LODISH DECLARATION
EXHIBIT E
PART 2 OF 2

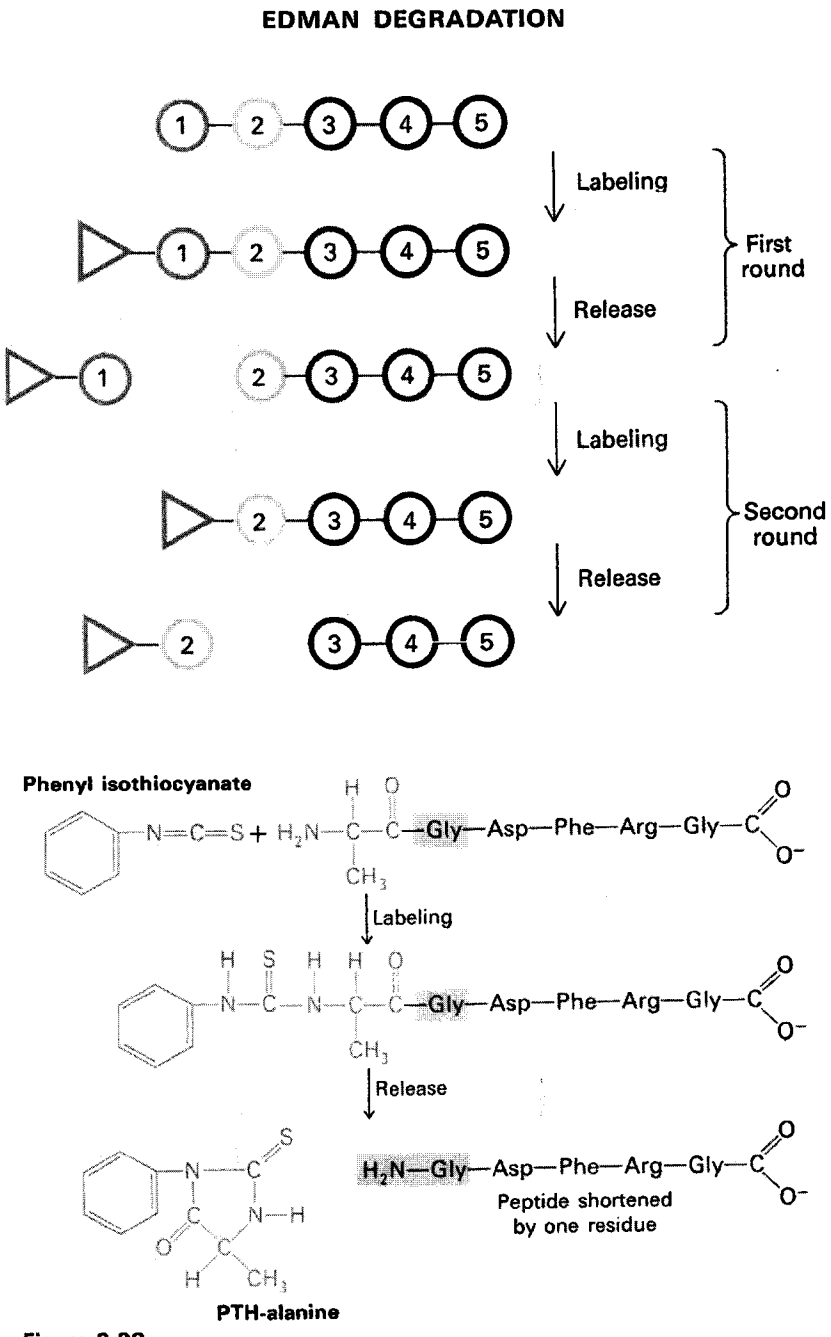
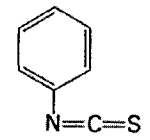


Figure 2-28
The Edman degradation. The labeled amino-terminal residue (PTH-alanine in the first round) can be released without hydrolyzing the rest of the peptide. Hence, the amino-terminal residue of the shortened peptide (Gly-Asp-Phe-Arg-Gly) can be determined in the second round. Three more rounds of the Edman degradation reveal the complete sequence of the original peptide.

graded in the acid-hydrolysis step. Pehr Edman devised a method for labeling the amino-terminal residue and cleaving it from the peptide without disrupting the peptide bonds between the other amino acid residues. The *Edman degradation* sequentially removes one residue at a time from the amino end of a peptide (Figure 2-28). *Phenyl isothiocyanate* reacts with the uncharged terminal amino group of the peptide to form a phenylthiocarbamoyl derivative. Under mildly acidic conditions, a cyclic derivative of the terminal amino acid is liberated, which leaves an intact peptide shortened by one amino acid. The cyclic compound is a phenylthiohydantoin (PTH) amino acid. The PTH-amino acid can be identified by chromatographic procedures. Furthermore, the amino acid composition of the shortened peptide:



Phenyl isothiocyanate

$$(\text{Arg, Asp, Gly}_2\text{, Phe})$$

can be compared with that of the original peptide:

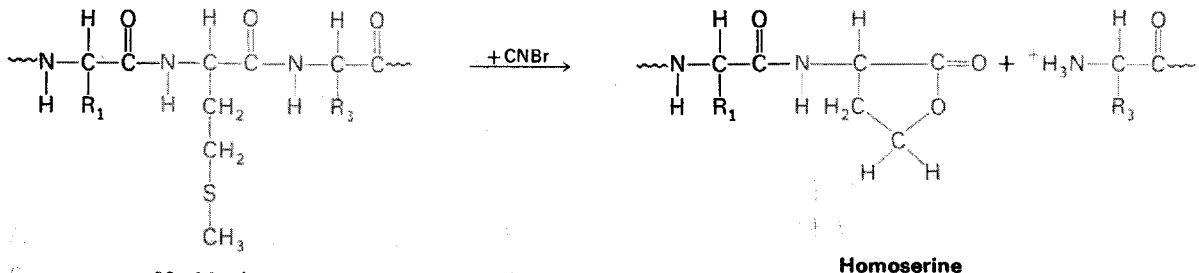
$$(\text{Ala, Arg, Asp, Gly}_2\text{, Phe})$$

The difference between these analyses is one alanine residue, which shows that alanine is the amino-terminal residue of the original peptide. The Edman procedure can then be repeated on the shortened peptide. The amino acid analysis after the second round of degradation is

$$(\text{Arg, Asp, Gly, Phe})$$

showing that the second residue from the amino end is glycine. This conclusion can be confirmed by chromatographic identification of PTH-glycine obtained in the second round of the Edman degradation. Three more rounds of the Edman degradation will reveal the complete sequence of the original peptide.

The experimental strategy for determining the amino acid sequence of proteins is to divide and conquer. A protein is *specifically cleaved into, smaller peptides* that can be readily sequenced by the Edman method. Specific cleavage can be achieved by chemical or enzymatic methods. For example, Bernhard Witkop and Erhard Gross discovered that cyanogen bromide (CNBr) splits the polypeptide chain only on the carboxyl side of methionine residues (Figure 2-29). A protein that has ten methionines will usually yield eleven



Methionine **Homoserine lactone**

Figure 2-29
Cyanogen bromide cleaves polypeptides on the carboxyl side of methionine residues.

peptides on cleavage with CNBr. Highly specific cleavage is also obtained with trypsin, a proteolytic enzyme from intestinal juice. Trypsin cleaves polypeptide chains on the carboxyl side of arginine and lysine residues (Figure 2-30). A protein that contains nine

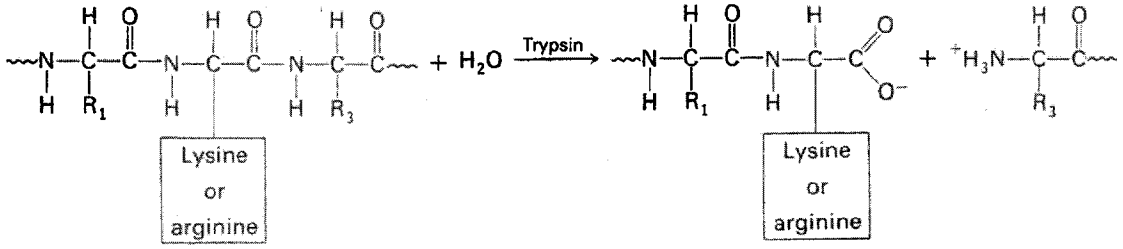


Figure 2-30
Trypsin hydrolyzes polypeptides on the carboxyl side of arginine and lysine residues.

lysines and seven arginines will usually yield seventeen peptides on digestion with trypsin. Each of these tryptic peptides, except for the carboxyl-terminal peptide of the protein, will end with either arginine or lysine. Several other ways of specifically cleaving polypeptide chains are given in Table 2-2.

