5
Conformational analysis and forces that determine protein structure

5-1 BASIC PROBLEMS OF PROTEIN STRUCTURE

One of the most fascinating and challenging problems of biochemistry is that of the physical basis for the markedly organized and specific structures of proteins. This problem takes on particular significance because the biological activity of these macromolecules is sensitive to variations in three-dimensional conformation. Because biological polymers generally can be unfolded (for example, by heat or urea treatment) and later refolded to the original state, it is reasonable to assume that, in the main, the conformations adopted by various biopolymers are the thermodynamically preferred ones. Some support for this assumption was obtained in a series of experiments by C. B. Anfinsen and colleagues (1961). These investigators demonstrated that, even when all four disulfide bonds of ribonuclease are broken and the protein is completely denatured by urea, the correct native structure is readopted upon removal of urea and reoxidation of the disulfide linkages.[§]

In this chapter, we discuss the various factors that determine protein conformation. We first consider inherent geometric factors such as the essentially fixed bond lengths and bond angles of the chain. Next we examine the restrictions on available conformations imposed by steric interactions. We also take up the refinement of this analysis with the use of more realistic potential functions. Finally, we look at other factors that are of extreme importance in determining protein structures.

[§] In spite of the results of this experiment, however, one must consider that native structures may represent local free energy minima, and that kinetic barriers may prevent attainment of conformations with even lower free energies.

These include the well-characterized hydrogen bond, as well as the less-understood "hydrophobic" interaction. If all of the above-mentioned factors were well enough understood, and if sufficient mathematical tools were available, then it should be possible to predict, for example, the three-dimensional structure of a protein from its amino acid sequence. This goal has not yet been achieved but, in view of recent progress, it is not completely unrealistic. (Further aspects of protein folding are discussed in Chapter 21.)

5-2 POLYPEPTIDE CHAIN GEOMETRIES

Figure 5-1 displays the α-L-polypeptide chain in its all-*trans* (planar zig-zag) form. Residues are indexed serially from 1 to *n*, starting from the N-terminus. Note that

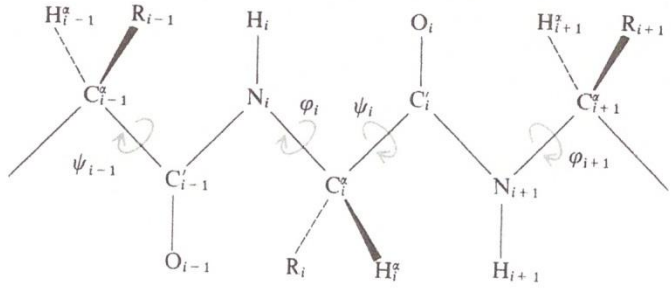
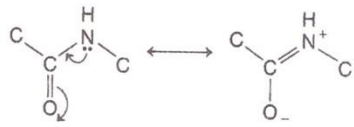


Figure 5-1

The α-L-polypeptide chain in the all-*trans* form.

similar atoms within the same residue are also distinguished; for example, the *i*th alpha-carbon is denoted C^{α}_i while the *i*th carbonyl carbon is C'_i. The individual amino acid units differ according to their R-group side chains. The various bond lengths and bond angles that characterize the chain backbone may be regarded as essentially fixed. Table 5-1 gives the generally accepted values for these geometric parameters.

An important aspect of polypeptide structure is that the amide bond usually occurs in the planar *trans* conformation. This can be rationalized to be a result of a resonance form that imparts double-bond character to the amide group.



The partial double-bond character shortens the amide bond by about 0.1 Å. As a result of the planar *trans* form of the amide group, the distance between successive alpha-carbon atoms is fixed at 3.8 Å. This means that the conformation of the chain as a whole is determined by the rotations about the N–C^{α} and C^{α}–C' bonds.

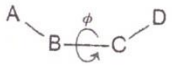
Table 5-1
Polypeptide chain geometries

Bond	Bond length (Å)	Bonds	Bond angle
$C^{\alpha}—C'$	1.53	$C^{\alpha}—C'—N$	113°
$C'—N$	1.32	$C'—N—C^{\alpha}$	123°
$N—C^{\alpha}$	1.47	$N—C^{\alpha}—C'$	110°
$C'=O$	1.24		
$N—H$	1.00		
$C^{\alpha}—C^{\beta}$	1.54		
$C^{\alpha}—H^{\alpha}$	1.07		

SOURCE: Data from V. Sasisekharan in *Collagen*, ed. N. Ramanathan (New York: Interscience, 1962), p. 39.

Internal rotation angles

In order to specify the three-dimensional conformation of a macromolecule, it is necessary to specify the internal rotation angles, or torsion angles. We follow the scheme of the IUPAC–IUB Commission on Biochemical Nomenclature (1970). Consider a four-atom system A–B–C–D:



Given fixed bond lengths and bond angles, the torsion angle ϕ describes the relationship between the C–D and A–B bonds. This angle is defined as the angle between the plane specified by bonds B–C and C–D and the plane specified by bonds A–B and B–C.

The torsion angle is, by convention, assigned a unique numerical value as follows. Figure 5-2a,b shows Newman and perspective projections of our four-atom system. A positive value of ϕ is assigned to the *right-hand rotation* necessary to bring the front atom (A or D) into an eclipsed position with the rear atom (D or A). For example, in the first and third illustrations of Figure 5-2a, right-hand rotation (viewed B to C) about B–C through the angle ϕ results in atom A eclipsing atom D. Conversely, in the second and fourth illustrations (viewed C to B), D is the front atom, and a right-hand rotation about C–B through the angle ϕ results in D eclipsing A. Thus, whether we look from B to C or from C to B, the same positive value for ϕ is obtained.

Alternatively, the same value of ϕ is obtained if we look from B to C and consider ϕ as the right-hand rotation of bond C–D (about B–C) that is required to bring atom D (the rear atom) out of eclipse by A (the front atom) to the designated position. We obviously can perform the equivalent operation by looking from C to B. Negative values of the torsion angle are *left-hand rotations* (Fig. 5-2b).

The value $\phi = 0^{\circ}$ corresponds to the eclipsed conformation in which the bonds

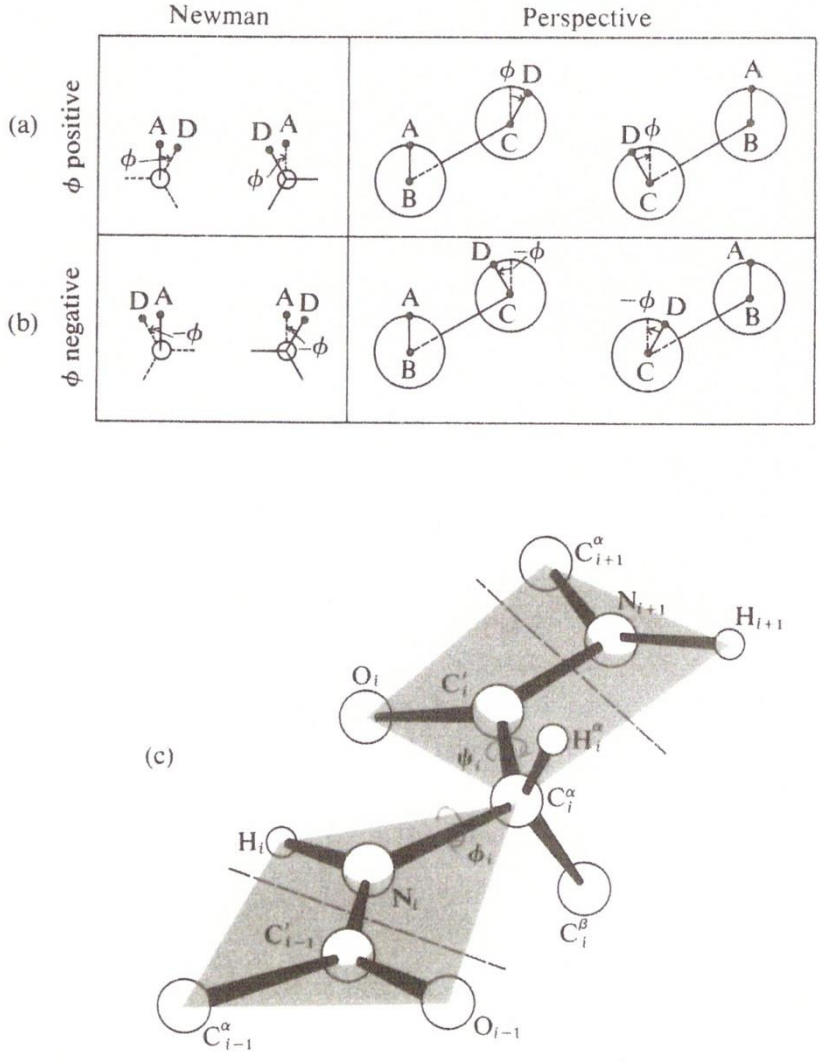
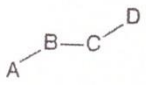


Figure 5-2

Torsion angles. (**a**) Newman and perspective diagrams illustrating positive values of the torsion angle ϕ for the four-atom (ABCD) system. (**b**) Similar diagrams illustrating negative values of ϕ. (**c**) Perspective drawing of the α-L-polypeptide chain, illustrating the torsion angles ϕ and ψ. The chain is drawn in the planar zig-zag (all-*trans*) form, for which $\phi = \psi = 180°$. The dashed lines delineate the extent of the *i*th amino acid residue. [After IUPAC–IUB Commission on Biochemical Nomenclature, *Biochemistry* 9:3471 (1970).]

