9

Secondary Structure of Proteins

INTRODUCTION

Proteins are made up of linear polypeptides that, after synthesis on a ribosome, fold up spontaneously to give a unique and biologically active three-dimensional structure. As was originally proposed by Linderstrom-Lang, three-dimensional structure can be considered at a number of levels. We have already discussed (Chap. 5) the *primary structure* of a protein: its amino acid sequence and various chemical characteristics such as disulfide bonds and covalently attached cofactors, carbohydrates, or other derivatives. In this chapter and in Chaps. 10 and 11 we consider secondary, tertiary (the other two levels originally described by Linderstrom-Lang), and quaternary structure.

Secondary structure may be defined as the local spatial organization of the polypeptide backbone without consideration of the side-chain conformations. As we will see, however, when considering the prediction of secondary structure from the amino acid sequence of the protein, the nature of the side chains in a particular region of polypeptide chain does influence whether a certain secondary structure is found. The secondary structure is defined by four basic categories: α helix, β strand (often associated into so-called "sheets"), β turn, and random coil.

The *tertiary structure* of a protein is defined as the packing of the foregoing secondary structural elements *within* a polypeptide chain into a three-dimensional structure. Although as just defined, a tertiary structural element should involve a single polypeptide chain, there are instances where an apparent tertiary structural element involves two or more polypeptide chains. As will be seen, the initial stages

of the folding of a polypeptide chain are dominated by local secondary structure in the nascent polypeptide.

Quaternary structure is the assembly of tertiary structural elements into an oligomeric form. Such an oligomer can be homologous (i.e., consisting of multiple copies of a single type of polypeptide chain; for example, *glutamate dehydrogenase*, which has six chemically identical polypeptide chains) or heterologous (i.e., consisting of two or more chemically distinct polypeptide chains; for example, *aspartate transcarbamoylase*, which is made up of 12 polypeptide chains, six each of two different types).

In this chapter we are concerned primarily with secondary structure (although, as will become evident, this cannot be done entirely without some consideration of tertiary structure) and the ways in which a protein may fold into its native structure. Within a single polypeptide chain it is possible that several distinct "domains" may occur. In such regions the tertiary structure of one part of the polypeptide appears to be independent of other domains. Such domains are often associated with particular functional regions of the protein. Although a wide variety of experimental evidence is discussed, much of the later part of the chapter focuses on whether or not the secondary structure of a protein can successfully be predicted from its primary sequence, and what insights such an endeavor gives to the basic mechanism of protein folding and to our understanding of the possible mechanisms of conformational changes.

PATHWAY OF PROTEIN FOLDING

The general mechanism of protein folding has been debated for many years, and the arguments fall into two categories. It has been suggested that the formation of a three-dimensional structure of a peptide chain is a totally thermodynamically controlled process. A wide variety of studies involving reversible denaturation support this contention. Studies of proteins such as staphylococcal nuclease have shown not only that complete activity can be recovered after denaturation, but that a number of measurable physical parameters, all of which reflect somewhat different aspects of the native conformation, show similar reversible acid-induced transitions. As we discuss in this chapter and in Chaps. 10 to 12, many physical parameters can be measured that reflect the conformation of a protein. Various spectroscopic properties of amino acid side chains or the polypeptide backbone structure can be followed as a protein undergoes a denaturation process. Staphylococcal nuclease shows a reversible acid-induced denaturation that has been followed by measurements of tryptophan or tyrosine fluorescence, which, as shown in Fig. 9-1, both reflect the same denaturation process.

Also shown in Fig. 9-1 is the correspondence of tyrosine absorbance changes with fluorescence changes. The same conformational transition has also been studied, in circular dichroism (CD) measurements, by following the molar ellipticity at 220 nm, which monitors various aspects of the secondary structure of the protein. Similarly, the overall hydrodynamic properties of the molecule, as measured with viscosity determinations, also show the transition. NMR measurements of the four histidine

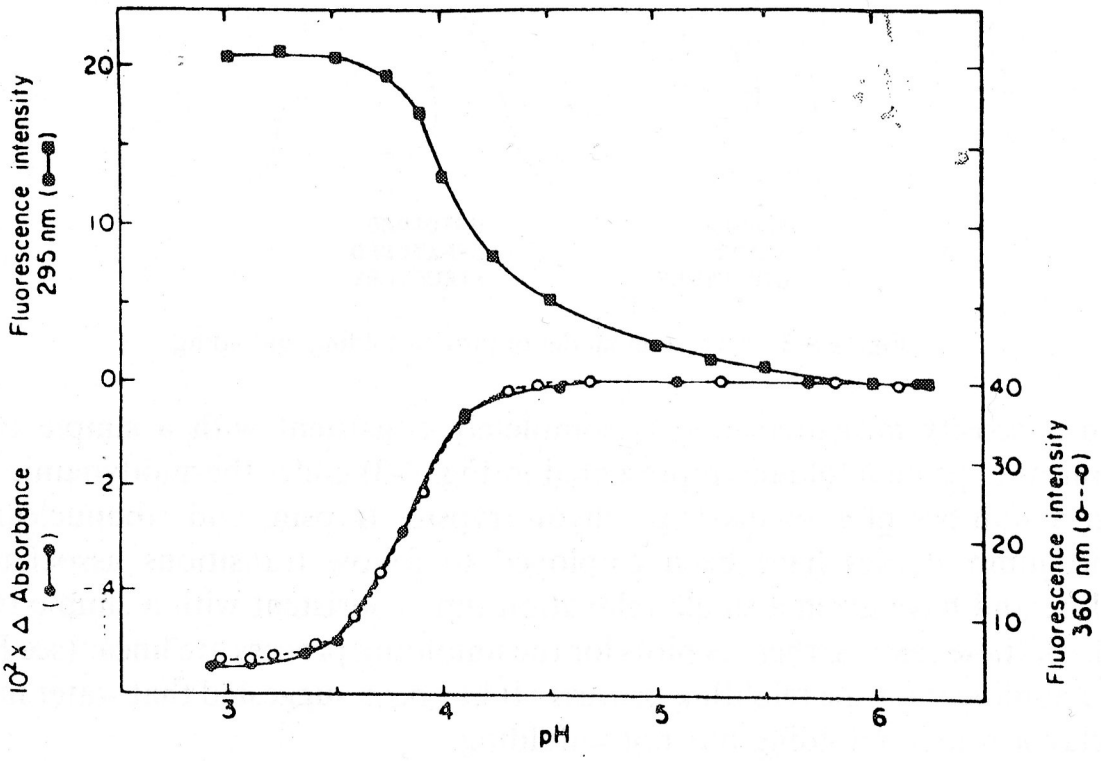

Figure 9-1 pH-induced denaturation of nuclease followed by tyrosine fluorescence (□), tryptophan fluorescence (○), or tyrosine absorbance (●) measurements. (From H. F. Epstein, A. N. Schechter, R. F. Chen, and C. B. Anfinsen, *J. Mol. Biol.*, *60*, 499–508. Copyright 1971 Academic Press, Inc., New York.)

resonances in the protein demonstrate the same transition. These various measurements are shown in Fig. 9-2.

The observation that this pH-induced transition is monitored by techniques reflecting individual amino acid side-chain environments (fluorescence and NMR), the state of the protein's secondary structure (CD) or the overall conformation of the

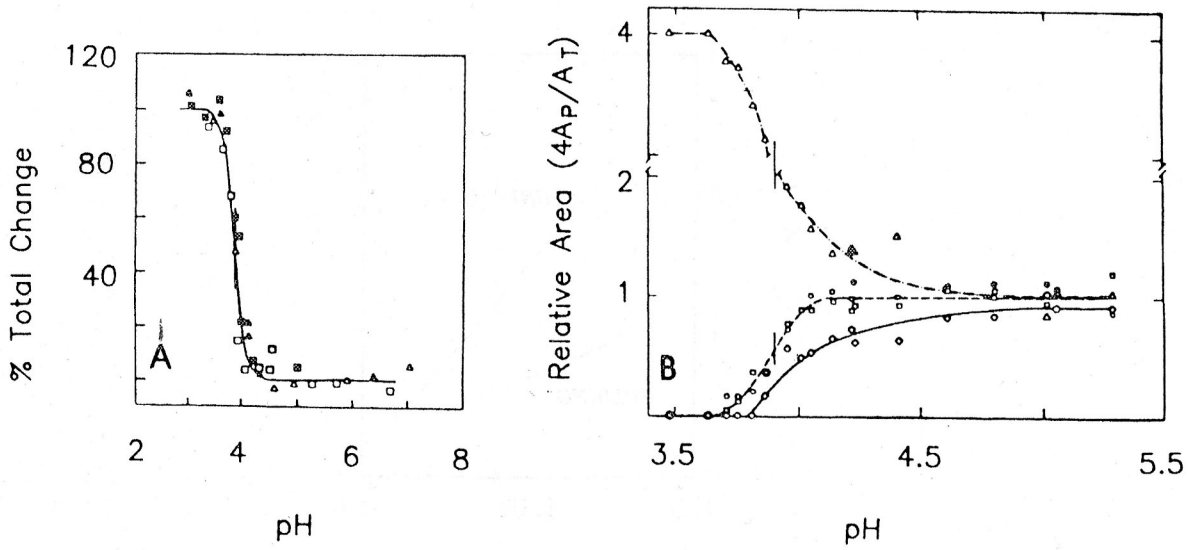

Figure 9-2 pH-induced denaturation of nuclease followed by (A) viscosity (□, ■) or CD measurements, (△, ▲) at 220 nm and (B) NMR measurements of the four histidines in the molecule. (○–○) H-1; (△–△) H-2; (□–□) H-3; (◇–◇) H-4.

