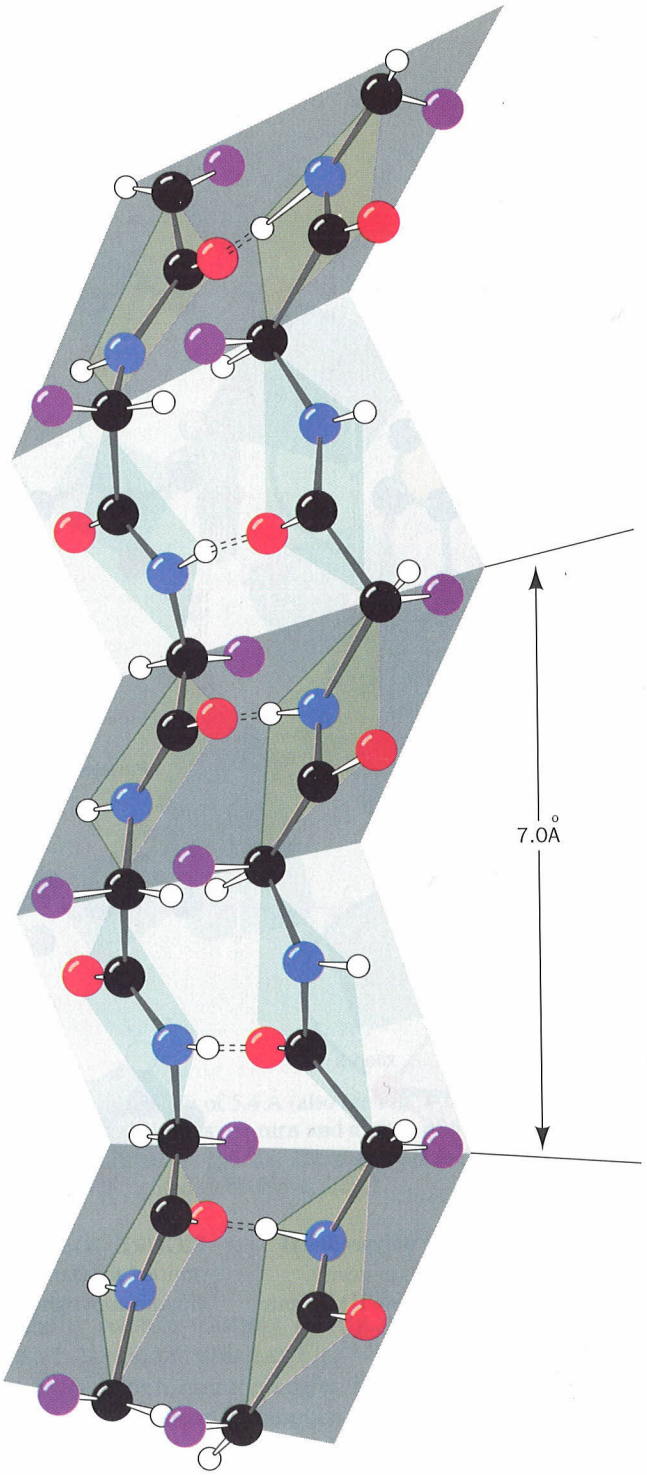


FIGURE 7-17. A two-stranded β antiparallel pleated sheet drawn so as to emphasize its pleated appearance. Dashed lines indicate hydrogen bonds. Note that the R groups (*purple balls*) on each polypeptide chain alternately extend to opposite sides of the sheet and that they are in register on adjacent chains. [Figure copyrighted © by Irving Geis.]

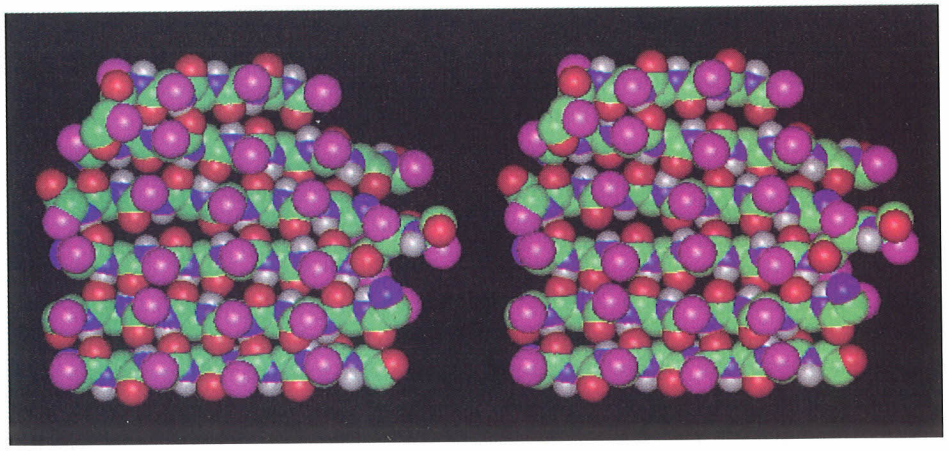


FIGURE 7-18. A stereo, space-filling representation of the six-stranded antiparallel β pleated sheet in jack bean concanavalin A as determined by X-ray crystal structure analysis. In the main chain, carbon atoms are green, nitrogen atoms are blue, oxygen atoms are red, and hydrogen atoms are white. R groups are represented by large purple balls. Instructions for viewing stereo drawings are given in the appendix to this chapter.

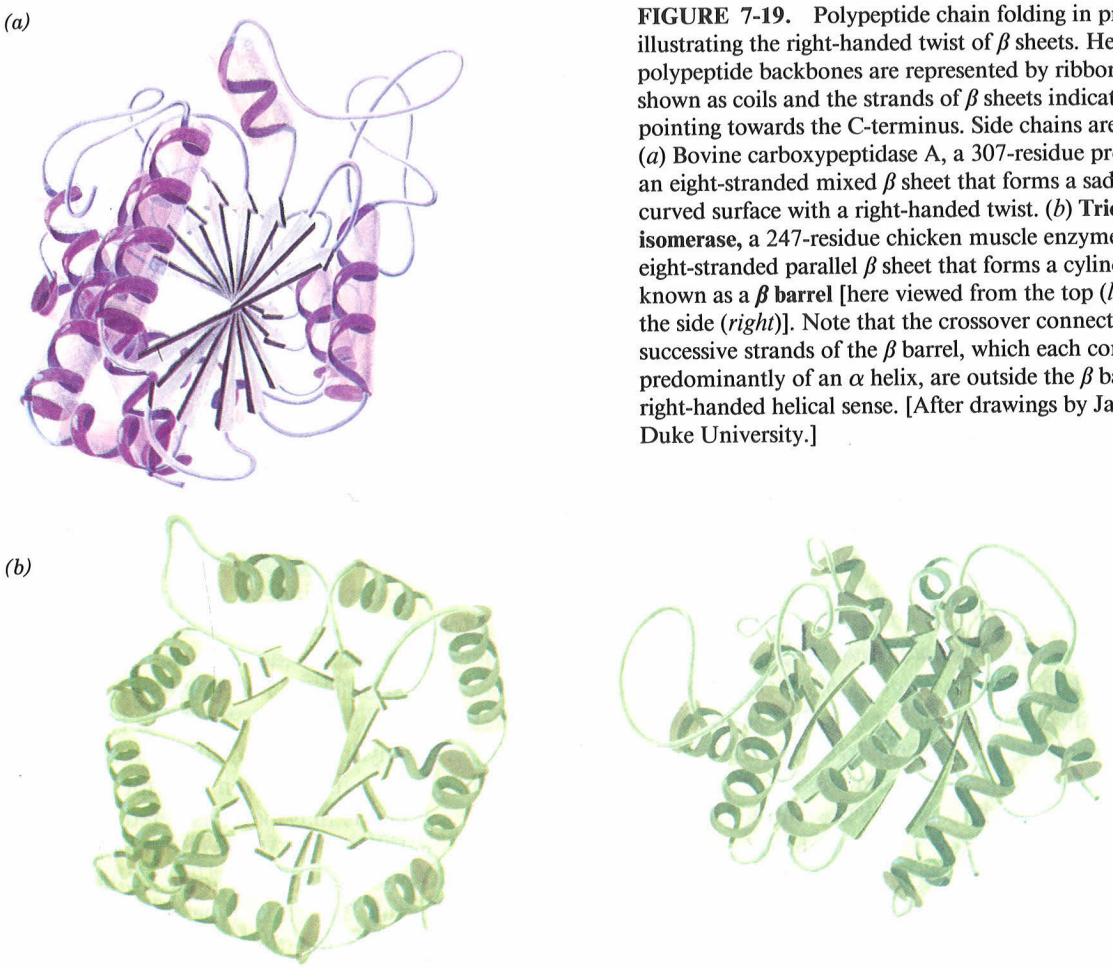


FIGURE 7-19. Polypeptide chain folding in proteins illustrating the right-handed twist of β sheets. Here the polypeptide backbones are represented by ribbons with α helices shown as coils and the strands of β sheets indicated by arrows pointing towards the C-terminus. Side chains are not shown. (*a*) Bovine carboxypeptidase A, a 307-residue protein, contains an eight-stranded mixed β sheet that forms a saddle-shaped curved surface with a right-handed twist. (*b*) **Triose phosphate isomerase,** a 247-residue chicken muscle enzyme, contains an eight-stranded parallel β sheet that forms a cylindrical structure known as a **β barrel** [here viewed from the top (*left*) and from the side (*right*)]. Note that the crossover connections between successive strands of the β barrel, which each consist predominantly of an α helix, are outside the β barrel and have a right-handed helical sense. [After drawings by Jane Richardson, Duke University.]

their polypeptide strands (e.g., Fig. 7-19). Such twisted β sheets are important architectural features of globular proteins since β sheets often form their central cores (Fig. 7-19). Conformational energy calculations indicate that a β sheet's right-handed twist is a consequence of nonbonded interactions between the chiral L-amino acid residues in the sheet's extended polypeptide chains. These interactions tend to give the polypeptide chains a slight right-handed helical twist (Fig. 7-19) which distorts and hence weakens the β sheet's interchain hydrogen bonds. A particular β sheet's geometry is thus the result of a compromise between optimizing the conformational energies of its polypeptide chains and preserving its hydrogen bonds.