A–B and C–D are *cis*. When these bonds are *trans*, as in



the torsion angle $\phi = 180°$ (or, equivalently, $\phi = -180°$). It is convenient to specify the values of ϕ as falling in the range $-180° \leqslant \phi \leqslant +180°$. In this way, the value of ϕ makes clear the relationship between enantiomeric conformations.

For the polypeptide backbone, with its fixed *trans* conformation for the amide bond, the important torsion angles are those for the N_i–C_i^α and the C_i^α–C_i' bonds. These angles are designated ϕ_i and ψ_i, respectively. For an α-L-polypeptide in the planar *trans* conformation (Fig. 5-2c), $\phi_i = \psi_i = 180°$ for all i.

(It should be noted that much of the earlier literature in the field is based on the convention that the reference state, $\phi_i = 0°$ and $\psi_i = 0°$, be assigned to the planar *trans* arrangement. Therefore, in reading this literature, 180° must be added to—or subtracted from—the numerical values of ϕ and of ψ in order for these values to correspond to the established convention used here. See Box 18-1 for further discussion, particularly with respect to conventions used in treating polymer chain statistics.)

The conformation of the polypeptide chain backbone as a whole is specified by enumerating the value of each ϕ_i and ψ_i. A helical form is generated when every $\phi_i = \phi_j$ and every $\psi_i = \psi_j$. For the right-handed alpha helix, $\phi = -57°$ and $\psi = -47°$. The ϕ, ψ coordinates of a variety of ordered forms of polypeptides are given in Table 5-2. These are all periodic, regular forms in which each rotation angle is the same for each residue. (Some of these structures are illustrated in Chapter 2.) Note that the pleated-sheet conformations are closest to the planar zig-zag structure shown in Figure 5-1.

Table 5-2

Approximate torsion angles for some regular structures

Structure	ϕ	ψ
Right-handed α helix [α-poly(L-alanine)]	−57°	−47°
Left-handed α helix	+57°	+47°
Parallel-chain pleated sheet	−119°	+113°
Antiparallel-chain pleated sheet [β-poly(L-alanine)]	−139°	+135°
Polyglycine II	−80°	+150°
Collagen (triple helix)	−51°, −76°, −45°	+153°, +127°, +148°
Poly(L-proline) I	−83°	+158°
Poly(L-proline) II	−78°	+149°

NOTE: For a fully extended chain, $\phi = \psi = +180°$. The imide bond is *cis* in polyproline I; it is *trans* in polyproline II.
SOURCE: After IUPAC–IUB Commission on Biochemical Nomenclature, *Biochemistry* 9:3471 (1970) [published simultaneously in *J. Biol. Chem.* 245:6489 (1970) and *J. Mol. Biol.* 52:1 (1970)].

In contrast to the periodic forms listed in Table 5-2, globular proteins (or random coils, of course) show ϕ, ψ coordinates varying considerably from residue to residue.

Ramachandran or steric contour diagrams

In theory, a chain molecule can adopt an essentially infinite variety of backbone conformations, each corresponding to a unique set of values for the various backbone rotation angles. However, many of these hypothetical conformations can be excluded from consideration on the basis of unfavorable steric overlaps. G. N. Ramachandran and colleagues (1963) were among the first to investigate this problem.

In studying steric interactions, we look first at overlaps between atoms whose distance of separation depends on a single rotation angle. For example, the distances C_i^α to O_i, C_i^α to N_{i+1}, or C_i^α to C_{i+1}^α depend solely upon the value of ψ_i. By determining the steric overlaps that occur as we vary ψ_i, we can exclude certain values of ψ_i. However, by examination of atomic scale models, it is immediately apparent that additional values of ψ_i can be eliminated because of overlaps between atoms that are separated by ϕ_i and ψ_i—for example, O_i and O_{i-1}. Because of these overlaps, which depend simultaneously upon the values of both ϕ_i and ψ_i, rotations about N_i–C_i^α and C_i^α–C_i' are said to be interdependent.

One can readily determine those values of the pair ϕ_i, ψ_i that are sterically allowed, by investigating contacts between all atoms whose distance of separation depends solely on these two rotation angles. The allowed domain of ϕ_i and ψ_i is *not* further reduced by interactions that are also simultaneously dependent on the values of the rotation angles in adjacent units, such as ϕ_{i-1} and ψ_{i-1}. For example, the distance of separation of O_i and R_{i-1} depends on ϕ_i, ψ_i, and ψ_{i-1}, but no contact between O_i and R_{i-1} is possible when ϕ_i and ψ_i are confined to their previously determined allowed domains. Thus, with amide groups fixed in the *trans* form, rotations within a given residue are *interdependent*, but they are sterically independent of rotations within neighboring residues.

At this point it is important to distinguish between short-range and long-range interactions. Those occurring between atoms or groups that are neighboring or close to each other in the sequence are termed short-range. This is the type discussed in the preceding paragraph. Long-range interactions are those between groups that are far apart in the sequence, and they occur only when the chain folding brings such groups into close proximity. This section is concerned largely with short-range effects, which are capable of giving considerable insight into protein conformation.

How can we best represent or diagram those values of ϕ_i and ψ_i that are sterically allowed for a given amino acid residue? This is accomplished by a map, in which ψ is the ordinate and ϕ is the abscissa. Lines then are drawn to enscribe the ϕ, ψ coordinates in which no unfavorable steric contacts occur.

Figure 5-3 is a steric diagram for a glycyl residue (R = H) situated within a polypeptide chain; Figure 5-4 is a similar diagram for alanine ($R = H_3$). The dark zones show normal contact radii for the various atom pairs, whereas the light zones

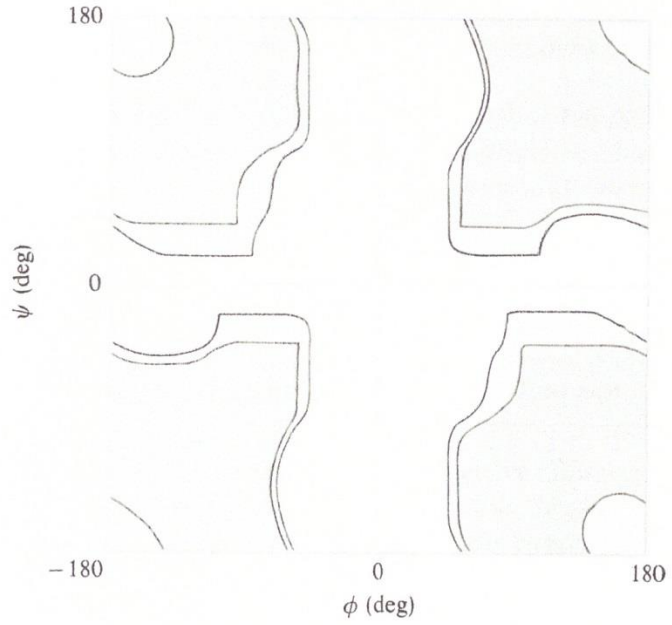


Figure 5-3

Steric contour diagram for a glycyl residue in a polypeptide chain. Dark zones show "normal" and light zones show "outer-limit" contours. [After G. N. Ramachandran et al., *Biophys. J.* 6:849 (1966).]

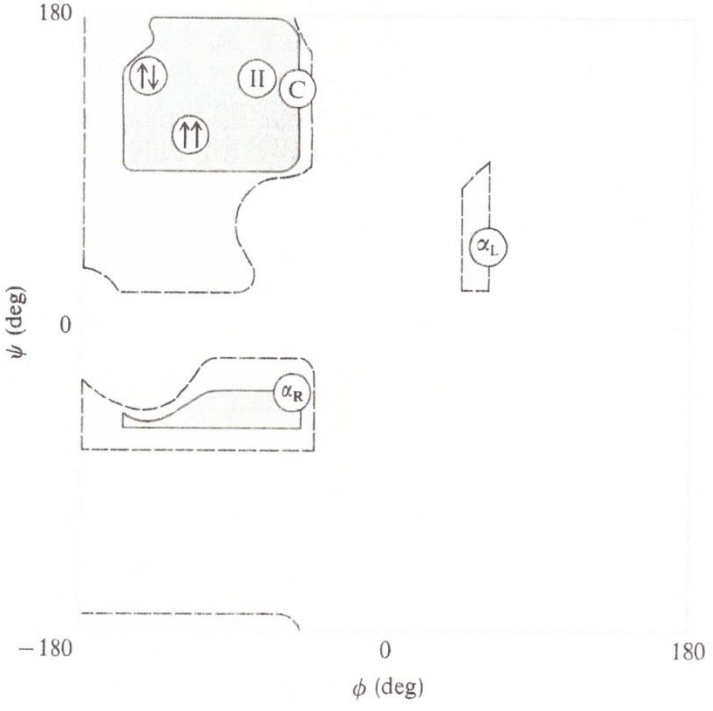


Figure 5-4

Steric contour diagram for an L*-alanyl residue in a polypeptide chain.* Dark zones show "normal" and light zones show "outer-limit" contours. Coordinates of right- and left-handed α helices (α), parallel (↑↑) and antiparallel (↑↓) pleated sheets, polyglycine II (II), and collagen (C) are denoted (see also Table 5-2). [After P. J. Flory, *Statistical Mechanics of Chain Molecules* (New York: Interscience, 1969), and G. N. Ramachandran et al., *J. Mol. Biol.* 7:95 (1963).]

show some of the shortest contact radii observed in crystal structures of relevent molecules; these somewhat shorter radii give the "outer limit" of the sterically allowed values.