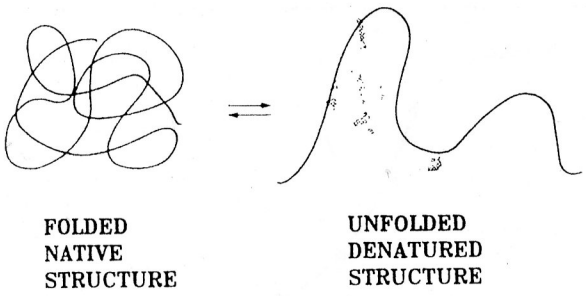

Figure 9-3 Two-state model of protein folding–unfolding.

protein (viscosity measurements) is completely consistent with a simple two-st mechanism of protein folding (represented in Fig. 9-3) under thermodynamic cont

In a number of proteins (e.g., chymotrypsin, trypsin, and ribonuclease) t perature-jump studies have been employed to follow transitions associated w unfolding and have given a single relaxation time consistent with a simple two-s model. In these cases Arrhenius plots for the unfolding process are linear (see Fig. 9 but are nonlinear for the refolding process. It has been suggested that water molec may play a role in refolding but not unfolding.

Although these examples support thermodynamic control of protein fold much contradictory evidence has been presented that favors the involvement of netic (i.e., path-dependent) processes in a variety of other proteins.

A number of proteins cannot be renatured to give active protein: Glutan dehydrogenase undergoes a two-stage dissociation in guanidine hydrochloride. intermediate concentrations of guanidine hydrochloride the hexamer dissociates trimer with little loss of native structure, as assessed by CD. At higher concen tions the trimer dissociates to a monomer form, with considerable loss of integ of tertiary structure. The reversibility of these induced transitions has been follo by measurements of the enzyme activity. Activity can be regained only by rem of guanidine hydrochloride when the dissociation and denaturation have proce

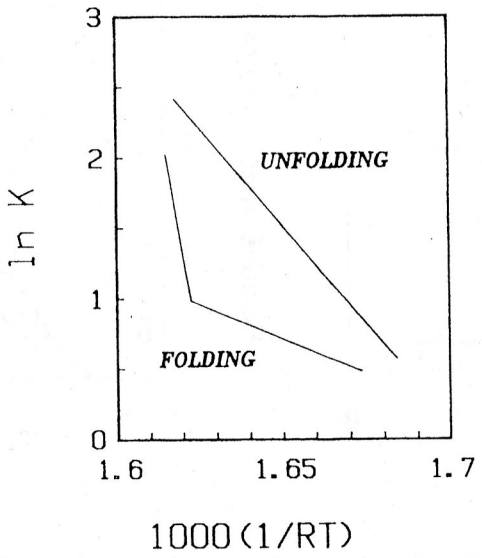

Figure 9-4 Arrhenius plots for protein folding and unfolding.

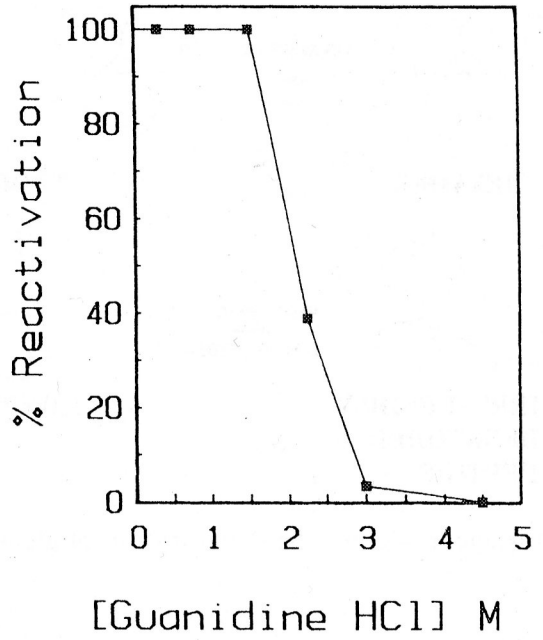

Figure 9-5 Renaturation of glutamate dehydrogenase after denaturation in guanidine hydrochloride.

no further than the trimer stage. When denatured polypeptide chains are formed, enzymatic activity cannot be regained, as shown in Fig. 9-5.

The amount of renaturation possible in this case can be correlated with the quaternary structure (as determined by light scattering) and status of the polypeptide chain conformation (as judged by CD), and this comparison (see Table 9-1) shows that the trimer can be renatured, but once the denaturation has reached the level of individual polypeptide chains, activity cannot be regained.

TABLE 9-1 Renaturation of glutamate dehydrogenase after guanidine hydrochloride denaturation

[Guanidine hydrochloride] (M)	Percent reactivation	Molecular state
4.5	0	100% monomer
3.0	3.4	95% monomer, 5% trimer
2.25	38.9	60% monomer, 40% trimer
1.5	100	100% trimer
0.75	100	85% trimer, 15% hexamer
0.3	100	10% trimer, 90% hexamer

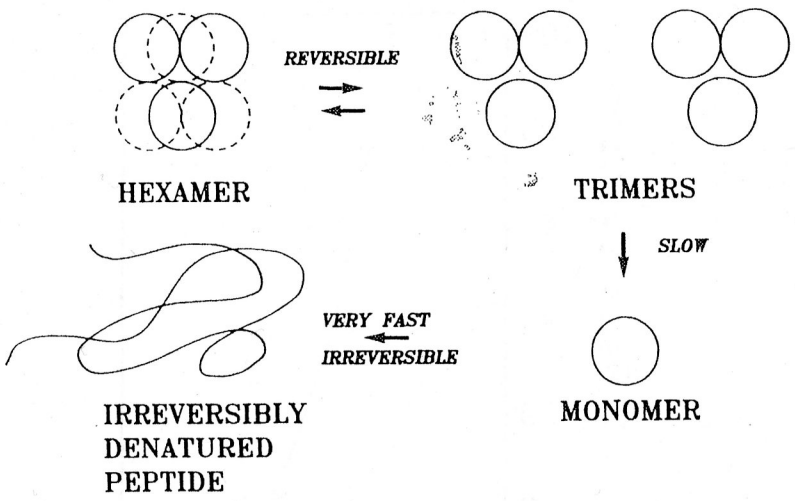

Figure 9-6 Multistage model for the denaturation of glutamate dehydrogenase.

These data are consistent, in the case of glutamate dehydrogenase, with a model (Fig. 9-6) showing a reversible dissociation of the hexamer to trimers and an irreversible dissociation of the trimer to monomers that rapidly lose "native" structure.

In similar types of experiments, enzymes such as muscle aldolase (which has four subunits) have been shown to regain a large portion (in the case of aldolase approximately 80%), but not always all, of their native activity, and that the *rate* of regain of activity is dependent on the presence of a substrate or ligand (fructose diphosphate in the case of aldolase). This fact suggests that the refolding and renaturation process is pathway dependent (i.e., kinetically controlled). In some proteins the presence of a ligand or cofactor is *necessary* for renaturation. Alkaline phosphatase requires the presence of zinc for refolding. With carbonic anhydrase, also a zinc-dependent enzyme, zinc is not required for refolding, but its presence does increase the rate.

There is a certain amount of evidence that protein folding can occur prior to the completion of synthesis. This suggests that *all* of the sequence may not be required to allow enzymatically active structures to form. During biosynthesis of a number of proteins, polysomes with attached partially complete, but immunologically competent, folded protein can be isolated by immunoprecipitation techniques. In some cases these partially complete but folded peptides also exhibit enzymatic activity. If the polypeptide chain folds during synthesis, one would expect to find evidence of *native* conformation—as judged by activity or antigenicity—only if the entire sequence is not required for native folding to occur. If this is true, one would also expect that isolated N-terminal regions of proteins might have some native conformation.