The **topology** (connectivity) of the polypeptide strands in a β sheet can be quite complex; the connecting links of these assemblies often consist of long runs of polypeptide chain which frequently contain helices (e.g., Fig. 7-19). The link connecting two consecutive antiparallel strands is topologically equivalent to a simple hairpin turn (Fig. 7-20*a*). However, tandem parallel strands must be linked by a cross-over connection that is out of the plane of the β sheet. Such crossover connections almost always have a right-handed

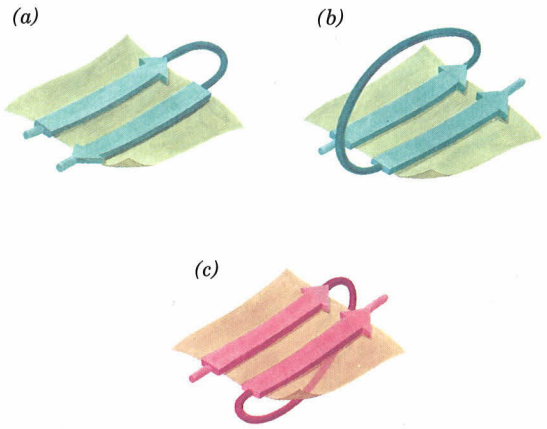


FIGURE 7-20. The connections between adjacent polypeptide strands in β pleated sheets: (*a*) The hairpin connection between antiparallel strands is topologically in the plane of the sheet. (*b*) A right-handed crossover connection between successive strands of a parallel β sheet. Nearly all such crossover connections in proteins have this chirality (see, e.g., Fig. 7-19*b*). (*c*) A left-handed cross-over connection between parallel β sheet strands. Connections with this chirality are rare. [After Richardson, J.S., *Adv. Protein Chem.* **34,** 290, 295 (1981).]

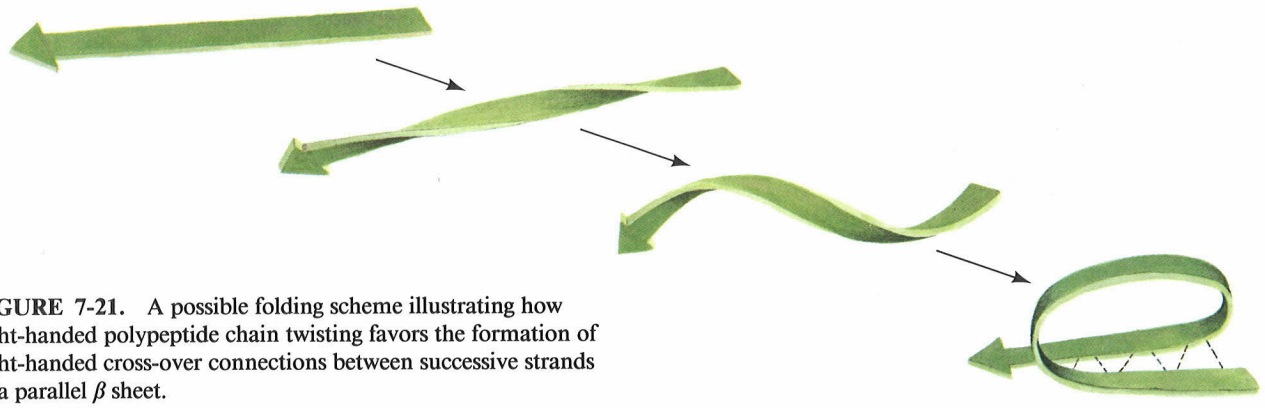


FIGURE 7-21. A possible folding scheme illustrating how right-handed polypeptide chain twisting favors the formation of right-handed cross-over connections between successive strands of a parallel β sheet.

helical sense (Fig. 7-20*b*), which is thought to better fit the β sheets' inherent right-handed twist (Fig. 7-21).

D. Nonrepetitive Structures

Regular secondary structures—helices and β sheets—comprise around half of the average globular protein. The protein's remaining polypeptide segments are said to have a **coil** or **loop conformation.** That is not to say, however, that these nonrepetitive secondary structures are any less ordered than are helices or β sheets; they are simply irregular and hence more difficult to describe. You should therefore not confuse the term coil conformation with the term **random coil,** which refers to the totally disordered and rapidly fluctuating set of conformations assumed by denatured proteins and other polymers in solution.

Globular proteins consist largely of approximately straight runs of secondary structure 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; indeed, they partially define these 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. 7-22). Both types of β bends are stabilized by a hydrogen bond although deviations from these ideal conformations often disrupt this hydrogen bond. Type I β bends may be considered to be distorted sections of 3_{10} helix. In Type II β bends, 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 β bend is often Pro since it can facilely assume the required conformation.

Almost all proteins of >60 residues contain one or more loops of 6 to 16 residues that are not components of helices or β sheets and whose end-to-end distances are <10 Å. Such

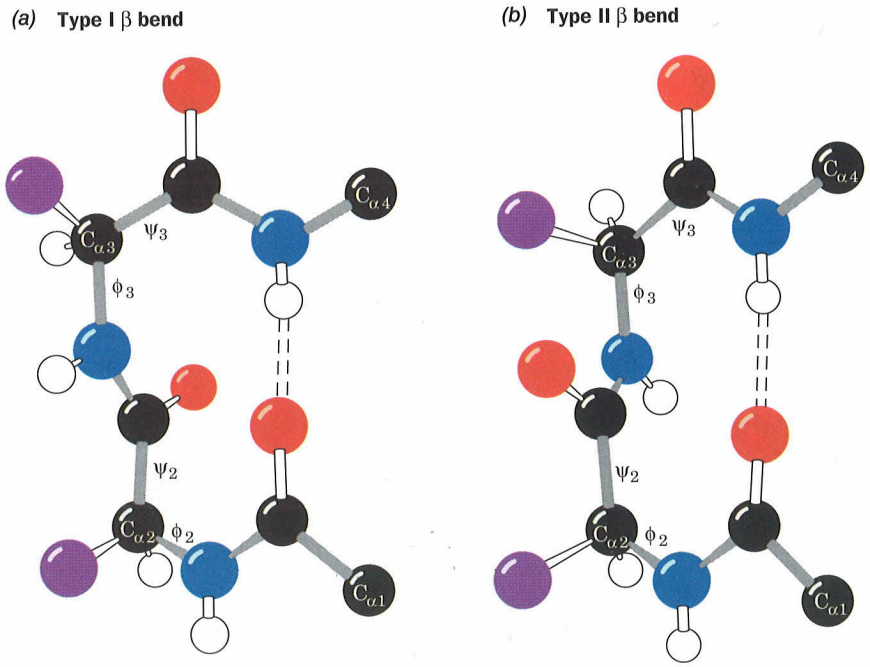


FIGURE 7-22. Reverse turns in polypeptide chains: (*a*) A Type I β bend, which has the following torsion angles:

$$\phi_2 = -60°, \quad \psi_2 = -30°,$$
$$\phi_3 = -90°, \quad \psi_3 = 0°.$$

(*b*) A Type II β bend, which has the following torsion angles:

$$\phi_2 = -60°, \quad \psi_2 = 120°,$$
$$\phi_3 = 90°, \quad \psi_3 = 0°.$$

Variations from these ideal conformation angles by as much as 30° are common. Hydrogen bonds are represented by dashed lines. [Figure copyrighted © by Irving Geis.]

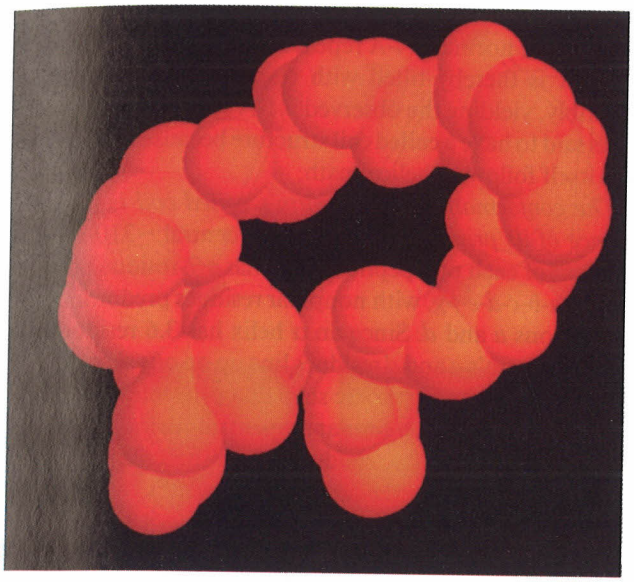


FIGURE 7-23. Space-filling representation of an Ω loop comprising residues 40 to 54 of cytochrome *c*. Only backbone atoms are shown; the addition of side chains would fill in the loop. [Courtesy of George Rose, Washington University School of Medicine.]