Table 5-3 lists, for various atom pairs, the minimal contact distances important for the generation of Figures 5-3 and 5-4. These contact distances are based on values found in crystallographic structure determinations.

Table 5-3
Minimum contact distances for important polypeptide atom pairs

Atom pair	Normal (Å)	Outer limit (Å)
C—C	3.2	3.0
C—O	2.8	2.7
C—N	2.9	2.8
C—H	2.4	2.2
O—O	2.8	2.7
O—N	2.7	2.6
O—H	2.4	2.2
N—N	2.7	2.6
N—H	2.4	2.2
H—H	2.0	1.9

SOURCE: O—O distances from C. Ramakrishnan and G. N. Ramachandran, *Biophys. J.* 5:909 (1965); other data from G. N. Ramachandran et al., *J. Mol. Biol.* 7:95 (1963).

The steric diagram for glycine is centrosymmetric—that is, symmetric with respect to any line that passes through the center ($\phi = 0°$, $\psi = 0°$). This is a consequence of the symmetry of the glycine residue. As might be anticipated, the *cis* configuration about either bond ($\phi = 0°$ or $\psi = 0°$) and configurations in the neighborhood of the vertical lines ($\phi = 0°$ or $\psi = 0°$) are sterically forbidden. In the case of alanine, the steric diagram is asymmetric, and the sterically allowed domain is significantly smaller. This smaller domain is due to the additional unfavorable contacts that involve the $-CH_3$ side chain. Steric diagrams for residues with longer side chains but that are not branched at the β-carbon—such as leucine, R = $-CH_2CH(CH_3)_2$—are essentially the same as Figure 5-4. A side chain that is branched at the β-carbon—such as valine, R = $CH(CH_3)_2$—does reduce somewhat the sterically permitted domain below the areas of Figure 5-4.

5-3 ESTIMATES OF POTENTIAL ENERGY

The steric diagrams of Figures 5-3 and 5-4 give an approximate indication of the allowed conformations of amino acid residues. They do not indicate, however, which parts of the sterically allowed domain are preferred—that is, they do not show

gradations in energy within the sterically permissible regions. In order to obtain an estimate of relative preference, it is necessary to compute the potential energy as a function of ϕ and ψ.

Methods for computing rotational potential functions are only semiempirical; therefore, energies computed by such methods are not to be viewed too literally. Nevertheless, when these methods are used in conjunction with configuration statistics (see Chapter 18), there usually is reasonable accord between calculated and observed chain dimensions; on the other hand, the use of "hard sphere" diagrams (such as Figs. 5-3 and 5-4) does not always give good agreement with experiment.

"Nonbonded" interactions

Consider first the attractive and repulsive interactions between atoms whose distance of separation is a function of ϕ and ψ. These nonbonded interactions give rise to an energy $E_{kl}(\phi_i, \psi_i)$, which is generally expressed as

$$E_{kl}(\phi_i, \psi_i) = a_{kl}/r_{kl}^m - c_{kl}/r_{kl}^6 \tag{5-1}$$

where the parameters a_{kl} and c_{kl} are characteristic of the atom pair or atomic group (e.g., CH_3) k and l. A typical plot of E_{kl} versus r_{kl} is given in Figure 5-5. The plot shows that there is no interaction at large distances but that, as the atoms approach each other, an attractive interaction sets in. This attractive interaction is soon overcome by a strong repulsion as the atoms begin to penetrate each other's atomic radii.

The sixth-power attractive term is the familiar London dispersion energy. The distribution of electron density around any nucleus undergoes rapid fluctuations that give rise to a separation between the center of electronic charge and the center

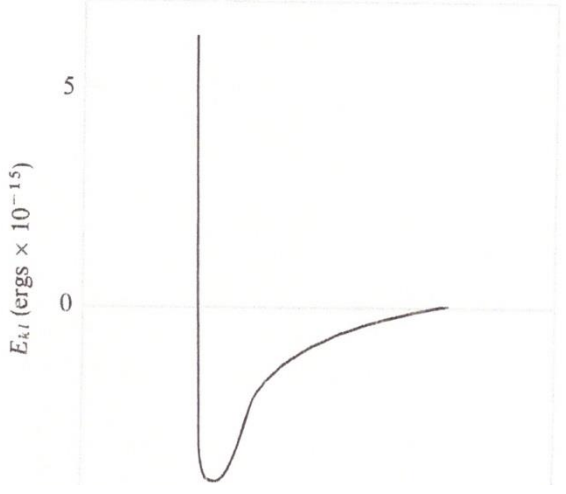


Figure 5-5

Nonbonded interaction energy E_{kl} between atoms or groups k and l plotted as a function of their separation. The example shown is for a pair of hydrogen molecules. [After W. J. Moore, *Physical Chemistry*, 3d ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1962), p. 715.]

of nuclear charge. This separation creates a transient dipole moment, which can resonate with that of a neighboring atom, so that the transient charges orient in an attractive configuration. The result is an induced-dipole–induced-dipole interaction, which accounts for the sixth-power dependence; the parameter c_{kl} is obtained from the atomic polarizabilities of atoms k and l, and the effective number of valence-shell electrons. (See Brant et al., 1967, and references therein.)

The repulsive term a_{kl}/r_{kl}^m in Equation 5-1 has no strict theoretical justification;[§] experimental results obtained with molecular beams suggest that m is in the range of 9 to 12. The parameter a_{kl} may be obtained by requiring the minimum in Figure 5-5 to occur at the sum of the atomic radii or by a variation of this procedure (see Brant et al., 1967). Some values for the atomic radii (r^0) that are used in polypeptide calculations are $r_H^0 = 1.2$ to 1.3 Å; $r_O^0 = 1.5$ to 1.6 Å; and $r_{CH_2}^0 = 1.95$ Å (the CH_2 group usually is treated as a single spherical entity).

Dipolar interactions

A second contribution to the energy producing rotations ϕ and ψ arises from the dipole interaction between adjacent amide groups. The amide-group dipole is approximately parallel to the N–H bond and points from N to H (Fig. 5-6). It has a

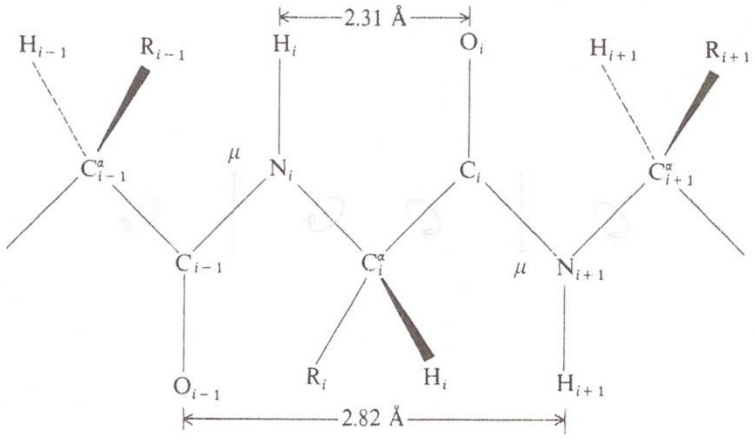


Figure 5-6

Representation of the amide-group dipole moment μ. In the fully extended chain, these moments alternate in direction along the chain. [After P. J. Flory, *Statistical Mechanics of Chain Molecules* (New York: Interscience, 1969), chap. VII.]

[§] An exponential repulsive potential also has been used. See, for example, Brant and Flory (1965).

magnitude of approximately 3.7 D. This is a relatively large moment; for example, the dipole moment of HCl is 1.03 D, that of CH_3Cl is 1.87 D, and that of HCN is 2.93 D. Therefore, we expect dipolar interactions to have a significant influence on polypeptide backbone conformation.

Clearly, the orientation of any two successive amide-group dipoles is determined by ϕ and ψ. The interaction energy E_d between two point dipoles $\boldsymbol{\mu}_A$ and $\boldsymbol{\mu}_B$ separated by the vector $\mathbf{r}$ is given by

$$E_d = \varepsilon^{-1}[\boldsymbol{\mu}_A \cdot \boldsymbol{\mu}_B/r^3 - 3(\boldsymbol{\mu}_A \cdot \mathbf{r})(\boldsymbol{\mu}_B \cdot \mathbf{r})/r^5] \tag{5-2}$$

where r is the scalar magnitude of $\mathbf{r}$, and ε is the dielectric constant. (See Box 5-1.) This expression is, of course, only approximate because it requires that the distance of separation of the amide dipoles be large compared to the separation of the partial charges of the dipole moment itself. To achieve a more accurate representation of the electrostatic interaction between adjacent amide dipoles, the energy may be computed by assigning partial charges to the appropriate atoms and then calculating the total electrostatic energy as a sum of charge–charge coulombic interactions. The energy then is computed from

$$E_d = \left(\sum_{ij} q_i q_j\right) \Big/ \varepsilon r_{ij} \tag{5-3}$$

where q_i and q_j are the partial charges on atoms i and j, and r_{ij} is their distance of separation. Partial charges must be chosen, of course, that will approximately reproduce the overall amide-group dipole moment as well as the accepted values for the individual N–H and C–O bond moments. These criteria are fulfilled by assigning partial charges of $-0.28e$ and $+0.28e$ to N and H, respectively, and $-0.39e$ and $+0.39e$ to O and C, respectively.