Table 2-2
Specific cleavage of polypeptides

Reagent	*Cleavage site*
Chemical cleavage	
Cyanogen bromide	Carboxyl side of methionine residues
Hydroxylamine	Asparagine–glycine bonds
2-Nitro-5-thiocyanobenzoate	Amino side of cysteine residues
Enzymatic cleavage	
Trypsin	Carboxyl side of lysine and arginine residues
Clostripain	Carboxyl side of arginine residues
Staphylococcal protease	Carboxyl side of aspartate and glutamate residues (glutamate only under certain conditions)

The peptides obtained by specific chemical or enzymatic cleavage are separated by chromatographic methods. The sequence of each purified peptide is then determined by the Edman method. At this point, the amino acid sequences of segments of the protein are known, but the order of these segments is not yet defined. The necessary additional information is obtained from what are called *overlap peptides* (Figure 2-31). An enzyme different from trypsin is

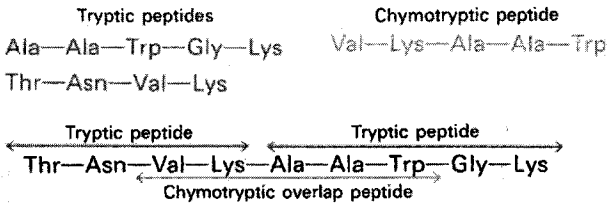


Figure 2-31
The peptide obtained by chymotryptic digestion overlaps two tryptic peptides, which thus establishes their order.

used to split the polypeptide chain at different linkages. For example, chymotrypsin cleaves preferentially on the carboxyl side of aromatic and other bulky nonpolar residues. Because these chymotryptic peptides overlap two or more tryptic peptides, they can be used to establish their order. The entire amino acid sequence of the protein is then determined.

These methods apply to a protein that consists of a single polypeptide chain without any disulfide bonds. Additional steps in the elucidation of sequence are necessary if a protein has disulfide bonds or more than one chain. For a protein made up of two or more polypeptide chains held together by noncovalent bonds, denaturing agents, such as urea or guanidine hydrochloride, are used to dissociate the chains. The dissociated chains must be separated before sequence determination can begin. For polypeptide chains that are covalently linked by disulfide bonds, as in insulin, oxidation with performic acid is used to cleave the disulfide bonds, yielding cysteic acid residues (Figure 2-32).

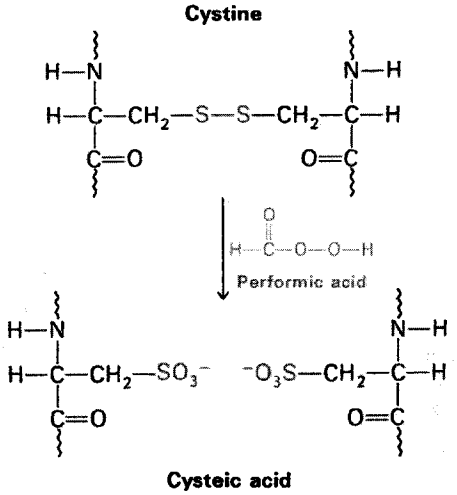


Figure 2-32
Performic acid cleaves disulfides.

Analyses of protein structures have been markedly accelerated by the development of the *sequenator,* an instrument for the automatic determination of amino acid sequence. A thin film of protein in a spinning cylindrical cup is subjected to the Edman degradation. The reagents and extracting solvents are passed over the immobilized film of protein, and the released PTH-amino acid is identified by high-pressure liquid chromatography. One cycle of the Edman degradation is carried out in less than two hours. The sequenator can determine the amino acid sequence of a polypeptide or a protein containing as many as a hundred residues.

CONFORMATION OF POLYPEPTIDE CHAINS

A striking characteristic of proteins is that they have well-defined three-dimensional structures. A stretched-out or randomly arranged polypeptide chain is devoid of biological activity, as will be discussed shortly. Function arises from *conformation,* which is the three-dimensional arrangement of atoms in a structure. Amino acid sequences are important because they specify the conformation of proteins.

In the late 1930s, Linus Pauling and Robert Corey initiated x-ray crystallographic studies of the precise structure of amino acids and peptides. Their aim was to obtain a set of standard bond distances

Angstrom (Å)—
A unit of length equal to 10^{-10} meter.
$1\ Å = 10^{-10}\ m = 10^{-8}\ cm = 10^{-4}\ \mu m = 10^{-1}\ nm$
Named after Anders J. Ångström (1814–1874), a spectroscopist.

and bond angles for these building blocks and then use this information to predict the conformation of proteins. One of their important findings was that *the peptide unit is rigid and planar*. The hydrogen of the substituted amino group is nearly always *trans* to the oxygen of the carbonyl group (Figure 2-33). There is no freedom of

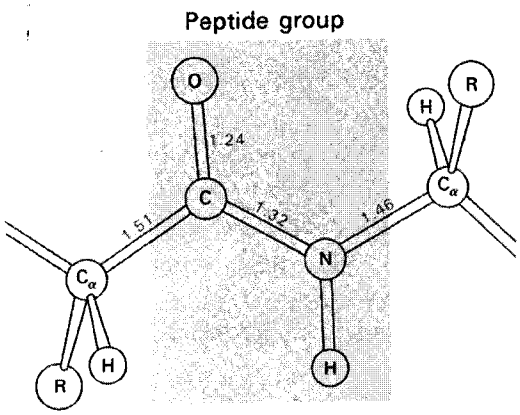


Figure 2-33
The peptide group is a rigid planar unit. Standard bond distances (in Å) are shown.

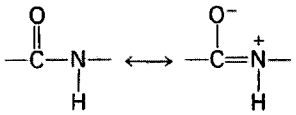


Figure 2-34
The peptide group is planar because the carbon–nitrogen bond has partial double-bond character.

rotation about the bond between the carbonyl carbon atom and the nitrogen atom of the peptide unit because this link has partial double-bond character (Figure 2-34). The length of this bond is 1.32 Å, which is between that of a C—N single bond (1.49 Å) and a C=N double bond (1.27 Å). In contrast, the link between the α-carbon atom and a carbonyl carbon atom is a pure single bond. The bond between an α-carbon atom and a nitrogen atom also is a pure single bond. Consequently, *there is a large degree of rotational freedom about these bonds on either side of the rigid peptide unit* (Figure 2-35). Rotations about these bonds are designated by the angles ψ and ϕ (Figure 2-36). The conformation of the main chain of a polypeptide is fully defined when ψ and ϕ for each amino acid residue are known.

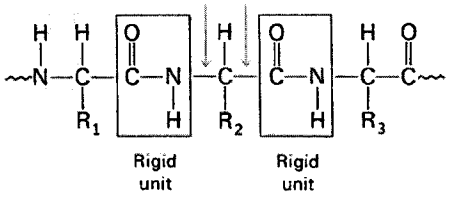


Figure 2-35
There is considerable freedom of rotation about the bonds joining the peptide groups to the α-carbon atoms.

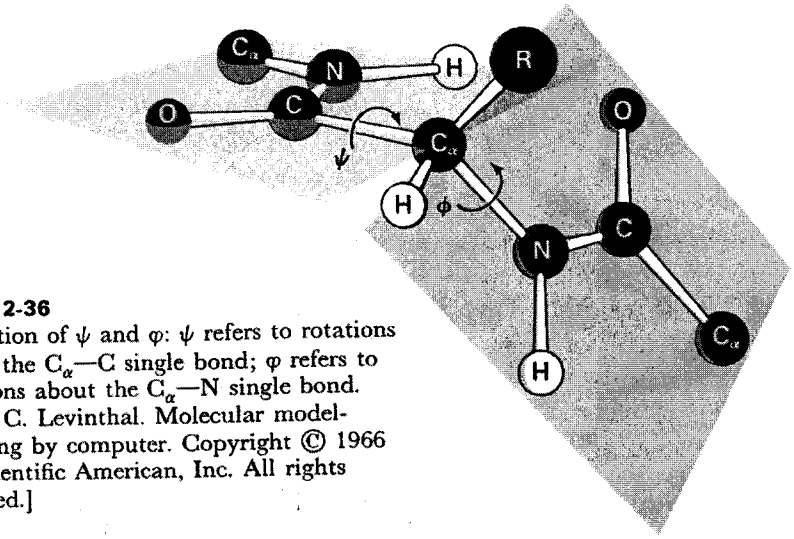


Figure 2-36
Definition of ψ and φ: ψ refers to rotations about the C_α—C single bond; φ refers to rotations about the C_α—N single bond. [After C. Levinthal. Molecular model-building by computer. Copyright © 1966 by Scientific American, Inc. All rights reserved.]

