10

Tertiary Structure of Proteins

INTRODUCTION

Many experimental approaches have been developed to examine the tertiary structure of proteins. Of these, x-ray crystallography has been the most powerful and influential in directing the way we think of tertiary structure and (considered in Chap. 12) changes in tertiary structure.

On the basis of the structure of myoglobin obtained from the x-ray crystallographic work of Kendrew, four generalizations were formulated to describe tertiary structure.

1. Proteins are compact structures having very small amounts of internal solvent molecules, which are present internally and presumed to have been trapped during the folding process.
2. Almost all the polar side chains in the protein are at the surface of the molecule, where they can interact with solvent and solute molecules in the bulk solvent. Any exception to this would indicate that a "nonsurface" polar group is involved in some internal function. In the case of myoglobin, for example, a histidine side chain is buried internally but is associated with the heme ring of the molecule.
3. All nonpolar residues, with the possible exceptions of glycine and alanine, are located in the interior of the molecule. Glycine and alanine, because of their "short" side chains, can be located at the surface.
4. All polar groups at the surface of the molecule, whether they are side-chain or main-chain C=O and N—H groups, have bound water molecules.

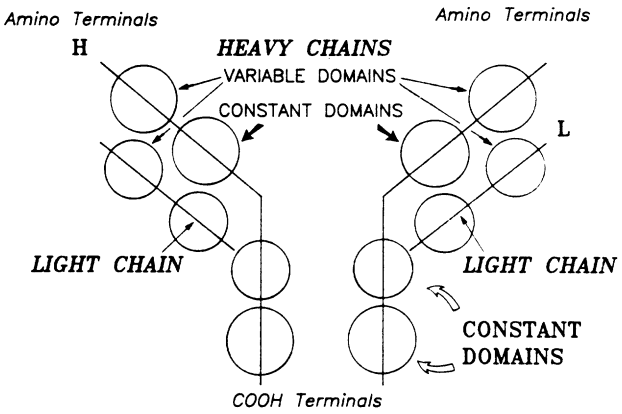


Figure 10-1 Schematic drawing of a typical antibody molecule composed of heavy (H) and light (L) chains.

When one considers that these "generalizations" were made on the basis of the examination of a single protein, it is surprising that numbers 1 and 4 have stood the test of time (and several hundred more crystal structures). The second and third generalizations, although essentially correct, have been modified in the light of further work.

Many proteins with molecular weights of more than about 16,000 Da have been shown to contain "domains" in which the polypeptide obeys the concept of a compact structure, with side chains forming a hydrophobic core and the peptide backbone wrapping around the outside. A typical "domain" containing protein is the IgG molecule, shown in a schematic representation in Fig. 10-1. In this molecule there are several domains, each containing contiguous segments of polypeptide chain, yet each quite separate physically from its adjacent domain. The domains in this instance are so distinct that they can be separated by proteolytic cleavage to give functionally active but separate molecules, each having different domains. The IgG molecule may be regarded as a somewhat anomalous protein, as its domains each contain two chemically separate polypeptide sequences. Although this is by no means unique, most protein domains consist of a single polypeptide chain.

The generalizations regarding the localization of polar and nonpolar side chains in proteins must be reassessed in light of the data shown in Table 10-1. These data were obtained by examination of the x-ray crystallographic structures of chymotrypsin, carboxypeptidase, and thermolysin. Although the observed distributions of the hydrophobic side chains of valine, leucine, isoleucine, and phenylalanine, and the charged side chains of lysine, arginine, aspartate, and glutamate do in general follow Kendrew's generalizations, it is quite apparent that among the "hydrophobic" residues, tyrosine, tryptophan, and proline contain anomalous members, while among the "hydrophilic" residues, serine and threonine also have "black sheep." Of the

TABLE 10-1 Side-chain environments in proteolytic enzymes

Residue	Location		
	Inside	Outside	Surface
Val, Leu, Ile, and Phe	100	43	57
Tyr and Trp	7	20	42
Gly and Ala	52	67	34
Pro	6	13	8
Ser and Thr	33	85	40
Asn and Gln	5	43	35
Lys and Arg	0	44	20
Glu and Asp	8	52	35

charged residues only glutamate and aspartate appear to be occasionally buried. As discussed in earlier chapters, the localization of these carboxyl groups in "nonpolar" regions of the protein leads to perturbed pK values. In some instances these perturbed carboxyl groups (e.g., Glu-35 in lysozyme) seem to have important biological functions, although in other proteins buried carboxyl groups do not appear to have direct biological function.

In this chapter we first consider some of the means available for obtaining tertiary-structural information. From there we examine ways of predicting tertiary structure, some aspects of the dynamic (as opposed to static) nature of protein structure, various aspects of the evolution of tertiary structure, and finally, the processes involved in the acquisition of the tertiary structure of a polypeptide chain.

METHODS OF STUDYING TERTIARY STRUCTURE

A variety of experimental approaches are available to examine different aspects of a protein's tertiary structure. Some of these have already been discussed in other contexts in earlier chapters and are referred to only briefly here. In many cases they give qualitative rather than precise structural information. Although of use in giving a general picture of the tertiary structure, the methods that provide such information are of more use to the protein chemist in providing evidence for, and a means of following, the various conformational changes a protein may undergo.

Solvent Perturbation

The spectral properties of a protein (we restrict the present discussion to UV absorbance and fluorescence) reflect the various environments of the chromophores or fluorophores that contribute to the overall absorbance or fluorescence. For example, a protein with several tryptophan residues usually shows a red shift in its

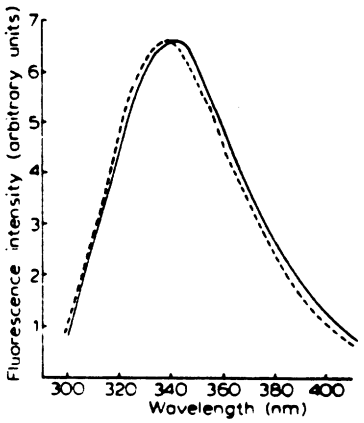


Figure 10-2 Effects of different excitation wavelengths on the emission properties of glyceraldehyde-3-phosphate dehydrogenase. [Reprinted with permission of J. E. Bell and K. Dalziel, *Biochim. Biophys. Acta*, *410*, 243–251 (1975).]

fluorescence emission if the exciting wavelength is increased from 280 nm to 290 nm (see Fig. 10-2).

This reflects the fact that the various tryptophan residues are in different environments with resulting different spectral properties. An examination of the different spectral properties of the chromophores or fluorophores thus provides information concerning the various environments (i.e., the tertiary structure of the protein) in which the residues are located. Without detailed information regarding the location of individual residues in the primary sequence, in terms of their contribution to the overall spectral properties, only qualitative information can be obtained: comments such as "most of the tryptophans in the protein appear to be in hydrophobic environments" are made. A detailed analysis of individual residues is possible if their spectral properties can be specifically modified. In bovine α-lactalbumin, for example, specific tryptophan residues can be oxidized by *N*-bromosuccinimide and, from the effects on the protein fluorescence, the fluorescence properties of individual tryptophan residues estimated. Experiments of this type have allowed the environment of Trp-118 in the molecule to be elucidated.

When tryptophan-118 in α-lactalbumin is modified, there is a far greater quenching of protein fluorescence observed using 290- or 300-nm excitation than using lower wavelengths (Table 10-2). This indicates that Trp-118 has a red-shifted absorbance spectrum and contributes more to the overall fluorescence upon excitation at 290 to 300 nm than the other three tryptophan residues in the protein, and therefore that Trp-118 is in an exposed environment. This is also reflected in the *blue* shift of the excitation spectrum of α-lactalbumin when Trp-118 is modified (Fig. 10-3C). As is also indicated in Fig. 10-3 and Table 10-2, modification of a second tryptophan residue (Trp-26) results in minimal changes in fluorescence, thus suggesting that Trp-26 has a much lower quantum yield than Trp-118.

TABLE 10-2 Quantum yields and quenching of fluorescence in α-lactalbumin modified by *N*-bromosuccinimide

Excitation wavelength (nm)	Quantum yield			Percent quenching	
	Native α-lactalbumin	α-Lactalbumin with 0.95 Trp modified	α-Lactalbumin with 1.6 Trp modified	α-Lactalbumin with 0.95 Trp modified	α-Lactalbumin with 1.6 Trp modified
270	0.033	0.0279	0.0265	15	20
280	0.033	0.0292	0.0281	12	15
290	0.033	0.0245	0.0239	26	28
300	0.033	0.0196	0.0192	41	42

In a similar vein, small molecules that perturb the environment of surface chromophores can be used to estimate the "exposure" of such residues to solvent. If perturbants of differing radii are used, a picture of the surface topography of the protein can be obtained. This method depends on resolving the difference spectrum obtained in the presence of the perturbant into its constituent tyrosine and tryptophan components. Figure 10-4 shows DMSO-induced difference spectra of model compounds.

The protein difference spectrum is then compared to difference spectra obtained using the same perturbant and model compounds. The following equations are used:

$$\Delta_{\lambda_1}(\text{protein}) = a\,\Delta_{\lambda_1}(\text{Trp}) + b\,\Delta_{\lambda_1}(\text{Tyr}) \tag{10-1}$$

$$\Delta_{\lambda_2}(\text{protein}) = a\,\Delta_{\lambda_2}(\text{Trp}) + b\,\Delta_{\lambda_2}(\text{Tyr}) \tag{10-2}$$

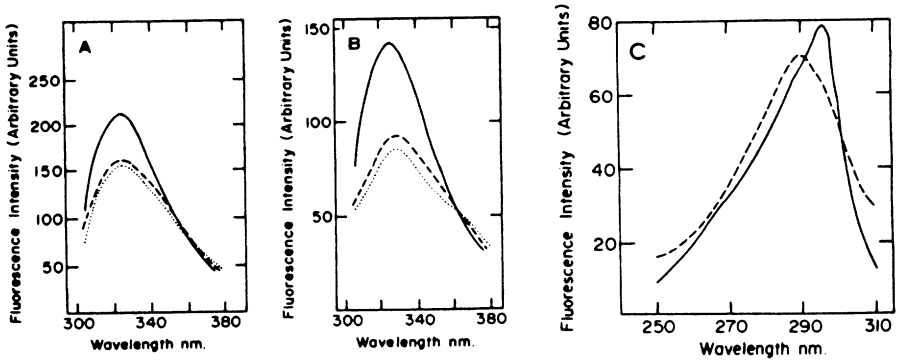


Figure 10-3 Fluorescence emission spectra: pH 7.0, 0.1 *M* NaCl, 10 *M* α-lactalbumin. (A) Excitation at 280 nm. (B) excitation at 300 nm; ——, Native α-lactalbumin; – – –, with 0.95 tryptophan residues modified by *N*-bromosuccinimide; · · ·, with 1.6 tryptophan residues modified by *N*-bromosuccinimide. (C) Fluorescence excitation spectra: emission monitored at 370 nm; ——, native α-lactalbumin; – – –, α-lactalbumin with 0.95 tryptophan residues modified. [From J. E. Bell, J. Castellino, I. P. Trayer, and R. L. Hill, *J. Biol. Chem.*, *250*, 7579–7585 (1975). Reprinted with permission of the copyright owner, The American Society of Biological Chemists, Inc., Bethesda Md.]

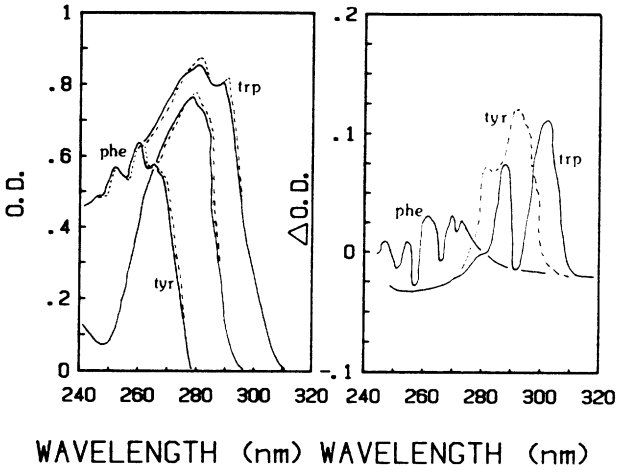


Figure 10-4 (A) Absorption spectra of the *N*-acetyl ethyl esters of tryptophan, tyrosine, and phenylalanine in water (——) and in 20% dimethoxysulfoxide (DMSO) (– – –); (B) DMSO-induced difference spectra for the *N*-acetyl ethyl esters of tryptophan, tyrosine, and phenylalanine.

where a and b are the apparent number of exposed tryptophan and tyrosine residues in the protein and the $\Delta_{\varepsilon\lambda}$ are the molar absorptivity differences of the protein and the free tryptophan and tyrosine model compounds at wavelength λ, the difference spectra maxima. Chromophore exposure is estimated by first neglecting the tyrosine contribution to the long-wavelength peak to obtain an approximate value of a.