Choosing a value for the dielectric constant ε is not a simple task. Although the macroscopic dielectric constant of water is about 80, the local microscopic constant should be considerably less. One expects the lines of force between two *adjacent* dipoles to pass, in large part, through the local backbone of the polymer itself, and not through the surrounding bulk water. Therefore, the high-frequency dielectric constant of solid amides and polyamides, to which contributions from the orientations of permanent dipoles are small, might be a more appropriate measure of the local dielectric constant along a polypeptide backbone in solution. The appropriate data suggest that a value in the range of $\varepsilon = 2$ to 5 might be a realistic estimate of the effective dielectric constant between two successive amide groups in a polypeptide chain.

To a good approximation, dipolar interactions between group dipoles that are second neighbors may be ignored because the dipolar interaction energy falls off rapidly with increasing distance of separation of the dipoles (see Eqn. 5-2). Moreover, at larger distances, the lines of electrostatic force may pass through some of the bulk solvent, thus raising the effective dielectric constant.

Intrinsic torsional potential

In addition to contributions examined above, a small hindrance to rotation also occurs because of an intrinsic torsional potential associated with single bonds. That is, apart from nonbonded or electrostatic interactions, the bond itself presents a barrier to rotation. Thus, the ϕ and ψ rotations each have intrinsic rotational hindrance potentials that are probably threefold (that is, three minima occur). These minima

Box 5-1 DERIVATION OF INTERACTION ENERGY BETWEEN TWO DIPOLES

The mutual potential energy of two point dipoles is not conceptually difficult to calculate. If we designate the respective dipole moments as $\boldsymbol{\mu}_A$ and $\boldsymbol{\mu}_B$, then the interaction energy is simply the potential energy of $\boldsymbol{\mu}_B$ in the electric field generated by $\boldsymbol{\mu}_A$ (or vice versa)—that is, $-\mathbf{E}_A \cdot \boldsymbol{\mu}_B$, where $\mathbf{E}_A$ is the electric field generated by the dipole $\boldsymbol{\mu}_A$. Therefore, the problem reduces to calculating $\mathbf{E}_A$.

For this purpose, we first consider the more general case of the field produced by an arbitrary charge distribution, and we illustrate the concept of the multipole expansion. Imagine a distribution of charges about an origin; the ith one has a charge $z_i e$ and has coordinates $\tilde{x}_i, \tilde{y}_i, \tilde{z}_i$. The electric potential Φ at some distance ($\mathbf{r} = \hat{\mathbf{i}}\tilde{x} + \hat{\mathbf{j}}\tilde{y} + \hat{\mathbf{k}}\tilde{z}$) from the origin is given by

$$\Phi = \sum\{z_i e/\varepsilon[(\tilde{x} - \tilde{x}_i)^2 + (\tilde{y} - \tilde{y}_i)^2 + (\tilde{z} - \tilde{z}_i)^2]^{1/2}\}$$

This expression may be expanded in a Taylor series in $\tilde{x}_i, \tilde{y}_i, \tilde{z}_i$ about the origin. The result is

$$\begin{aligned}\Phi = (1/\varepsilon)\{&\sum(z_i e/r) + \sum[z_i e\tilde{x}_i(\partial X^{-1/2}/\partial \tilde{x}_i)] + \sum[z_i e\tilde{y}_i(\partial X^{-1/2}/\partial \tilde{y}_i)] \\ &+ \sum[z_i e\tilde{z}_i(\partial X^{-1/2}/\partial \tilde{z}_i)] + \tfrac{1}{2}\sum[z_i e\tilde{x}_i^2(\partial^2 X^{-1/2}/\partial \tilde{x}_i^2)] \\ &+ \sum[z_i e\tilde{x}_i\tilde{y}_i(\partial^2 X^{-1/2}/\partial \tilde{x}_i \partial \tilde{y}_i)] + \cdots\}\end{aligned}$$

where $X = (\tilde{x} - \tilde{x}_i)^2 + (\tilde{y} - \tilde{y}_i)^2 + (\tilde{z} - \tilde{z}_i)^2$, and all the derivatives are to be evaluated at $\tilde{x}_i = \tilde{y}_i = \tilde{z}_i = 0$.

Note that the first term is simply the electrostatic potential at the point r for the situation in which all the charges $z_i e$ are concentrated at the origin (potential = [net charge]/εr. The following terms correct for the fact that the charges are distributed about the origin. The first three terms contain the $\tilde{x}$, $\tilde{y}$, and $\tilde{z}$ components of the dipole moment $\boldsymbol{\mu}$ of the distribution. That is,

$$\mu_{\tilde{x}} = \sum z_i e\tilde{x}_i; \qquad \mu_{\tilde{y}} = \sum z_i e\tilde{y}_i; \qquad \mu_{\tilde{z}} = \sum z_i e\tilde{z}_i$$

The next nine terms involve the nine components of the quadrapole moment $\mathbf{Q}$ of the distribution, where

$$Q_{\tilde{x}\tilde{x}} = \sum z_i e\tilde{x}_i^2; \qquad Q_{\tilde{y}\tilde{y}} = \sum z_i e\tilde{y}_i^2; \qquad Q_{\tilde{z}\tilde{z}} = \sum z_i e\tilde{z}_i^2; \qquad Q_{\tilde{x}\tilde{y}} = \sum z_i e\tilde{x}_i\tilde{y}_i; \ldots$$

The nine components of the quadrapole moment make up the quadrapole moment tensor.

occur at rotation angles of (approximately) 60°, 180°, and 300°. The barrier heights between minima are believed to be fairly small, on the order of 1 kcal mole^{-1}. The torsional energy $E_{\text{tor}}(\phi_i, \psi_i)$ thus can be represented as

$$E_{\text{tor}}(\phi_i, \psi_i) = (E_\phi^0/2)(1 + \cos 3\phi) + (E_\psi^0/2)(1 + \cos 3\psi) \tag{5-4}$$

where E_ϕ^0 and E_ψ^0 are the barrier heights associated with ϕ and ψ rotations, respectively.

Using the relationships for the components of the dipole and quadrapole moments, and noting that $(\partial X^{-1/2}/\partial \tilde{x}_i)_{\tilde{x}_i = 0} = -\partial r^{-1}/\partial \tilde{x}$ (and so on for derivatives with respect to $\tilde{y}_i$ and $\tilde{z}_i$), we are able to write the series expansion of Φ as

$$\begin{aligned}\Phi = (1/\varepsilon)[&(\textstyle\sum z_i e)/r - \mu_{\tilde{x}}(\partial r^{-1}/\partial \tilde{x}) - \mu_{\tilde{y}}(\partial r^{-1}/\partial \tilde{y}) - \mu_{\tilde{z}}(\partial r^{-1}/\partial \tilde{z}) \\ &+ \tfrac{1}{2}Q_{\tilde{x}\tilde{x}}(\partial^2 r^{-1}/\partial \tilde{x}^2) + \tfrac{1}{2}Q_{\tilde{y}\tilde{y}}(\partial^2 r^{-1}/\partial \tilde{y}^2) + \tfrac{1}{2}Q_{\tilde{z}\tilde{z}}(\partial^2 r^{-1}/\partial \tilde{z}^2) \\ &+ Q_{\tilde{x}\tilde{y}}(\partial^2 r^{-1}/\partial \tilde{x}\partial \tilde{y}) + \cdots]\end{aligned}$$

This is the multipole expansion of the potential Φ; the terms containing the components of the dipole moment make up the dipole potential, and the second-derivative terms are the quadrapole potential.

Assume now that the charge distribution contains equal numbers of opposite charges (so that $\sum z_i e = 0$) displaced a small distance from the origin to give a dipole moment $\boldsymbol{\mu}_A$ (for example, the amide dipole is equivalent to two equal and opposite unit charges displaced a distance of 3.7 Å from each other). In the expansion of Φ, the leading nonvanishing term is the dipole potential, which (for small displacements of charges) dominates over the quadrapole potential, which involves second-power terms in the displacements. Noting that $\partial r^{-1}/\partial \tilde{x} = -\tilde{x}/r^3$, we obtain

$$\Phi = (\mu_{A\tilde{x}}\tilde{x} + \mu_{A\tilde{y}}\tilde{y} + \mu_{A\tilde{z}}\tilde{z})/\varepsilon r^3$$

where $\mu_{A\tilde{x}}$, $\mu_{A\tilde{y}}$, and $\mu_{A\tilde{z}}$ are the components of $\boldsymbol{\mu}_A$. The field $\mathbf{E}_A$ is

$$\begin{aligned}\mathbf{E}_A &= -\nabla\Phi \\ &= -\nabla\boldsymbol{\mu}_A \cdot \mathbf{r}/\varepsilon r^3 \\ &= -\boldsymbol{\mu}_A/\varepsilon r^3 - (\boldsymbol{\mu}_A \cdot \mathbf{r}/\varepsilon)\nabla(1/r^3)\end{aligned}$$

The potential energy of interaction with the dipole $\boldsymbol{\mu}_B = \hat{\mathbf{i}}\mu_{B\tilde{x}} + \hat{\mathbf{j}}\mu_{B\tilde{y}} + \hat{\mathbf{k}}\mu_{B\tilde{z}}$ located at the distance $\mathbf{r}$ from $\boldsymbol{\mu}_A$ is simply $-\mathbf{E}_A \cdot \boldsymbol{\mu}_B$, which after simplification can be written as

$$E_d = (\boldsymbol{\mu}_A \cdot \boldsymbol{\mu}_B/\varepsilon r^3) - 3(\boldsymbol{\mu}_A \cdot \mathbf{r})(\boldsymbol{\mu}_B \cdot \mathbf{r})/\varepsilon r^5$$

This is the result given in Equation 5-2.