PERIODIC STRUCTURES: THE ALPHA HELIX, BETA PLEATED SHEET, AND COLLAGEN HELIX

Can a polypeptide chain fold into a regularly repeating structure? To answer this question, Pauling and Corey evaluated a variety of potential polypeptide conformations by building precise molecular models of them. They adhered closely to the experimentally observed bond angles and distances for amino acids and small peptides. In 1951, they proposed two periodic polypeptide structures, called the α helix and β pleated sheet.

The *α helix* is a rodlike structure. The tightly coiled polypeptide main chain forms the inner part of the rod, and the side chains extend outward in a helical array (Figures 2-37 and 2-38). The α

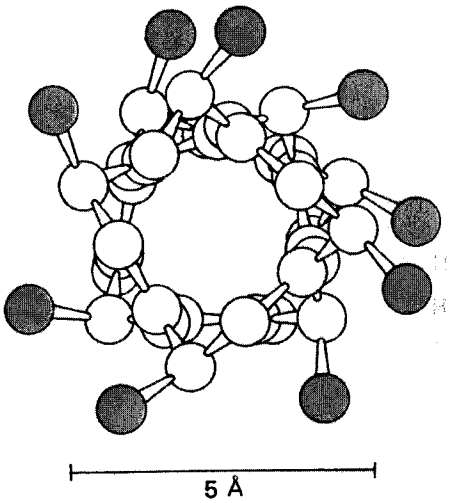
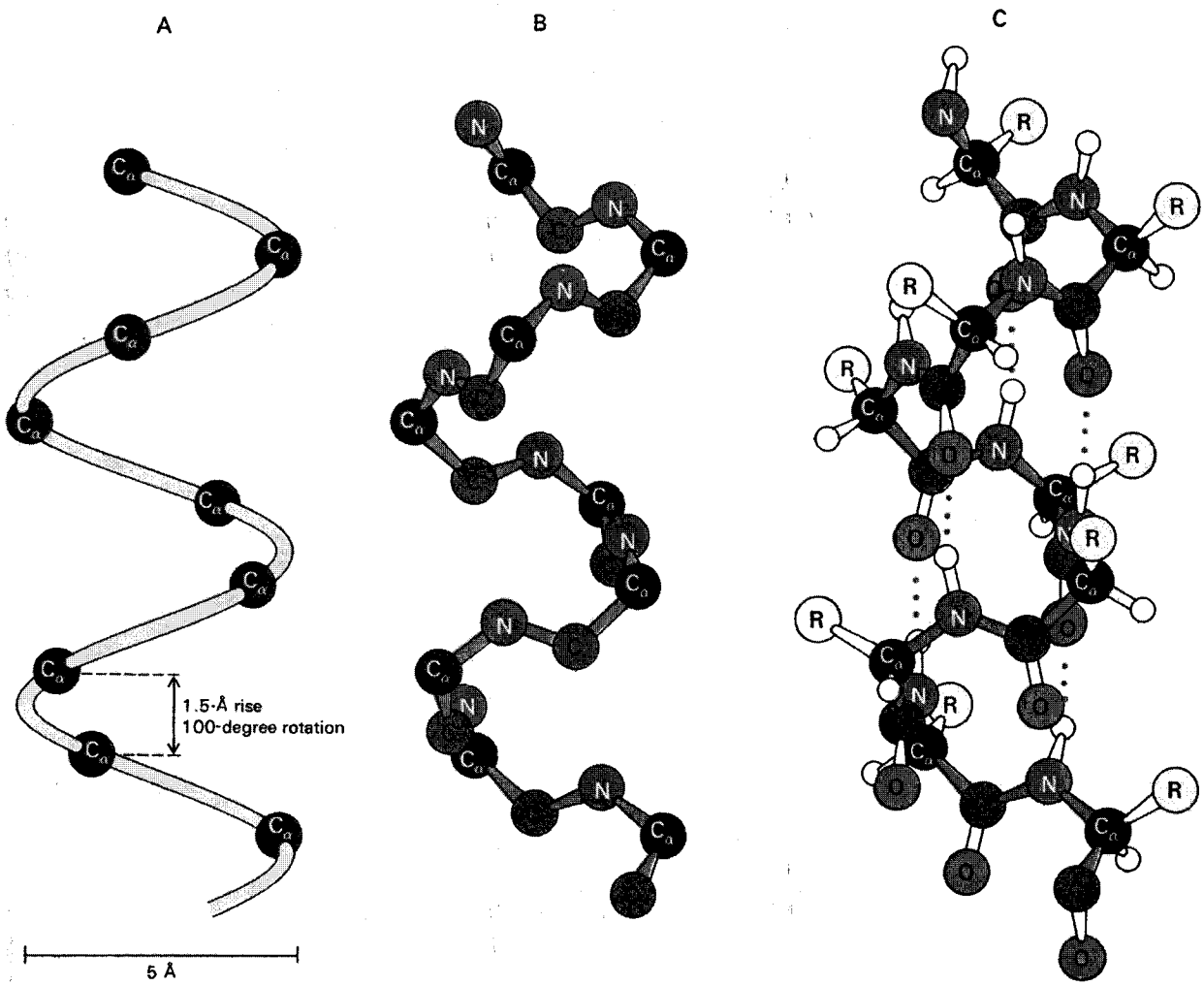


Figure 2-38
Cross-sectional view of an α helix. Note that the side chains (shown in green) are on the outside of the helix. The van der Waals radii of the atoms are larger than shown here; hence there is almost no free space inside the helix.

Figure 2-37
Models of a right-handed α helix: (A) only the α-carbon atoms are shown on a helical thread; (B) only the backbone nitrogen (N), α-carbon (C_α), and carbonyl carbon (C) atoms are shown; (C) entire helix. Hydrogen bonds (denoted in part C by · · · in red) between NH and CO groups stabilize the helix.



helix is stabilized by hydrogen bonds between the NH and CO groups of the main chain. The CO group of each amino acid is hydrogen bonded to the NH group of the amino acid that is situated four residues ahead in the linear sequence (Figure 2-39). Thus,

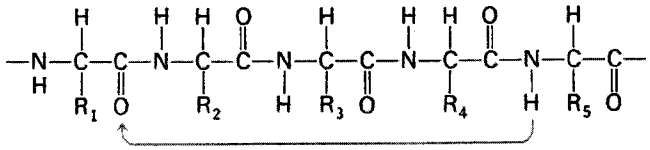


Figure 2-39
In the α helix, the NH group of residue n is hydrogen bonded to the CO group of residue $(n - 4)$.

> "When we consider that the fibrous proteins of the epidermis, the keratinous tissues, the chief muscle protein, myosin, and now the fibrinogen of the blood all spring from the same peculiar shape of molecule, and are therefore probably all adaptations of a single root idea, we seem to glimpse one of the great coordinating facts in the lineage of biological molecules."
>
> K. BAILEY, W. T. ASTBURY, AND K. M. RUDALL
> *Nature*, 1943

all the main-chain CO and NH groups are hydrogen bonded. Each residue is related to the next one by a translation of 1.5 Å along the helix axis and a rotation of 100°, which gives 3.6 amino acid residues per turn of helix. Thus, amino acids spaced three and four apart in the linear sequence are spatially quite close to one another in an α helix. In contrast, amino acids two apart in the linear sequence are situated on opposite sides of the helix and so are unlikely to make contact. The pitch of the α helix is 5.4 Å, the product of the translation (1.5 Å) and the number of residues per turn (3.6). The screw-sense of a helix can be right-handed (clockwise) or left-handed (counterclockwise); the α helices found in proteins are right-handed.

The α-helix content of proteins of known three-dimensional structure is highly variable. In some, such as myoglobin and hemoglobin, the α helix is the major structural motif. Other proteins, such as the digestive enzyme chymotrypsin, are virtually devoid of α helix. The single-stranded α helix discussed above is usually a rather short rod, typically less than 40 Å in length. A variation of the α-helical theme is used to construct much longer rods, extending to 1000 Å or more. Two or more α helices can entwine around each other to form a cable. Such *α-helical coiled coils* are found in several proteins: keratin in hair, myosin and tropomyosin in muscle, epidermin in skin, and fibrin in blood clots. The helical cables in these proteins serve a mechanical role in forming stiff bundles of fibers.