In experiments with staphylococcal nuclease, which has 149 residues, three fragments result from tryptic cleavage: residues 1 to 5, residues 6 to 49, and residues 50–149. When separated they have no enzymatic activity and no detectable helix structure (as judged by CD measurements). When the fragments are mixed, approximately 8% of the expected native activity is regained. The mixture has approximately 10% α helix compared to the 18% estimated for the native protein. This would argue against the expectation of finding some evidence of native structure in the isolated

N-terminal regions of a polypeptide, and thus favors the idea of protein folding being governed by thermodynamic criteria.

Antigenic Detection of Protein Folding

As indicated previously, antibodies to native structure have been used to detect the existence of such native structure in fragments or during synthesis. In general, antibodies can be made to both native or unfolded, denatured structure. It is possible to show that an antibody reacts with the *native* conformation of a peptide by showing that ligands which stabilize the native conformation do not inhibit the interaction of antibody with the protein but in an equilibrium situation where both unfolded and folded protein exist, increase the interaction. Although antibodies have found uses in establishing the existence of native structure, studies that have attempted to use them (to either the native or denatured states of a protein) to follow the kinetics of folding or unfolding have been criticized on the grounds that the antibody will act to stabilize or destabilize one or the other component of an equilibrium between the folded and unfolded protein's form.

Mechanism for Protein Folding

From the preceding discussion it appears that some proteins may fold in accordance with the concept of folding being governed only by thermodynamic criteria, while others fold by kinetically controlled pathways. It may not be coincidence that most of the proteins that can be reversibly denatured successfully are globular proteins of relatively small size, while those that cannot be renatured, or which require the presence of co-ligands for refolding, are larger and often multi-subunit allosteric proteins. The folding process itself cannot be completely random since it would take too long relative to the overall life of the protein. A relatively small protein of, for example, 100 amino acids has approximately 10^{40} possible conformations, depending on its primary structure. Small proteins generally fold on a time scale of seconds. The molecular motions involved in the folding of a polypeptide chain occur on a nanosecond-to-picosecond time scale, suggesting that at most about 10^{11} conformations could be randomly screened during the folding process.

It seems likely that in most instances protein folding is a kinetically governed process and that nucleation events may direct the pathway. The scheme shown in Fig. 9-7 outlines a *suggested* process by which a polypeptide chain acquires its native conformation and is proposed purely as a basis for considering the possible importance of secondary structure in directing protein folding.

In this figure the early events are the *transitory* formation of small regions of formal secondary structure: The lifetimes of these regions vary with their individual stability. As regions of the polypeptide chain with areas of local secondary structure randomly interact with one another, some interactions lead to stabilization of these structures, whereas others lead to destabilization. These interactions, which of course can occur between widely separated regions of the primary sequence, result in new regions of secondary structure being formed and the refining of existing regions. The

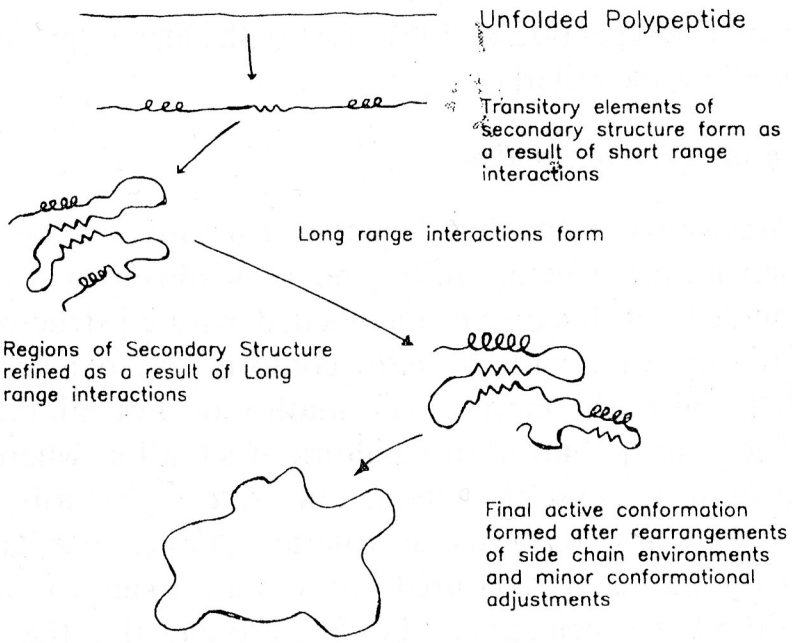

Figure 9-7 Proposed pathway for protein folding.

formation, via such long-range interactions, of a *primal* binding site for a ligand offers a mechanism for the frequently observed ligand requirement for folding or ligand enhancement of the rate of folding. Regions of transitory secondary structure stabilized by such long-range interactions become *nucleation* sites that now direct subsequent folding of the remainder of the molecule. It is probable that a variety of local conformational rearrangements happen after the formation of the principal elements of secondary structure in the protein.

In small proteins this process occurs rapidly and in all likelihood in a highly cooperative manner, giving rise to the simple two-state kinetics often observed. In larger proteins, which may have more than one domain or may contain subunits, a more complex process exists. The initial stages are undoubtedly the same, however, as the various domains independently form local nucleation centers and acquire some type of tertiary structure, interactions between the nascent domains occur and can have further stabilizing or destabilizing effects on elements of secondary structure within the domains. A similar situation holds with subunits: subunit–subunit interactions may have significant effects on the tertiary and secondary structure of the individual subunits. Subunit structures with high energies of interaction would especially be expected to suffer from secondary and tertiary rearrangements as a result of subunit–subunit interactions, which may explain the difficulty in renaturing such proteins.

This suggests that it may not be possible to refold a denatured protein to the same conformation as that obtained during biosynthesis. During biosynthesis local domains may form (these can certainly be detected antigenically) and some formed early in synthesis direct (or at least influence) the folding of other regions. This is a quite different situation from that found in a renaturation experiment, where all the domains in the molecule are simultaneously folding and as a result have somewhat different effects on each other, indicating the possibility that different final conforma-

tions are arrived at in the two cases. Some proteins may not refold after denaturation as a result of the fact that they undergo post-translational (and hence post-folding) modification. Such modifications may involve proteolytic processing or covalent derivatization of certain amino acid side chains which could affect folding pathways.

During this discussion we have emphasized that folding occurs at two stages: an early stage involving local regions of secondary structure, which is dominated by the nature of the polypeptide chain in local regions, and a later stage where interactions between local regions of secondary structure occur as the tertiary structure is being built up. The first stage depends on short-range interactions between near-neighbor residues, while the second involves long-range interactions. Each stage is important in the *final secondary structure* of the protein as well as, of course, in the tertiary structure.

FORMAL SECONDARY STRUCTURE

A polypeptide chain consists of a series of amino acids chemically linked by peptide bonds. The peptide bond, illustrated in Fig. 9-8, can be considered as a planar structure. Although usually drawn as in Fig. 9-8, the planarity can be construed as being imposed by the partial double-bond character of the amide group resulting from resonance stabilization. Given this planarity it is evident that the conformation of a peptide can be described by two angles, the ϕ angle and the ψ angle, which are those describing the rotations of two planar peptide bonds on either side of an α-carbon atom in the polypeptide. These angles are indicated in Fig. 9-8. To describe the backbone conformation of a polypeptide chain, all one needs is a description of the

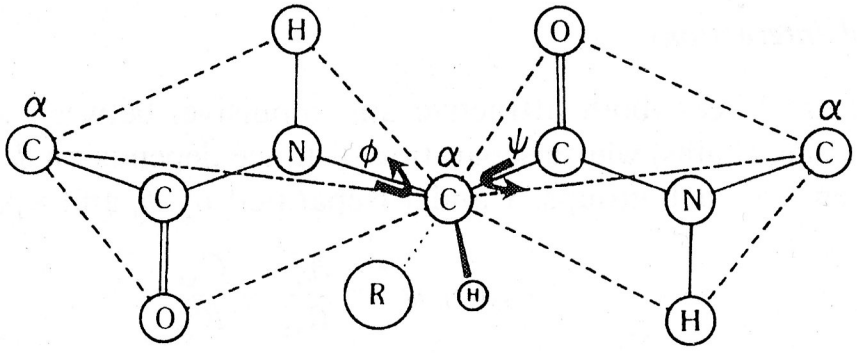

Figure 9-8 Peptide bonds joining three α-carbon atoms. ———, Planar nature of the rotations about the ϕ and ψ angles; – – –, virtual bonds connecting the α-carbons.

ϕ and ψ angles for each successive α-carbon atom in the chain. Because of steric considerations, certain of these angles are not possible: ϕ and ψ are *interdependent*. This holds only for those angles about a particular α-carbon atom. The ϕ and ψ angles of a given α-carbon are *not* affected by those angles of other α-carbon atoms within the peptide.