Ω loops (so named because they have the necked-in shape of the Greek upper case letter omega; Fig. 7-23), which may contain reverse turns, 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 an important role in biological recognition processes.

Many proteins have regions that are truly disordered. Extended, charged surface groups such as Lys side chains or the N- or C-termini of polypeptide chains are good examples: They often wave around in solution because there are few forces to hold them in place (Section 7-4). Sometimes entire polypeptide chain segments are disordered. Such segments may have functional roles, such as the binding of a specific molecule, so they may be disordered in one state of the protein (molecule absent) and ordered in another (molecule bound). This is one mechanism whereby a protein can interact flexibly with another molecule in the performance of its biological function.

2. FIBROUS PROTEINS

Fibrous proteins are highly elongated molecules whose secondary structures are their dominant structural motifs. Many fibrous proteins, such as those of skin, tendon, and bone, function as structural materials that have a protective, connective, or supportive role in living organisms. Others, such as muscle and ciliary proteins, have motive functions. In this section, we shall discuss structure–function relationships in four common and well-characterized fibrous proteins: keratin, silk fibroin, collagen, and elastin (muscle and ciliary proteins are considered in Section 34-3). The structural simplicity of these proteins relative to those of globular proteins (Section 7-3) makes them particularly amenable to understanding how their structures suit them to their biological roles.

Fibrous molecules rarely crystallize and hence are usually not subject to structural determination by single-crystal X-ray structure analysis (Section 7-3A). Rather than crystallizing, they associate as fibers in which their long molecular axes are more or less parallel to the fiber axis but in which they lack specific orientation in other directions. The X-ray diffraction pattern of such a fiber, Fig. 7-24, for example, contains little information, far less than would be obtained if the fibrous protein could be made to crystallize. Consequently, the structures of fibrous proteins are not known in great detail. Nevertheless, the original X-ray studies of proteins were carried out in the early 1930s by William Astbury on such easily available protein fibers as wool and tendon. Since the first X-ray crystal structure of a protein was not determined until the late 1950s, these fiber studies constituted the first tentative steps in the elucidation of the structural principles governing proteins and formed much of the experimental basis for Pauling's formulation of the α helix and β pleated sheet.

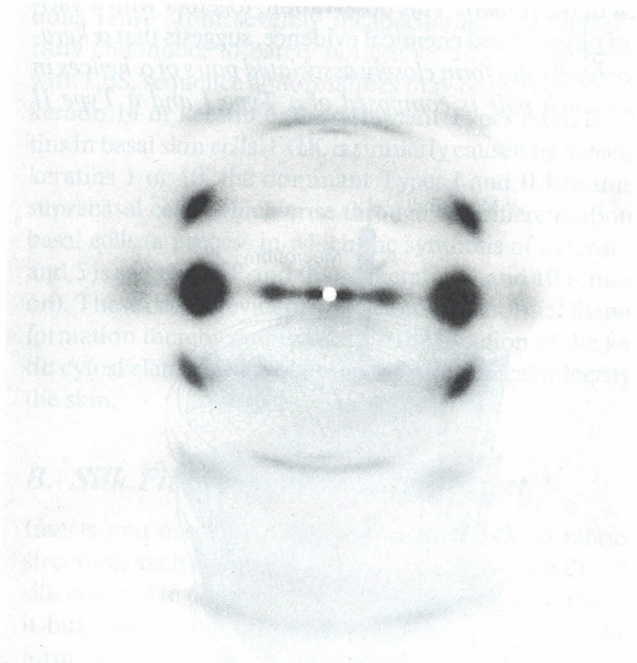


FIGURE 7-24. The X-ray diffraction photograph of a fiber of *Bombyx mori* silk obtained by shining a collimated beam of monochromatic X-rays through the silk fiber and recording the diffracted X-rays on a sheet of photographic film placed behind the fiber. The photograph has only a few spots and thus contains little structural information. [From March, R.E., Corey, R.B., and Pauling, L., *Biochim. Biophys. Acta* **16,** 5 (1955).]

A. α Keratin — A Helix of Helices

Keratin is a mechanically durable and chemically unreactive protein occurring in all higher vertebrates. It is the principal component of their horny outer epidermal layer, comprising up to 85% of the cellular protein, 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 have around 30 keratin variants which are expressed in a tissue-specific manner and which are classified as belonging to families of relatively acidic (Type I) and relatively basic (Type II) polypeptides. Keratin filaments, which form the intermediate filaments of skin cells (Section 1-2A), must contain at least one member of each type.

Electron microscopic studies indicate that hair, which is composed mainly of α keratin, consists of a hierarchy of structures (Figs. 7-25 and 7-26). A typical hair is ~20 μm in diameter and is constructed from dead cells, each of which contains packed **macrofibrils** (~2000 Å in diameter) that are oriented parallel to the hair fiber (Fig. 7-25). The macrofibrils are constructed from **microfibrils** (~80 Å wide) that are cemented together by an amorphous protein matrix of high sulfur content.

Moving to the molecular level, the X-ray diffraction pattern of α keratin resembles that expected for an α helix (hence the name α keratin). Yet, α keratin exhibits a 5.1-Å spacing rather than the 5.4-Å distance corresponding to the pitch of the α helix. This observation, together with a variety of physical and chemical evidence, suggests that *α keratin polypeptides form closely associated pairs of α helices in which each pair is composed of a Type I and a Type II keratin chain twisted in parallel into a left-handed coil (Fig. 7-26a).* The normal 5.4-Å repeat distance of each α helix in the pair is thereby tilted with respect to the axis of this assembly, yielding the observed 5.1-Å spacing. This assembly is said to have a **coiled coil** structure because each α helix axis 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 on one side of the α helix to form a hydrophobic strip that promotes its lengthwise association with a similar strip on another such α helix (Fig. 7-27; hydrophobic residues, as we shall see in Section 7-4C, have a strong tendency to associate). Indeed, the slight discrepancy between the 3.6 residues per turn of a normal α helix and the ~3.5 residue repeat of α keratin's hydrophobic strip is responsible for the coiled coil's coil. The resulting 18° inclination of the α helices relative to one another permits their contacting side chains to interdigitate efficiently, thereby greatly increasing their favorable interactions. Coiled coils, as we shall see, occur in several globular proteins as well as in other intermediate filament proteins.

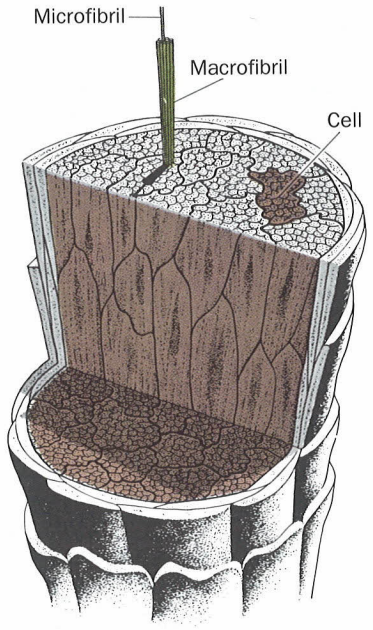


FIGURE 7-25. The macroscopic organization of hair. [Figure copyrighted © by Irving Geis.]

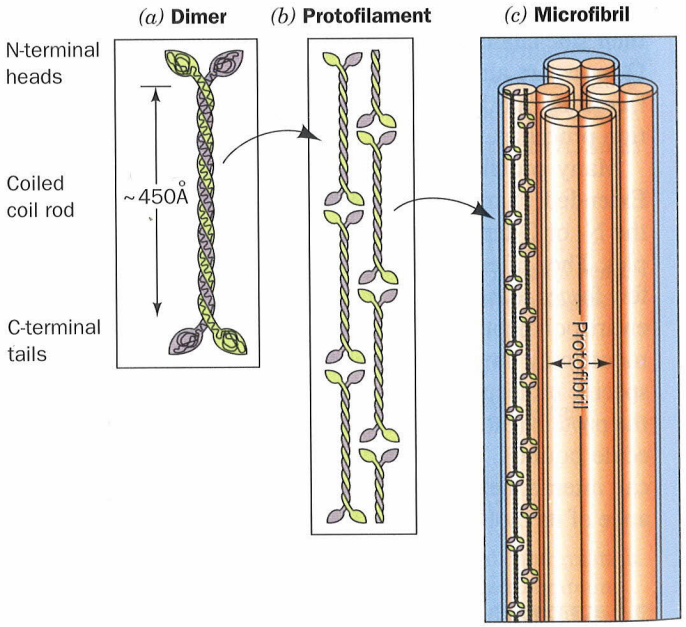


FIGURE 7-26. The structure of α keratin. (*a*) The central ~310 residues of one polypeptide chain each of Types I and II α keratins associate in a dimeric coiled coil. The conformations of the polypeptides' globular N- and C-terminal domains are unknown. (*b*) Protofilaments are formed from two staggered and antiparallel rows of associated head-to-tail coiled coils. (*c*) The protofilaments dimerize to form a protofibril, four of which form a microfibril. The structures of these latter assemblies are poorly characterized.