The structure of the α helix was deduced by Pauling and Corey six years before it was actually to be seen in the x-ray reconstruction of the structure of myoglobin. *The elucidation of the structure of the α helix is a landmark in molecular biology because it demonstrated that the conformation of a polypeptide chain can be predicted if the properties of its components are rigorously and precisely known.*

In the same year, Pauling and Corey discovered another periodic structural motif, which they named the β pleated sheet (β because it was the second structure they elucidated, the α helix having been the first). The *β pleated sheet* differs markedly from the α helix in that it is a sheet rather than a rod. The polypeptide chain in the β pleated sheet is almost fully extended (Figure 2-40) rather than

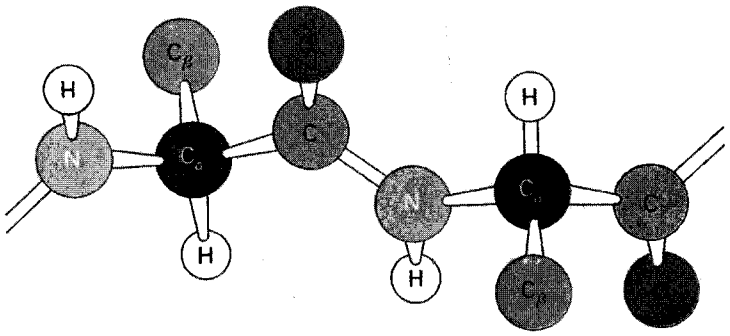


Figure 2-40
Conformation of a dipeptide unit in a β pleated sheet. The polypeptide chain is almost fully stretched out.

being tightly coiled as in the α helix. The axial distance between adjacent amino acids is 3.5 Å, in contrast with 1.5 Å for the α helix. Another difference is that the β pleated sheet is stabilized by hydrogen bonds between NH and CO groups in *different* polypeptide strands, whereas in the α helix the hydrogen bonds are between NH and CO groups in the *same* polypeptide chain. Adjacent strands in a β pleated sheet can run in the same direction (*parallel β sheet*) or in opposite directions (*antiparallel β sheet*). For example, silk fibroin consists almost entirely of stacks of antiparallel β sheets (Figure 2-41). Such β-sheet regions are a recurring structural motif in many proteins. Structural units comprising from two to five parallel or antiparallel β strands are especially common.

The *collagen helix,* a third periodic structure, will be discussed in detail in Chapter 9. This specialized structure is responsible for the high tensile strength of collagen, the major component of skin, bone, and tendon.

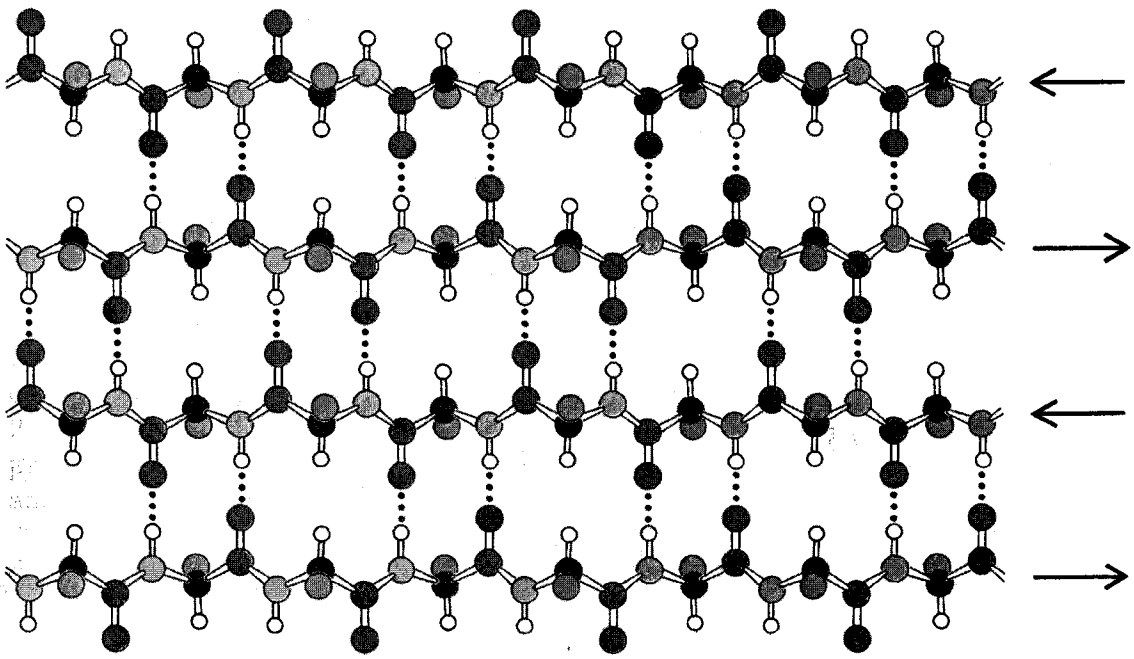


Figure 2-41
Antiparallel β pleated sheet. Adjacent strands run in opposite directions. Hydrogen bonds between NH and CO groups of adjacent strands stabilize the structure. The side chains (shown in green) are above and below the plane of the sheet.

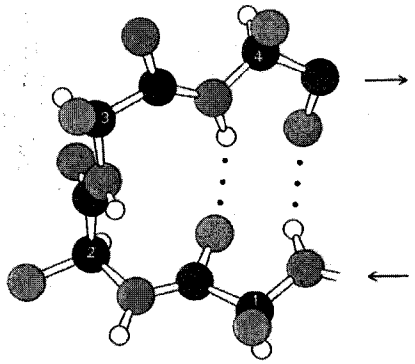


Figure 2-42
Structure of a β-turn. The CO group of residue 1 of the tetrapeptide shown here is hydrogen bonded to the NH group of residue 4, which results in a hairpin turn.

POLYPEPTIDE CHAINS CAN REVERSE DIRECTION BY MAKING β-TURNS

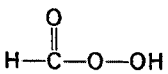
Most proteins have compact, globular shapes due to frequent reversals of the direction of their polypeptide chains. Analyses of the three-dimensional structures of numerous proteins have revealed that many of these chain reversals are accomplished by a common structural element called the *β-turn*. The essence of this hairpin turn is that the CO group of residue n of a polypeptide is hydrogen bonded to the NH group of residue $(n + 3)$ (Figure 2-42). Thus, a polypeptide chain can abruptly reverse its direction.

LEVELS OF STRUCTURE IN PROTEIN ARCHITECTURE

In discussing the architecture of proteins, it is convenient to refer to four levels of structure. *Primary structure* is simply the sequence of amino acids and location of disulfide bridges, if there are any. The primary structure is thus a complete description of the covalent connections of a protein. *Secondary structure* refers to the steric relationship of amino acid residues that are close to one another in the linear sequence. Some of these steric relationships are of a regular kind, giving rise to a periodic structure. The α helix, the β pleated sheet, and the collagen helix are examples of secondary structure. *Tertiary structure* refers to the steric relationship of amino acid residues that are far apart in the linear sequence. It should be noted that the dividing line between secondary and tertiary structure is arbitrary. Proteins that contain more than one polypeptide chain display an additional level of structural organization, namely *quaternary structure*, which refers to the way in which the chains are packed together. Each polypeptide chain in such a protein is called a *subunit*. Another useful term is *domain*, which refers to a compact, globular unit of protein structure. Many proteins fold into domains having masses that range from 10 to 20 kdal. The domains of large proteins are usually connected by relatively flexible regions of polypeptide chain.

AMINO ACID SEQUENCE SPECIFIES THREE-DIMENSIONAL STRUCTURE

Insight into the relation between the amino acid sequence of a protein and its conformation came from the work of Christian Anfinsen on ribonuclease, an enzyme that hydrolyzes RNA. Ribonuclease is a single polypeptide chain consisting of 124 amino acid residues (Figure 2-43). It contains four disulfide bonds, which can be irreversibly oxidized by *performic acid* to give cysteic acid residues