Total energy as the sum of individual contributions

Our goal now is to calculate the total energy as a function of ϕ_i and ψ_i for the ith amino acid residue situated within a polypeptide chain. As ϕ_i and ψ_i are varied, the distance between all atoms whose separation depends solely on ϕ_i and/or ψ_i must be calculated. For each atom pair k and l, the nonbonded energy E_{kl} is computed according to Equation 5-1, and the sum over all atom pairs gives the total nonbonded interaction energy. In addition, the distance of separation between the two dipoles on either side of the ith α-carbon atom depends on ϕ_i and ψ_i, so that the dipolar energy must be calculated for every value of ϕ_i and ψ_i (Eqn. 5-2 or 5-3). Finally, the contribution of the torsional potential must also be computed for every value of ϕ_i and ψ_i (Eqn. 5-4). The total rotational potential $E(\phi_i, \psi_i)$ thus is given by

$$E(\phi_i, \psi_i) = \sum_{k,l} [E_{kl}(\phi_i, \psi_i) + E_{\mathrm{d}}(\phi_i, \psi_i) + E_{\mathrm{tor}}(\phi_i, \psi_i)] \tag{5-5}$$

where the summation extends over all atom pairs k and l whose distance of separation depends on ϕ_i and/or ψ_i. Equation 5-5 is, therefore, the basis for the calculation of conformational energies of polypeptide chain molecules.[§]

5-4 RESULTS OF POTENTIAL-ENERGY CALCULATIONS

Glycyl and L-alanyl residues

The results of energy estimates calculated according to Equation 5-5 are shown in Figure 5-7 for a glycyl residue and in Figure 5-8 for an alanyl residue, both situated within a polypeptide chain. The energy contour zones are drawn at intervals of 1 kcal mole^{-1} relative to the minima marked by X. In the case of glycyl (Fig. 5-7), two symmetrically disposed minima occur at about $\phi = -80°$, $\psi = +90°$, and $\phi = +80°$, $\psi = -90°$. There are four distinct regions on the glycyl map in which the residue energy is less than 5 kcal mole^{-1}. These regions are comparable to the "sterically allowed" regions shown in Figure 5-3.

In the case of the L-alanyl residue (Fig. 5-8), three distinct low-energy regions are designated by roman numerals. The lowest points of regions I and II are close to the coordinates of the right-handed and left-handed alpha helices, respectively (see Table 5-2).[§§] It is clear, however, that the minimum within region I is lower than that of region II. The energy map, therefore, provides a rational basis for the well-known preference of L-polypeptides for the right-handed α-helical conformation.

[§] This equation ignores special solvent interactions that depend on conformation.

[§§] Hydrogen-bonding interactions between peptide units separated by three pairs of ϕ, ψ angles play a significant role in stabilizing the α-helical conformations near the minima within regions I and II. These interactions have *not* been included in the calculations of Figures 5-7 and 5-8, and consequently these diagrams should be viewed as applicable to situations where the solvent or other factors render such interactions unlikely.

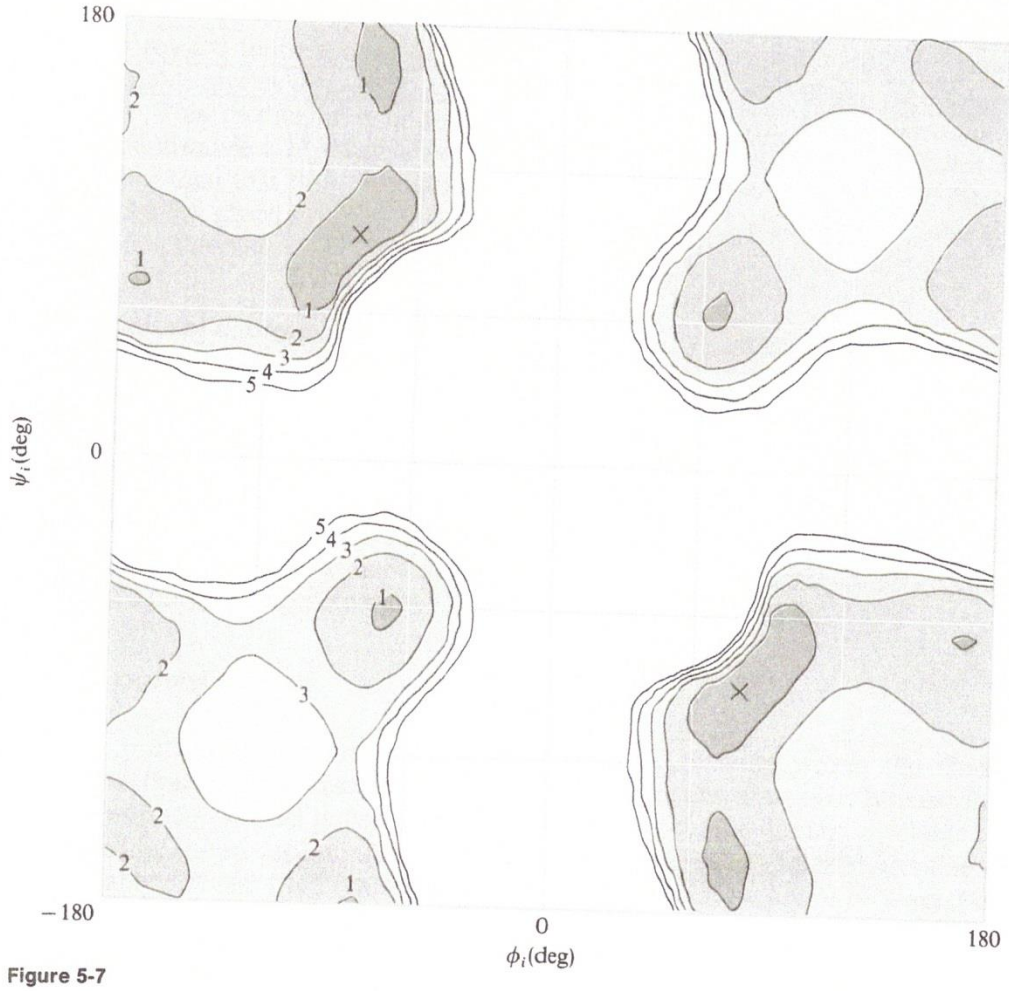


Figure 5-7

Energy contour diagram for a glycyl residue in a polypeptide chain calculated from Equation 5-5. Changes in shading show 1-kcal intervals; X indicates positions of lowest energy. [After D. A. Brant and P. J. Flory, *J. Mol. Biol.* 23:47 (1967).]

Importance of the dipolar interaction

Region III is the lowest-energy portion of the energy diagram of Figure 5-8, with the 2 kcal zone enscribing a relatively large area around the minumum (X). This region encompasses the extended forms of the residue, including the well-characterized β-pleated sheet structures (see Table 5-2). However, it should be mentioned that, energetically speaking, the chief reason for the preference of region III over region I arises from the amide dipolar interactions. If the dipolar interactions are omitted, region I becomes preferred over region III. The necessity for including the dipolar