Rearranging Eq. (10-1) gives

$$a \simeq \frac{\Delta_{\lambda_1}(\text{protein})}{\Delta_{\lambda_1}(\text{Trp})} \tag{10-3}$$

Thus an approximate value for b is obtained by rearranging Eq. (10-2).

$$b \simeq \frac{\Delta_{\lambda_2}(\text{protein}) - a\,\Delta_{\lambda_2}(\text{Trp})}{\Delta_{\lambda_2}(\text{Tyr})} \tag{10-4}$$

The values of a and b may then be refined by using Eqs. (10-1) and (10-2) reiteratively.

Limited Proteolysis

Susceptibility to proteolysis has often been used to study conformational changes. Altered susceptibility of a protein in the presence of a ligand suggests that the protein has undergone a conformational change that alters the susceptibility of proteolytically cleavable bonds. In such experiments decreased proteolysis could simply reflect physical protection of a susceptible bond from the added protease. Usually, these studies are used only when *increased* proteolysis results from the conformational change.

Because of the availability of a wide variety of proteases with defined specificities and the increased sensitivity of the detection of protein fragments (either by HPLC methods or by gel electrophoresis methods with such sensitive detection as that afforded by silver staining), it has become possible to use limited proteolysis to obtain defined information concerning the environment of various peptide bonds in a protein and changes in their environment induced by ligands or protein–protein interactions. This requires the *limited* proteolysis of the protein being studied such that only a single peptide bond is cleaved per protein molecule, ensuring that only peptide bonds that are exposed in the native protein are studied. Two experimental situations can be envisaged: (1) the primary sequence of the protein is known and information concerning the exposure of various peptide bonds to solvent is sought, and (2) the primary sequence may or may not be known, but information about conformational changes is sought.

The second situation is simply a matter of examining the peptide maps of the *partially* cleaved protein obtained in the presence and absence of the conformational perturbant and looking for differences, and will not be dealt with in more detail.

In the first situation, however, there is the potential for obtaining quite detailed information on the exposure of various bonds to solvent (i.e., the added protease). Two approaches can be employed.

1. After limited proteolysis, the peptide map is obtained. From each cleavage event (only one per individual polypeptide chain) two fragments result; thus for each available cleavage site there are two fragments. Isolation and identification of the various fragments, in comparison with the primary sequence of the protein, allows identification of each susceptible bond.

2. The second varient requires a chemical modification of the original protein prior to proteolysis. If the protein can be reversibly denatured, it may be possible to N-terminal label the peptide with, for example, radioactive cyanate. If the protein does not undergo such reversible denaturation, a *specific* chemical modification (e.g., use of a radioactive site-specific reagent or a radioactive general modification reagent to label a single reactive sulfhydryl or lysine residue with a radioactive derivative) suffices. The modification is used as a *reference point* after proteolysis to assist in the identification of fragments. Fragments are detected after separation by their radioactivity, that is, only one of each pair of fragments is detected. If N-terminal labeling has been possible, the size of the fragment may be sufficient to indicate which susceptible bond has been cleaved. If an internal label has been used, the detected fragments fall into one of two classes, as outlined in Fig. 10-5. After limited cleavage and separation of labeled fragments, the C-terminal and the N-terminal residues of the fragments are determined to indicate which of the two classes each fragment falls into. Once this has been established, the identification of the cleavage sites is the same as with N-terminal-labeled proteins.

In summary, limited proteolysis readily gives information about conformational changes and in conjunction with the primary sequence can give precise information

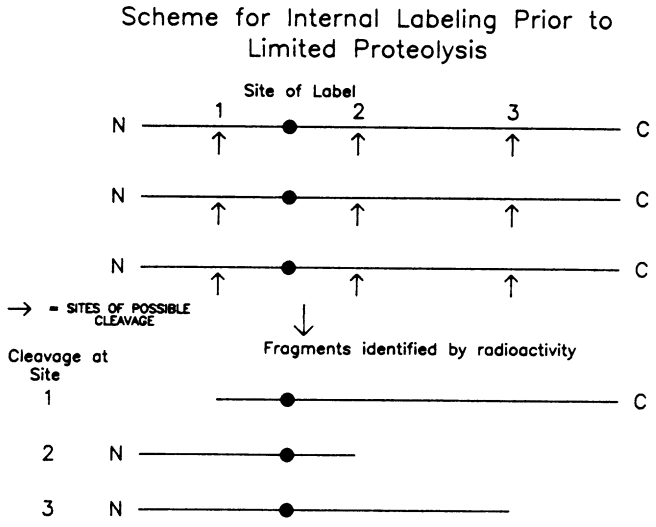


Figure 10-5 Scheme for internal labeling prior to limited proteolysis.

regarding exposed peptide linkages in the native structure and whether or not particular regions are involved in conformational changes.

X-ray Crystallography

Clearly, x-ray crystallography is a most powerful tool in determining the tertiary structure of a protein. Later in this chapter and in Chaps. 11 and 12 we consider in some detail the *type* of information that it gives about tertiary structure. At this point, however, we examine some of the problems, experimental approaches, and questions that are encountered when x-ray crystallography of a protein is attempted. There is no discussion of the technical and theoretical aspects.

Let us begin with a quote from Karplus: "The x-ray structure of a protein is an average structure of many molecules obtained over a long period of time. As such, the probability that a single molecule has the structure, as determined, is essentially zero". We use this quotation not to discredit x-ray crystallography but to give focus to the various questions a protein chemist or enzymologist must consider when examining a protein's crystal structure.

The major question that must be addressed is whether or not the crystal structure resembles that of the protein in solution. Several comments must be made. The conditions necessary to crystallize proteins often are very nonphysiological (such as high salt concentrations, the presence of organic solvents, or high/low pH) and usually require long periods of time (often, 1 to 2 weeks, which may limit the type of protein that can successfully be crystallized, as it must be stable for these time periods). Related proteins, for example α- and γ-chymotrypsin, have essentially identical amino acid sequences but may crystallize quite differently. The tertiary structures obtained for these chymotrypsins are, however, with the exception of a few surface residues,

essentially identical, suggesting that the crystallization conditions do not affect the final elucidated structure.

It is often possible to examine the *activity* of the protein while it is in its crystal lattice. Clearly, there are many factors that would be expected to affect the activity of a protein in a crystal lattice, such as diffusion of substrates or products; these can be taken into account. Ribonuclease, carboxypeptidase, chymotrypsin, papain, and alcohol dehydrogenase have all been shown to retain activity in the crystal form. When attempts to correct for diffusional effects are made, it appears that at least with chymotrypsin and papain, the crystal forms have their expected activity. With alcohol dehydrogenase and carboxypeptidase, however, approximately a 200-fold decrease in activity in the crystal form is estimated, which may be attributed to an overall decrease in protein motility, even though there is no major tertiary structural differences between the solution and the crystal structures.

This brings up the question of resolution. In small-molecule crystal structures resolution to about 0.01 Å is possible. This level is *not* achievable in proteins: There is an inherent limitation arising from the fact that protein molecules do not align as precisely in crystal structures as do small molecules. This seems to be due in part to the purity of the proteins and in part to the solvent content of the crystals. In crystals with high solvent content, and consequently less intermolecular contact, the crystal gives diffraction patterns showing a greater level of random disorder, suggesting that unrestrained surface residues may have multiple conformations.

Finally, we must consider the effects of the *heavy atom derivatives* that are used to solve the crystal structure. This is usually achieved in one of three ways.

Trial-and-Error Approach. Crystallization either occurs from a solution containing a heavy atom salt, usually platinum chloride, or the crystal, once formed, is soaked in a solution containing a heavy atom salt, with the hope that a specific derivative will be formed. Platinum chloride seems to bind to exposed methionine or histidine residues in many proteins, and this approach is often successful.

Specific Chemical Modifications. Amino acid side chains in the protein can often be specifically modified with a chemical reagent containing a heavy atom. Mercury-based reagents (e.g., *p*-hydroxymercuribenzoate or *S*-mercuric, *N*-dansyl cysteine) are often employed to modify sulfhydryl groups and introduce a heavy atom. Reagents containing iodine are often useful, and the iodination of tyrosine residues or of amino groups (with iodo derivatives of phenylisothiocyanate) has been done.

Heavy Metal Substitution. In many cases a naturally occurring metal ion in a metalloprotein can be removed and replaced with a heavy metal. The zinc in carbonic anhydrase can be removed and the zinc-free enzyme dialyzed against mercuric acetate, which results in Hg replacing Zn. Calcium ions in proteins can often be replaced with barium, as in the case of staphylococcal nuclease. Various proteins, which on the basis of activity have no metal requirement, may in fact have specific metal binding sites (often for calcium or magnesium) that can be substituted for by a heavy metal ion.

Heavy metal derivatives of proteins may, of course, cause conformational alterations that can be reflected in the determined crystal structure. Activity measurements in solution in the presence of whatever chemical modification or substituted metal are used can give an indication of whether the heavy metal ion produces such altered states.

Chemical Methods

In Chap. 7 a wide variety of chemical modification reagents were discussed, and general mention of their use in assessing availability of particular amino acid side chains for modification was made. In addition, several instances of chemical cross-linking were considered briefly in the context of tertiary structure determination.

In general, the fact that a particular amino acid side chain can be chemically modified by a water-soluble modification reagent implies that the residue is solvent accessible (i.e., exposed at the surface of the protein). Information concerning residues in the primary sequence that are exposed in the tertiary structure can thus be obtained. In conjunction with secondary- and tertiary-structure prediction, chemical modification data can give meaningful insight into the tertiary structure. It is important that the modification is to a single residue, so that tertiary-structural changes as a result do not lead to additional modification events, which may be misinterpreted in terms of exposure in the original tertiary structure. Table 10-3 details some of the chemically modifiable residues in bovine glutamate dehydrogenase, together with their predicted secondary and tertiary structures. In each case, where defined secondary structure is thought to be present, the tertiary-structural assessment indicates a hydrophilic (i.e., solvent exposed) character (the ? in Table 10-3 indicates that the modified residue ap-

TABLE 10-3 Analysis of chemical modification in relationship to predicted secondary structure

Residue	Predicted secondary structure	Possible amphiphilic nature
Lysine-27	α Helix	Hydrophilic
Lysine-105	α Helix	Hydrophilic
Cysteine-115	β Sheet	Hydrophilic
Lysine-126	Random	—
Lysine-154	α Helix	Hydrophilic
Methionine-169	Random	—
Tyrosine-190	β Sheet	?
Tyrosine-262	Random	—
Lysine-269	Random	—
Cysteine-319	β sheet (N terminal)	?
Lysine-333	Random	—
Lysine-358	α Helix	Hydrophilic
Tyrosine-407	β Sheet (N terminal)	?
Lysine-420	α Helix	?

pears at the hydrophilic end of a possible amphiphilic structure). This type of analysis gives confidence in assigning modifiable residues to hydrophilic environments.

Chemical cross-linking studies of the type described in Chap. 7 can also, with the use of cross-linking reagents of differing lengths and specificities, give much valuable information, either of a direct nature or in conjunction with prediction studies, regarding tertiary structure.

Solution Methods

In Chap. 4 it was indicated that various approaches for the determination of molecular weight depend on assumptions concerning the *shape* of the protein whose molecular weight is being evaluated. The converse of this is, of course, that if the molecular weight of the protein is known, such methods give information on the overall shape of the protein. Although such approaches as sedimentation, gel filtration, and light scattering do not yield precise information concerning the surface topography of a protein, they can give valuable data on the overall shape,. usually in terms of an axial ratio. Such methods are often sensitive enough to detect differences in shape under different conditions, and as such can be used to follow conformational changes. These approaches are discussed in more detail in Chap. 12.