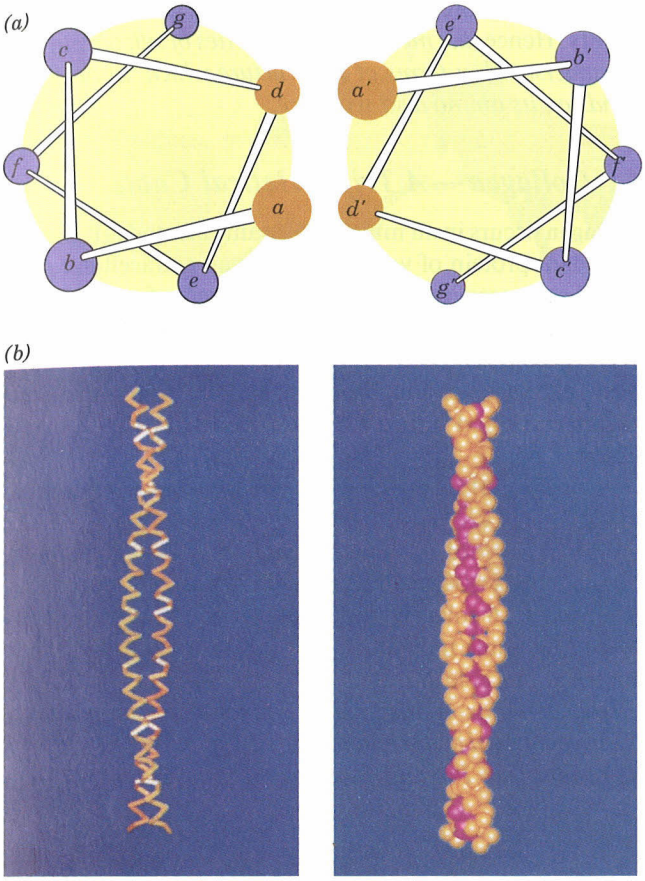


FIGURE 7-27. The two-stranded coiled coil. (*a*) View down the coil axis showing the interactions between the nonpolar edges of the α helices. The α helices have the pseudorepeating heptameric sequence *a-b-c-d-e-f-g* in which residues *a* and *d* are predominantly nonpolar. [After McLachlan, A.D. and Stewart, M., *J. Mol. Biol.* **98,** 295 (1975).] (*b*) Side view in which the polypeptide backbone is represented in skeletal (*left*) and space-filling (*right*) forms. Note the interlocking of the contacting nonpolar side chains (shown as red spheres) in the space-filling model. [Courtesy of Carolyn Cohen, Brandeis University.]

The higher order substructure of α keratin is poorly understood. The N- and C-terminal domains of each polypeptide probably have a flexible conformation and facilitate the assembly of the coiled coils into ~30-Å-wide protofilaments. These are thought to consist of two staggered antiparallel rows of head-to-tail aligned coiled coils (Fig. 7-26*b*). Two such protofilaments are thought to comprise a ~50-Å-wide protofibril, four of which, in turn, form a microfibril (Fig. 7-26*c*).

α Keratin is rich in Cys residues which form disulfide bonds that cross-link adjacent polypeptide chains. This accounts for its insolubility and resistance to stretching, two of α keratin's most important biological properties. 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 any forces tending to deform them. The disulfide bonds can be reductively cleaved with mercaptans (Section 6-1B). Hair so treated can be curled and set in a "permanent wave" by application of an oxidizing agent which reestablishes the disulfide bonds in the new "curled" conformation. Although the insolubility of α keratin prevents most animals from digesting it, the clothes moth larva, which has a high concentration of mercaptans in its digestive tract, can do so to the chagrin of owners of woolen clothing.

The springiness of hair and wool fibers is a consequence of the coiled coil's tendency to untwist when stretched and to recover its original conformation when the external force is relaxed. After some of its disulfide bonds have been cleaved, however, an α keratin fiber can be stretched to over twice its original length by the application of moist heat. In this process, as X-ray analysis indicates, the α helical structure extends with concomitant rearrangement of its hydrogen bonds to form a β pleated sheet. β Keratin, such as that of feathers, exhibits a similar X-ray pattern in its native state (hence the β sheet).

Keratin Defects Result in a Loss of Skin Integrity

The inherited skin diseases **epidermolysis bullosa simplex (EBS)** and **epidermolytic hyperkeratosis (EHK)** are characterized by skin blistering arising from the rupture of epidermal basal cells (Fig. 1-14*d*) and suprabasal cells, respectively, as caused by mechanical stresses that normally would be harmless. Symptomatic variations in these conditions range from severely incapacitating, particularly in early childhood, to barely noticeable. In families afflicted with EBS, sequence abnormalities may be present in either keratin 14 or keratin 5, the dominant Types I and II keratins in basal skin cells. EHK is similarly caused by defects in keratins 1 or 10, the dominant Types I and II keratins in suprabasal cells (which arise through the differentiation of basal cells, a process in which the synthesis of keratins 14 and 5 is switched off and that of keratins 1 and 10 is turned on). These defects evidently interfere with normal filament formation thereby demonstrating the function of the keratin cytoskeleton in maintaining the mechanical integrity of the skin.

B. Silk Fibroin — A β Pleated Sheet

Insects and arachnids (spiders) produce **silk** to fabricate structures such as cocoons, webs, nests, and egg stalks. The silk is stored in aqueous solution in the gland that produces it but, during spinning, is converted to a water-insoluble form. Most silks consist of the fibrous protein **fibroin** and a gummy amorphous protein named **sericin** that cements the fibroin fibers together. An adult moth emerging from its sealed cocoon secretes a protease **(cocoonase)** that digests sericin, thereby enabling the moth to push the fibroin filaments aside and escape from the cocoon. In the preparation of silk cloth, which consists only of fibroin, the sericin is removed by treatment with boiling soap solution.

FIGURE 7-28. The domestic silkworm *Bombyx mori* in the process of constructing its cocoon. [Hans Pfletschinger/Peter Arnold, Inc.]

Silk fibroin from the cultivated larvae (silkworms) of the moth *Bombyx mori* (Fig. 7-28) exhibits an X-ray diffraction pattern (Fig. 7-24) indicating that *its polypeptide chains form antiparallel β pleated sheets in which the chains extend parallel to the fiber axis.* Sequence studies have shown that long stretches of the chain are comprised of the six-residue repeat

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

This sequence forms β sheets with its Gly side chains extending from one surface and its Ser and Ala side chains extending from the other surface (as in Fig. 7-17). The β sheets thereby stack to form a microcrystalline array such that layers of contacting Gly side chains from neighboring sheets alternate with layers of contacting Ser and Ala side chains (Fig. 7-29). This structure, in part, accounts for silk's mechanical properties. *Silk fibers are strong but only slightly extensible because appreciable stretching would require breaking the covalent bonds of its nearly fully extended polypeptide chains.* Yet, the fibers are flexible because neighboring β sheets associate only through relatively weak van der Waals forces.

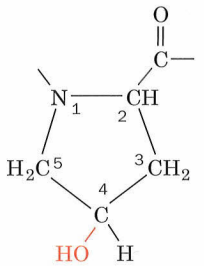
Although large segments of silk fibroin have the repeating hexameric amino acid sequence, it also has regions in which bulky residues such as Tyr, Val, Arg, and Asp occur. These residues distort and therefore disorder the microcrystalline array drawn in Fig. 7-29. Silk fibers are consequently composed of alternating crystalline and amorphous regions. The amorphous regions are largely responsible for the extensibility of the silk fibers. Silks from different species have different proportions of bulky side groups and therefore different mechanical properties. The greater a fibroin's proportion of bulky groups, the less its elasticity (the ability to resist deformation and to recover its original shape when the deforming forces are removed) and the greater its extensibility. Hence, *the mechanical properties of silk fibroin can be understood in terms of its structure which, in turn, depends on its amino acid sequence.*

C. Collagen—A Triple Helical Cable

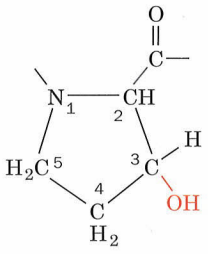
Collagen occurs in all multicellular animals and is the most abundant protein of vertebrates. It is an extracellular protein that is organized into insoluble fibers of great tensile strength. This suits collagen to its role as the major stress-bearing component of **connective tissues** such as bone, teeth, cartilage, tendon, ligament, and the fibrous matrices of skin and blood vessels. Collagen occurs in virtually every tissue.

A single molecule of Type I collagen has a molecular mass of ~285 kD, a width of ~14 Å, and a length of ~3000 Å. It is composed of three polypeptide chains. Mammals have at least 30 genetically distinct polypeptide chains comprising 16 collagen variants that occur in different tissues of the same individual. The most prominant of these are listed in Table 7-2.