Performic acid

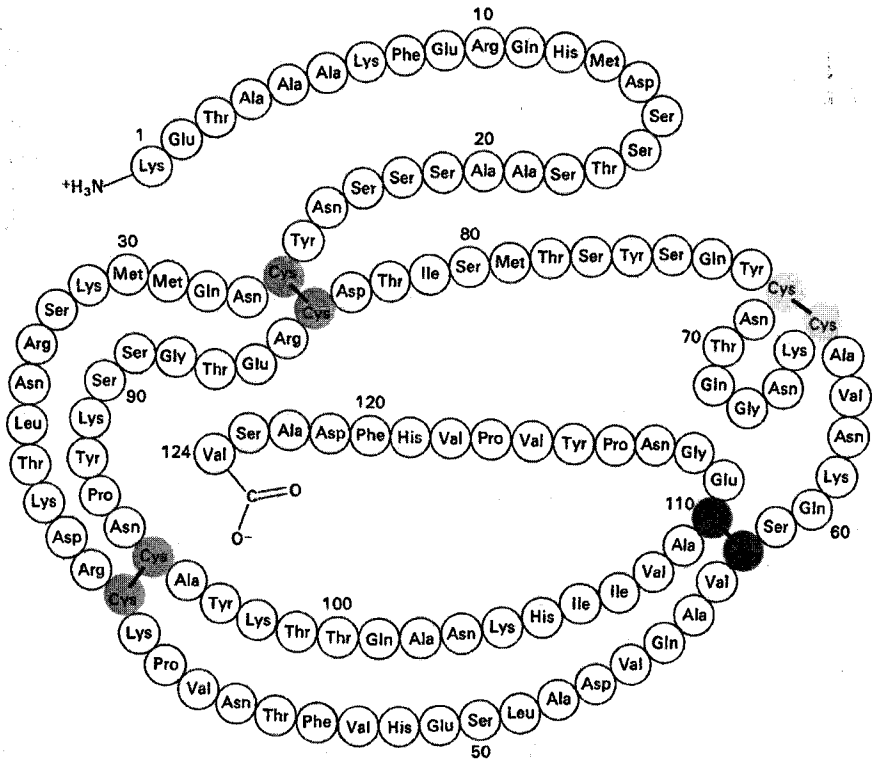


Figure 2-43
Amino acid sequence of bovine ribonuclease. The four disulfide bonds are shown in color. [After D. G. Smyth, W. H. Stein, and S. Moore. *J. Biol. Chem.* 238(1963):277.]

(Figure 2-32). Alternatively, these disulfide bonds can be cleaved reversibly by reducing them with a reagent such as *β-mercaptoethanol,* which forms mixed disulfides with cysteine side chains (Figure 2-44). In the presence of a large excess of β-mercap-

$HO—CH_2—CH_2—SH$
β-Mercaptoethanol

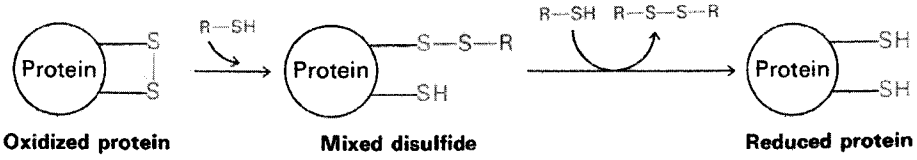


Figure 2-44
Reduction of the disulfide bonds in a protein by an excess of a sulfhydryl reagent such as β-mercaptoethanol.

toethanol, the mixed disulfides also are reduced, so that the final product is a protein in which the disulfides (cystines) are fully converted into sulfhydryls (cysteines). However, it was found that ribonuclease at 37°C and pH 7 cannot be readily reduced by β-mercaptoethanol unless the protein is partially unfolded by denaturing agents such as *urea* or *guanidine hydrochloride.* Although the mecha-

$H_2N—C(=O)—NH_2$
Urea

$H_2N—C(=NH_2^+\ Cl^-)—NH_2$
Guanidine hydrochloride

nism of action of these denaturing agents is not fully understood, it is evident that they disrupt noncovalent interactions. Polypeptide chains devoid of cross-links usually assume a *random-coil* conformation in 8 M urea or 6 M guanidine HCl, as evidenced by physical properties such as viscosity and optical rotary spectra. When ribonuclease was treated with β-mercaptoethanol in 8 M urea, the product was a fully reduced, randomly coiled polypeptide chain *devoid of enzymatic activity*. In other words, ribonuclease was *denatured* by this treatment (Figure 2-45).

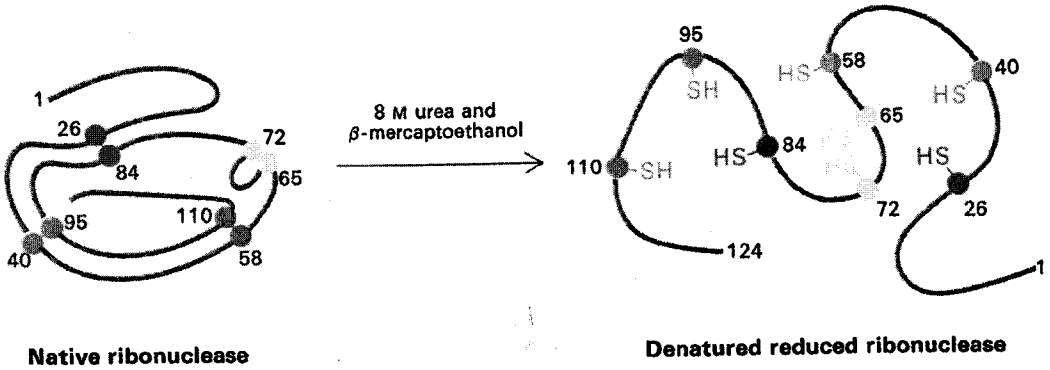


Figure 2-45
Reduction and denaturation of ribonuclease.

Anfinsen then made the critical observation that the denatured ribonuclease, freed of urea and β-mercaptoethanol by dialysis, slowly regained enzymatic activity. He immediately perceived the significance of this chance finding: the sulfhydryls of the denatured enzyme became oxidized by air and the enzyme spontaneously refolded into a catalytically active form. Detailed studies then showed that nearly all of the original enzymatic activity was regained if the sulfhydryls were oxidized under suitable conditions (Figure 2-46). All of the measured physical and chemical properties of the refolded enzyme were virtually identical with those of the native enzyme. These experiments showed that *the information needed to specify the complex three-dimensional structure of ribonuclease is contained in its amino acid sequence*. Subsequent studies of other proteins have

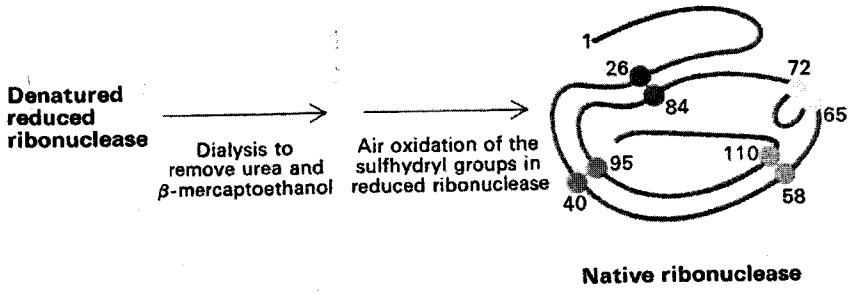


Figure 2-46
Renaturation of ribonuclease.

established the generality of this principle, which is a central one in molecular biology: *sequence specifies conformation.*

A quite different result was obtained when reduced ribonuclease was reoxidized while it was still in 8 M urea. This preparation was then dialyzed to remove the urea. Ribonuclease reoxidized in this way had only 1% of the enzymatic activity of the native protein. Why was the outcome of this experiment different from the one in which reduced ribonuclease was reoxidized in a solution free of urea? The reason is that wrong disulfide pairings were formed when the random-coil form of the reduced molecule was reoxidized. There are 105 different ways of pairing eight cysteines to form four disulfides; only one of these combinations is enzymatically active. The 104 wrong pairings have been picturesquely termed "scrambled" ribonuclease. Anfinsen then found that "scrambled" ribonuclease spontaneously converted into fully active, native ribonuclease when trace amounts of β-mercaptoethanol were added to the aqueous solution of the reoxidized protein (Figure 2-47). The added

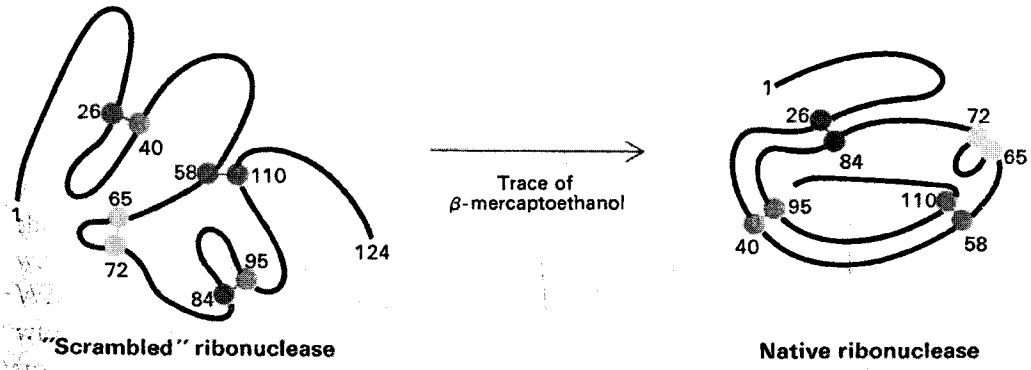


Figure 2-47
Formation of native ribonuclease from "scrambled" ribonuclease in the presence of a trace of β-mercaptoethanol.