A potentially very useful way of obtaining tertiary-structural information (at the level of the separation between points in a protein molecule) is fluorescence resonance energy transfer (see Figs. 10-6 and 10-7). If several *specific* sites can be used for such measurements, distances between specific points on a protein can be estimated. Such

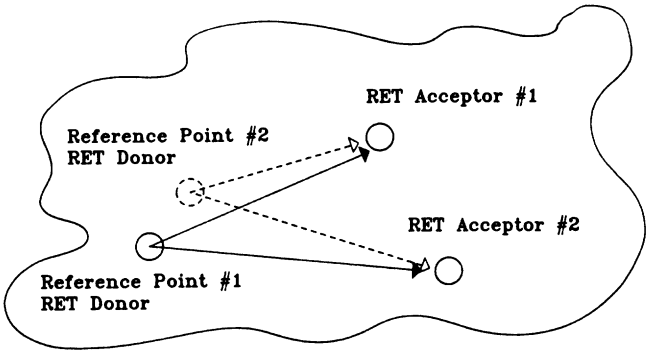
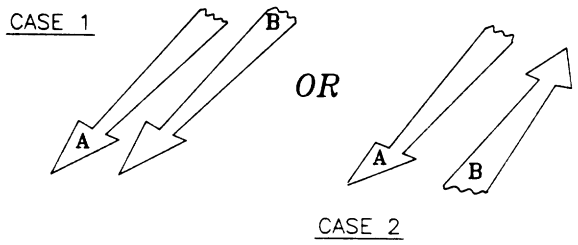


Figure 10-6 General resonance energy transfer measurements used to obtain information about the relative location of various fluorescence donors and acceptors that might be introduced into a protein.



EXPERIMENTAL PROTOCOL

1. Add RET Donor @ point A
2. Add RET Acceptor @ point B
3. Measure distance between A & B

RESULTS distinguish between case 1 and case 2

Figure 10-7 Use of fluorescence energy transfer measurements to resolve specific questions with regard to the orientation of elements of tertiary structure.

distances can, for example, be used to estimate the separation of specific binding sites on a protein or the distance of a specific site from a metal-ion binding site. In addition, they (as with chemical cross-linking data) can be useful in the prediction of tertiary structure by restricting the number of possible arrangements of secondary structure that have to be considered.

PREDICTION OF TERTIARY STRUCTURE

The prediction of the tertiary structure of a protein from its amino acid sequence alone, although in theory possible, is probably a practical impossibility. Some workers have attempted such feats based on thermodynamic calculations but have, in large measure, been frustrated by the magnitude of the task. (Such attempts have, however, given much insight into the dynamic nature of protein tertiary structure.) We concern ourselves here with the prediction of what we shall term *tertiary structure packing elements*. The basis for this approach comes from the model of protein folding described in Chap. 9. Any region of *secondary* structure within the tertiary structure of the protein is packed according to the general principles espoused by Kendrew and summed up in the phrase "hydrophobics in, hydrophilics out." A particular α helix or β sheet is oriented in the tertiary structure of the molecule such that it obeys this simple rule. If some idea of the amphiphilic nature of particular elements of secondary structure can be obtained, some idea is also obtained of how such secondary structure may be oriented in the tertiary structure of the protein.

With β sheets there is a distinct "sidedness" of the secondary structure. Shown in Fig. 10-8 is a representation of the predicted β-sheet regions of glutamate dehydrogenase. Alternating residues in each region of secondary structure lie on opposite

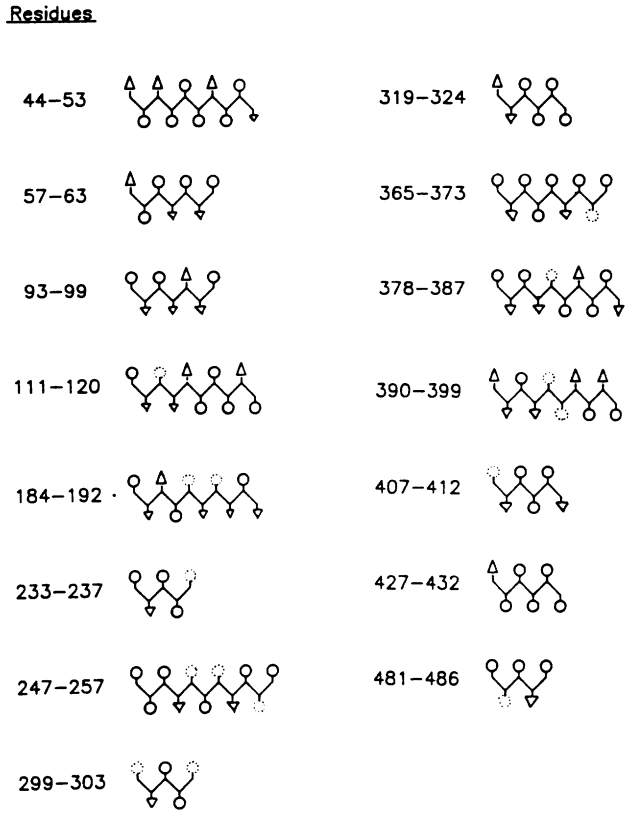


Figure 10-8 Depiction of β sheets predicted for glutamate dehydrogenase.

sides of the sheet, and depending on the nature of the side chain, the sheet may acquire a hydrophobic or a hydrophilic side.

In Fig. 10-8 hydrophobic residues are represented as circles and hydrophilic residues as triangles. Obviously, many of these β-sheet regions have clearly hydrophobic or hydrophilic sides. In other instances, a predicted sheet region has one-half of the sheet (in terms of the linear sequence) hydrophobic. In such a case it is be inferred that the sheet might extend from the inside of the tertiary structure toward the solvent-exposed exterior. β Sheets with a hydrophobic side are expected to have the hydrophobic side facing inward, with the more exposed, hydrophilic side oriented toward the surface of the tertiary structure. Sheet regions with a completely hydrophobic nature are internally located.

α-Helical regions of secondary structure can be assessed in terms of their regions of hydrophobicity by means of "Edmundson wheels"; several are shown in Fig. 10-9 for predicted helical regions of glutamate dehydrogenase. These wheels represent a

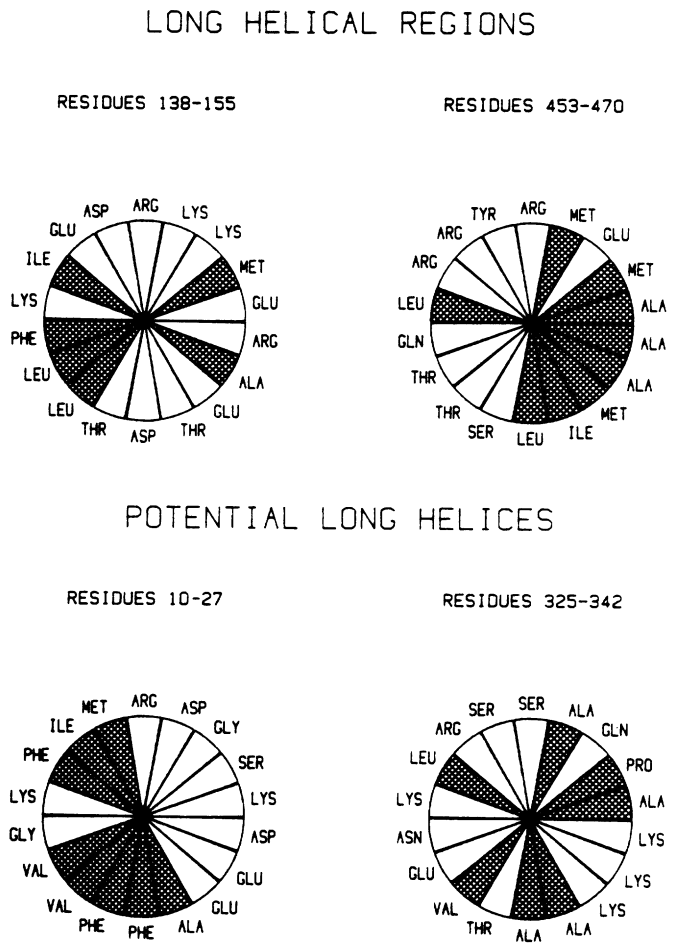


Figure 10-9 Edmundson wheels for several regions of α helix predicted for glutamate dehydrogenase.

projection, onto a flat surface, of an α helix, looking down on the axis of the helix. The construction of such a projection is shown in Fig. 10-10, where the number of each segment of the "wheel" represents the position of the residue in the linear sequence.

In the final representations, illustrated in Fig. 10-9, the hydrophobic side chains are shown as shaded segments and the hydrophilic side chains as open segments. Not only do these representations allow an assessment of stabilizing hydrophobic ridges along a predicted helix, but they also show the amphiphilic nature of the helices. As with the β-sheet tertiary-structure packing elements, α helices can thus be oriented with regard to whether or not they are likely to face internally or externally within the tertiary structure of the protein.

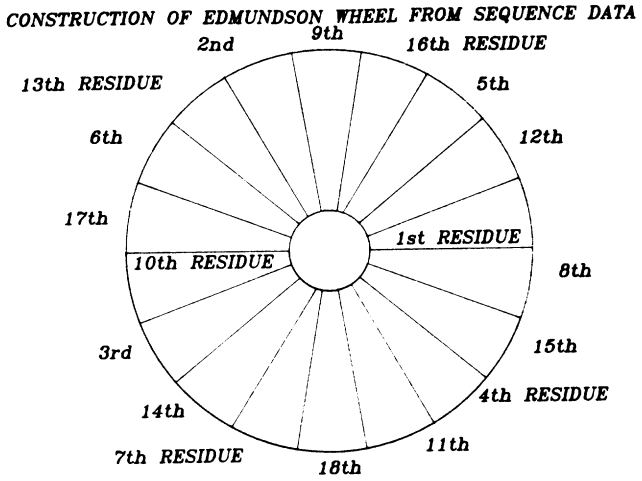


Figure 10-10 Scheme for Edmundson wheels: prediction from sequence.

An analysis of such tertiary-structural packing elements in conjunction with the results of chemical modification, chemical cross-linking, fluorescence resonance energy transfer, and active (or binding)-site labeling studies can give considerable insight into the overall tertiary structure of a protein.

A more precise conception can be obtained by further consideration of the nature of the interaction between packing elements in the formal tertiary structure. Strands of β sheet can interact to give a parallel or an antiparallel sheet. From an analysis of β-sheet regions of proteins whose crystal structure has been elucidated, several generalizations can be made.

1. In antiparallel sheet structures, branched side chains are found next to unbranched side chains to give better packing.
2. In parallel sheet structures, like-type side chains are found next to one another, either branched or unbranched.

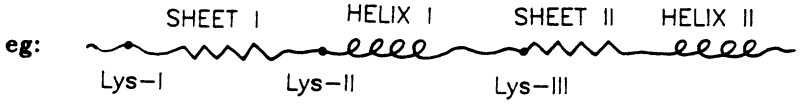
Similarly, in proteins where α helices must pack with one another, it is found that the various helices have "rows" and "ridges" and that they pack by their intercalation.

Although we have not discussed β turns in any detail in the context of tertiary structure, it is found, again by examination of known crystal structures, that β-turn regions usually lie at the surfaces of proteins and that the side chains of the residues involved usually point in the same direction.

In summary, although we stated at the start of this section that the prediction of the tertiary structure of a protein based only on its amino acid sequence remains a practical impossibility, it does seem that a judicious application of some of the principles of tertiary-structural packing elements and their interactions, together with chemical and biophysical experimental results, does allow a tertiary-structural model

SEQUENCE

Secondary Structure Prediction

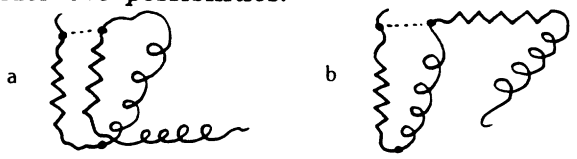


CHEMICAL CROSS–LINKING

Find Lys–I cross–linked to Lys–III

This places restrictions on possible conformations

Consider two possibilities:



EXAMINE PACKING ELEMENTS:

Look for hydrophobic surfaces

For Example,possible interactions might include:

SHEET I – SHEET II
HELIX I – SHEET II
HELIX I – HELIX II

IF RESULTS SHOW :

1. That sheets I & II both have strong hydrophobic sides– consistent with "a"
2. If helix I and sheet I both have hydrophobic sides,could favor "b".