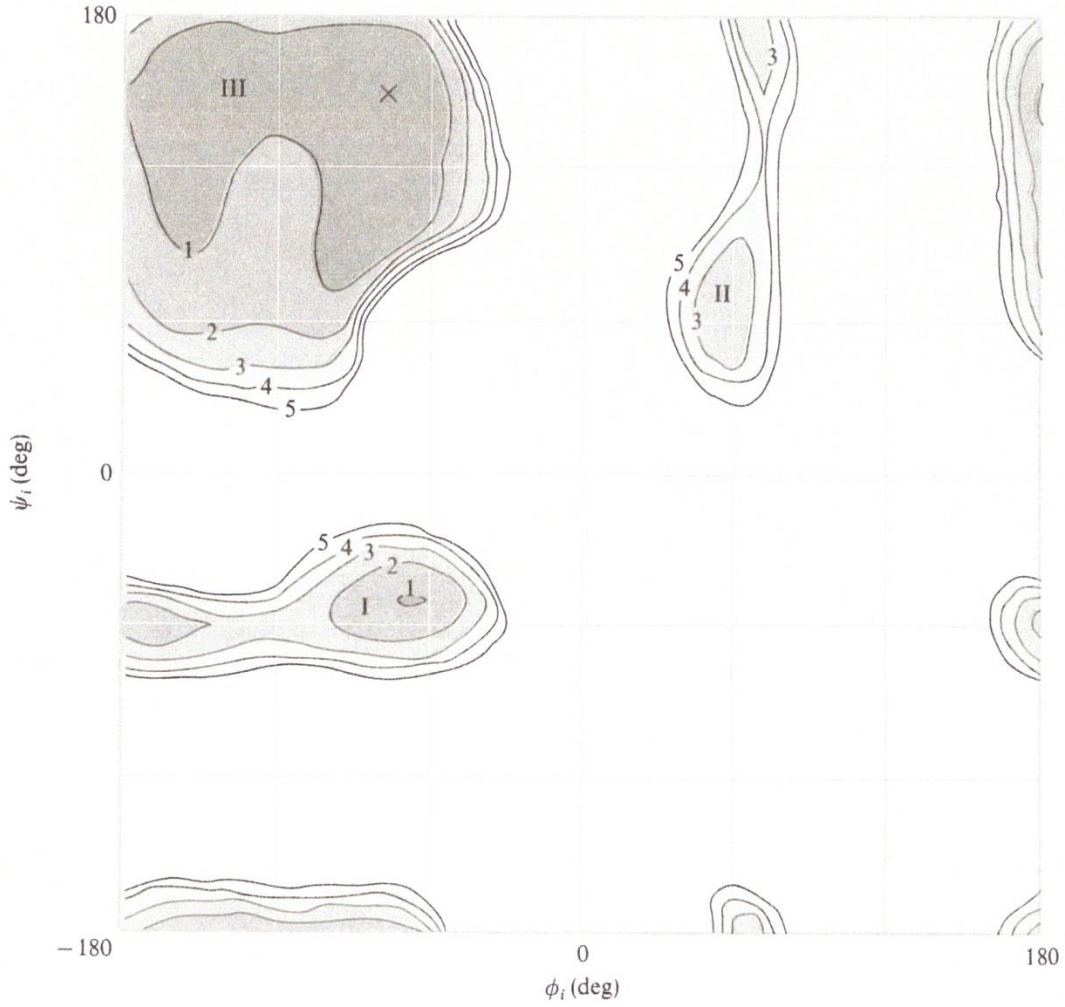


Figure 5-8

Energy contour diagram for an L*-alanyl residue in a polypeptide chain* calculated from Equation 5-5 [After D. A. Brant and P. J. Flory, *J. Mol. Biol.* 23:47 (1967).]

interactions in the calculations has experimental support from measurements of the random-coil dimensions of polyalanine-like chains. These measurements have shown that the backbone is biased toward extended conformations to a degree that is explained only by taking into account the electrical interactions of the backbone. (This point is discussed in greater detail in Chapter 18.)

Consequences of glycyl and L-alanyl conformational energies

The energy maps for glycine and alanine have some interesting implications. Because of their symmetrically disposed low-energy regions, glycine residues are naturally more "flexible" than alanine or alaninelike amino acids (those bearing side chains).

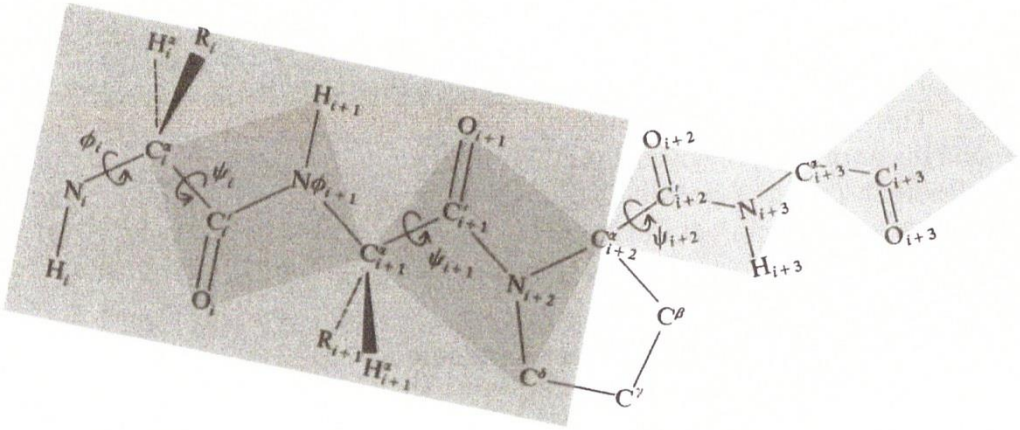


Figure 5-9

Schematic illustration of a polypeptide chain containing an isolated proline residue. [After P. R. Schimmel and P. J. Flory, *J. Mol. Biol.* 34:105 (1968).]

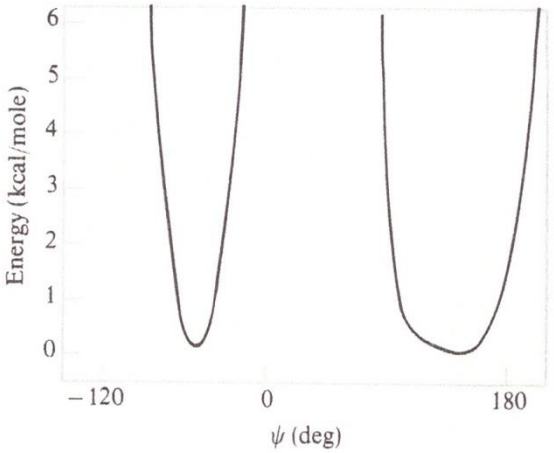


Figure 5-10

Plot of conformational energy versus ψ for an isolated proline within a polypeptide chain. [After P. R. Schimmel and P. J. Flory, *J. Mol. Biol.* 34:105 (1968).]

Certainly the relative ease with which the residue can adopt the compact $\psi \cong -5$ conformation makes it a desirable candidate for accommodating turns and bends i a polypeptide chain.

Conformational energy of the residue preceding L-proline

We have mentioned that rotations ϕ_i, ψ_i of a residue within a polypeptide chain ar independent of rotations within neighbors $i-1$ and $i+1$. The steric interactions c residue i with its predecessor $i-1$ are, therefore, independent of the nature of th predecessor because only the carbonyl group (C=O) of the preceding residue i involved. However, in the case of the succeeding residue $(i+1)$, it matters whethe

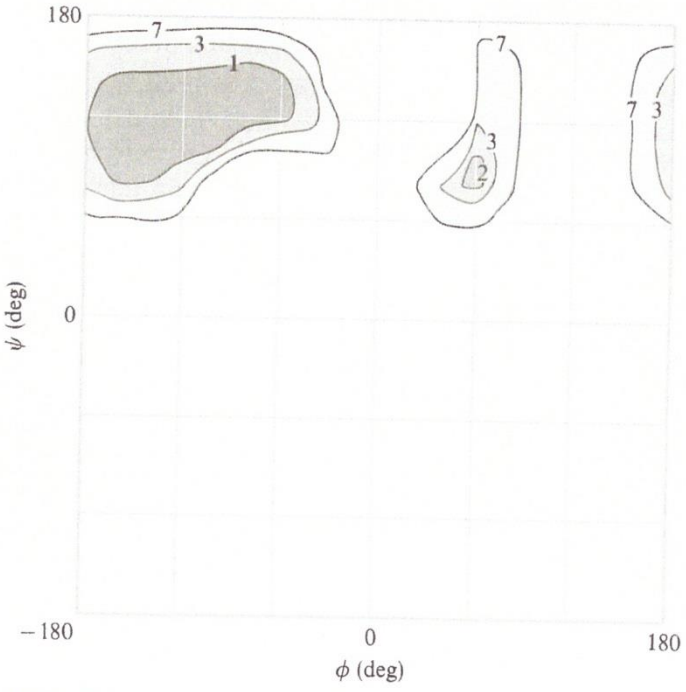


Figure 5-11

Energy contour diagram for an L*-alanyl residue that is succeeded by proline.* [After P. R. Schimmel and P. J. Flory, *J. Mol. Biol.* 34:105 (1968).]

or not the $(i + 1)$th residue is proline; the usual interactions of the ith residue with the amide hydrogen of the $(i + 1)$th are replaced by steric conflicts involving the atoms of the pyrrolidine ring, particularly the $-CH_2-$ group attached to the imide nitrogen. Therefore, the rotational freedom of a residue succeeded by proline may be expected to be more restricted than if it preceded one of the other residues.

Figure 5-11 is an energy diagram for an L-alanyl residue succeeded by a L-prolyl residue. Comparison of Figure 5-11 with Figure 5-8 reveals that the proline residue has significantly curtailed the domain accessible to alanine. Region I is completely inaccessible, with conformations throughout the entire range of $-180° < \psi < 60°$ now involving unfavorable steric overlaps. These added unfavorable interactions especially involve contacts between the $-CH_3$ side group of alanine with the $-CH_2$ group attached to the imido nitrogen.

When glycine precedes an L-proline residue, the general aspects of its energy map are not greatly altered from that shown in Figure 5-7. All four sterically allowed regions remain intact. The energy in the vicinity of $\psi = 0°$ is raised owing to overlaps involving the pyrrolidine ring.

Conformational constraints on the residue preceding proline

Region I is not accessible for residues that bear a side chain and are succeeded by proline. This fact has an especially significant consequence: the right-handed α-helical conformation is no longer possible for those residues. On the other hand, even though its nitrogen cannot participate in a hydrogen bond, the proline residue can adopt the right-handed α-helical conformation (see Fig. 5-10). Therefore, a proline can occur at the beginning of an α-helical sequence, but it prevents the preceding residue from adopting the helical conformation if that residue is not a glycine. Thus, the fact that the proline nitrogen cannot participate in a hydrogen bond is not the primary cause for the ability of this residue to disrupt α-helical sequences, although this reason is often given as the basis for proline's helix-disrupting tendencies.

5-5 EXPERIMENTALLY OBSERVED VALUES FOR ROTATION ANGLES

In the preceding sections we have considered the energetically allowed conformations of amino acid residues situated within polypeptide chains. The energy functions that were used to calculate the energy diagrams are only semiempirical, of course. The question naturally arises as to whether or not the observed conformations of amino acid residues agree with the general predictions of these calculations. This question can best be answered by considering the ϕ, ψ rotation angles of residues in the various crystal structures of peptides, polypeptides, and proteins.