The steric effects on ϕ and ψ angles are conveniently represented in a *Ramachandran plot*, which shows energy contours in a plot of ϕ versus ψ. Figure 9-9 shows

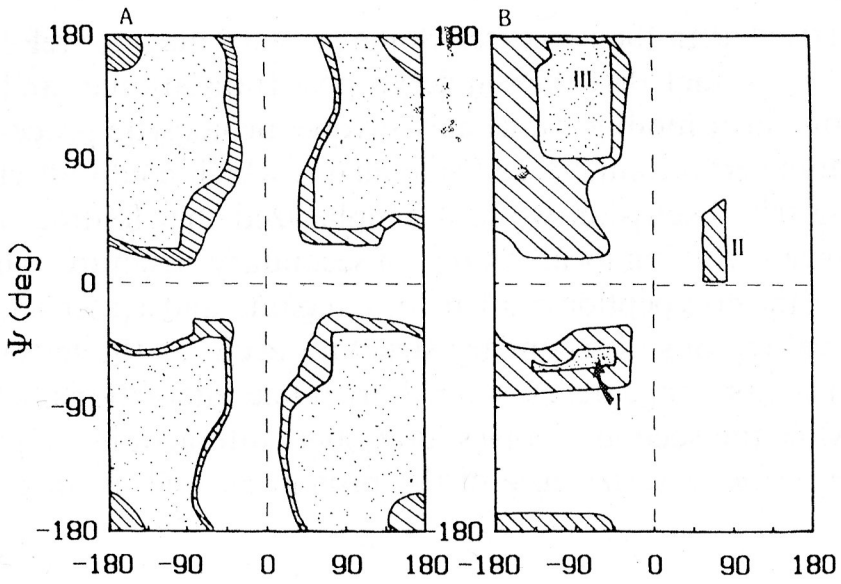

Figure 9-9 Ramachandran plots for (A) glycine and (B) alanine in peptides indicating allowable configurations (hatched areas) and areas of "contact" (stippled areas). Other areas are not accessible for steric reasons.

such plots for glycine or alanine in a peptide. The plot for glycine is essentially symmetrical, whereas alanine, due to unfavorable side-chain contacts, has some additional areas of conformational space restricted. Onto diagrams such as these, energy contours can be superimposed that indicate within the sterically allowable areas which ones are energetically favored. Three types of interactions must be considered in assessing the energy of a particular conformation.

Nonbonded Interactions

These are forces (both attractive and repulsive) between atoms or groups of atoms (e.g., side chains) whose separation distance depends on ϕ and ψ. The energy, E_{xy}, between two such groups, x and y (separated by ϕ_i and ψ_i), is given in

$$E_{xy}(\phi_i, \psi_i) = \frac{A_{xy}}{R_{xy}} - \frac{C_{xy}}{R_{xy}^6} \tag{9-1}$$

where A and C are parameters characteristic of the groups involved and R is the distance of separation. At large values of R there is no interaction, but as R decreases, first an attractive force, and subsequently a repulsive force, operate. The repulsive force becomes significant as the groups penetrate each others atomic radii.

Dipolar Interactions

As might be expected, the amide groups in the peptide bonds have dipole moments, which are oriented approximately parallel to the N—H bond in the direction toward the H. Since the amide dipole is quite large, dipolar interactions play a significant role in backbone conformation. The energy of interaction, E_d, between

two dipoles U_p and U_q separated by the vector $\mathbf{r}$ is given by Eq. (9-2), and, since it arises from the orientation of two adjacent amide groups in the peptide, is governed by ϕ and ψ angles,

$$E_d = \mathscr{E}^{-1}\left[\frac{U_p U_q}{r^3} - \frac{3(U_p \cdot \mathbf{r})(U_q \cdot \mathbf{r})}{r^5}\right] \tag{9-2}$$

where r is the scalar magnitude of $\mathbf{r}$ and $\mathscr{E}$ is the dielectric constant. The energy E_d is usually computed from Eq. (9-3) using partial charges and summing the charge–charge interactions to give the total electrostatic energy,

$$E_d = \frac{\sum_{xy} Q_x Q_y}{\mathscr{E} R_{xy}} \tag{9-3}$$

where Q is the partial charge and R is the distance of separation. The dielectric constant, $\mathscr{E}$, is much less than the dielectric constant of water since the dipolar interaction passes through the protein. Dipolar interactions longer than nearest-neighboring amide groups are not usually considered since the energy of interaction decreases rapidly with distance.

Intrinsic Torsional Potential

Earlier we discussed the planar nature imposed on the peptide bond by its partial double-bond nature. The inference is that the single bonds to the α-carbons allow free rotation. In reality this is not true, and as a result, ϕ and ψ rotations have an intrinsic rotational hinderence with an associated torsional energy, $E_{\text{tor}}(\phi_i, \psi_i)$, given by

$$E_{\text{tor}}(\phi_i, \psi_i) = \frac{E^0\phi}{2}(1 + \cos 3\phi) + \frac{E^0\psi}{2}(1 + \cos 3\psi) \tag{9-4}$$

where $E^0\phi$ and $E^0\psi$ are the energy barriers associated with the ϕ and ψ rotations.

The total energy, $E(\phi_i, \psi_i)$, is simply the sum of these contributions and is given by

$$E(\phi_i, \psi_i) = \sum_{xy}\left[E_{xy}(\phi_i, \psi_i) + E_d(\phi_i, \psi_i) + E_{\text{tor}}(\phi_i, \psi_i)\right] \tag{9-5}$$

Using Eq. (9-5), energy contour diagrams analogous to the Ramachandran plots can be obtained. Figure 9-10 illustrates such contour maps for glycine and alanine in general detail. As in the Ramachandran plots, glycine is symmetrical, whereas the alanine plot is not. With alanine three low-energy regions are indicated (I → III). Regions I and II correspond to right- and left-handed helices, and from the diagram it is apparent that the right-handed (region I) is favored over the left-handed (region II). Region III is clearly the overall lowest-energy region and corresponds to an extended form of the residue that includes the β-strand conformation. If dipolar interaction energies are omitted from the calculations, region I would have an overall lower energy than III. This emphasizes the importance of dipolar interactions involving "hydrophobic" residues in governing a tendency to be associated with β-strand

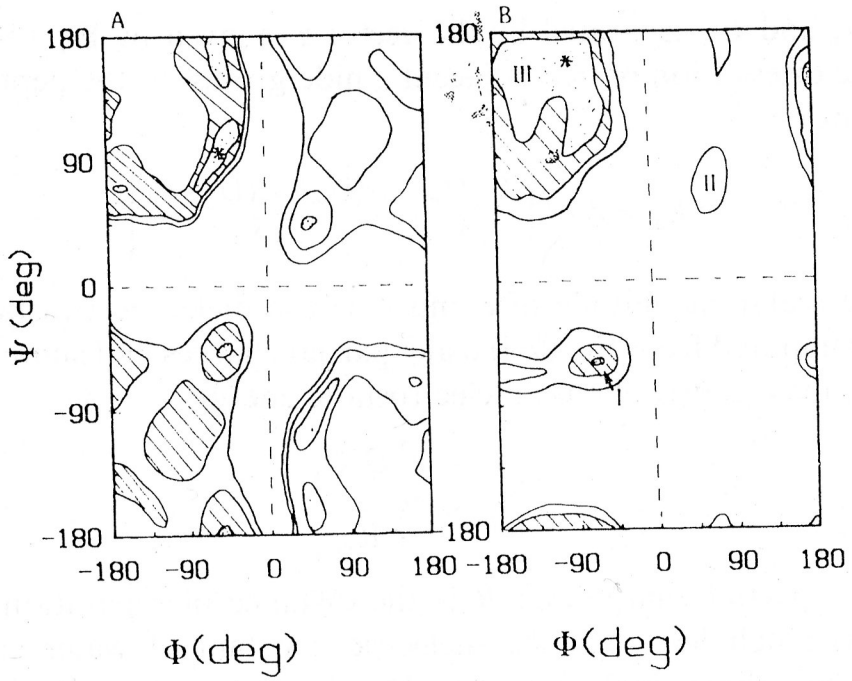

Figure 9-10 General detail of energy contour maps: (A) glycine; (B) alanine. The overall energy minimum is indicated as *.

conformations. As the microscopic dielectric constant decreases (in a hydrophobic environment), the energy of dipolar interaction increases.

Although we have considered only glycine and alanine in this discussion, alanine can be considered as representative of most amino acids containing a side chain. From Figs. 9-9 and 9-10 it is clear that glycine is far more flexible than other residues. The one other residue that should be considered is proline. Proline, because of the rigid pyrrolidine ring, has a fixed ϕ angle of about $-60°$, and the conformational energy depends only on the ψ angle. A plot of energy versus ψ has two minima, at $\psi = -55°$ and $\psi = 145°$, which fall into regions I and III in the plot (of alanine) shown in Fig. 9-10. These two allowed conformations correspond to a relatively compact form and an extended form, and are energetically quite similar. The compact form ($\psi = -55°$) is ideal for accommodating a turn or a bend in a chain. Perhaps the most important point regarding proline residues, however, is the effect they have on the conformational energy map of the residue preceding them. With the exception of a glycine, residues preceding proline have energy maps lacking region I, which involves unfavorable steric overlaps involving, in the case of alanine, the methyl side chain and the CH_2 group attached to the imido nitrogen. The effect of this is to deny a residue preceding proline the conformational space associated with a right-handed α helix. Proline itself can adopt this conformation, but it prevents the preceding residue from doing so. As a result, proline may occur at the start of a helix but rarely within a helix.