INTRODUCE R.E.T COUPLE:

If acceptor – donor pair can be introduced at known residues in helix I & II, then distance measurements could favor "a" or "b".

FURTHER CROSS–LINKING & CHEMICAL MODIFICATION:

Use photo–activatable reagent specific for lysine – look for other cross–links. General Chem.Mod gives exposed residues.

FINAL TESTS:

Once a predicted structure is obtained it can be tested using site directed mutagenesis to alter key residues in secondary structure and examine effects on function & structure:correct model gives prediction

Figure 10-11 Scheme for building a tertiary structure model based on prediction and experimental evidence.

of a protein to be formulated. A simplified scheme for how such an endeavor might proceed is given in Fig. 10-11.

As indicated, a variety of biochemical and biophysical experimental methods can be used to obtain a relatively low resolution picture of the tertiary structure. Although such pictures cannot compare to the detailed information obtained from, for example x-ray crystallography, they can contribute significantly to the state of knowledge about the tertiary structure of a given protein.

It is not unusual to have available a primary structure from which secondary-structure predictions can be made. The limitations of such predictions were discussed in Chap. 9 with the principal criticism involving the lack of any knowledge of tertiary structure and the effects that might have on secondary structure. It is necessary to examine environmental effects of experimentally determined facets of tertiary structure on the predicted secondary structure: For example, does the interaction of a helix with a particular region of sheet lead to increased or decreased stability of either element?

Although such combinatorial approaches to examining the secondary and tertiary structure of a protein are low resolution, they can give insight into the types of structures associated with the binding sites and surfaces of a protein and lead to an increased understanding of its tertiary structure. More important, it is possible that such approaches permit the design of experiments to further test postulated relationships between regions of the protein.

of a protein to be formulated. A simplified scheme for how such an endeavor might proceed is given in Fig. 10-11.

As indicated, a variety of biochemical and biophysical experimental methods can be used to obtain a relatively low resolution picture of the tertiary structure. Although such pictures cannot compare to the detailed information obtained from, for example x-ray crystallography, they can contribute significantly to the state of knowledge about the tertiary structure of a given protein.

It is not unusual to have available a primary structure from which secondary-structure predictions can be made. The limitations of such predictions were discussed in Chap. 9 with the principal criticism involving the lack of any knowledge of tertiary structure and the effects that might have on secondary structure. It is necessary to examine environmental effects of experimentally determined facets of tertiary structure on the predicted secondary structure: For example, does the interaction of a helix with a particular region of sheet lead to increased or decreased stability of either element?

Although such combinatorial approaches to examining the secondary and tertiary structure of a protein are low resolution, they can give insight into the types of structures associated with the binding sites and surfaces of a protein and lead to an increased understanding of its tertiary structure. More important, it is possible that such approaches permit the design of experiments to further test postulated relationships between regions of the protein.

DYNAMIC ASPECTS OF TERTIARY STRUCTURE

General Experimental Evidence for the Dynamic Nature of Proteins

For many years there was discussion as to whether proteins were best regarded as essentially rigid structures (suggested by early ideas of enzyme mechanisms and given some credence by x-ray crystallographic structures of proteins) or as "floppy bodies" with large amounts of conformational mobility. In terms of enzyme mechanisms it was pointed out that an enzyme needed conformational flexibility so that it could lower the free-energy level of a transition state for an enzyme–substrate complex via a set of conformational fluctuations; furthermore, transient fluctuations of structure were inevitable even at thermodynamic equilibrium. Direct evidence of such conformational fluctuations; on a nanosecond time scale were provided by the fluorescence quenching experiments of Lakowicz and Weber. They demonstrated that all regions of a protein molecule were accessible to collisional quenchers such as oxygen, even though the crystal structure showed no possible mechanism for such quenching: the protein *must* undergo rapid structural fluctuations, allowing direct, collisional quenching to occur.

We can consider the *time scale* of various events involved with protein structure and function, as shown in Table 10-4. The majority of these processes occur at the surface of a protein molecule and involve relaxation of bound water molecules. A

TABLE 10-4 Time scale of events

Determinant	Time (s)
Protein surface	
Bound-water relaxation	10^{-9}
Side-chain rotational correlation	10^{-10}
Ionization of side chains	10^{-7}–10^{-9}
Protein conformations	
Local motion	10^{-8}–10^{-9}
Isomerization	10^{-2}–10^{-7}
Folding–unfolding	10^{2} –1
Enzyme substrate reactions	
Encounter rate	Diffusion control
Lifetime of transition state	10^{-10}
Metal-ion coordination events	10^{-6}–10^{-9}
ES conformational changes	10^{-2}–10^{-4}
Covalent ES intermediate lifetimes	10^{-2}–10^{-4}

variety of different experimental approaches have been used to give some insight into the dynamic nature of protein tertiary structures, and here we briefly consider several types.

Cryogenic Experiments. If at room temperature a protein is in a dynamic state of flux between a number of energy equivalent conformations (i.e., has a dynamic structure), it should be possible to trap various structures by lowering the temperature. This has been examined by studying the kinetics of ligand binding to hemoglobin at low temperature. It is found that nonexponential rebinding of ligand occurs after photo-dissociation, indicating that there exists a continuous spectrum of activation energies due to the existence of many conformational states in which the thermal energy, at these low temperatures, is less than the kinetic energy barrier between the states. At room temperatures the thermal energy is such that these many states are rapidly interconverted and the rebinding of ligand occurs as a simple, single exponential process.

Hydrogen-Exchange Kinetics. In a protein there exists many protons which exchange if exposed to tritiated water. In an unstructured polypeptide the NH protons of the peptide bond exchange very rapidly. In a protein in its native structure these same protons exchange at a much slower rate than in the unfolded form, but they do still exchange. Although many other side-chain protons in a native protein can exchange with solvent protons, their rate is too fast to be measured. In some cases, buried side-chain protons exchange at rates slow enough for measurement.

As with the fluorescence quenching experiments referred to earlier, an examination of the x-ray crystallographic structure of a protein does *not* reveal a mechanism by which such exchanges can take place with internal NH protons. In effect, hydrogen exchange samples the solvent accessibility to interconverting forms of a protein.

There are a number of ways that this can be followed experimentally. For the NH group of the peptide bonds the simplest is to follow the infrared amide II band, which differs for NH and NT (or ND). Where the chemical shift of a group whose proton exchanges is known, NMR spectroscopy gives information about exchange rates. This approach, however, can be applied only when the resonance is shifted away from the main protein signal. In a crystal lattice the positions of H and D (or T) can be determined by neutron diffraction. For such experiments proteins are crystallized from aqueous solution, and deuterated solvent is perfused into the crystal. Neutron diffraction studies give information about *which* protons can be exchanged, although it must be recognized that the information obtained is relevant to exchanges that can take place *within the constraints* of the crystal lattice.

Several nonspectroscopic methods of following hydrogen exchange are available. In general, they depend on the rapid separation of solvent from protein and direct estimation of tritium in either solvent or protein. Such separations can be achieved by freeze-drying to obtain the solvent and/or protein, or by rapid dialysis or gel filtration.

As indicated earlier, hydrogen-exchange kinetics reflect the conformation and accessibility of exchangeable protons to solvent (the most slowly exchanging protons are often those which are hydrogen bonded). There are two basic mechanisms by which such dynamic exchange can take place: local unfolding and exchange from a folded conformation. It should be noted that in most proteins there is a core of protons that exchange at rates too slow to be measured when temperatures are reduced below those where continual thermal unfolding occurs. This suggests that some exchanges do not occur from the folded conformation but require the dynamic structure of the protein. It is also generally observed that any effect causing an increased stability of the protein to denaturation results in *decreased* exchange rates.

Two mechanisms by which solvent can have access to buried regions of the protein allow exchanges to take place. These are summarized in Fig. 10-12. In the first, there exists a conformation of the protein in which certain regions are totally inaccessible, and exchange of protons in that region cannot occur unless there is a reversible, localized cooperative unfolding of an internally hydrogen-bonded segment. In the second mechanism there may exist a number of smaller-amplitude motions, any of which individually would not allow exchange to take place, but which together lead to a finite probability that solvent can progressively penetrate into the exchange site.

Evidence for *local unfolding* as the mechanism of such exchanges has come from studies of oxygen binding to hemoglobin. When oxygen binds there is an *increase* in the exchange rates of some, but not all, of the protons that could be exchanged in deoxyhemoglobin, thus allowing preferential labeling of protons that exchange slowly in deoxyhemoglobin but rapidly in oxyhemoglobin. This differentially labeled species allows the transfer kinetics of these protons from the rapidly to the slowly exchanging type of proton to be followed. It is found that the transfer is approximately first order, indicating that the process results from a perturbation of a local unfolding equilibrium as a result of oxygen binding.

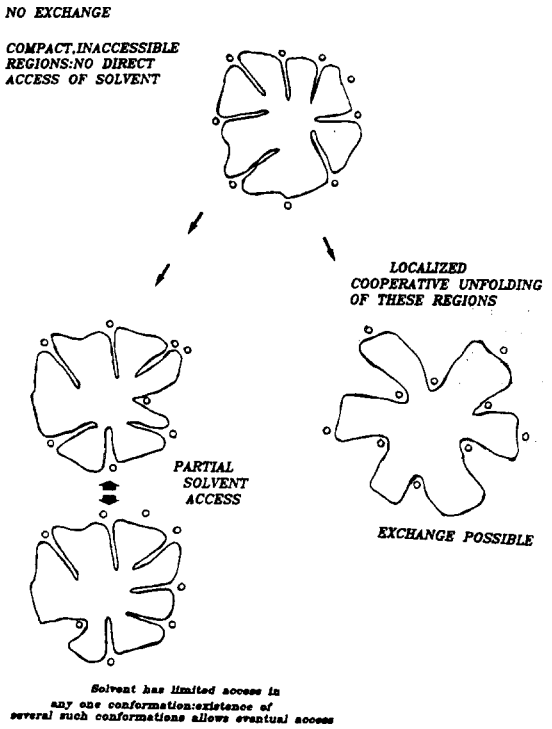


Figure 10-12 Possible mechanisms for the exchange of protons from regions of a protein that appear inaccessible.

Hydrogen exchange is a very sensitive probe for induced conformational changes. An increase in exchange rate reflects an *increased* internal accessibility and mobility and a *decreased* conformational stability. In lysozyme, the activation energy for exchange increases after binding of an inhibitor that decreases the exchange rate. It is also observed that binding of the inhibitor stabilizes the protein against thermal unfolding. In general, decreased exchange rates have correlated well with increased stability to temperature or urea denaturation, while increased exchange rates have correlated with increased susceptibility to proteolytic cleavage.

Models for the Dynamic Nature of Polypeptide Structure

The period between 1965 and 1975 can be described as the decade of the rigid molecule. Replicas of double-helical DNA and various protein molecules dominated many textbooks and much of the thinking. Results from high-resolution x-ray crystallography strongly influenced protein characterization. The intricate and detailed drawings of proteins that developed led to the image of a protein having each of its atoms fixed in position. When x-ray data made visible cases of conformational

change, such as ligand or substrate binding, the conclusions proposed were abrupt transitions between otherwise static structures.