A word of caution must be voiced, however. Several factors can cause or magnify a discrepancy between the predicted and observed conformations. One arises from the fact that the diagrams given in Figures 5-3, 5-4, 5-7, 5-8, 5-10, and 5-11 are based on the assumption of fixed bond angles and bond lengths. A variation of a few degrees in a bond angle, which is not unreasonable, can cause some alterations in the shapes of the "allowed" regions; a small change in bond lengths can similarly alter the diagrams. In addition, the accuracy to which the ϕ, ψ coordinates are determined in a protein crystal, for example, is subject to an error of at least several degrees.

Lysozyme as an example

With these limitations in mind, we can pursue the question of how well the observed conformations agree with the energy diagrams. For an example, we consider the residues in crystalline lysozyme that occur in nonhelical sequences. (Residues within sequences experience contributions to their conformational energy that are not included in our calculations.[§])

[§] For example, hydrogen bonds occur between groups separated by three ϕ, ψ pairs in a α-helix. These bonds, however, are not necessarily favored in aqueous environment, where water molecules can compete for hydrogen-bonding sites within the protein. In a folded protein, of course, the chain itself can form pockets of nonpolar environments, which can allow internal hydrogen bonding within the protein.

Figure 5-12a is an energy contour map for a residue bearing a CH_2–R′ side chain at the alpha-carbon, and situated within a polypeptide chain; the individual conformations of the appropriate residues are represented by points. (The energy diagram is identical, of course, to the one for alanine given in Figure 5-8.)

It is clear that almost all of the residues fall within areas that are not excessively high in energy. In fact, 22 of the 61 residues plotted in Figure 5-12a fall within the 1 kcal $mole^{-1}$ zone surrounding the minimum X. In general, of course, the relative number of residues within a particular region depends on the energy as well as the effective area of that region, according to the principles of statistical mechanics. We might qualitatively expect, therefore, that the relative abundancies of the three regions should follow the order III > I > II. This is, in fact, what is observed. However, it is also observed that more residues appear to fall in the relatively high-energy saddle region between regions I and III than in region II. This apparent discrepancy can be rationalized, according to the limitations discussed above.

The degree to which certain residues do not conform to the calculated preferred conformations can also serve as an index of the extent and magnitude of forces that have not been accounted for. Steric or other kinds of interactions between residues that are far apart in the sequence, solvent–protein interactions, and forces of crystal packing can presumably be strong enough to encourage a residue to adopt a conformation that is at the fringes of or outside of an "allowed" domain. Hence, some of the discrepancies between calculated and observed conformations may not reflect any shortcoming in the calculations or in the crystallographic data, but rather reflect the magnitude of these compensatory forces.

Figure 5-12b is an energy contour map for a glycyl residue (compare Figure 5-7) situated within a polypeptide chain, with the glycine residue conformations of lysozyme represented by points. Only eleven residue conformations are plotted. Of these, two fall in excessively high-energy areas. The remaining residues appear to scatter throughout the "allowed" regions, and show no particular tendency to cluster within the very lowest (1 kcal) contours.

Other proteins, as well as various peptide structures, have been subjected to comparisons of the predicted and observed ϕ, ψ coordinates of the individual residues. By and large, the residues in these structures fall within the "allowed" domains of the energy maps. Thus, in trying to predict the conformation of a given protein it is reasonably safe to assume that the residues will adopt conformations within the predicted "allowed" regions. However, the latitude of choice of conformations within the permissible areas is broad enough to allow for a great number of polymer conformations. For this reason, attention must be turned to the rather specialized forces that aid in stabilizing the highly ordered conformations of native protein molecules. It is these forces that determine, from the sterically permitted conformations, the one (or ones) that is thermodynamically preferred by a significant margin.

A number of specialized forces and interactions occurring in proteins are major determinants of the native three-dimensional structure. These include the covalent cross-linking of chain segments via disulfide bridges, noncovalent interactions such as the hydrogen bond, ionic interactions, "hydrophobic" interactions, and other solvent–macromolecule interactions. These forces, taken together, are sufficiently strong and specific to bias the molecule toward a highly preferred conformation.

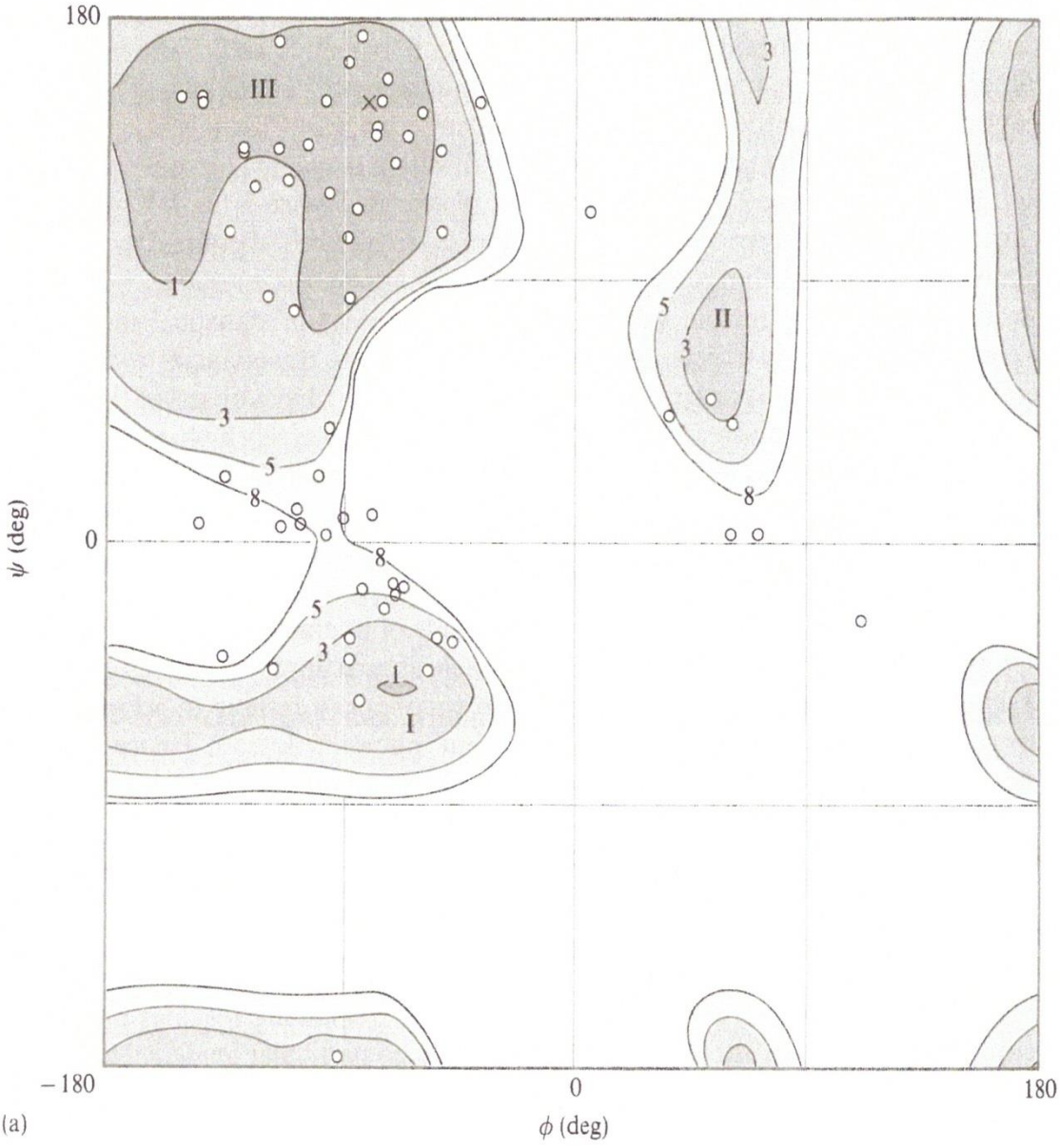
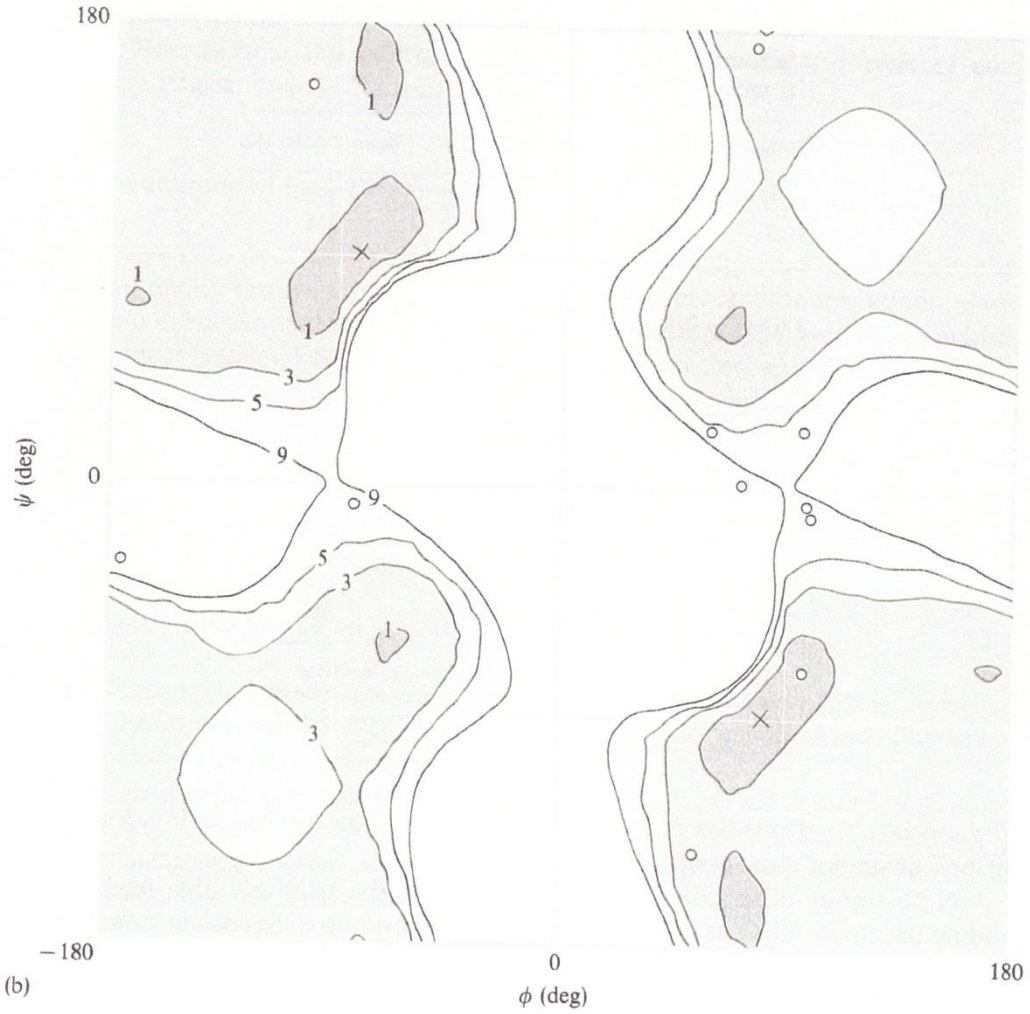