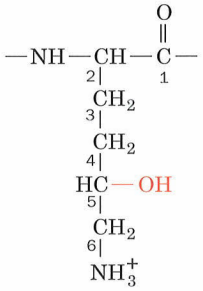
Collagen has a distinctive amino acid composition: Nearly one third of its residues are Gly; another 15 to 30% of its residues are Pro and ***4-hydroxyproline (Hyp).***



4-Hydroxyprolyl residue (Hyp)



3-Hydroxyprolyl residue



5-Hydroxylysyl residue (Hyl)

3-Hydroxyproline and **5-hydroxylysine (Hyl)** also occur in collagen but in smaller amounts. Radioactive labeling experiments have established that these nonstandard hydroxylated amino acids are not incorporated into collagen during polypeptide synthesis: If ^{14}C-labeled 4-hydroxyproline is administered to a rat, the collagen synthesized is not radioactive, whereas radioactive collagen is produced if the

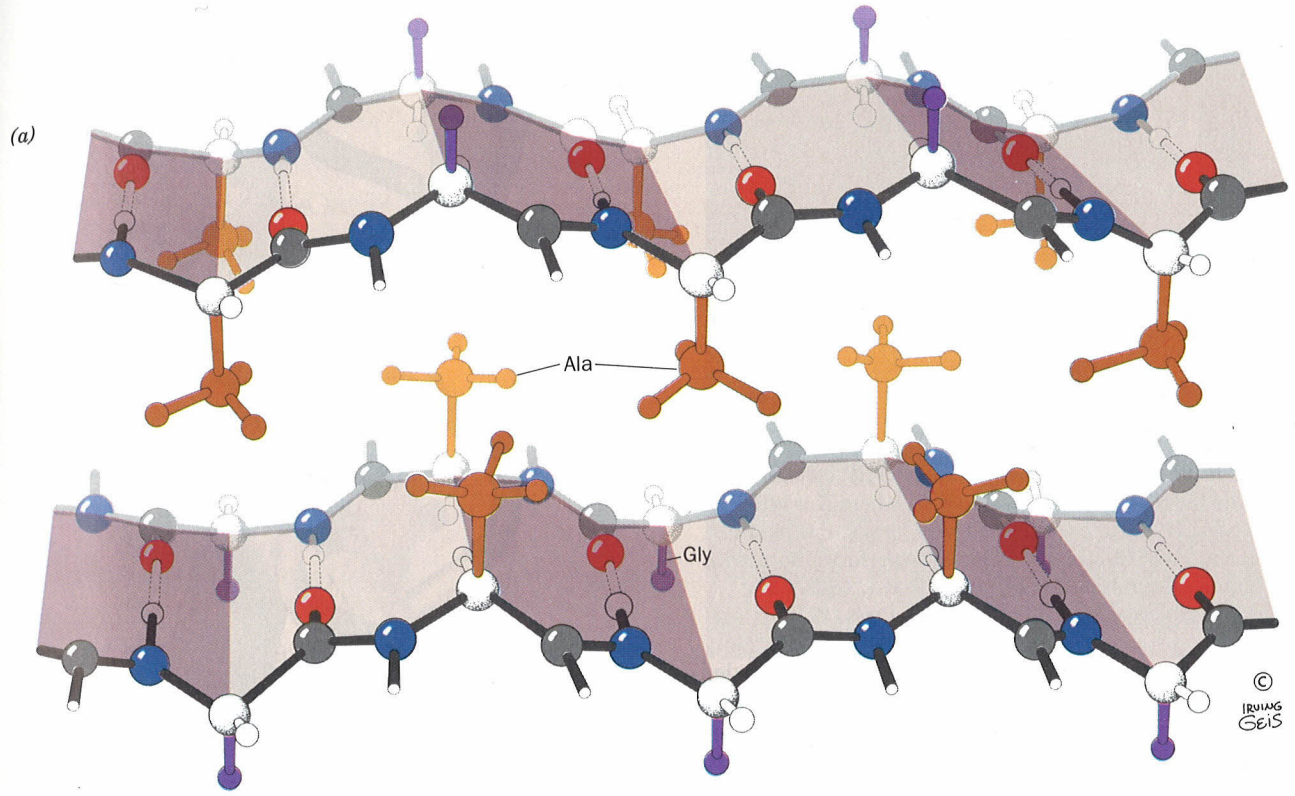


FIGURE 7-29. The three-dimensional architecture of silk fibroin. (*a*) Silk's alternating Gly and Ala (or Ser) residues extend to opposite sides of a given β sheet so that the Ala side chains extending from one β sheet efficiently nestle between those of the neighboring sheet and likewise for the Gly side chains. (*b*) Gly side chains from neighboring β sheets are in contact as are those of Ala and Ser. The intersheet spacings consequently have the alternating values 3.5 and 5.7 Å. [Figure copyrighted © by Irving Geis.]

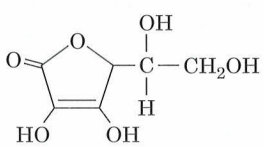
TABLE 7-2. THE MOST ABUNDANT TYPES OF COLLAGEN

Type	Chain Composition	Distribution
I	$[\alpha 1(I)]_2\alpha 2(I)$	Skin, bone, tendon, blood vessels, cornea
II	$[\alpha 1(II)]_3$	Cartilage, intervertebral disk
III	$[\alpha 1(III)]_3$	Blood vessels, fetal skin

Source: Eyre, D.R., *Science* **207,** 1316 (1980).

rat is fed ^{14}C-labeled proline. The hydroxylated residues appear after the collagen polypeptides are synthesized, when certain Pro residues are converted to Hyp in a reaction catalyzed by the enzyme **prolyl hydroxylase.**

Hyp confers stability upon collagen, probably through intramolecular hydrogen bonds that may involve bridging water molecules. If, for example, collagen is synthesized under conditions that inactivate prolyl hydroxylase, it loses its native conformation (denatures) at 24°C, whereas normal collagen denatures at 39°C (denatured collagen is known as **gelatin**). Prolyl hydroxylase requires **ascorbic acid (vitamin C)**



Ascorbic acid (vitamin C)

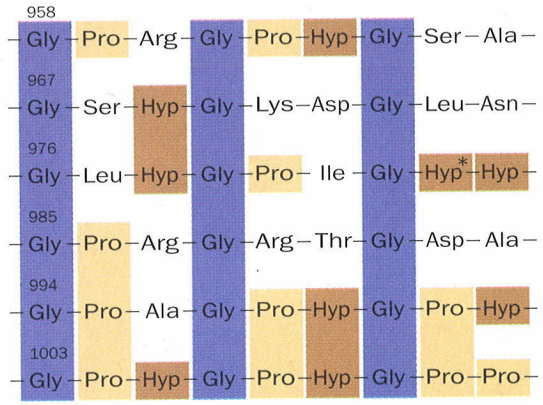


FIGURE 7-30. The amino acid sequence at the C-terminal end of the triple helical region of the bovine α1(I) collagen chain. Note the repeating triplets Gly-X-Y, where X is often Pro and Y is often Hyp. Here Gly is shaded in purple, Pro in tan, and Hyp and Hyp* (3-hydroxyPro) in brown. [From Bornstein, P. and Traub, W., *in* Neurath, H. and Hill. R.L. (Eds.), *The Proteins* (3rd ed.), Vol. 4, *p.* 483, Academic Press (1979).]

to maintain its enzymatic activity. In the vitamin C deficiency disease **scurvy,** the collagen synthesized cannot form fibers properly. This results in the skin lesions, blood vessel fragility, and poor wound healing that are symptomatic of scurvy.

The amino acid sequence of bovine collagen a1(I), which is similar to that of other collagens, consists of monotonously repeating triplets of sequence Gly-X-Y over a continuous 1011-residue stretch of its 1042-residue polypeptide chain (Fig. 7-30). Here X is often Pro and Y is often Hyp. The restriction of Hyp to the Y position stems from the specificity of prolyl hydroxylase. 5-Hydroxylysine is similarly restricted to the Y position.

The high Gly, Pro, and Hyp content of collagen suggests that its polypeptide backbone conformation resembles those of the polyglycine II and polyproline II helices (Fig. 7-15). X-Ray and model building studies indicate that *collagen's three polypeptide chains, which individually resemble polyproline II helices, are parallel and wind around each other with a gentle, right-handed, ropelike twist to form a triple-helical structure (Fig. 7-31).* 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 (Fig. 7-30). It also requires that the three polypeptide chains be staggered so that the Gly, X, and Y residues from the three chains occur at similar levels (Fig. 7-32). The staggered 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. The bulky and relatively inflexible Pro and Hyp residues confer rigidity on the entire assembly.

Collagen's well-packed, rigid, triple helical structure is responsible for its characteristic tensile strength. As with the twisted fibers of a rope, the extended and twisted polypep-

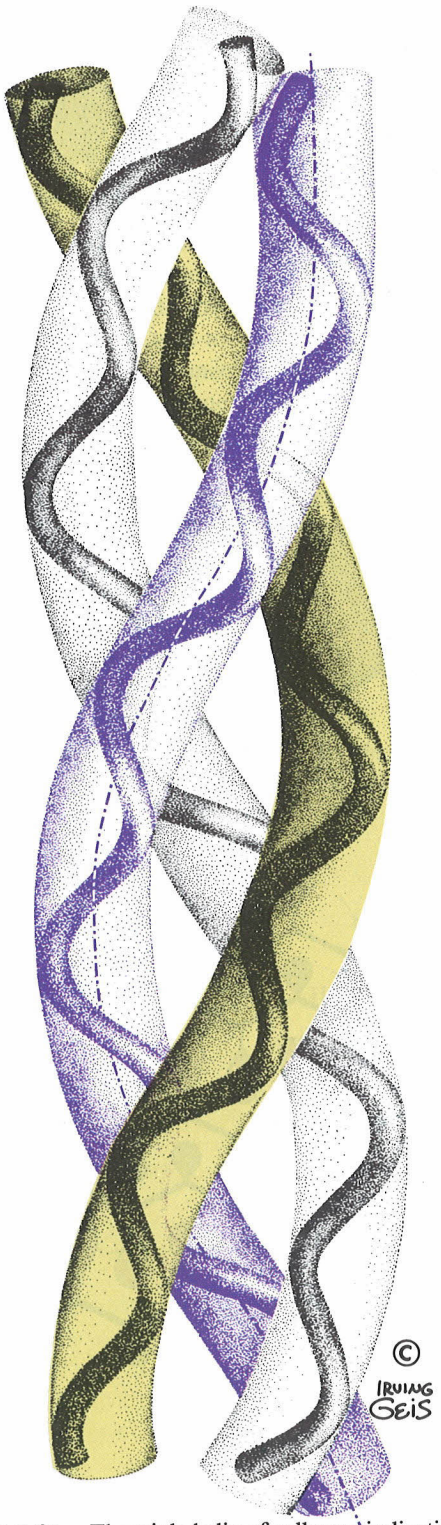


FIGURE 7-31. The triple helix of collagen, indicating how the left-handed polypeptide helices are twisted together to form a right-handed superhelical structure. Ropes and cables are similarly constructed from hierarchies of fiber bundles that are alternately twisted in opposite directions. An individual polypeptide helix has 3.3 residues per turn and a pitch of 10.0 Å (in contrast to polyproline II's 3.0 residues per turn and pitch of 9.4 Å; Fig. 7-15). The collagen triple helix has 10 Gly-X-Y units per turn and a pitch of 86.1 Å. [Figure copyrighted © by Irving Geis.]