Today the static picture of protein structure is replaced by the view of proteins as delicately balanced dynamic systems. It is now realized that the atomic positions determined by x-ray diffraction studies represent the average equilibrium geometry of the molecule; these atoms exhibit fluidlike motions of sizable amplitudes around their average positions. The growing importance of protein dynamics has led to more accurate interpretations of protein function not possibly described by a static viewpoint. Some examples are the following:

1. The functional interactions of flexible ligands with their binding sites, which often require conformational adjustments in both ligand and binding protein.
2. Structural changes in binding proteins, which regulate activity through induced-fit and allosteric effects.
3. The chemical transformations of substrates by enzyme, which involve atomic displacements in the enzyme–substrate complexes. The mechanisms and rates of such transformations are sensitive to dynamic properties of these systems.
4. Electron transfer processes, which depend strongly on vibronic coupling and fluctuations that alter the distances between the donor and acceptor.
5. Various structural transitions that sometimes alter points of contact between proteins and other molecules which affect the activities of (a) contractile proteins such as myosin, (b) other enzymes, and (c) antibodies, as well as proteins involved in (d) membrane transport and (e) genetic regulation. Overall, it appears that understanding the functions of proteins requires an investigation of the dynamics of structural fluctuations and their relation to activity and conformational change.

In this section we discuss in detail two methods used to study specific motions that occur in proteins. The first presents a simplified model of a globular protein and the second presents a simplified model of a polypeptide chain, more specifically an α helix.

Globular proteins exhibit a variety of motions that can be classified according to their amplitude, energy, or time scale (Tables 10-4 to 10-6). Such structural flexibility implies that the secondary bonds are strong enough to fit different conforma-

TABLE 10-5 Classification of internal motions of globular proteins

Scales of motion (300 K)
Amplitude: 0.01–100 Å
Energy: 0.1–100 kcal
Time: 10^{-15}–10^{3} s
Types of motion
Local: atom fluctuations, side-chain oscillations, loop and "arm" displacements
Rigid body: helices, domains, subunits
Large scale: opening fluctuation, folding and unfolding
Collective: elastic-body modes, coupled atom fluctuations, soliton and other nonlinear motional contributions

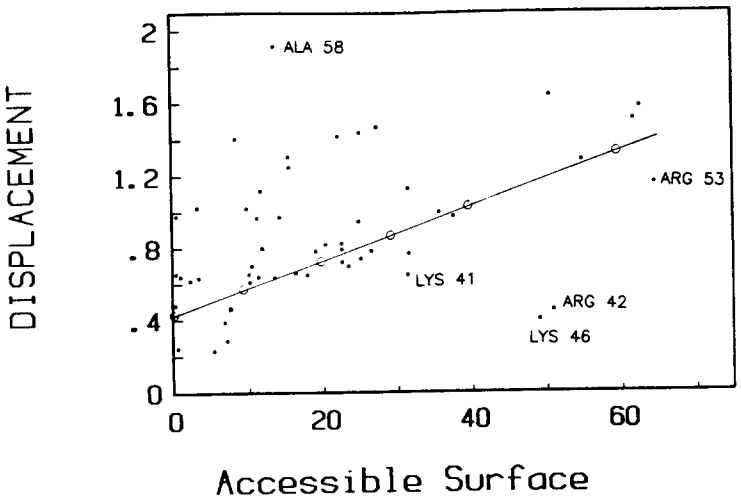


Figure 10-17 Plot of displacement versus accessible surface area. (Reprinted with permission from: P. K. Ponnuswamy and R. Bhaskaran, *Int. J. Pep. Protein Res., 24,* 168–179. Copyright 1984 Munksgaard International Publishers Ltd., Copenhagen, Denmark.)

Peptide Models: The Helix–Coil Transition

The helix–coil transition has several roles: Frequently, it nucleates the folding of globular proteins and has also been implicated in the promotion of certain hormone–receptor interactions. Despite its importance, it is difficult to evaluate the dynamics of this transition, as the time scale limits analysis. Whereas most dynamic simulation methods for polypeptide chains utilize a time span of 10^{-12} to 10^{-9} second, the helix–coil transition requires a range of 10^{-9} to 10^{-6} second. To overcome this, the following model for the polypeptide chain, in the aqueous solution, was proposed. The simplified character of the chain permits the calculation of overall peptide energies and the forces acting on individual residues.

Theory. In the simplified model, each amino acid residue is depicted by a soft sphere with a volume comparable to that of the corresponding amino acid. A consideration of a variety of x-ray structures suggests that with an α helix this approximation is adequate. Each sphere is defined as a single interaction center, R_i, which is connected to other interaction centers by virtual bonds, as shown in Fig. 10-18.

The centroids of the interaction centers are placed at the C^β positions of the detailed structures. These centers are linked by virtual bonds and corresponding virtual dihedral angles, which replace the ϕ and ψ dihedral angles about the C^α. The virtual dihedral angle Φ_i is defined in terms of the residue centers i, $i + 1$, $i + 2$, $i + 3$; $\Phi_i = 0$ for the eclipsed conformation. Each virtual dihedral angle has an associated torsional potential that represents the average sum of nonbonded interactions between the atoms of near-neighbor residues. Reference to the detailed model of the standard

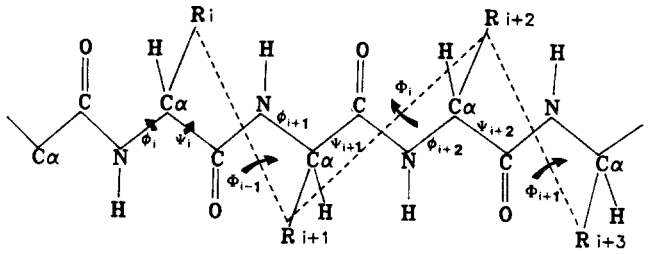


Figure 10-18 Geometry of the simplified polypeptide model compared with that of the atomic model. The dashed lines are virtual bonds connecting the interaction centers R_i.

α-helix geometry, presented by Pauling et al., leads to the following conditions for the simplified chain: b, equilibrium bond length = 5.14 Å, θ_0, equilibrium bond angle = 87.2°, and Φ_α, characteristic α-helix dihedral angle = 38.3°. Levitt determined a 30° width for α-helix region; thus the range $25° < \Phi < 55°$ defines the α-helix boundary for the simplified chain.

In the model, torsional motions about the virtual bonds depict large-scale conformational changes. The energy function representing the chain defines a balance between interactions that stabilize either the helical conformation or the coil conformations. This energy function is obtained by averaging inter-residue interactions over all possible local atomic configurations within each residue; thus the model incorporates the separate time scales for local and overall chain motions. The reduced number of degrees of freedom allows rapid calculation of the energy and forces, which in turn permits significantly longer dynamic simulations than a more detailed model. Although it does not accurately represent detailed atomic interactions, such as side-chain hydrogen bonding, the model does describe overall structure flexibility directed by simple packing effects, such as steric and hydrophobic interactions, within the polypeptide–solvent system. The lengthened time scale of these fluctuations causes rapid averaging of localized motions, such as side-chain internal rotations. The dynamic simulation applied to the model simulates the overall chain motions for periods of several hundred nanoseconds. While the method appropriately simulates local unfolding and folding of proteins and their secondary structural elements, such as helix–coil transitions, it could not accurately simulate folding of an entire protein, which requires a longer time scale and more detailed atomic interactions.

To study dynamics theoretically, it is necessary to understand the potential-energy surface, the energy of the system as a function of the atomic coordinates. This potential energy is used to determine the relative stabilities of different possible structures of the system. The forces that act on the atoms of the system are derived from the first derivative of the energy potential with respect to the atom positions. These forces are then used to calculate dynamic properties of the system by solving equations of motion that determine the change in atomic positions over time. Most of the motions occurring at ordinary temperatures leave bond lengths and bond an-

gles of polypeptide chains near their equilibrium values, which remain relatively constant throughout the protein; thus the standard dimensions of the α helix proposed by Pauling in 1951 were used to determine the dimensions in the simplified model. It is the contacts among nonbonded atoms that are significant in the potential energy of the higher structures.

The following energy function approximates the energy potential, reflecting the average *over the polypeptide degrees* of freedom omitted in the simplified model and *over degrees of freedom* of the solvent molecules. The function is a sum of six distinct interactions:

$$E = E_{\text{bond}} + E_{\theta} + E_{\Phi} + E_{\text{sol}} + E_{\text{ev}} + E_{\alpha} \tag{10-17}$$

E_{bond} and E_{θ} represent the bond and bond-angle interactions, respectively;

$$E_{\text{bond}} = \sum k_b(b - b_0)^2 \tag{10-18}$$

and is the summation over all virtual bonds; and

$$E_{\theta} = \sum k_{\theta}(\theta - \theta_0)^2 \tag{10-19}$$

and is the summation over all virtual bond angles.

k_b and k_{θ} are force constants and, as indicated earlier, $b_0 = 5.14$ Å and $\theta_0 = 87.2°$. E_{bond} and E_{θ} are harmonic functions, which implies that they represent the vibrational energies of oscillating, springlike systems.

E_{Φ} represents the energy potential of near-neighbor atomic interactions, averaged over all dihedral angles consistent with a given Φ value. Recall that the simplified polypeptide model has only one backbone rotational degree of freedom per residue (Φ_i), while the detailed atomic model has 2, ϕ, and ψ. The position of the interaction center, R_{i+3}, is essentially determined by ψ_{i+1} and ϕ_{i+2} through their sum, $\psi_{i+1} + \phi_{i+2}$; therefore, Φ_i depends on ψ_{i+1} and ϕ_{i+2}. Variations in the difference $\psi_{i+1} - \phi_{i+2}$ correspond to rotations of the plane of the amide group between R_{i+1} and R_{i+2}; such rotations minimally affect the overall chain direction. Thus this energy potential or mean force is averaged over amide plane orientations as well as side-chain and solvent molecule configurations. It is assumed that the local reorientations of the amide planes are rapid compared to the variations in Φ, which reflect the slow changes in the overall shape of the chain.

E_{sol} represents interactions between residues separated by three or more virtual bonds. More specifically, E_{sol} represents the contribution to the energy potential of net attractive effects, from van der Waals interactions, excluded volume effects, and solvent interactions, and is given by

$$E_{\text{sol}} = \sum \sigma g(r_{ij}), \qquad i > j \tag{10-20}$$

where r_{ij} is the distance between interaction centers i and j, and $g(r)$ is a sigmoidal function which varies from $g(0) = 1$ to $g(r) = 0$ for r_{ij} 9 Å (Fig. 10-19). This summation is taken over centers separated by three or more bonds. It represents the free energy of attractive interactions that occur upon transfer of a residue from coil or aqueous surroundings, to helix or more hydrophobic surroundings.

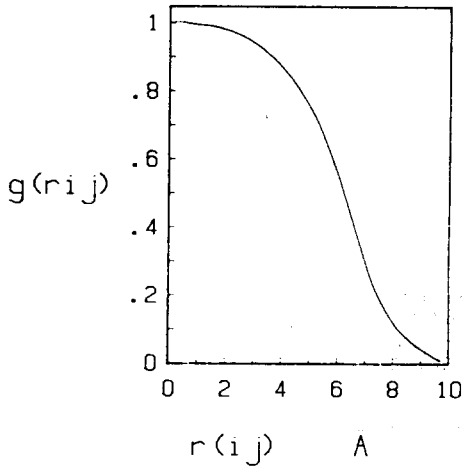


Figure 10-19 Dependence of $g(r_{ij})$ on r_{ij}.

E_{ev} represents the excluded-volume term. This free-energy contribution, which is a purely repulsive potential, opposes the change in attractive interactions introduced by E_{sol}. E_{sol} implies that once the coil residue relinquishes its freedom of mobility and commits itself to helix structure, it gains hydrophobic stabilization energy. However, its presence among the other residues eliminates a crevice of "free space" and depletes this latent stabilization source. So the E_{ev} term is needed to prevent the equation from overemphasizing the stabilization energy and thereby to prevent the excessive stability of "globules" (i.e., overlapped nonbonded residues).

E_α represents helix-stabilization energy and is given by

$$E_\alpha = \sum A_\Phi f(\Phi) \tag{10-21}$$

It is summed over all the virtual dihedral angles. The function $f(\Phi)$ is bell shaped; it reaches its maximum value of unity at $\Phi = 40°$; it diminishes to 0 if $\Phi > 55°$ or $\Phi < 25°$, the conditions when Φ is outside the α-helical range. The coefficient A_Φ for each dihedral angle depends on adjacent dihedral angles, as demonstrated in Fig. 10-20.