Figure 5-12

Observed residue coordinates plotted on conformational energy maps. Points show observed coordinates of residues in crystalline egg-white lysozyme occurring in nonhelical sequences. Conformational energy maps are calculated for residues situated within polypeptide chains. (**a**) L-Alanyl-like residues. (**b**) Glycyl residues. [After D. A. Brant and P. R. Schimmel, *Proc. Natl. Acad. Sci. USA* 58:428 (1967).]

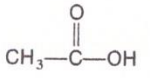
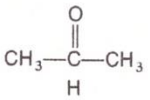


(b)

5-6 HYDROGEN BONDING

A hydrogen bond generally is said to exist between a donor molecule D–H and an acceptor A when there is evidence that the two molecules associate in a fashion specifically involving the hydrogen atom of the donor. It is not hard to marshall spectroscopic evidence for the existence of a hydrogen bond in particular situations where it might be suspected to form. For example, frequency shifts of infrared and Raman bands are readily interpreted in terms of hydrogen-bond formation. Another direct physical method that can be used for detecting hydrogen bonds is proton magnetic resonance (see Chapter 9). On the other hand, evidence for strong intermolecular associations also can be obtained by comparing rather ordinary macroscopic properties such as freezing and boiling points of pure liquids, and their heats

Table 5-4
Comparison of melting and boiling points for molecules of similar size

Compound	Melting point (K)	Boiling point (K)
H_2O	273	373
H_2S	190	211
	290	391
	178	330
CH_3CH_2OH	156	351
$CH_3CH_2CH_3$	83	231
$CH_3—O—CH_3$	135	249
CH_3NH_2	181	267
CH_3CH_3	101	185

SOURCE: Data from G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond* (San Francisco: W. H. Freeman and Company, 1960), p. 36; and *Handbook of Chemistry and Physics*, 39th ed. (Cleveland, Ohio: Chemical Rubber Pub. Co., 1958).

of vaporization. Table 5-4 compares melting points and boiling points for several groups of similar sized substances.

Note among other comparisons, for example, the relatively high freezing and boiling points for H_2O as compared to H_2S. The striking difference in these macroscopic properties clearly points to a strong intermolecular association existing in water. Consider also the boiling points of three closely similar aromatic liquids: the boiling point of chlorobenzene is 132°, that of *o*-chlorophenol is 175°, and that of *p*-chlorophenol is 217°. Can you rationalize these differences?

It is experimentally observed that a hydrogen bond between D–H and A can form when the D and A atoms are strongly electronegative. In most cases, the three atoms also must be essentially colinear when forming the bond. But what is the physical basis for the hydrogen bond's strength? A simple and early view holds that the bond is essentially electrostatic in nature and forms as a result of attractive coulombic interactions. The small charge separations responsible for the coulombic interaction are envisioned as existing before bond formation. This view is supported by the fact that the strongest hydrogen bonds form when D and A are fluorine atoms and the next strongest when they are oxygen atoms, with nitrogen being somewhat weaker. However, this simple electrostatic viewpoint cannot account for all experimental observations—for example, the lack of any relationship between the dipole moment and the hydrogen-bond strength of a base. It is clear, therefore, that a more

sophisticated quantum mechanical description is appropriate, and many have been offered. Nevertheless, the electrostatic model remains useful for making some qualitative predictions and rationalizations.

Water's competition for hydrogen-bonding sites on a protein

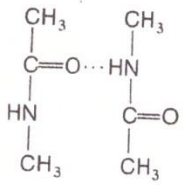
The hydrogen bonds of most importance for protein structure are those between the backbone amide nitrogens and carbonyl oxygens. These hydrogen bonds play a major role in stabilizing the various helical and ordered polypeptide conformations (such as the α-helix and β-pleated sheet structures), and they also play a prominent role in stabilizing ordered segments of folded protein molecules. Consequently it was assumed by many investigators that hydrogen bonds must play the dominant role in determining protein conformation.

It was eventually pointed out that water—the solvent for native proteins—can readily participate in hydrogen-bond formation. Consequently, the more pertinent question involves the free energy change for the process

$$\text{NH}\cdots\text{H}_2\text{O} + \text{OH}_2\cdots\text{O=C} \rightleftharpoons \text{NH}\cdots\text{O=C} + \text{released } \text{H}_2\text{O} \qquad (5\text{-}6)$$

It is not clear what the free energy change should be for this process, especially because the liberated water molecules probably participate in some kind of hydrogen bonding with other water molecules.

The general problem was carefully investigated by I. M. Klotz and J. S. Franzen (1962), using infrared spectroscopy. These investigators studied the association of *N*-methylacetamide (NMA), a good analog for the peptide backbone. The NMA molecule can form an amide hydrogen bond as follows:



where the amide group is in the *trans* conformation. The equilibrium between monomer (M) and dimer (D) was studied, and the data were analyzed in terms of the degree of association (α), which is defined as

$$\alpha = 2(\text{D})/[(\text{M}) + 2(\text{D})] = \text{fraction of molecules in dimeric state}$$

The parameter α was measured in three different solvents—carbon tetrachloride, dioxane, and water—so that the effect of solvent could be directly tested. Figure 5-13 shows the results as plots of α versus the *N*-methylacetamide (NMA) concentration. It is strikingly apparent that the association of NMA differs markedly in the three

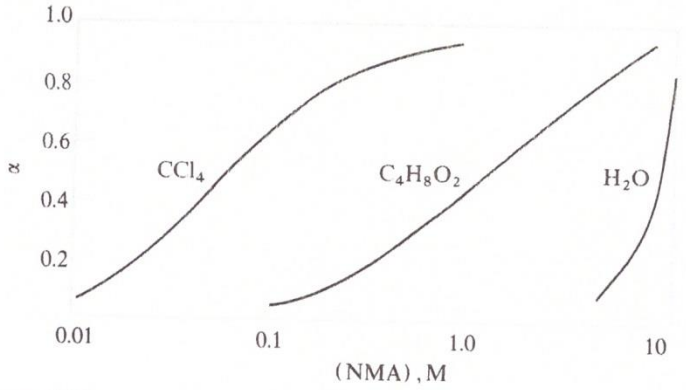


Figure 5-13

Association of N-methyl acetamide (NMA) in carbon tetrachloride, dioxane, and water. [After I. M. Klotz and J. S. Franzen, *J. Am. Chem. Soc.* 84:3461 (1962).]

solvents. The strongest association occurs in carbon tetrachloride, which cannot form any hydrogen bonds; a far weaker association occurs in water, an obviously strong competitor for the hydrogen-bonding sites on NMA. Dioxane, which has two oxygens that can accept a hydrogen bond, is intermediate between carbon tetrachloride and water.

The temperature dependence of the association in the three solvents also was measured in order to assess the thermodynamic parameters of the association (Table 5-5). It is clear that the association of NMA is thermodynamically preferred in carbon tetrachloride, but not in water.

These results clearly demonstrate the importance of the solvent in determining whether the change in state will be favorable. How do these results affect our understanding of the strength of hydrogen-bonding interactions in proteins? The interior of a protein is largely hydrocarbon in nature, so that the measurements carried out in carbon tetrachloride might be most relevant for a process in which a hydrogen

Table 5-5

Thermodynamics of amide hydrogen-bond formation by NMA at 25°C

Solvent	ΔH^0 (kcal mole^{-1})	ΔS^0 (cal deg^{-1} mole^{-1})	ΔG^0 (kcal mole^{-1})
CCl_4	−4.2	−11	−0.92
Dioxane	−0.8	−4	0.39
H_2O	0.0	−10	3.1

NOTE: The standard state is 1 mole liter^{-1}.
SOURCE: Data from I. M. Klotz and J. S. Franzen, *J. Am. Chem. Soc.* 84:3461 (1962).