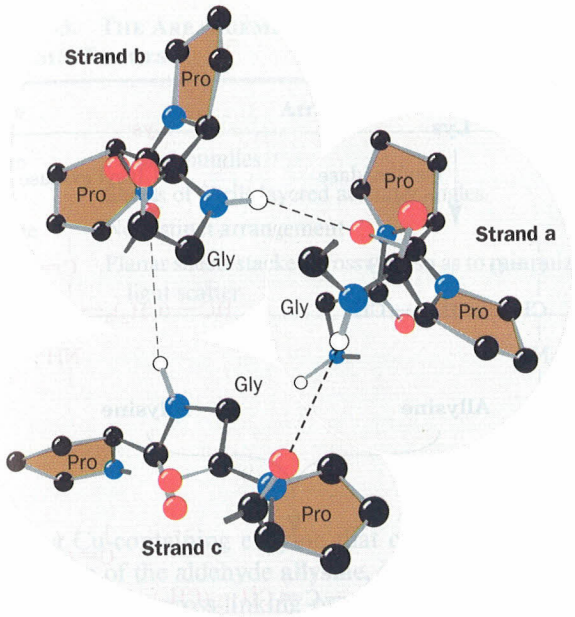


FIGURE 7-32. A projection down the triple helix axis of the collagen-like polymer $(Gly\text{-}Pro\text{-}Pro)_n$ as viewed from its carboxyl end. The residues in each chain, Gly-X-Y, are vertically staggered such that a Gly, an X, and a Y residue from different chains are on the same level along the helix axis. The dashed lines represent hydrogen bonds between each Gly N—H group and the oxygen of the succeeding X residue on a neighboring chain. Every third residue on each chain must be Gly because there is no room near the helix axis for the side chain of any other residue. The bulky pyrrolidine side chains (*brown*) of the Pro residues are situated on the periphery of the triple helix where they are sterically unhindered. [After Yonath, A. and Traub, W., *J. Mol. Biol.* **43,** 461 (1969).]

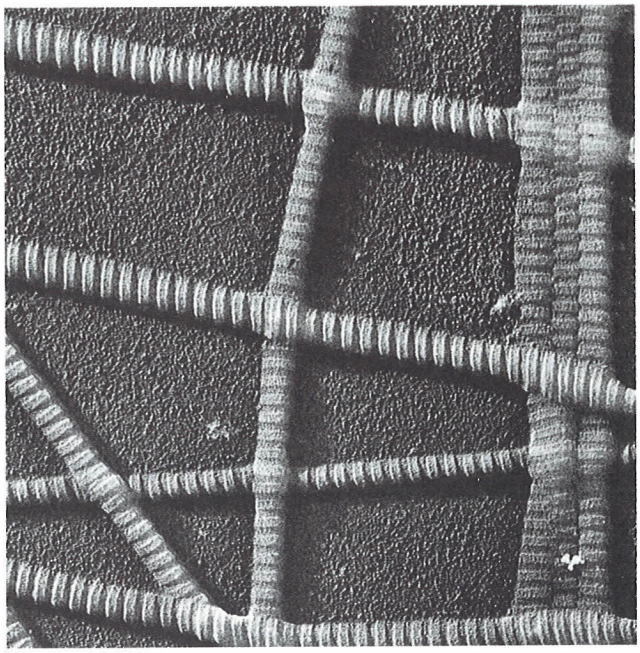


FIGURE 7-33. An electron micrograph of collagen fibrils from skin. [Courtesy of Jerome Gross, Massachusetts General Hospital.]

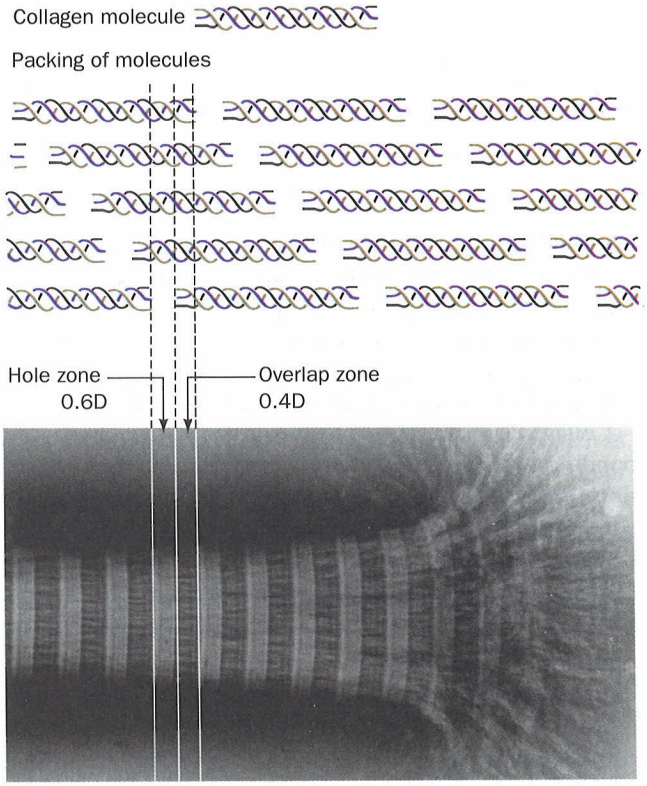


FIGURE 7-34. The banded appearance of collagen fibrils in the electron microscope arises from the schematically represented staggered arrangement of collagen molecules (*above*) that results in a periodically indented surface. *D*, the distance between cross striations, is ~680 Å, so the length of a 3000-Å-long collagen molecule is 4.4*D*. [Courtesy of Karl A. Piez, Collagen Corporation.]

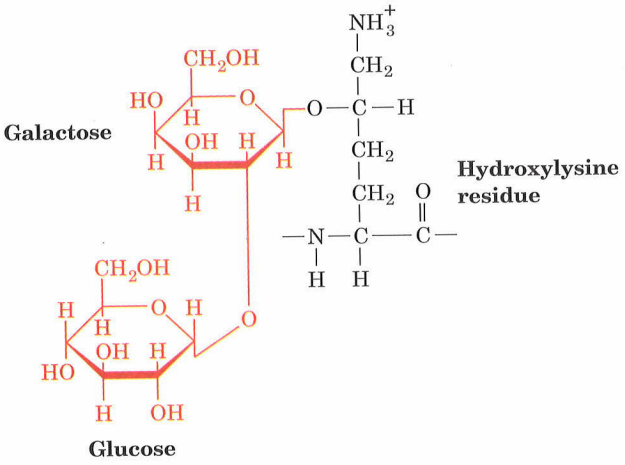
tide chains of collagen convert a longitudinal tensional force to a more easily supported lateral compressional force on the almost incompressible triple helix. This occurs because the oppositely twisted directions of collagen's polypeptide chains and triple helix (Fig. 7-31) prevent the twists from being pulled out under tension (note that successive levels of fiber bundles in ropes and cables are likewise oppositely twisted). The successive helical hierarchies in other fibrous proteins exhibit similar alternations of twist directions, for example, keratin (Section 7-2A) and muscle (Section 34-3A).

Collagen Is Organized into Fibrils

Types I, II, III, V, and XI collagens form distinctive banded fibrils (Fig. 7-33) that are mostly, if not entirely, composed of several different types of collagens. These fibrils have a periodicity of 680 Å and a diameter of 100 to 2000 Å depending on the types of collagen they contain and their tissue of origin (the other collagen types form different sorts of aggregates such as networks; we will not discuss them further). Computerized model building studies indicate that collagen molecules are laterally organized in a precisely staggered array (Fig. 7-34). The darker portions of

the banded structures correspond to the 400-Å "holes" on the surface of the fibril between head-to-tail aligned collagen molecules. Structural and energetic considerations suggest that the conformations of individual collagen molecules, much like those of individual α helices and β sheets, are but marginally stable. The driving force for the assembly of collagen molecules into a fibril is apparently provided by the added hydrophobic interactions within the fibrils in a manner analogous to the packing of secondary structural elements to form a globular protein (Section 7-3B).