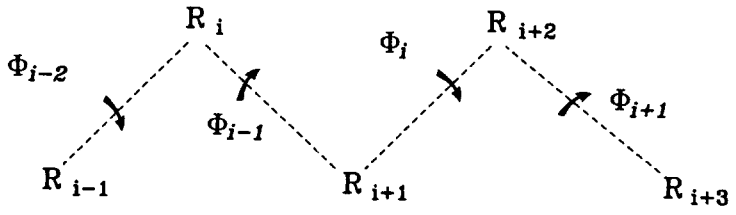


Figure 10-20 Schematic representation of virtual dihedral angles.

$A_{\Phi_i} = -6.0$ kcal/mol if Φ_{i-1}, Φ_{i-2}, and Φ_{i+1} are in the α-helix range. This value reflects the difficulty of nucleating a coil sequence in the interior of an α helix. $A_{\Phi_i} = -1.4$ kcal/mol if Φ_{i-1}, Φ_{i-2}, and Φ_1 are in the α-helix range, but Φ_{i+1} is not. Thus the α-helix stabilization term contributes up to -1.4 kcal/mol upon the addition of a residue to the helix. This term includes free-energy changes associated with the formation of a backbone hydrogen bond and the freezing of amide-plane rotational motions. Various calorimetric and pH titration studies of helix–coil transitions suggest that hydrogen-bond formation provides an estimated -1.5 kcal/mol of free-energy change, the larger of the contributions. The entropy reduction upon freezing the amide-plane orientation is less than 2 eu, which corresponds to a free-energy increase at 25°C of less than 0.6 kcal/mol. Therefore, the stabilization energy used appears to be of a reasonable magnitude. $A_{\Phi_i} = 0$ if I_i is not in the helix interior or at the helix-coil interface.

Stochastic Dynamics. The method of stochastic dynamics is used to study the dynamics of the helix–coil transition. First, the system is divided into two parts: One part serves as a heat bath for the other part whose dynamics are to be examined; this situation could be a protein in a solvent or a portion of protein within the surrounding protein. Second, this method assumes that the displacement of the dynamic part is analogous to molecular diffusion in a liquid or solid. Third, the energy potential is determined. This potential-energy function corresponds to the free energy of displacement of the elements being studied due to surrounding bath atoms; in our simplified model, these bath atoms are both solvent molecules and other chain residue atoms, while the dynamic element is a specific residue. The motion of the chain is largely determined by the time variation of its nonbonded interactions with the neighboring atoms. Finally, forces are derived from the energy potential which speed or slow the motion of the chain in a given direction. The chain motion is described by a set of Langevin equations of motion. The forces are plugged into these equations and the solutions define a new chain position. For example, the Langevin equation for a particle in one dimension is given by

$$m\frac{d^2x}{dt^2} = F(x) - f\frac{dx}{dt} + R(t) \tag{10-22}$$

where m is the mass of the particle, x the position of the particle, t the time, $m(d^2x/dt^2)$ the acceleration of the particle, and $F(x)$ represents the force on the particle derived from the energy potential. The remaining terms represent effective bath forces: $f(dx/dt)$ is the average frictional force caused by the motion of the particle relative to its surroundings, f the friction coefficient, and $R(t)$ represents any remaining randomly fluctuating forces not categorized by the mean force: electrostatic, dipole influences. For local denaturations such as helix–coil transition, the acceleration term is neglected in comparison to the others. This motion has no inertial character and imitates Brownian motions.

To study the dynamics of the helix–coil transition, the internal Brownian motion of a 15-residue chain in solvent surroundings is computer simulated. The system of

Brownian particles, placed in their initial configuration, is simulated by space trajectories. Each particle is moved along a trajectory calculated from the appropriate chain diffusion equation. This equation includes diffusional, hydrodynamic, and interparticle interactions as well as a weighted sum of all other forces acting on the Brownian particles. The trajectories are composed of successive displacements taken over a short time (Δt). These displacement values are derived from the appropriate equation of motion. To minimize the number of time steps required to calculate a trajectory, Δt is chosen as large as possible provided that the forces on the particles remain nearly constant during each step. Each successive configuration of the chain is selected from a probability distribution, which is the short-time solution of the chain diffusion equation, with the previous configuration as an initial condition. In this particular study, hydrodynamic interactions among the residues are neglected. The chain is initially in an all-helical configuration. The first 11 residues are held fixed in space, while the last four residues (R_{15}, R_{14}, R_{13}, and R_{12}) are allowed to move. The virtual dihedral angles Φ_{12}, Φ_{11}, Φ_{10}, and Φ_9 correspond to the mobile residues, respectively.

Figure 10-21 shows the components of the potential energy for residue 12 as a function of Φ_9. Results for the other mobile residues are nearly identical. Note the

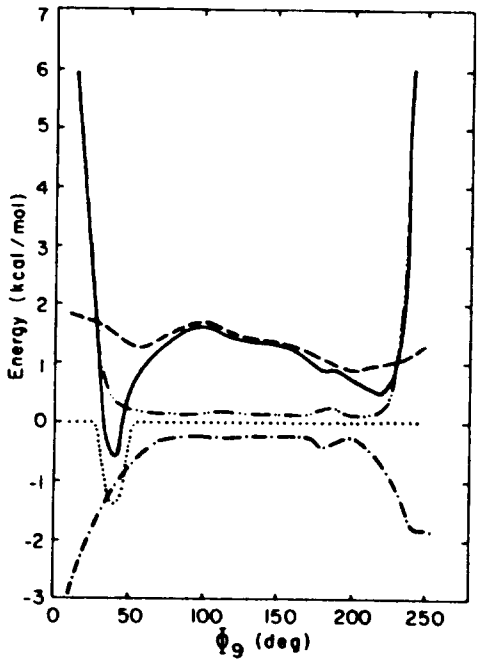


Figure 10-21 Potential of mean force and its components for a residue at the helix–coil interface for E(——), E_{Φ}(– – –), E_{sol}(–·–), E_{ev}(–···–), and E(· · ·). (Reprinted with permission from: J. A. McCammon and S. H. Northrup, *Biopolymers*, *19*, 2033–2045. Copyright 1980 John Wiley & Sons, Inc., New York.)

TABLE 10-9 Maximum relaxation times, $1/\tau_{max}$, and the rate constants, k_f, of the helix growth process at various temperatures[a] for poly(α-L-glutamic acid)

T(°C)	$1/\tau_{max}(10^5\ s^{-1})$	$k_f(10^7\ s^{-1})$
15	2.9 ± 0.2	2.4 ± 0.2
25	3.1 ± 0.2	2.6 ± 0.2
35	3.5 ± 0.2	2.9 ± 0.2
45	4.1 ± 0.3	3.4 ± 0.2

[a] In the estimations of k_f, a σ value of 3×10^{-3} was used.

free-energy barrier of approximately 1 kcal/mol per residue for the helix–coil transition. Other studies support the validity of this calculated value. Researchers used the electric field pulse (EFP) apparatus, which applies an electric field density and detects changes in electric conductivity, to follow the lifetime of the helix–coil transition. Generally, lifetime is defined as the time required for an atom or group of atoms in a high-energy state to return to a more stable, lower-energy state. In the case of the helix–coil transition, the lifetime defines the time required for a residue to leave its coil state and enter the helix state, leading to helix growth. This lifetime is expressed by

$$\frac{1}{\tau} = k_f[(s' - 1)^2 + 4\sigma] \tag{10-23}$$

where k is the forward rate constant for helix growth, s' the equilibrium constant for helix growth, and σ the nucleation parameter. The temperature dependence of these parameters is shown in Table 10-9.

The rate constant k_f is related to the activation enthalpy ΔH and entropy ΔS by

$$k_f = \frac{kT}{h} e^{-(\Delta H^\circ - T\Delta S^\circ)/RT} \tag{10-24}$$

The calculated values are

$$\Delta H = 1.5 \pm 0.4 \text{ kcal/mol}$$

$$\Delta S = -20 \pm 2 \text{ kcal/mol}$$

These values support the activation energy calculated from the energy potential.

Three independent simulations were performed, each for a total length of 12 ns. The time histories of the last three dihedral angles, Φ_{12}, Φ_{11}, and Φ_{10}, are shown in Fig. 10-22; the time history for Φ_9, which showed relatively little mobility, is omitted from the figure for clarity. Points are plotted at intervals of 120 ps.

Since the helix–coil transition occurs at the end of a single helical section, Φ_i is either in the α-helical range (when $\Phi_{i-1}, \Phi_{i-2}, \ldots, \Phi_1$, and Φ_{i+1} are in the α-helical range) or at the helix–coil interface (when $\Phi_{i-1}, \Phi_{i-2}, \ldots, \Phi_1$ are in the α-helix range

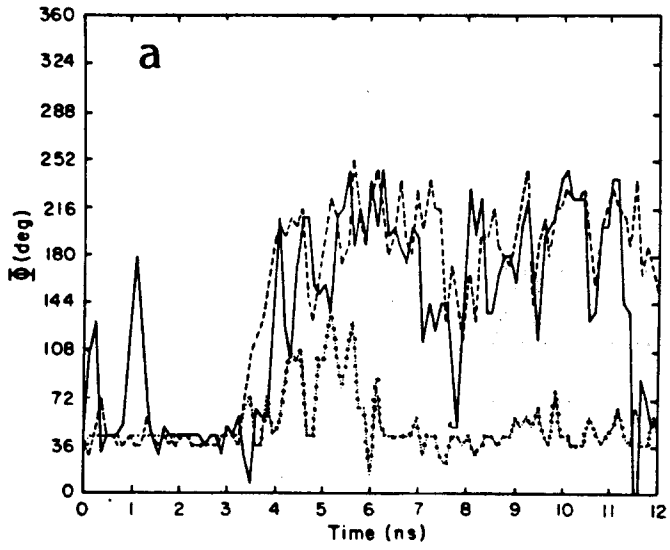


Figure 10-22A

but Φ_{i+1} is not). Nucleation at a second site along the helix chain would be rare on the time scale of the given simulations. Significant helix unwinding occurs in each simulation; thus the mobility of the residues markedly increases at the end of the chain. When two or more residues leave the helix they do so sequentially, although this is not always apparent from Fig. 10-22 due to the 120-ps intervals. After un-

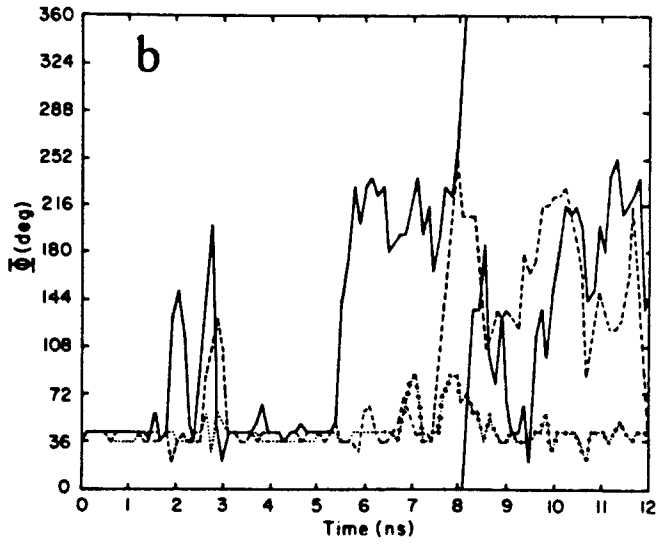


Figure 10-22B

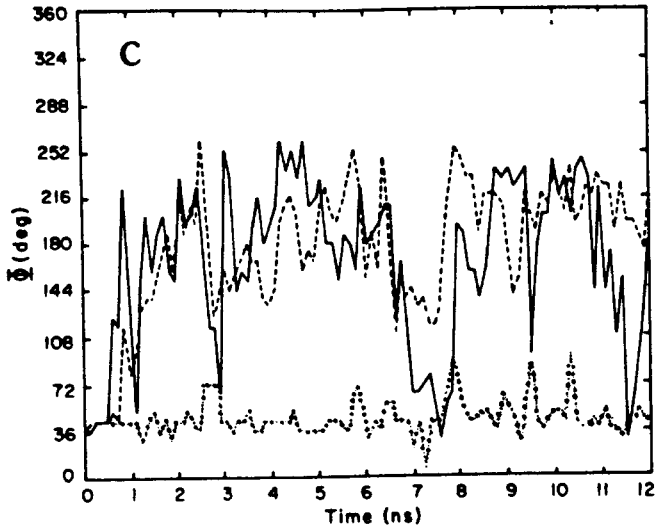


Figure 10-22C

Figure 10-22 (A) Dihedral angle histories during the first helix-unwinding simulation: Φ_{12}(——), Φ_{11}(– – –), and Φ_{10}(· · ·). (B) Dihedral angle histories during the second helix-unwinding simulation: Φ_{12}(——), Φ_{11}(– – –), and Φ_{10}(· · ·). (C) Dihedral angle histories during the third helix-unwinding simulation: Φ_{12}(——), Φ_{11}(– – –), and Φ_{10}(· · ·). (Reprinted with permission from: J. A. McCammon and S. H. Northrup, *Biopolymers*, *19*, 2033–2045. Copyright 1980 John Wiley & Sons, Inc., New York.)

winding, Φ_{12} and Φ_{11} tend to remain out of the helix range; however, Φ_{12} drifts to smaller values while Φ_{11} returns to the helix near the end of the second simulation. In conclusion, the final study results suggest:

1. R_{15} unwinds from the helix readily and exhibits large fluctuations after R_{14} has also unwound.
2. R_{14} unwinds less readily than R_{15} and exhibits smaller fluctuations after unwinding.
3. R_{13} exhibits occasional transient departures from the helix, even after the last two residues are unwound. R_{13} seems to be in limbo between the two states.