On a purely mechanical basis it is possible to describe α-helix, β-strand, and β-turn secondary structures in terms of the ϕ and ψ angles of adjacent residues in the peptide. If four or more consecutive residues have ϕ and ψ angles within 40° of $(-60°, -50°)$ the region of peptide is in a right-handed α helix. If three or more

residues have ϕ and ψ angles within 40° of (−120°, 110°) or (−140°, 135°), the structure is a β strand (either parallel or antiparallel): β turns consist of four consecutive residues where the polypeptide chain folds back on itself by about 180°, and they can be either right-handed or left-handed.

α Helix

Figure 9-11 shows several representations of α helices. The α helix has approximately 3.6 residues per turn and is stabilized to a large extent by hydrogen bonds formed within the backbone of the chain between amide protons and carbonyl oxygens.

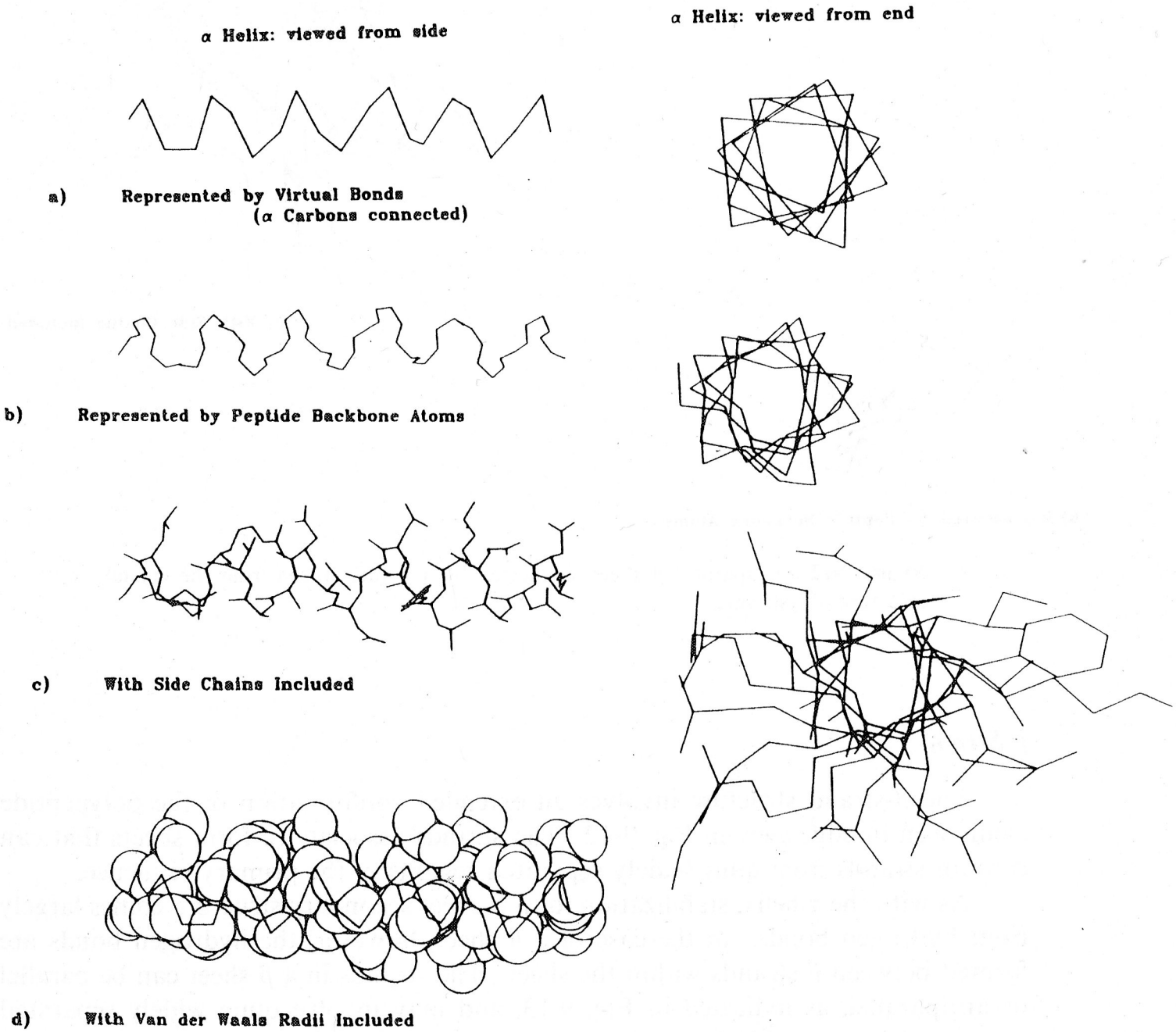

Figure 9-11 An α helix, taken from the crystal structure of lysozyme, is shown as a side view and as an end view.

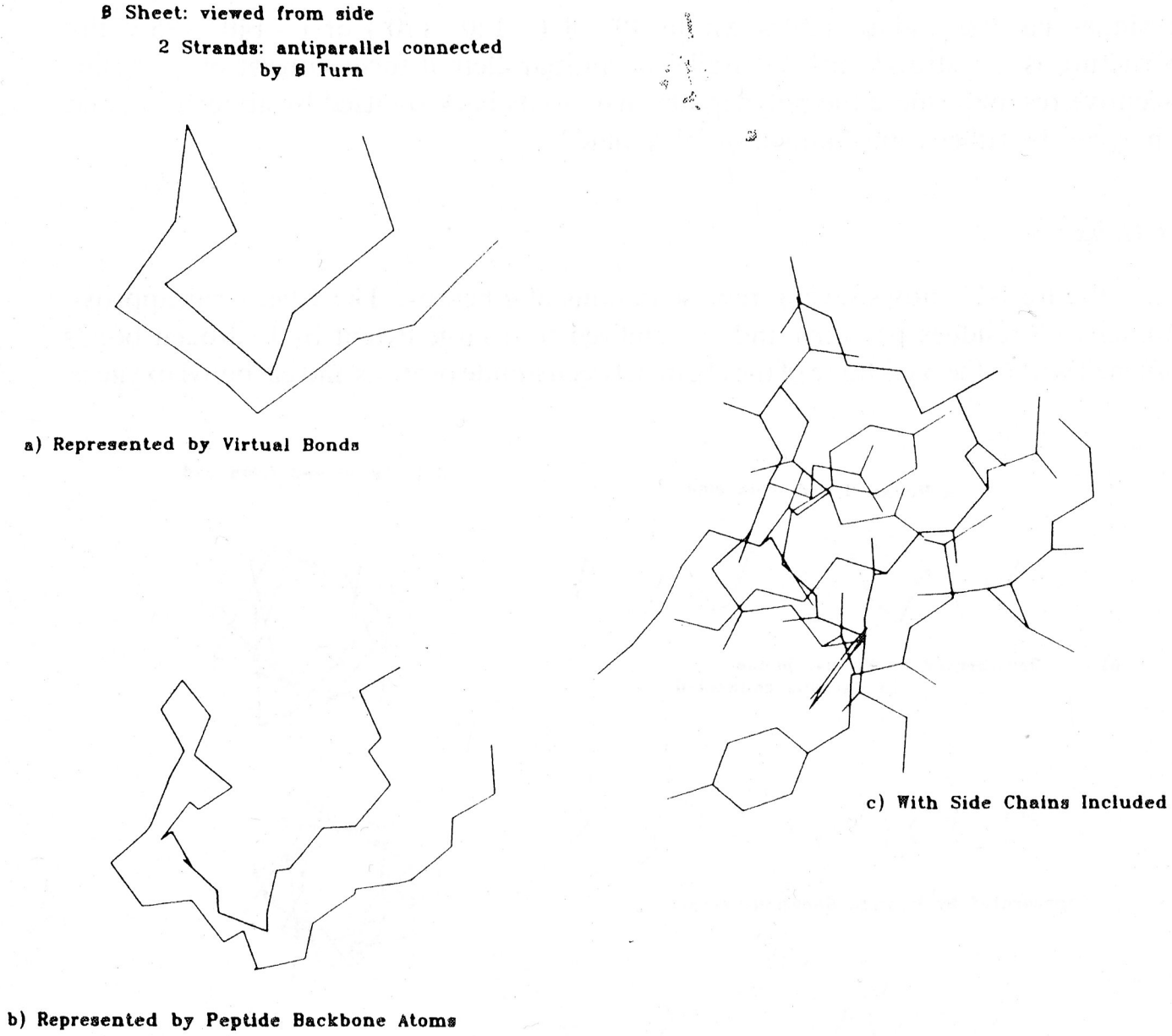

Figure 9-12 Antiparallel β sheet, connected by a β turn, taken from the crystal structure of lysozyme.