β-mercaptoethanol catalyzed the rearrangement of disulfide pairings until the native structure was regained, which took about ten hours. This process was driven entirely by the decrease in free energy as the "scrambled" conformations were converted into the stable, native conformation of the enzyme. *Thus, the native form of ribonuclease appears to be the thermodynamically most stable structure.*

Anfinsen (1964) wrote:

> It struck me recently that one should really consider the sequence of a protein molecule, about to fold into a precise geometric form, as a line of melody written in canon form and so designed by Nature to fold back upon itself, creating harmonic chords of interaction consistent with biological function. One might carry the analogy further by suggesting that the kinds of chords formed in a protein with scrambled disulfide bridges, such as I mentioned earlier, are dissonant, but that, by giving an opportunity for rearrangement by the addition of mercaptoethanol, they modulate to give the pleasing harmonics of the

native molecule. Whether or not some conclusion can be drawn about the greater thermodynamic stability of Mozart's over Schoenberg's music is something I will leave to the philosophers of the audience.

PROTEINS FOLD BY THE ASSOCIATION OF α-HELICAL AND β-STRAND SEGMENTS

How are the harmonic chords of interaction created in the conversion of an unfolded polypeptide chain into a folded protein? One possibility a priori is that all possible conformations are searched to find the energetically most favorable form. How long would such a random search take? Consider a small protein with 100 residues. If each residue can assume three different conformations, the total number of structures is 3^{100}, which is equal to 5×10^{47}. If it takes 10^{-13} seconds to convert one structure into another, the total search time would be $5 \times 10^{47} \times 10^{-13}$ seconds, which is equal to 5×10^{34} seconds, or 1.6×10^{27} years! Note that this length of time is a minimal estimate because the actual number of possible conformations per residue is greater than three and the time that it takes to change from one conformation into another is probably considerably longer than 10^{-13} seconds. Clearly, it would take much too long for even a small protein to fold by randomly trying out all possible conformations to determine which one is energetically best.

How, then, do proteins fold in a few seconds or minutes? The answer is not yet known, but a plausible hypothesis is that *small stretches of secondary structure serve as intermediates in the folding process.* According to this model, short segments (~15 residues) of an unfolded polypeptide chain flicker in and out of their native α-helical or β-sheet form. These transient structures find each other by diffusion and stabilize each other by forming a complex (Figure 2-48). For example, two α helices, two β strands, or an α helix and a β strand may come together. These αα, ββ, and αβ complexes, which are called *folding units,* then act as nuclei to stabilize other flickering elements of secondary structure. This model is supported by several lines of experimental evidence. The first is that the tendency of a polypeptide to adopt a regular secondary structure depends to a large degree on its amino acid composition. The formation of an α helix is favored by glutamate, methionine, alanine, and leucine residues, whereas β-sheet formation is enhanced by valine, isoleucine, and tyrosine residues. Second, the transition from a random coil to an α helix can occur in less than a microsecond. Thus, short segments of secondary structure can be formed very rapidly. Third, the postulated folding units (αα, ββ, and αβ complexes) are in fact major elements in protein structure. The challenge now is to directly detect and identify the evanescent intermediates in folding and recreate the pathway that gives form and function to the polypeptide chain.

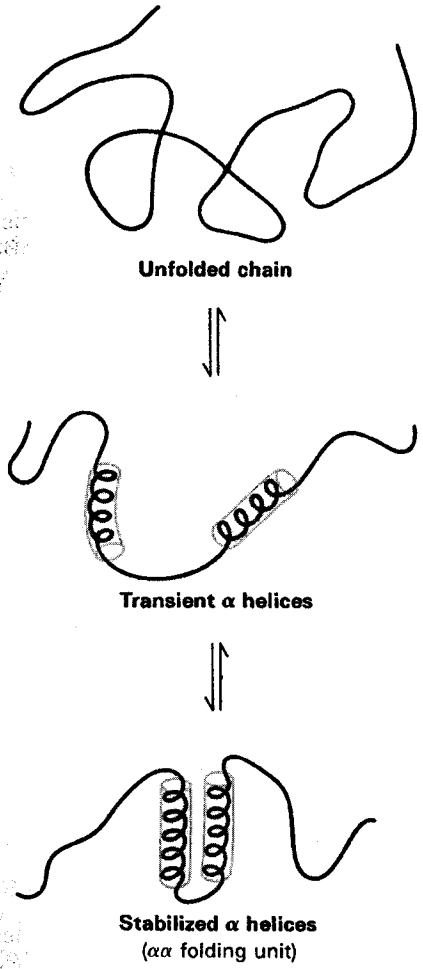


Figure 2-48
Postulated step in protein folding. Two segments of an unfolded polypeptide chain transiently become α helical. These helices are then stabilized by the formation of a complex between the two segments.

SUMMARY

Proteins play key roles in nearly all biological processes. All enzymes, the catalysts of chemical reactions in biological systems, are proteins. Hence, proteins determine the pattern of chemical transformations in cells. Proteins mediate a wide range of other functions, such as transport and storage, coordinated motions, mechanical support, immune protection, excitability, and the control of growth and differentiation.

The basic structural units of proteins are amino acids. All proteins in all species from bacteria to humans are constructed from the same set of twenty amino acids. The side chains of these buildings blocks differ in size, shape, charge, hydrogen-bonding capacity, and chemical reactivity. They can be grouped as follows: (a) aliphatic side chains—glycine, alanine, valine, leucine, isoleucine, and proline; (b) hydroxyl aliphatic side chains—serine and threonine; (c) aromatic side chains—phenylalanine, tyrosine, and tryptophan; (d) basic side chains—lysine, arginine, and histidine; (e) acidic side chains—aspartic acid and glutamic acid; (f) amide side chains—asparagine and glutamine; and (g) sulfur side chains—cysteine and methionine.

Many amino acids, usually more than a hundred, are joined by peptide bonds to form a polypeptide chain. A peptide bond links the α-carboxyl group of one amino acid and the α-amino group of the next one. In some proteins, a few side chains are cross-linked by disulfide bonds, which result from the oxidation of cysteine residues. A protein consists of one or more polypeptide chains. Each kind of protein has a unique amino acid sequence that is genetically determined. The amino acid sequence of a protein is elucidated in the following way. First, its amino acid composition is determined by ion-exchange chromatography of an acid hydrolysate of the protein. The amino-terminal residue is identified by using an end-group reagent such as dansyl chloride. Second, the protein is specifically cleaved into small peptides. For example, trypsin hydrolyzes proteins on the carboxyl side of lysine and arginine residues. Third, the amino acid sequence of these peptides is then determined by the Edman technique, which successively removes the amino-terminal residue. Finally, the order of the peptides is established from the amino acid sequences of overlap peptides.

The critical determinant of the biological function of a protein is its conformation, which is defined as the three-dimensional arrangement of the atoms of a molecule. Three regularly repeating conformations of polypeptide chains are known: the α helix, the β pleated sheet, and the collagen helix. Short segments of the α helix and the β pleated sheet are found in many proteins. An important principle is that the amino acid sequence of a protein specifies its three-dimensional structure, as was first shown for ribonuclease. Reduced, unfolded ribonuclease spontaneously forms the correct

disulfide pairings and regains full enzymatic activity when oxidized by air following removal of mercaptoethanol and urea. Proteins fold by the association of short polypeptide segments that transiently adopt α-helical or β-sheet forms.

SELECTED READINGS

WHERE TO START

Dickerson, R. E. and Geis, I., forthcoming. *Proteins: Structure, Function, and Evolution.* Benjamin. [A lucid and well-illustrated introduction to proteins.]

Moore, S., and Stein, W. H., 1973. Chemical structures of pancreatic ribonuclease and deoxyribonuclease. *Science* 180:458–464.

Anfinsen, C. B., 1973. Principles that govern the folding of protein chains. *Science* 181:223–230.

BOOKS ON PROTEIN CHEMISTRY

Cantor, C. R., and Schimmel, P. R., 1980. *Biophysical Chemistry.* Freeman. [Chapters 2 and 5 in Part I and Chapters 20 and 21 in Part III give an excellent account of the principles of protein conformation.]

Cooper, T. G., 1977. *The Tools of Biochemistry.* Wiley. [A valuable guide to experimental methods in protein chemistry and other facets of biochemistry. Principles of procedures are clearly presented.]

Neurath, H., and Hill, R. L., (eds.), 1976. *The Proteins* (3rd ed.). Academic Press. [A multivolume treatise that contains many fine articles.]

Haschemeyer, R. H., and Haschemeyer, A. E. V., 1973. *Proteins: A Guide to Study by Physical and Chemical Methods.* Wiley-Interscience.