Since the helix–coil transition free-energy barriers are similar for all the residues, their reduced mobility in the chain interior indicates that these residues have smaller effective diffusion constants. Unwinding of an interior residue requires simultaneous displacement of residues in the coil, so that larger frictional forces are involved. In addition, the coil region does not move as a rigid body, but its torsional motions are correlated to minimize dissipative effects. For example, examine the large displacements of Φ_{10} during the first simulation. With Φ_{11} and Φ_{12} in extended

formations, the positive correlation of $\Delta\Phi_{10}$ and $\Delta\Phi_{11}$ tends to minimize the displacements of residues 14 and 15 when residue 13 moves, while the negative correlation of $\Delta\Phi_{10}$ and $\Delta\Phi_{12}$ also tends to minimize the displacement of residue 15 when residue 13 moves. Thus the movements seem to complement each other and appear controlled even though they are governed by diffusional force.

EVOLUTION OF TERTIARY STRUCTURE

The conservation of amino acid residues involved in a particular catalytic function is a well-documented phenomenon (e.g., the serine proteases). Similarly, homologies between sequences of similar proteins from various sources have been used to establish evolutionary relationships, as seen in the cytochrome *c* system. Comparison of sequences of superoxide dismutase from bacterial and animal sources has also given further insight into the evolution of mitochondria in animal systems. Such homology in the primary structure of proteins is not entirely unexpected.

However, where no particular catalytic function needs to be conserved, but only a given secondary structure such as a region of helix or pleated sheet, many more types of amino acid replacement are possible, and the conformation of a particular region of polypeptide chain may be conserved even though the sequences are not. For instance, as is indicated in Tables 10-10 and 10-11, a variety of amino acids can be interchanged in a helical region without disrupting the helical structure. Similarly, if a hydrophobic region on a protein needs to be conserved for, say, interaction with another macromolecule or a hydrophobic ligand, a number of amino acid residues can provide essentially similar hydrophobic properties. Yet in terms of specific catalytic functions, the possibilities of replacement are far fewer—threonine and serine may be interchanged in some instances, but the properties of histidine are unique.

In this section we discuss the structural homology within the class of proteins that bind adenine nucleotides—in particular the dehydrogenases; and we examine this structural homology (or in some cases the lack of particular elements of structural homology) in terms of the known enzymology of the systems. Most of the work discussed is x-ray crystallographic, but some sequence work is also included.

TABLE 10-10 Effects of residues on helix structures

Residues that stabilize helix structures: Alanine, leucine, phenylalanine, tyrosine, tryptophan, cysteine, methionine, histidine, asparagine, glutamine, valine
Residues that destabilize helix structures: Serine, isoleucine, threonine, glutamate, aspartate, lysine, arginine, glycine
Residues that break helix structures: Proline, hydroxyproline

TABLE 10-11 Roles of residues in protein and enzyme structure and function

Amino acid residue	Possible roles in protein and enzyme structure and function
Arginyl	Hydrophilic; electrostatic interactions
Lysyl	Hydrophilic; electrostatic interactions; attachment of prosthetic group or cofactor in amide bond; interacting to form a Schiff's base; ligand to metal ion
Histidyl	Hydrophilic or hydrophobic (depending on ionization); electrostatic interactions; proton transfer; ligand to metal ion; hydrogen bonding; acceptor in transfer reactions
Glutamyl, aspartyl	Hydrophilic; electrostatic interactions; proton transfer; ligand to metal ion; covalent linking in ester or amide through γ-carboxyl
Glutaminyl	Hydrophilic; hydrogen bonding
Asparaginyl	Hydrophilic; hydrogen bonding
Seryl	Hydrogen bonding; nucleophile; covalent linkage of OH in esters
Threonyl	Hydrogen bonding; nucleophile; covalent linkage of OH in esters; hydrophobic
Glycyl	Lock of side chain permits flexibility of folding and cross-hydrogen bonding
Alanyl, valyl lecucyl, isoleucyl, phenylalanyl	Hydrophobic interactions; determinants of steric and conformation specificity, e.g., numerous alanyl residues favor α-helix formation while numerous valyl or isoleucyl in sequence tend to inhibit formation of such features
Tyrosyl	Hydrophobic; hydrogen bonding; proton transfer; electrostatic interactions; at high pH; ligand to metal ions
Tryptophanyl	Hydrophobic; hydrogen bonding
Cysteinyl	Nucleophile; acyl acceptor; hydrogen bonding; ligand to metal ions
Cystyl	Cross-linking through disulfide bonds
Methionyl	Hydrophobic; hydrogen bonding to S (?); ligand to metal ions
Prolyl	Interrupts α and β structures allowing irregular conformation; hydrophobic

Ways of Assessing Structural Homology

Structural homology between two proteins (or peptides) is assessed by estimating the distances between equivalent backbone atoms of the molecule from their atomic coordinates. We will not discuss in any detail the mathematics of this procedure, but just qualitatively describe a few relevant aspects. Obviously, when two protein structures with closely homologous sequences or the same protein in two conformations are being compared, it is a fairly simple matter to decide which are the equivalent backbone atoms.

When there is no prior knowledge of atomic equivalence it is necessary to establish any structural homology purely on the basis of the tertiary structures. In practice this is usually done by taking a small region of, say, three to four peptides from one of the molecules and comparing it systematically with similar-sized blocks from the second molecule. The mean distance between supposedly corresponding atoms and the scatter from the mean is calculated. If a given group of atoms from molecule II has a scatter of less than, for example, 1 A from molecule I, the two blocks can be equivalenced. If the scatter is too high, the block of atoms for comparison is slid one residue down the chain on molecule I, and the procedure is repeated.

In the cases of the dehydrogenases that we will talk about, lactate dehydrogenase (LDH), malate dehydrogenase (MDH), horse liver alcohol dehydrogenase (ADH), and muscle glyceraldehyde-3-phosphate dehydrogenase (GA-3P-DH), the C atoms that were used in the previous type of procedure were selected from the common hydrogen-bonded scheme within the parallel pleated-sheet region of each of these enzymes. It has been found that when the hydrogen bonds of two different parallel pleated-sheet regions are aligned in an equivalent manner with the polarity of the strands being the same, the corresponding βC atoms are on the same side of the sheet. As a result, the alignment of residues within parallel pleated sheets is relatively simple once the positions of the βC atoms and the hydrogen-bonding scheme have been determined.

Dehydrogenases

Although there are very marked sequence correlations between dehydrogenases of similar function from different species—for instance, yeast and muscle GA-3P-DH, yeast and horse liver ADH, and bovine and chicken GDH—there are no really significant sequence similarities between dehydrogenases with different functions—for example, L-ADH and GDH. However, the crystallographic structures of this class of enzymes have revealed a very interesting pattern of structure–function correlations.

The three-dimensional structure of dogfish LDH was the first of the dehydrogenase structures to be obtained at high enough resolution to follow the chain in its entirety. In the original structure 331 amino acids were used, although it is now known that only 329 residues are present. The outstanding feature of the LDH structure was the region of the molecule that bound the pyridine nucleotide coenzyme, which consisted of residues 24 to 162 and appeared to be arranged in a series of β-pleated sheets connected by either helical regions or loops of polypeptide chain. In LDH there are six parallel strands of pleated sheet (labeled βA $\rightarrow$ βF in Fig. 10-23, which is a schematic representation of the coenzyme binding site). The βA strand consists of five residues, 24 to 28; the B strand, four residues, 49 to 52; the C strand, five residues, 92 to 96; the E strand, four residues, 135 to 138; and the F strand, also four residues, 159 to 162.

Subsequent to the structure of LDH, those of liver ADH (at 2.9 Å resolution) and lobster GA-3P-DH (at 3.0 Å resolution), as well as the structure of MDH at 3.0-Å resolution, were reported. In each the coenzyme binding domain was found to consist of a series of β-pleated sheets, as was the case with LDH.

Table 10-12 gives the sequence location and the number of residues in each of those six pleated-sheet regions in these three enzymes. Although the coenzyme binding domain is in the first half of the molecule in LDH and GA-3P-DH, it is in the second half in L-ADH. Also shown are the root-mean-square deviations between equivalent atoms, which gives an estimate of the structural homology between the compared regions. All three possible comparisons are shown in Table 10-12: ADH with LDH, ADH with GA-3P-DH, and LDH with GA-3P-DH. The distances between the 26 equivalent αC atoms used in the pleated-sheet regions are no greater than those found

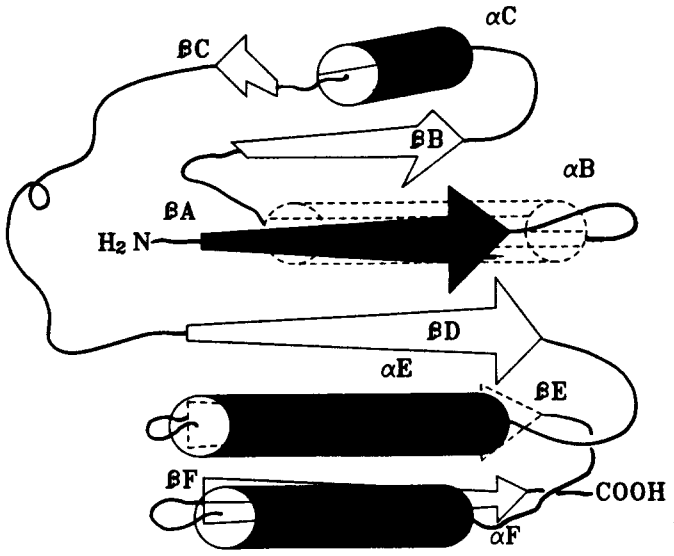


Figure 10-23 Schematic representation of the structure of the coenzyme binding domain in dehydrogenases.

in a comparison of identical structures such as subtilisin BPN and subtilisin novo. In fact, the errors in the measurements of atomic coordinates on which these distance estimates are based are about the same magnitude. The angle of twist of the pleated sheet (about 100°) is constant for each of the dehydrogenases, whereas pleated sheets have been found to have twists varying from almost 0° for concanavalin A to about 220° for carbonic anhydrase.

Despite this marked structural homology there is very little, if any, significant homology in the sequences of these three proteins in this nucleotide binding region. Over the entire domain, comprising approximately 120 residues, there are only five

TABLE 10-12 Correspondence between equivalent αC atoms used in the alignment of the nucleotide binding regions of different dehydrogenases

Structural element	Residue numbers			Root-mean-square deviation between equivalent atoms (Å)		
	ADH 1	LDH 2	GA-3P-DH 3	1–2	1–3	2–3
βA	195–199 (5)	24–28 (5)	3–7 (5)	0.97	0.94	1.05
βB	219–222 (4)	49–52 (4)	28–34 (4)	1.33	1.54	0.89
βC	238–241 (4)	78–81 (4)	71–74 (4)	1.10	2.61	2.04
βD	263–267 (5)	92–96 (5)	90–94 (5)	1.12	1.32	1.30
βE	289–292 (4)	135–138 (4)	116–119 (4)	0.87	1.85	2.04
βF	313–316 (4)	159–162 (4)	142–145 (4)	1.34	1.17	1.21

residues common to all three sequences. In particular, three of these residues, Gly-28, Gly-33, and Asp-53 (in the LDH sequence), lie in the central region of the nucleotide binding domain and have particular functions in coenzyme binding. The outer regions of the domain, in particular α-helix IF and the loop between βC and βD, have significantly different structures in the three dehydrogenases; it appears that for correct coenzyme binding there are no stringent requirements for the outer regions of the domain.