β Strand

The β-strand structure involves an extended conformation of the polypeptide chain seen in edge view in Fig. 9-12. The strands associate to form sheets that can contain strands from quite widely separated regions of the primary sequence.

As with the α helix, stabilization for a β-sheet secondary structure comes largely from hydrogen bonds. In the case of a β sheet, however, the hydrogen bonds are formed between β strands within the sheet. The strands in a β sheet can be parallel or antiparallel, as indicated in Fig. 9-13, and may involve quite widely separated regions of polypeptide.

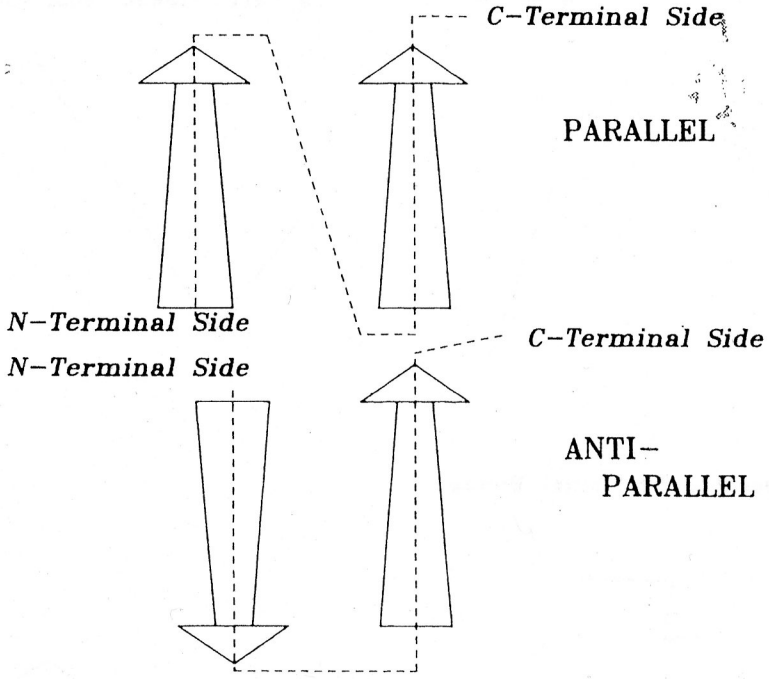

Figure 9-13 Arrangements of strands in β sheets: (A) parallel; (B) antiparallel.

β Turn

Figure 9-14 shows two views of an observed β-turn structure. As with the other formal secondary structures we have considered, much of the stability for this structure comes from the formation of hydrogen bonds between amide protons and carbonyl oxygens. Because of their conformational energies, discussed earlier, proline residues are often found in turn regions, usually in the second or third position.

Each of these structures is taken from the actual secondary structure observed in lysozyme and is shown in several representations. When the α carbons are connected by "virtual" bonds the simplest representation of each structure is obtained. Tracing the peptide-bond backbone also gives a simplified but clear view of each structure. This view might be considered the most appropriate. Inclusion of the side chains tends to obscure the basic element of the secondary structure, particularly with the β turn. Several important features of these secondary structure elements are demonstrated by Figs. 9-11 to 9-14.

The end view of the α helix (Fig. 9-11) clearly illustrates the "sided" nature of a helix: The amino acid side chains form "ridges" and "grooves" along the length of the helix. This helix feature is apparent only by considering the structure with side chains.

The β sheet (Fig. 9-12) shows a clear "twist" to its structure, and from examining the different views of the β turn it is apparent that it is not planar. As might be expected, the nature of the amino acid side chains in these two structural elements affects the twist of the sheet and the orientation of the turn.

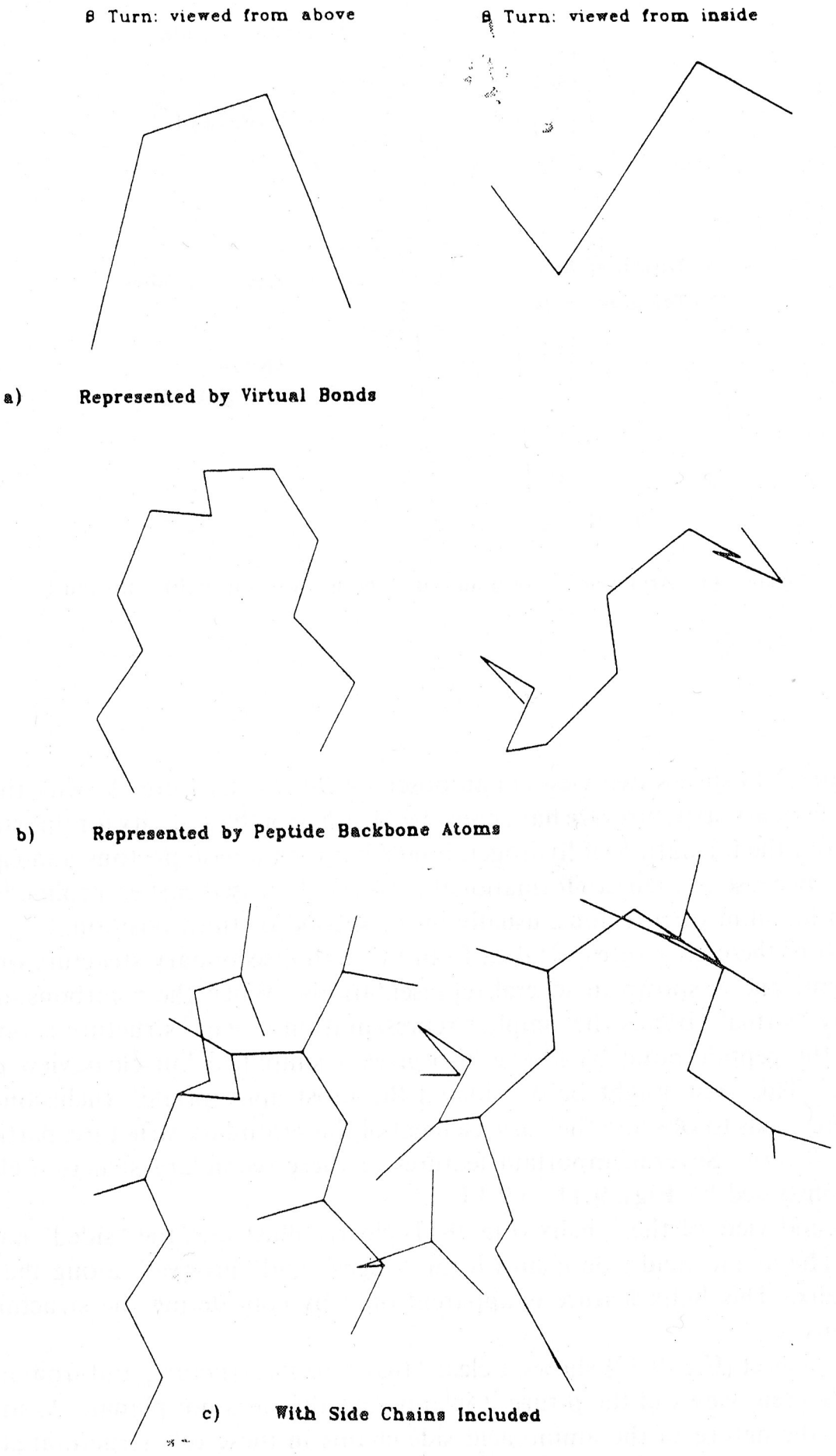

Figure 9-14 A β turn, taken from the crystal structure of lysozyme, shown from the top and from inside the turn.

PREDICTION OF SECONDARY STRUCTURE

Because the prediction of secondary structure is based largely on the character of the amino acids that are found in nature to be in certain of the secondary structures, it is informative to consider briefly how procedures for these predictions from primary sequence have been developed and applied. Many of the problems are similar to those the protein itself must encounter during the folding process! Predictive methods are based on the probability that a particular type of amino acid residue is found in a certain type of secondary structure. These data are obtained in one of two ways: In the first, probabilities are obtained by examining the crystal structures of known proteins and counting the number of times particular residues appear in α helices, β strands, or β turns. Alternatively, polymers of single amino acids are used and their secondary structure determined. From the tendency of such polymers and various copolymers to form α helix and β sheet, an assessment of the contribution of individual residues to these structures can be made. In some instances it is useful to keep information on *where* in the type of secondary structure these residues appear most frequently, as this can be helpful in defining starting points and termination points for the type of secondary structure. Table 9-2 gives such information concerning probabilities of residues appearing in α helices, β strands, and β turns.