Hirs, C. H. W., and Timasheff, S. N., (eds.), 1977. *Enzyme Structure,* Part E. Methods in Enzymology, vol. 47. Academic Press. [An excellent collection of authoritative articles on amino acid analysis, end-group methods, chemical and enzymatic cleavage, peptide separation, sequence analysis, chemical modification, and peptide synthesis.]

Schultz, G. E., and Schirmer, R. H., 1979. *Principles of Protein Structure.* Springer-Verlag.

AMINO ACID SEQUENCE DETERMINATION

Konigsberg, W. H., and Steinman, H. M., 1977. Strategy and methods of sequence analysis. *In* Neurath, H., and Hill, R. L., (eds.), *The Proteins* (3rd ed.), vol. 3, pp. 1–178. Academic Press.

Niall, H., 1977. Automated methods for sequence analysis. *In* Neurath, H., and Hill, R. L., (eds.), *The Proteins* (3rd ed.), vol. 3, pp. 179–238. Academic Press.

Needleman, S. B., (ed.), 1977. *Advanced Methods in Protein Sequence Determination.* Springer-Verlag.

Dayhoff, M. O., (ed.), 1972. *Atlas of Protein Sequence and Structure.* National Biomedical Research Foundation, Silver Springs, Maryland. [A valuable compilation of all known amino acid sequences. This atlas is revised every few years.]

COVALENT MODIFICATION OF PROTEINS

Uy, R., and Wold, F., 1977. Post-translational covalent modification of proteins. *Science* 198:890–896.

Hershko, A., and Fry, M., 1975. Post-translational cleavage of polypeptide chains: role in assembly. *Ann. Rev. Biochem.* 44:775–797.

Glazer, A. N., Delange, R. J., and Sigman, D. S., 1975. *Chemical Modification of Proteins.* North-Holland.

FOLDING OF PROTEINS

Creighton, T. E., 1978. Experimental studies of protein folding and unfolding. *Prog. Biophys. Mol. Biol.* 33:231–297.

Baldwin, R. L., 1975. Intermediates in protein folding reactions and the mechanism of protein folding. *Ann. Rev. Biochem.* 44:453–475.

Chou, P. Y., and Fasman, C. D., 1978. Empirical predictions of protein conformation. *Ann. Rev. Biochem.* 47:251–276.

Sternburg, M. J. E., and Thornton, J. M., 1978. Prediction of protein structure from amino acid sequence, *Nature* 271:15–20.

Karplus, M., and Weaver, D. L., 1976. Protein-folding dynamics. *Nature* 260:404–406.

Levitt, M., and Chothia, C., 1976. Structural patterns in globular proteins. *Nature* 261:552–558.

Chothia, C., 1975. Structural invariants in protein folding. *Nature* 254:304–308.

Anfinsen, C. B., 1964. On the possibility of predicting tertiary structure from primary sequence. *In* Sela, M., (ed.), *New Perspectives in Biology,* pp. 42–50. American Elsevier.

REVIEW ARTICLES

Advances in Protein Chemistry.
Annual Review of Biochemistry.
Annual Review of Biophysics and Bioengineering.

APPENDIX
Acid-Base Concepts

Ionization of Water

Water dissociates into hydronium (H_3O^+) and hydroxyl (OH^-) ions. For simplicity, we refer to the hydronium ion as a hydrogen ion (H^+) and write the equilibrium as

$$H_2O \rightleftharpoons H^+ + OH^-$$

The equilibrium constant K_{eq} of this dissociation is given by

$$K_{eq} = \frac{[H^+][OH^-]}{[H_2O]} \qquad (1)$$

in which the terms in brackets denote molar concentrations. Because the concentration of water (55.5 M) is changed little by ionization, expression 1 can be simplified to give

$$K_w = [H^+][OH^-] \qquad (2)$$

in which K_w is the ion product of water. At 25°C, K_w is 1.0×10^{-14}.

Note that the concentrations of H^+ and OH^- are reciprocally related. If the concentration of H^+ is high, then the concentration of OH^- must be low, and vice versa. For example, if $[H^+] = 10^{-2}$ M, then $[OH^-] = 10^{-12}$ M.

Definition of Acid and Base

An acid is a proton donor. A base is a proton acceptor.

$$\text{Acid} \rightleftharpoons H^+ + \text{Base}$$

$$\underset{\text{Acetic acid}}{CH_3\text{—}COOH} \rightleftharpoons H^+ + \underset{\text{Acetate}}{CH_3\text{—}COO^-}$$

$$\underset{\text{Ammonium ion}}{NH_4^+} \rightleftharpoons H^+ + \underset{\text{Ammonia}}{NH_3}$$

The species formed by the ionization of an acid is its conjugate base. Conversely, protonation of a base yields its conjugate acid. Acetic acid and acetate ion are a conjugate acid-base pair.

Definition of pH and p*K*

The pH of a solution is a measure of its concentration of H^+. The pH is defined as

$$pH = \log_{10}\frac{1}{[H^+]} = -\log_{10}[H^+] \qquad (3)$$

The ionization equilibrium of a weak acid is given by

$$HA \rightleftharpoons H^+ + A^-$$

The apparent equilibrium constant K for this ionization is

$$K = \frac{[H^+][A^-]}{[HA]} \qquad (4)$$

The pK of an acid is defined as

$$pK = -\log K = \log\frac{1}{K} \qquad (5)$$

Inspection of equation 4 shows that *the pK of an acid is the pH at which it is half-dissociated.*

Henderson-Hasselbalch Equation

What is the relationship between pH and the ratio of acid to base? A useful expression can be derived starting with equation 4. Rearrangement of that equation gives

$$\frac{1}{[H^+]} = \frac{1}{K}\frac{[A^-]}{[HA]} \qquad (6)$$

Taking the logarithm of both sides of equation 6 gives

$$\log\frac{1}{[H^+]} = \log\frac{1}{K} + \log\frac{[A^-]}{[HA]} \qquad (7)$$

Substituting pH for $\log 1/[H^+]$ and pK for $\log 1/K$ in equation 7 yields

$$pH = pK + \log\frac{[A^-]}{[HA]} \qquad (8)$$

which is commonly known as the Henderson-Hasselbalch equation.

The pH of a solution can be calculated from equation 8 if the molar proportion of A^- to HA and the pK of HA are known. Consider a solution of 0.1 M acetic acid and 0.2 M acetate ion. The pK of acetic acid is 4.8. Hence, the pH of the solution is given by

$$\text{pH} = 4.8 + \log \frac{0.2}{0.1} = 4.8 + \log 2$$
$$= 4.8 + 0.3 = 5.1$$

Conversely, the pK of an acid can be calculated if the molar proportion of A^- to HA and the pH of the solution are known.

Buffering Power

An acid-base conjugate pair (such as acetic acid and acetate ion) has an important property: it resists changes in the pH of a solution. In other words, it acts as a *buffer*. Consider the addition of OH^- to a solution of acetic acid (HA):

$$HA + OH^- \rightleftharpoons A^- + H_2O$$

A plot of the dependence of the pH of this solution on the amount of OH^- added is called a *titration curve*. Note that there is an inflection point in the curve at pH 4.8, which is the pK of acetic acid.

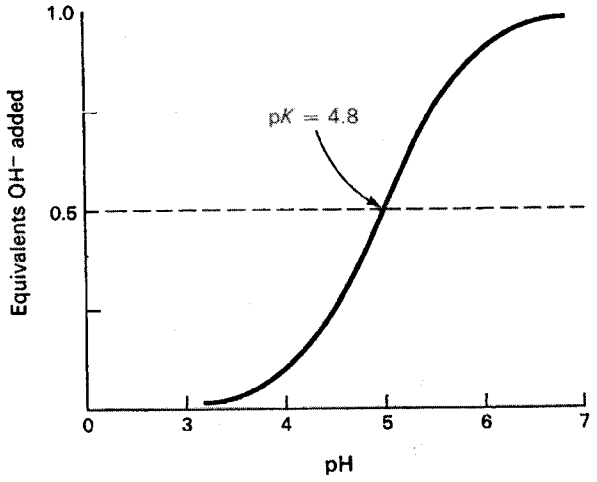


Figure 2-49
Titration curve of acetic acid.

In the vicinity of this pH, a relatively large amount of OH^- produces little change in pH. In general, a weak acid is most effective in buffering against pH changes in the vicinity of its pK value.

pK Values of Amino Acids

An amino acid such as glycine contains two ionizable groups: an α-carboxyl group and a protonated α-amino group. As base is added, these two groups are titrated (Figure 2-50). The pK of the α-COOH group is 2.3, whereas that of the α-NH_3^+ group is 9.6. The pK values of these groups in other amino acids are similar. Some amino acids, such as aspartic acid, also contain an ionizable side chain. The pK values of ionizable side chains in amino acids range from 3.9 (aspartic acid) to 12.5 (arginine).