Collagen contains covalently attached carbohydrates in amounts that range from ~0.4 to 12% by weight, depending on the collagen's tissue of origin. The carbohydrates, which consist mostly of glucose, galactose, and their disaccharides, are covalently attached to collagen at its 5-hydroxylysyl residues by specific enzymes.



Although the function of carbohydrates in collagen is unknown, the observation that they are located in the "hole" regions of the collagen fibrils suggests that they are involved in directing fibril assembly.

Collagen Fibrils Are Covalently Cross-Linked

Collagen's insolubility in solvents that disrupt hydrogen bonding and ionic interactions is explained by the observation that it is both intramolecularly and intermolecularly covalently cross-linked. The cross-links cannot be disulfide bonds, as in keratin, because collagen is almost devoid of Cys residues. Rather, they are derived from Lys and His side chains in reactions such as those in Fig. 7-35. **Lysyl**

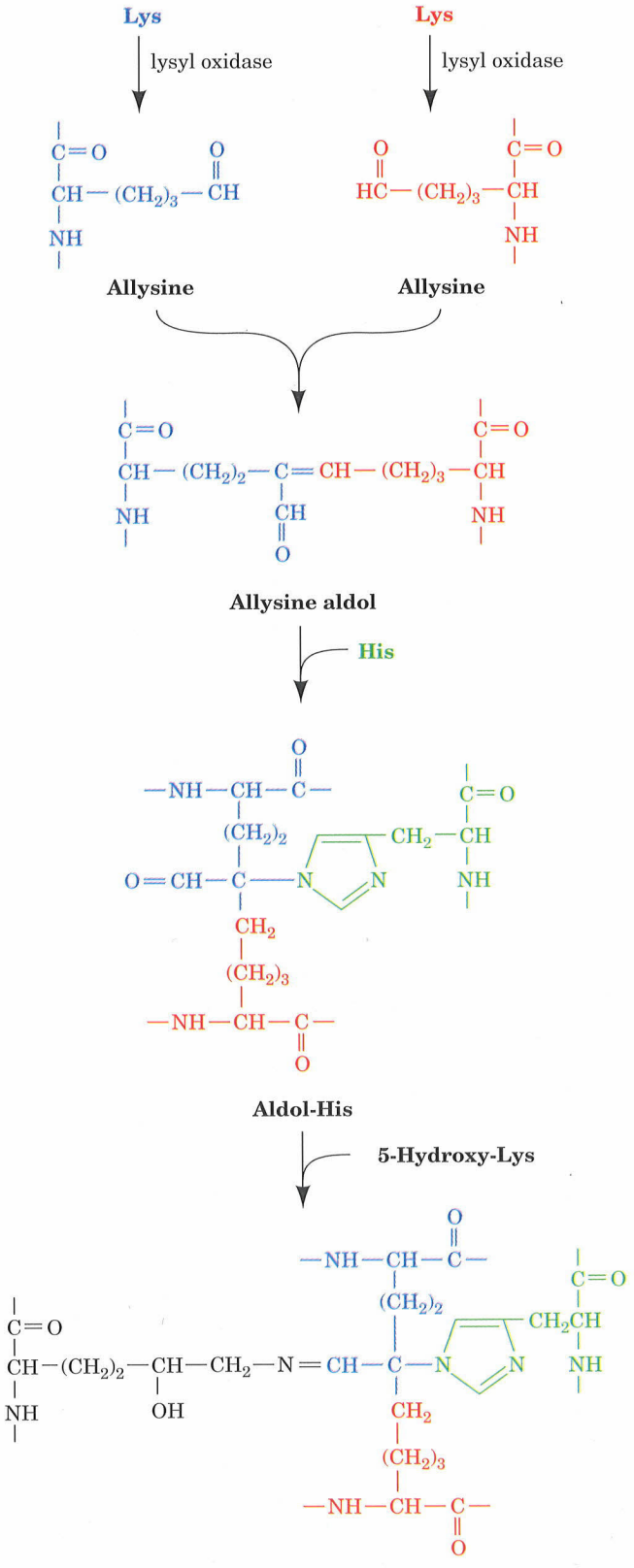


FIGURE 7-35. A biosynthetic pathway for cross-linking Lys, 5-hydroxylysyl, and His side chains in collagen. The first step in the reaction is the lysyl oxidase–catalyzed oxidative deamination of Lys to form the aldehyde allysine. Two such aldehydes then undergo an aldol condensation to form **allysine aldol.** This product can react with His to form **aldol histidine.** This, in turn, can react with 5-hydroxylysine to form a Schiff base (an imine bond), thereby cross-linking four side chains.

TABLE 7-3. THE ARRANGEMENT OF COLLAGEN FIBRILS IN VARIOUS TISSUES

Tissue	Arrangement
Tendon	Parallel bundles
Skin	Sheets of fibrils layered at many angles
Cartilage	No distinct arrangement
Cornea	Planar sheets stacked crossways so as to minimize light scatter

oxidase, a Cu-containing 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, instead, tend to occur near the N- and C-termini of the collagen molecules.

The importance of cross-linking to the normal functioning of collagen is demonstrated by the disease **lathyrism** which occurs in humans and other animals as a result of the regular ingestion of seeds from the sweet pea *Lathyrus odoratus.* The symptoms of this condition are serious abnormalities of the bones, joints, and large blood vessels which are caused by an increased fragility of the collagen fibers. The causative agent of lathyrism, **β-aminopropionitrile,**

$$N{\equiv}C{-}CH_2{-}CH_2{-}NH_3^+$$

β-Aminopropionitrile

inactivates lysyl oxidase by covalently binding to its active site. This results in markedly reduced cross-linking in the collagen of lathrytic animals.

The degree of cross-linking of the collagen from a particular tissue increases with the age of the animal. This is why meat from older animals is tougher than that from younger animals. In fact, individual molecules of collagen (called **tropocollagen**) can only be extracted from the tissues of very young animals. Collagen cross-linking is not the central cause of aging, however, as is demonstrated by the observation that lathyrogenic agents do not slow the aging process.

The collagen fibrils in various tissues are organized in ways that largely reflect the functions of the tissues (Table 7-3). Thus tendons (the "cables" connecting muscles to bones), skin (a tear-resistant outer fabric), and cartilage (which has a load-bearing function) must support stress in predominantly one, two, and three dimensions, respectively, and their component collagen fibrils are arrayed accordingly. How collagen fibrils are laid down in these arrangements is unknown. However, some of the factors guiding collagen molecule assembly are discussed in Sections 30-5A and B.

Collagen Defects Are Responsible for a Variety of Human Diseases

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. 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. All known amino acid changes within Type I collagen's triple helical region result in abnormalities, indicating that the structural integrity of this region is essential for proper collagen function.

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 or lysyl oxidase. One group of at least 10 different collagen deficiency diseases, the **Ehlers–Danlos syndromes,** are all characterized by hyperextensibility of the joints and skin. The "India-rubber man" of circus fame had an Ehlers–Danlos syndrome. Many degenerative diseases exhibit collagen abnormalities in the affected tissues. Examples of such tissues are the cartilage in **osteoarthritis** and the fibrous **atherosclerotic plaques** in human arteries.

D. Elastin—A Nonrepetitive Coil

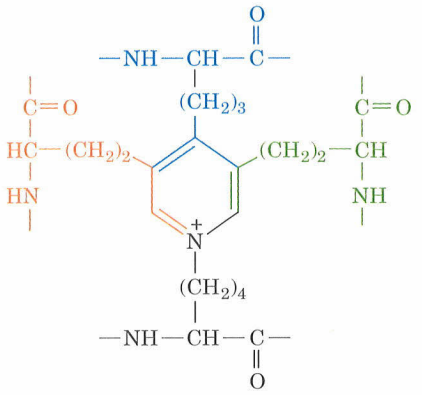
Elastin is a protein with rubberlike elastic properties whose fibers can stretch to several times their normal length. It is the principle protein component of the elastic yellow connective tissue that occurs in the lungs, the walls of large blood vessels such as the aorta, and elastic ligaments such as those in the neck. The inelastic white connective tissue of tendons contains only a small amount of elastin. The hyperextensibility of the joints and skin characteristic of certain collagen deficiency diseases results from the loss of rigidity ordinarily conferred by collagen coupled with the normal presence of elastin.

Elastin, like silk fibroin and collagen, has a distinctive amino acid composition. It consists predominantly of small, nonpolar residues: It is one-third Gly, over one-third Ala + Val, and is rich in Pro. However, it contains little hydroxyproline, no hydroxylysine, and few polar residues.

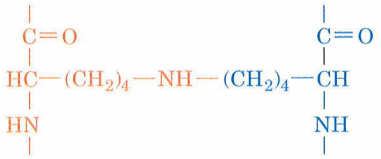
Elastin forms a three-dimensional network of fibers that exhibit no recognizable periodicity in the electron microscope. Furthermore, according to X-ray analyses, the fibers are devoid of regular secondary structure.