Several generalities can be drawn from Table 9-2: (1) the charged residues are unfavorable for β-strand formation, and three of them (Asp, His, Arg) are also α-helix

TABLE 9-2 Conformational parameters for amino acid residues appearing in α helix, β strand, and β turn based on frequencies of occurrence in proteins of known structure

Residue	α Helix	β Strand	β Turn
Glu	1.51	0.37	0.74
Met	1.45	1.05	0.6
Ala	1.42	0.83	0.66
Leu	1.21	1.30	0.59
Lys	1.16	0.74	1.01
Phe	1.13	1.38	0.6
Gln	1.11	1.10	0.98
Trp	1.08	1.37	0.96
Ile	1.08	1.60	0.47
Val	1.06	1.70	0.50
Asp	1.01	0.54	1.46
His	1.00	0.87	0.95
Arg	0.98	0.93	0.95
Thr	0.83	1.19	0.96
Ser	0.77	0.75	1.43
Cys	0.70	1.19	1.19
Tyr	0.69	1.47	1.14
Asn	0.67	0.89	1.56
Pro	0.57	0.55	1.52
Gly	0.57	0.75	1.56

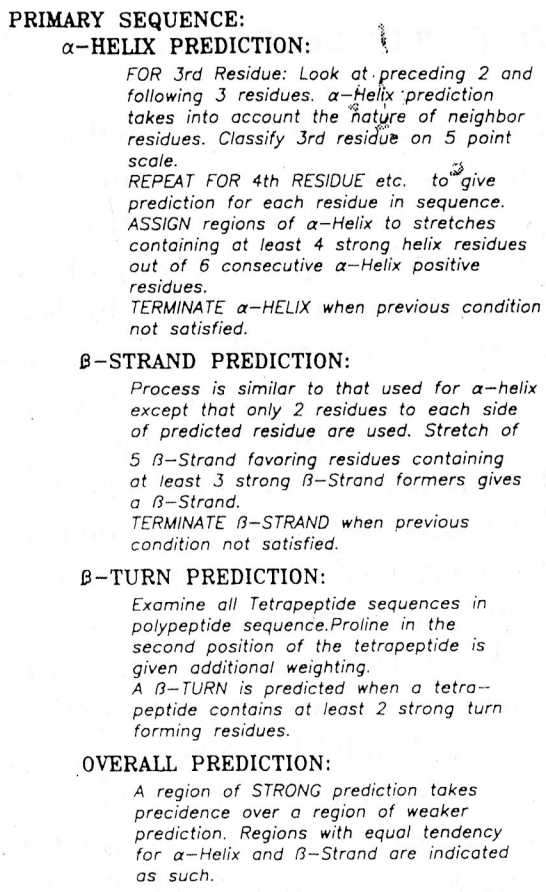

Figure 9-15 Protocol used for the prediction of secondary structures in proteins.

indifferent; (2) residues that tend to break α helices (Pro, Gly, Asn, Tyr) also tend to be residues with high probability of appearing in β turns, and (3) residues with a strong tendency to be in β strands are rarely found in β turns.

This type of information has been applied to secondary-structure prediction in a variety of ways. One of the more successful is outlined in Fig. 9-15. In this predictive scheme the influence on neighboring residues is taken into account in attempting to assign a propensity of each residue in a peptide to be in an α helix, a β strand, or a β turn. Each type of secondary structure is "predicted" independently and the final "prediction" based on a comparison not only of the "strength" of the prediction but also on the predictions for adjacent residues. For example, it is quite possible for a region of peptide to contain a residue that has a high probability of being in either a β strand or an α helix; if the neighboring residues are predominantly helical, this weights the final choice between β strand and α helix for the prediction. Finally, regions of secondary structure are predicted based on certain "nucleation" rules. For an α helix to be indicated six adjacent helical residues must be present, for a β strand to be indicated five adjacent strand residues must be present, and for a β turn two residues, of a tetrapeptide sequence, must be indicated as strong turn formers. With β-turn predictions, weighting is given to proline in the *second* position in the turn.

Any such predictive scheme is limited by several factors: (1) the data base from which the probability values of the individual amino acids are taken, and (2) the lack

of consideration of long-range interactions.

Little can be done about the second point, of course, but the first gives a way of optimizing predictions for members of a class of proteins. Instead of using a random selection of proteins as the data base, a considerable improvement in prediction accuracy (which for general proteins is around 65%) can be achieved if related proteins only are used. In many cases this is possible; for example, by optimizing the parameters using known dehydrogenase secondary structures the accuracy of prediction using schemes such as that shown in Fig. 9-15 is of the order of 92 to 95%, which leads to increased confidence in the prediction of an unknown dehydrogenase.

Although secondary-structure predictions can only be confirmed after the three dimensional structure of the protein is known, it is possible to "test" predictions in terms of the percentage of the protein in α-helix or β-strand structures, where such gross parameters can be experimentally measured.

Perhaps the most useful information that can be obtained from secondary-structure predictions, at least in terms of protein chemistry, is some idea of potential nucleation sites in folding or regions of conformational flexibility. Potential nucleation sites may be regions of the polypeptide chain where strong predictions for a particular secondary structure are made. As with nucleation, prediction is based largely on short-range interactions. When protein sequences are used in prediction schemes, some regions end up having essentially equivalent probability of being helical or strand. Such regions probably represent areas of the protein where long-range interactions contribute the deciding influence with respect to formal secondary structure. Since these long-range interactions are imposed by the tertiary structure of the protein, such ambiguous areas may be regions involved in conformational changes within the protein, because small changes in long-range interactions may cause such an "ambivalent" region to switch from one formal secondary structure to another.

EXPERIMENTAL ASSESSMENT OF SECONDARY STRUCTURE

X-ray crystallography gives a clear picture of the various types of secondary structure present in proteins, although it is not a readily accessible method for determining the amounts and types of formal secondary structure. The only other approach that gives some idea of the amount of secondary structure present in a molecule is circular dichroism (CD). Figure 9-16 shows CD spectra for poly-L-lysine in an α-helical, a β-strand, and a random conformation.

From these spectra it is evident that the longer-wavelength edge of the negative spectrum observed for the helical form is somewhat characteristic of α helix. CD measurements at 222 nm have been used to "quantitate" the amount of helix present in a protein. The approach is largely correlative, using crystal structure α-helix contents to calibrate the dependence of the measured CD at 222 nm on helix content. In an adaptation of this approach, CD spectra between 200 and 245 nm have been used to estimate both α-helix and β-strand content, using computer fitting to equation

$$\mathrm{CD}_{\mathrm{meas}} = f_\alpha \mathrm{CD}_\alpha + f_\beta \mathrm{CD}_\beta + f_r \mathrm{CD}_r \qquad (9\text{-}6)$$

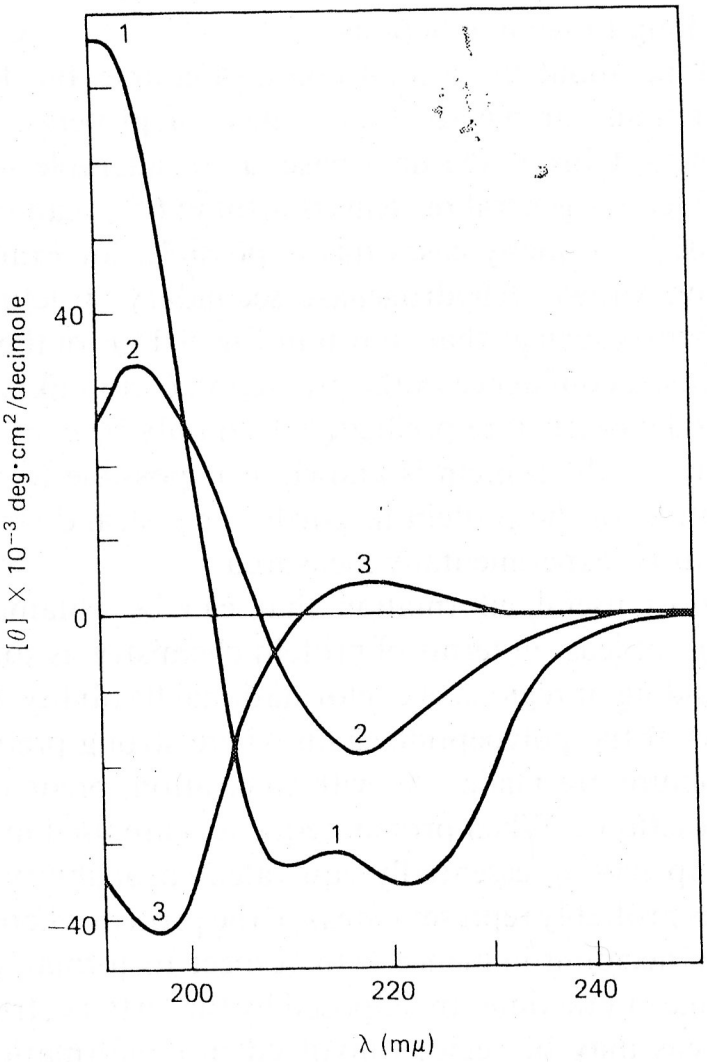

Figure 9-16 CD spectra of poly-L-lysine in various conformations. Curve 1, 100% helix; curve 2, 100% β; curve 3, 100% random oil. (Reprinted with permission from: N. Greenfield and G. D. Fasman, *Biochemistry*, *8*, 4108–4116. Copyright 1969 American Chemical Society, Washington, D. C.)

where f_α, f_β, and f_r are the fractions of helix, strand, and random coil, and CD_α, CD_β, and CD_r are appropriate constants calculated from proteins with known structures.