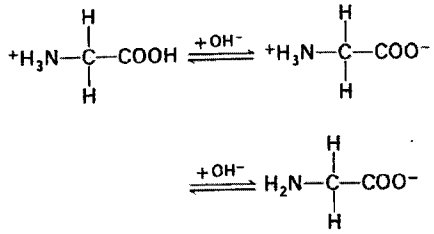


Figure 2-50
Titration of the ionizable groups of an amino acid.

Table 2-3
pK values of some amino acids

Amino acid	pK values (25°C)		
	α-COOH group	α-NH_3^+ group	Side chain
Alanine	2.3	9.9	
Glycine	2.4	9.8	
Phenylalanine	1.8	9.1	
Serine	2.1	9.2	
Valine	2.3	9.6	
Aspartic acid	2.0	10.0	3.9
Glutamic acid	2.2	9.7	4.3
Histidine	1.8	9.2	6.0
Cysteine	1.8	10.8	8.3
Tyrosine	2.2	9.1	10.9
Lysine	2.2	9.2	10.8
Arginine	1.8	9.0	12.5

After J. T. Edsall and J. Wyman. *Biophysical Chemistry* (Academic Press, 1958), ch. 8.

PROBLEMS

1. The following reagents are often used in protein chemistry:

CNBr	Dansyl chloride
Urea	6 N HCl
β-Mercaptoethanol	Ninhydrin
Trypsin	Phenyl isothiocyanate
Performic acid	Chymotrypsin

Which one is the best suited for accomplishing each of the following tasks?

(a) Determination of the amino acid sequence of a small peptide.
(b) Identification of the amino-terminal residue of a peptide (of which you have less than 10^{-7} g).
(c) Reversible denaturation of a protein devoid of disulfide bonds. Which additional reagent would you need if disulfide bonds were present?
(d) Hydrolysis of peptide bonds on the carboxyl side of aromatic residues.
(e) Cleavage of peptide bonds on the carboxyl side of methionines.
(f) Hydrolysis of peptide bonds on the carboxyl side of lysine and arginine residues.

2. What is the pH of each of the following solutions?

(a) 10^{-3} N HCl.
(b) 10^{-2} N NaOH.
(c) A mixture of equal volumes of 0.1 M acetic acid and 0.03 M sodium acetate.
(d) A mixture of equal volumes of 0.1 M glycine and 0.05 M NaOH.
(e) A mixture of equal volumes of 0.1 M glycine and 0.05 M HCl.

3. What is the ratio of base to acid at pH 4, 5, 6, 7, and 8 for an acid with a pK of 6?

4. Tropomyosin, a muscle protein, is a two-stranded α-helical coiled coil. The protein's mass is 70 kdal. The average residue is about 110 dal. What is the length of the molecule?

5. Anhydrous hydrazine has been used to cleave peptide bonds in proteins. What are the reaction products? How might this technique be used to identify the carboxyl-terminal amino acid?

6. The amino acid sequence of human adrenocorticotropin, a polypeptide hormone, is

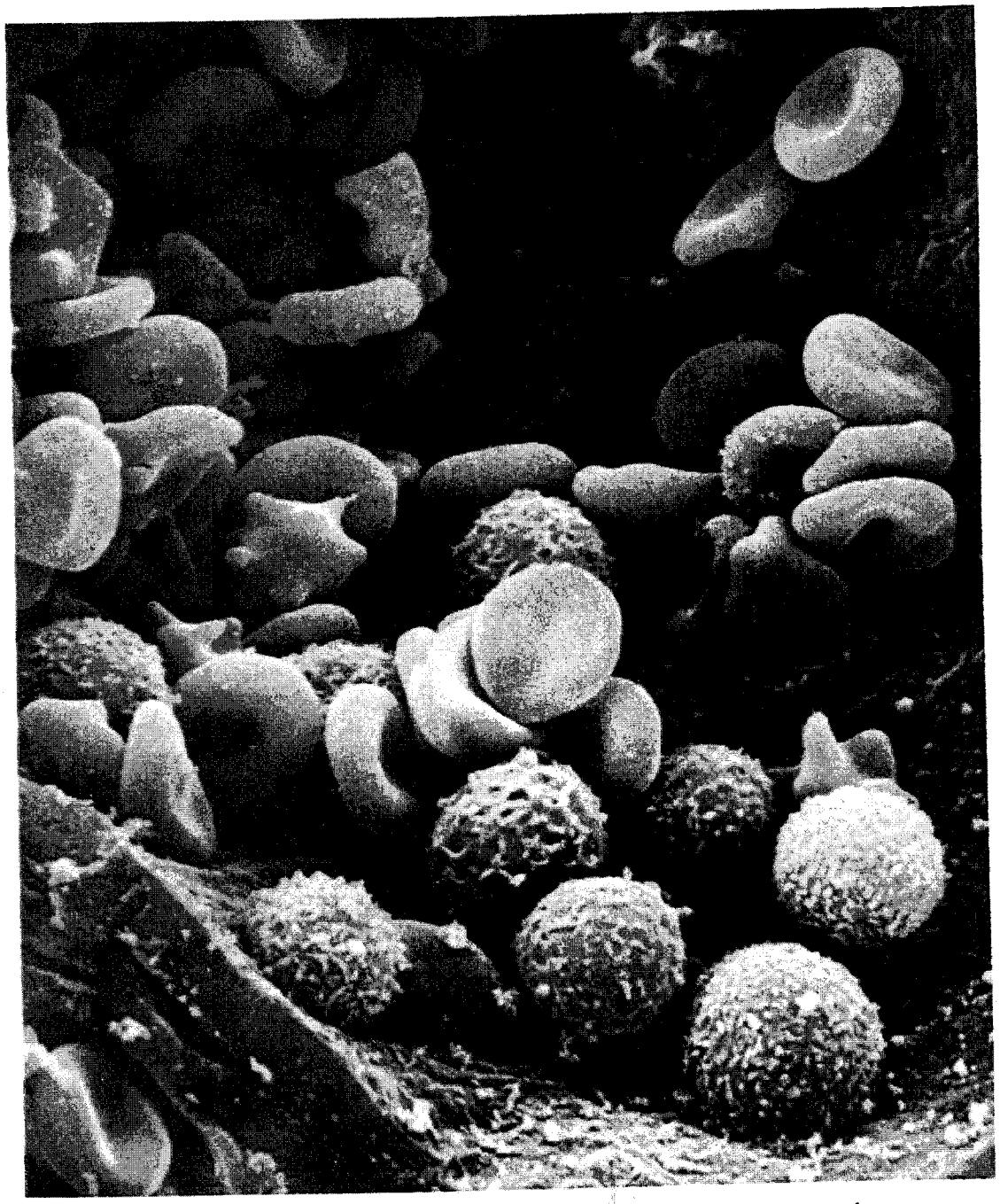
Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val-Gly-Lys-Lys-Arg-Arg-Pro-Val-Lys-Val-Tyr-Pro-Asp-Ala-Gly-Glu-Asp-Gln-Ser-Ala-Glu-Ala-Phe-Pro-Leu-Glu-Phe

(a) What is the approximate net charge of this molecule at pH 7? Assume that its side chains have the pK values given in Table 2-3 (p. 40) and that the pKs of the terminal $-NH_3^+$ and —COOH groups are 7.8 and 3.6, respectively.
(b) How many peptides result from the treatment of the hormone with cyanogen bromide?

7. Ethyleneimine reacts with cysteine side chains in proteins to form *S*-aminoethyl derivatives. The peptide bonds on the carboxyl side of these modified cysteine residues are susceptible to hydrolysis by trypsin. Why?

8. An enzyme that catalyzes disulfide-sulfhydryl exchange reactions has been isolated. Inactive "scrambled" ribonuclease is rapidly converted into enzymatically active ribonuclease by this enzyme. In contrast, insulin is rapidly inactivated by this enzyme. What does this important observation imply about the relation between the amino acid sequence and the three-dimensional structure of insulin?

For additional problems, see W. B. Wood, J. H. Wilson, R. M. Benbow, and L. E. Hood, *Biochemistry: A Problems Approach* (Benjamin, 1974), ch. 2; and R. Montgomery and C. A. Swenson, *Quantitative Problems in the Biochemical Sciences,* 2nd ed. (Freeman, 1976), chs. 6, 7, and 8.



Scanning electron micrograph showing erythrocytes (biconcave-shaped) and leucocytes (rounded) in a small blood vessel. [From *Tissues and Organs* by R. G. Kessel and R. H. Kardon. W. H. Freeman and Company. Copyright © 1979.]