These three invariant residues just mentioned each have a particular functional significance. If residue 28 (in LDH) has a β-carbon atom, this position would overlap with the C_2 position in the ribose of the coenzyme, preventing the adenosine part of the coenzyme from binding in the correct orientation. There appears to be an absolute requirement for a glycine residue in this position.

One of the carboxyl oxygens of Asp-53 forms a hydrogen bond to the oxygen of the adenine ribose, which is apparently important in coenzyme binding. It might be noted that in NADP the extra phosphate group is attached to the oxygen and the hydrogen bond cannot then form. This interaction may well be important in determining the specificity for NAD of these dehydrogenases.

The remaining invariant glycine, Gly-33, although not involved in the coenzyme binding area, does seem to be important for structural reasons. A β-carbon atom in this position would point from helix αB toward the pleated-sheet and would thus interfere with the main chain of the sheet.

So far we have discussed the structural similarities between various dehydrogenases. There are, of course, differences and we can attempt to relate these differences to some aspects of their enzymology.

In LDH (and MDH), a 20-residue chain connects the D and E strands of the β-pleated sheet. This loop undergoes a marked change when the coenzyme, or various adenine-containing derivatives, are bound. In L-ADH only three residues are used to join these strands. Similarly, in GA-3P-DH this loop region is absent. It has been suggested that in LDH (or MDH) the main trigger for the conformational change is an interaction between the adenosine phosphate and an arginine residue in this loop region. Since in L-ADH and GA-3P-DH this region is missing, it was suggested that conformational changes induced by coenzyme binding in these proteins may involve the nicotinamide moiety. Very recent studies with rabbit muscle GA-3P-DH have demonstrated that the conformational change in this enzyme associated with the negative cooperativity in coenzyme binding does indeed require the nicotinamide moiety. Unlike the conformational changes observed in LDH or MDH, ADPR does *not* induce the conformational change. Since cooperative phenomena have been reported in GA-3P-DH as well as ADH, it may be that the *loss* of this loop region between the βD and the βE regions of the pleated sheet is related to the evolution of cooperative phenomena in the dehydrogenases.

Related Proteins

So far we have discussed the dehydrogenases whose crystal structures have been determined. There are, however, a number of other proteins that bind adenine nucleotides whose crystal structures have also been determined.

The crystal structure of flavodoxin, a protein that binds flavin mononucleotide, indicates that the nucleotide binding site is similar in structure to that used to bind a nucleotide in LDH or the other dehydrogenases. This structural homology can be seen in diagonal plots (Fig. 10-24) representing distances between α-carbon atoms as a function of residue number for flavodoxin and LDH. The contour plots represent given αC–αC distances and uniquely reflect the structure of the protein backbone. Clearly, there is quite a marked similarity in the conformation of the FMN binding site in flavodoxin and either half of the dinucleotide binding domain of LDH. Also, from these diagonal plots it can be seen that the remainder of the LDH molecule bears *no* structural resemblance to the coenzyme binding domain.

Recently, similar nucleotide binding structures have been recognized in phosphoglycerate kinase and adenylate kinase. A structure very similar to the nucleotide binding domain of flavodoxin has also recently been identified in rhodanese, a protein whose biological function has not yet been completely established, but which is known to bind FAD and FMN as well as NAD.

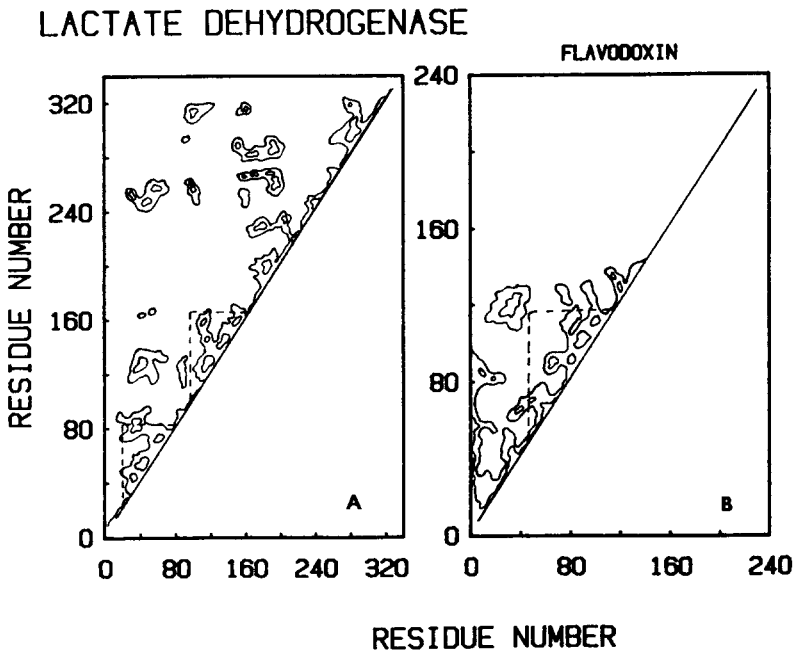


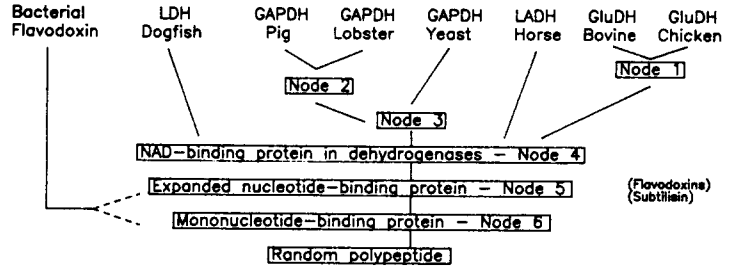
Figure 10-24 (A) Diagonal plot for LDH representing distances between C atoms. Only major contours are shown. Different parts of the structure have been identified along the diagonal. The first two structural domains between which there is less contact are marked in a way that emphasizes the comparison between them. (B) Diagonal plot for flavin. The marked domain in the central part of the polypeptide chain has a structural and functional similarity to each of the first two domains of LDH.

It also appears possible that a structure similar to the nucleotide binding region of the dehydrogenases is present to a lesser extent in some other proteins. For instance, there is a region in subtilisin, residues 121 to 181, that in some ways resembles the mononucleotide binding domain of LDH, residues 22 to 81, which contains βA, βB, and βC regions of the coenzyme domain. Note that in this region the hydrophobic pocket which binds the aromatic side chain of the subtilisin substrate, between Gly-127 and Gly-154, is in a similar position to the hydrophobic pocket that binds adenine in LDH, containing Val-27 and Val-54.

Evolutionary Significance

The retention of an essentially identical fold, despite the wide variety of sequences, demonstrates that there is a stability of tertiary structure even in cases where there is no apparent conservation of primary structure. That these observations are a chance occurrence is unlikely, as such a large number of amino acid residues are involved. Also, it seems unlikely that these large structural similarities are the result of convergent evolution from different precursor molecules, especially as alternative structures of nucleotide binding proteins are known to exist: for instance, ribonuclease or staphylococcal nuclease. Thus it seems that this class of nucleotide binding proteins represents an example of divergent evolution.

If this is the case, a relative estimate of the time elapsed since divergence from a common ancestor can be made using as a measure of evolutionary distance the



MBC/C over all comparisons	Largest observed MBC/C	Smallest observed MBC/C	Time scale yrs ago x 10^9
0.03	0.03	0.03	0.3
0.43	0.43	0.43	0.6
0.59	0.59	0.59	1.2
1.14	1.21	1.00	1.5 – 3.2
1.26 1.14	1.43 1.30	1.08 1.00	3.2 – 4.5
1.33	1.53	1.19	3.2 – 4.5
1.51			4.0 – 4.5

Figure 10-25 Evolutionary tree with observed minimum base changes per codon (MBC/C) and possible time scale derived primarily from fossil data.

minimum base change per codon. In the cases of the nucleotide binding proteins where the homology is based on a structural similarity in a given domain, Rossmann and co-workers have obtained an evolutionary tree for that region of each of the considered proteins that is involved in nucleotide binding. This is given in Fig. 10-25.

One point which might be made is that if one considers pig and yeast glyceraldehyde-3-phosphate dehydrogenase, the observed minimum base change per codon in the nucleotide binding domain is 0.59, while for the remainder of the molecules the value is only 0.35. Presumably the differing functions of the two domains within a single polypeptide chain imposes different rates of evolution.

One might also speculate that the dehydrogenases and related nucleotide binding proteins evolved by a process of gene fusion; that is, there is one gene for the nucleotide binding domain common to all dehydrogenases and related proteins that has fused with one gene for a substrate binding domain different for each of these proteins. This is a particularly attractive hypothesis when one remembers that in the case of phosphoglycerate kinase, the substrate binding and nucleotide binding domains are clearly separated in space as well as in sequence.

ACQUISITION OF TERTIARY STRUCTURE

The discussion thus far has been aimed at giving a picture of the tertiary structure of a protein as being in a dynamic state. From this we can rationalize the differing mechanisms of protein folding examined in Chap. 10. With small proteins the dynamic nature of the final tertiary structure permits the folding process to be essentially pathway independent and the "final" structure can be reached rapidly. With proteins, where there is evidence for pathway-dependent folding, the dynamic nature of the process, as well as the final conformation, offers a mechanism by which a substrate, for example, can stabilize an intermediate and "direct" the folding process.

Protein unfolding–folding is often considered in terms of a two-stage process for small proteins that can be renatured after denaturation. This implies that there are no stable intermediates in the process. As more sophisticated approaches for detecting such intermediates become available it is apparent that even in small proteins such as bovine pancreatic trypsin inhibitor (BPTI) or ribonuclease (RNase), various intermediates can be identified. In BPTI, the fully reduced protein appears completely unfolded: When refolding occurs it is found that of 15 possible species containing only a single disulfide, only two are significantly populated, and they are in rapid equilibrium with one another. The major species involves the Cys-30–Cys-51 disulfide bond found in the native protein. Disulfide bonds also play a significant role in stabilizing some protein structures. In RNase, the correct disulfide bonds are formed when the denatured protein is renatured. However, when four residues are removed from the C-terminal end, the correct disulfide bonds are not reformed.

In many proteins when renaturation is attempted it is found that regain of the native structure is a multiphasic process. In cytochrome c, refolding after guanidine

hydrochloride–induced denaturation can be followed by monitoring tryptophan fluorescence (which in the native protein is quenched by heme interaction), absorbance at 287 nm (from the five tyrosines in the molecule), or by absorbance at 695 nm (which monitors the heme ligation in the native protein). Over short time scales these three monitors of conformation reflect similar processes. However, at longer time scales more processes are observed with the absorbance measurements at 287 nm than with the tryptophan fluorescence, suggesting that slower processes are occurring, affecting tyrosine environments in the protein. Proline peptide bonds can undergo slow cis-trans isomerizations that appear to be important for these slower processes in cytochrome c refolding. This cis-trans equilibrium constant is near 1 and is affected by the local environment in terms of the primary sequence. Proline isomerization has a characteristic activation energy around 20 kcal/mol, and its pH dependence is governed by the presence of ionizable groups near the peptide bond. Such isomerizations may well be responsible for the slow "refinements" of the folded structure of a protein observed in many instances.

In this chapter and Chap. 9 we examined aspects of secondary and tertiary structure, emphasizing the dynamic nature of protein structure together with the importance of long-range interactions in governing final form. These long-range interactions play a pivotal role in the folding pathway as well as influencing local secondary structure. In addition, as will be discussed in Chaps. 11 and 12, they are intimately involved in the assembly of quaternary structure in oligomeric proteins and in the mechanism of conformational changes in allosteric proteins.