Application of such measurements to proteins with unknown structure can yield a useful approximation of the possible amounts of α helix and β strand. Such measurements with glutamate dehydrogenase give approximately 40% helix and 13% strand. These can be compared with estimates made by secondary-structure predictions using the approaches described earlier, which give approximately 24% each of helix and strand structures. The dilemma lies in the fact that both approaches are subject to unknown errors. The measured CD spectra are probably best used to indicate protential changes in secondary structure for a particular molecule rather than to give amounts of particular types of secondary structure.

CD spectral measurements have also been used to follow denaturation profiles of proteins. Figure 9-17 shows the pH-induced unfolding of tropomyosin followed by CD at 222 nm as well as by absorbance measurements at 295 nm, which have essentially similar profiles, consistent with two-state folding–unfolding, as discussed earlier.

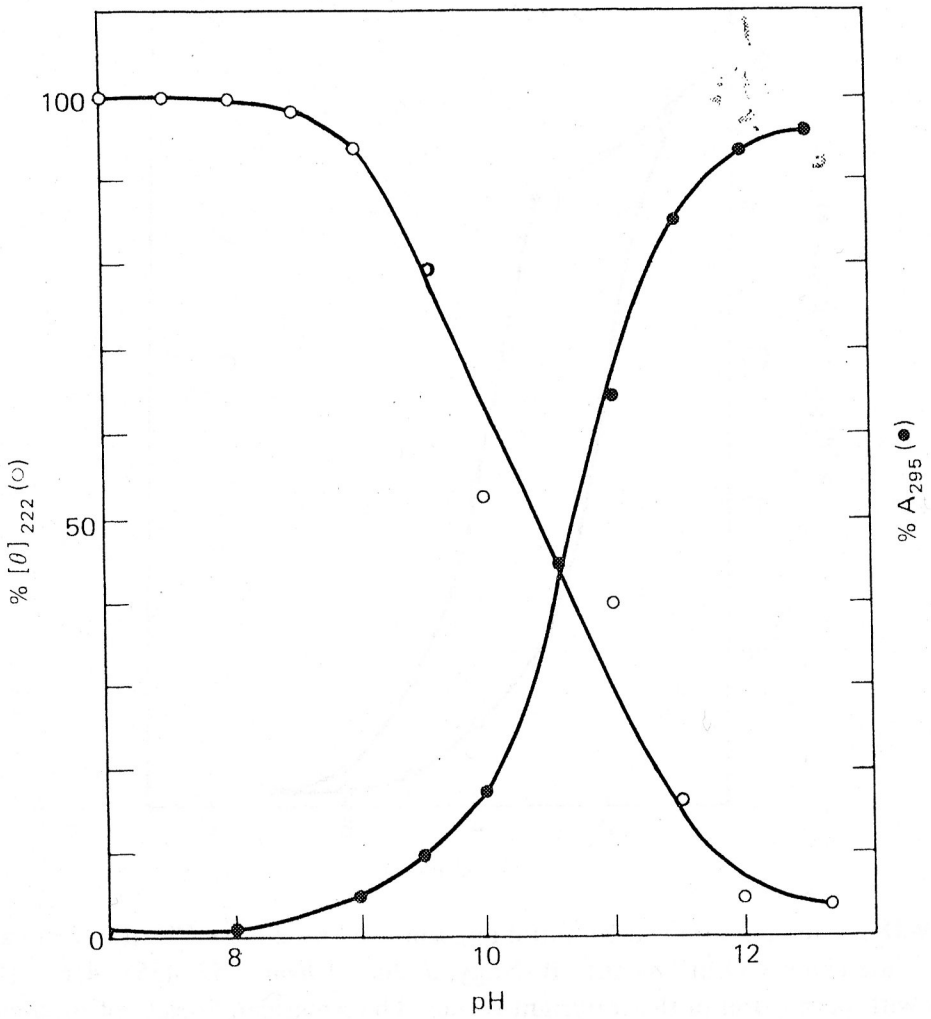

Figure 9-17 Unfolding of tropomyosin followed by CD measurements at 222 nm (○) and absorbance (●) measurements. B. Nagy, *J. Biol. Chem.*, *252*, 4557–4563. (Reprinted with permission of the copyright owner, The American Society of Biological Chemists, Inc., Bethesda, Md.)

On the other hand, in many instances, following α helix (by CD 222 nm measurements) and aromatic residue environments (by CD measurements at 284 nm) gives quite different profiles. Figure 9-18 shows the urea-induced unfolding of a protein, where clearly the requirements of a simple two-state model are not met, since measurements at 222 nm (helix) give a quite different concentration dependence from measurements at 284 nm, where the signal arises from tyrosine side chains.

In summary, CD measurements for the assessment of secondary structure have found some uses: α-helical secondary structure can be experimentally estimated in solution with some degree of confidence, and estimates of helical content based on CD 222-nm measurements are often used to support amounts of protein α helix estimated by secondary-structure prediction. Because of the underlying spectrum of β strand, however, such measurements can be regarded as qualitative at best. They are of considerable use, however, in examining *changes* in secondary protein structure induced by denaturation (as we have discussed here) or induced by ligand binding (as we discuss in Chap. 12).

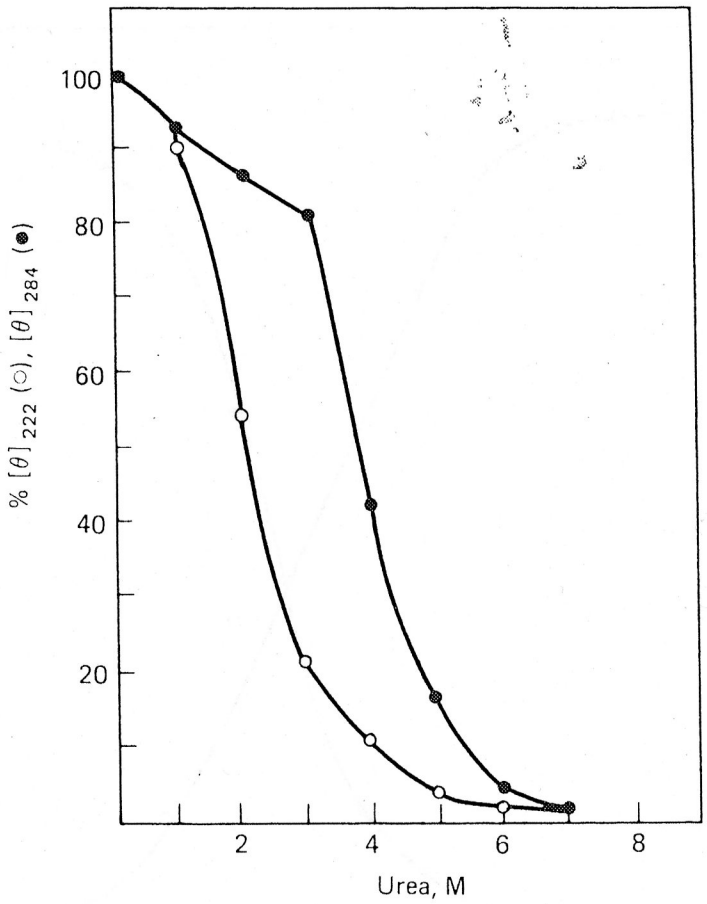

Figure 9-18 Urea-induced protein unfolding followed by α-helix CD at 222 nm and tyrosine side-chain CD at 284 nm. B. Nagy, *J. Biol. Chem.*, *252*, 4557–4563. (Reprinted with permission of the copyright owner, The American Society of Biological Chemists, Inc., Bethesda, Md.)

In this chapter we have examined various aspects of the secondary structure of proteins. Although much insight into the folding pathway and tertiary-structure assembly is obtained by considering secondary structure, perhaps the most important consideration involves the mobility inherent in protein structure. As developed in detail in Chaps. 10 to 12, the stabilities and mobility of formal secondary structure, dominated by rotation about the dihedral angles of the peptide bonds, are intimately involved in the acquisition of tertiary structure and mechanisms of conformational changes in proteins.