The covalent cross-links in elastin are formed by allysine aldol, which also occurs in collagen (Fig. 7-35), and the

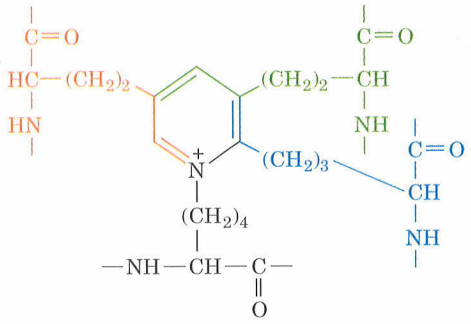
compounds **lysinonorleucine, desmosine,** and **isodesmosine.**



Desmosine



Lysinonorleucine



Isodesmosine

Lysinonorleucine, which likewise occurs in collagen, results from the reduction of the Schiff base (imine bond) formed by the condensation of a Lys side chain with that of allysine (Fig. 7-35). Desmosine and isodesmosine are unique to elastin and are responsible for its yellow color; they result from the condensation of three allysine and one lysine side chains. The primary structure of elastin consists of alternating hydrophobic segments, thought to be responsible for the protein's elastic properties, and Lys-rich segments that contain the cross-links. The Lys residues usually occur in pairs, which suggests that a given desmosine or isodesmosine serves to cross-link two rather than four elastin polypeptides.

3. GLOBULAR PROTEINS

Globular proteins comprise a highly diverse group of substances that, in their native state, exist as compact spheroidal molecules. Enzymes are globular proteins as are transport and receptor proteins. In this section we consider the tertiary structures of globular proteins. However, since most of our detailed structural knowledge of proteins, and thus to a large extent their function, has resulted from X-ray crystal structure determinations of globular proteins and, more recently, from their nuclear magnetic resonance (NMR) structure determinations, we begin this section with a discussion of the capabilities and limitations of these powerful techniques.

A. Interpretation of Protein X-Ray and NMR Structures

X-Ray crystallography is a technique that directly images molecules. X-Rays must be used to do so because, according to optical principles, the uncertainty in locating an object is approximately equal to the wavelength of the radiation used to observe it (both X-ray wavelengths and covalent bond distances are ~1.5 Å; individual molecules cannot be seen in a light microscope because visible light has a minimum wavelength of 4000 Å). There is, however, no such thing as an X-ray microscope because there are no X-ray lenses. Rather, a crystal of the molecule to be visualized is exposed to a collimated beam of monochromatic X-rays and the consequent diffraction pattern is recorded on photographic film (Fig. 7-36) or by a radiation counter. The intensities of the diffraction maxima (darkness of the spots on the film) are then used to construct mathematically the three-dimensional image of the crystal structure through methods that are outside the scope of this text. In

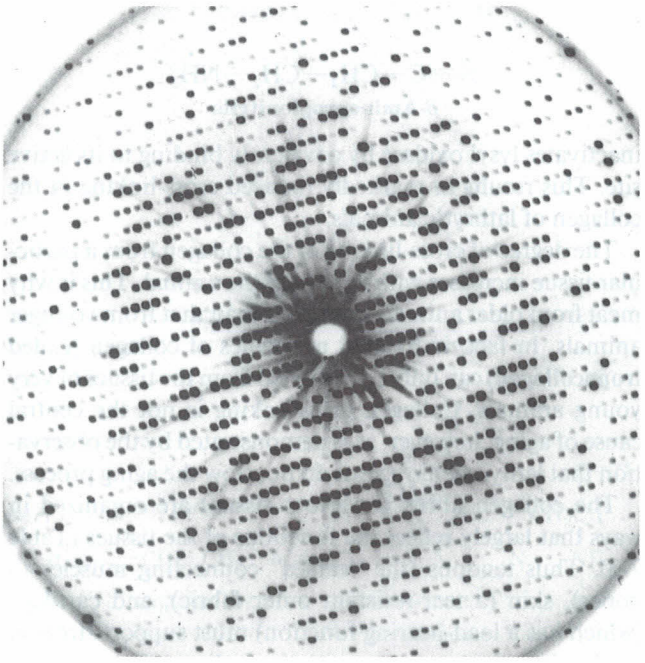


FIGURE 7-36. An X-ray diffraction photograph of a single crystal of sperm whale myoglobin. The intensity of each diffraction maximum (the darkness of each spot) is a function of the myoglobin crystal's electron density. The photograph contains only a small fraction of the total diffraction information available from a myoglobin crystal. [Courtesy of John Kendrew, Cambridge University.]

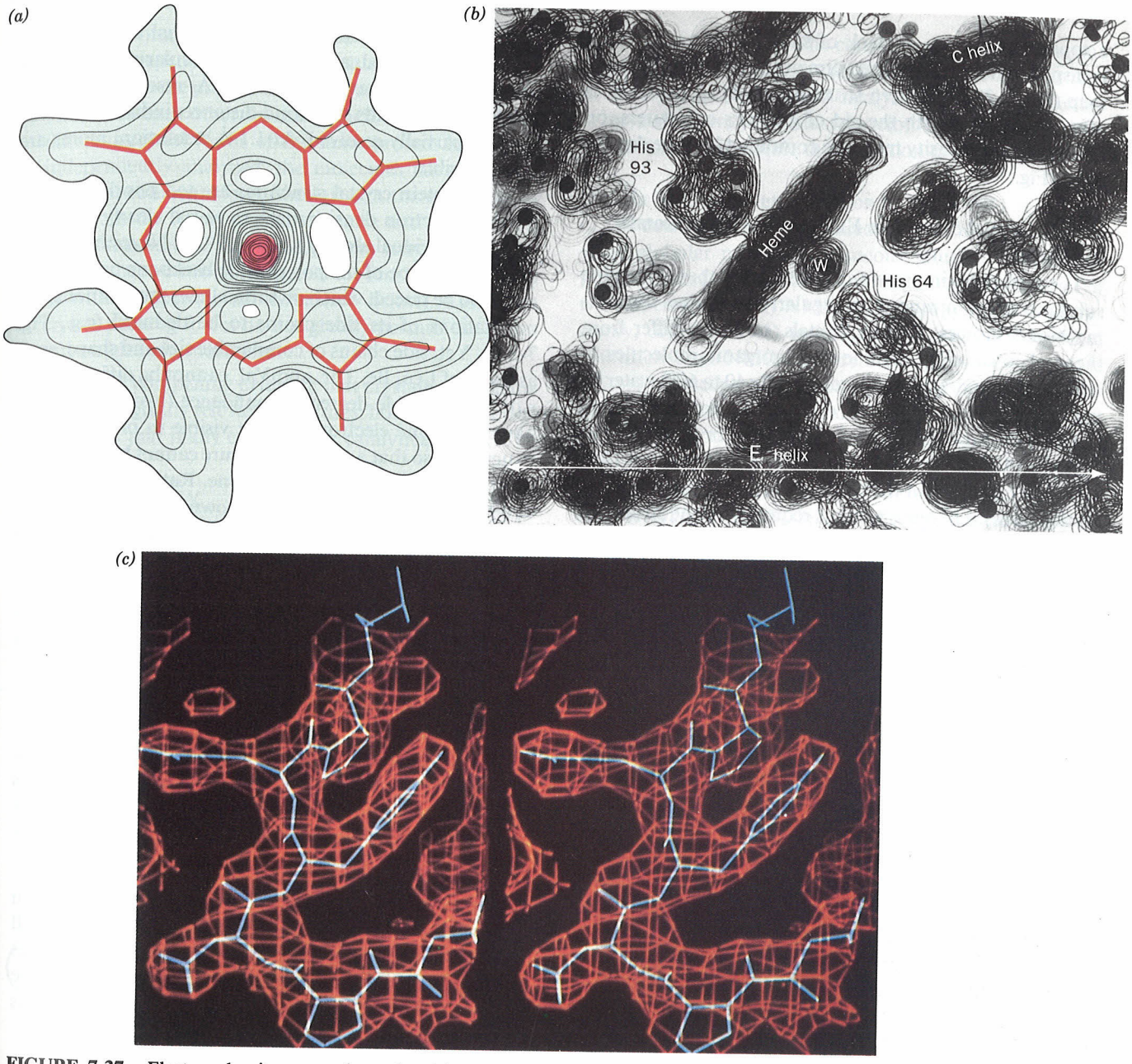


FIGURE 7-37. Electron density maps of proteins. (*a*) A section through the 2.0-Å-resolution electron density map of sperm whale myoglobin, which contains the heme group (*red*). The large peak at the center of the map represents the electron-dense Fe atom. [After Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., and Shore, V.C., *Nature* **185,** 434 (1960).] (*b*) A portion of the 2.4-Å-resolution electron density map of myoglobin constructed from a stack of contoured transparencies. Dots have been placed at the positions deduced for the nonhydrogen atoms. The heme group is seen edge-on together with its two associated His residues and a water molecule, W. An α helix, the so-called E helix (Fig. 7-12), extends across the bottom of the map. Another α helix, the C helix, extends into the plane of the paper on the upper right. Note the hole along its axis. [Courtesy of John Kendrew, Cambridge University, U.K.] (*c*) A portion of the 3.0-Å-resolution electron density map of a human rhinovirus (the cause of the common cold, Section 32-2C) contoured in three dimensions on a graphics computer and shown in stereo. Only a single contour level (*orange*) is shown, together with an atomic model of the corresponding polypeptide segment (*white*). Instructions for viewing stereo diagrams are given in the appendix to this chapter. [Courtesy of Michael Rossmann, Edward Arnold, and Gerrit Vriend, Purdue University.]

what follows, we discuss some of the special problems associated with interpreting the X-ray crystal structures of proteins.

X-Rays interact almost exclusively with the electrons in matter, not the nuclei. An X-ray structure is therefore an image of the electron density of the object under study. Such **electron density maps** may be presented as a series of parallel sections through the object. On each section, the electron density is represented by contours (Fig. 7-37*a*) in the same way that altitude is represented by the contours on