Koide, T., Odani, S., & Ono, T. (1982) *FEBS Lett. 141*, 222–224.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Leung, L. L. K., Harpel, P. C., Nachman, R. L., & Rabellino, E. M. (1983) *Blood 62*, 1016–1021.
Leung, L. L. K., Nachman, R. L., & Harpel, P. C. (1984) *J. Clin. Invest. 73*, 5–12.
Levine, R. L., & Federici, M. M. (1982) *Biochemistry 21*, 2600–2606.
Lijnen, H. R., Hoylaerts, M., & Collen, D. (1980) *J. Biol. Chem. 255*, 10214–10222.
Lijnen, H. R., Hoylaerts, M., & Collen, D. (1983a) *J. Biol. Chem. 258*, 3803–3808.
Lijnen, H. R., Rylatt, D. B., & Collen, D. (1983b) *Biochim. Biophys. Acta 742*, 109–115.
Miles, E. W. (1977) *Methods Enzymol. 47*, 431–442.
Morgan, W. T. (1978) *Biochim. Biophys. Acta* 535, 319–333.
Morgan, W. T. (1981) *Biochemistry 20*, 1054–1061.
Morgan, W. T., & Muller-Eberhard, U. (1972) *J. Biol. Chem. 247*, 7181–7187.
Morgan, W. T., & Muller-Eberhard, U. (1976) *Arch. Biochem. Biophys. 176*, 431–441.
Morgan, W. T., & Smith, A. (1984) *Protides Biol. Fluids 31*, 23–26.
Morris, J. P., Blatt, S., Powell, J. R., Strickland, D. K., & Castellino, F. S. (1981) *Biochemistry 20*, 4811–4816.
Pajot, P. (1976) *Eur. J. Biochem. 63*, 263–269.
Rylatt, D. B., Sia, D. Y., Mundy, J. P., & Parish, C. R. (1981) *Eur. J. Biochem. 119*, 641–646.
Segrest, J. P., & Jackson, R. L. (1972) *Methods Enzymol. 28*, 54–63.
Skoza, L., & Mohos, S. (1976) *Biochem. J. 159*, 457–462.
Soininen, R., & Ellfolk, N. (1973) *Acta Chem. Scand. 27*, 35–46.
Spencer, R. L., & Wold, F. (1969) *Anal. Biochem. 32*, 185–190.

Theory for the Folding and Stability of Globular Proteins†

Ken A. Dill
Departments of Pharmaceutical Chemistry and Pharmacy, University of California, San Francisco, California 94143
Received May 3, 1984

ABSTRACT: Using lattice statistical mechanics, we develop theory to account for the folding of a heteropolymer molecule such as a protein to the globular and soluble state. Folding is assumed to be driven by the association of solvophobic monomers to avoid solvent and opposed by the chain configurational entropy. Theory predicts a phase transition as a function of temperature or solvent character. Molecules that are too short or too long or that have too few solvophobic residues are predicted not to fold. Globular molecules should have a largely solvophobic core, but there is an entropic tendency for some residues to be "out of place", particularly in small molecules. For long chains, molecules comprised of globular domains are predicted to be thermodynamically more stable than spherical molecules. The number of accessible conformations in the globular state is calculated to be an exceedingly small fraction of the number available to the random coil. Previous estimates of this number, which have motivated kinetic theories of folding, err by many tens of orders of magnitude.

Few heteropolymers are both globular and soluble. Proteins are the principal exception. Globularity and solubility cannot be achieved with any random sequence of monomers; certain principles of structure and function must be obeyed (Kauzmann, 1959; Fisher, 1964; Flory, 1969; Tanford, 1968; Brandts, 1968; Edsall, 1968; Edsall & McKenzie, 1983; Lifschitz, 1968; Volkenstein, 1970; Richards, 1977; Klapper, 1971, 1973). A molecule with too many solvophilic residues will prefer solvation to globularity. Molecules with too many solvophobic residues will aggregate, as occurs with oil in water. In addition, globularity requires that the chain can pack well in the condensed state. Typical globular proteins have densities approaching those of crystalline hydrocarbons and amino acids and compressibilities a factor of 20 smaller than liquid hydrocarbons, nearly equal to those of some metals, and they contain less than 3 vol % of internal water or cavities (Richards, 1974, 1977; Klapper, 1971, 1973; Chothia, 1975; Kuntz & Kauzmann, 1974; Connolly, 1981; Gavish et al., 1983; Sturtevant, 1977; Nemethy et al., 1981). The importance of packing also follows from the fact that evolution conserves residue size and shape (Schultz & Schirmer, 1979). But high density comes at a high price; enormous configurational entropy must be overcome to achieve it. The fact that most enzymes are condensed suggests that catalytic function may require the high density state. It is a reasonable hypothesis that this is due to the requirement that the atoms of the active site have relatively invariant spatial positions during a significant fraction of the time required to attract and hold the substrate for the catalytic act. In this regard, the primary molecular mechanism for maintaining relative spatial invariance, i.e., for reducing the amplitude of out-of-phase internal thermal motion, is that of steric constraint, which is achieved, as in the solid state, through high density packing. In the solid state, incident thermal energy may be distributed in modes of motion whose spatial wavelengths are larger than the size of the active site, and thus, this thermal energy may be exchanged with the protein through relatively nondisruptive rigid body motions of the active site.

For those sequences which satisfy the requirements above, it follows that relatively few spatial conformations are available in the globular state, their number being limited by (i) the

† Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this work and to the National Institute of General Medical Sciences.

0006-2960/85/0424-1501$01.50/0

degree to which water–insoluble residues must cluster to avoid solvent contact and (ii) the degree to which the volume occupied by some chain segments is unavailable for others and prohibits their conformational freedom. These constitute the primary interactions responsible for the folding of globular proteins; other interactions such as those due to ionic and hydrogen bonding, while not negligible, are of secondary importance, for they are little changed in the transition from the random coil to the globular states (Kauzmann, 1959; Schellman, 1955; Tanford, 1962, 1968, 1970; Brandts, 1968; Privalov, 1979; Kyte & Doolittle, 1982). The principal purpose of the present work is to calculate the free energies of the molecular configurations of heteropolymers and to identify those states whose free energies are smallest. We adapt lattice statistical mechanics for this purpose.

THEORY

Consider an ensemble of all possible spatial conformations of a linear heteropolymer molecule of given monomer sequence. Each molecule contains n monomers of two types: $n_h = n\phi_h$ of the monomers are solvophobic and $n_p = n\phi_p = n(1 - \phi_h)$ are solvophilic. For the calculation of thermodynamic state functions characterizing the folding of a molecule from the random coil to the globular state, we are at liberty to construct a fictitious intermediate state such that folding is characterized by the following two-step process: (I) condensation whereby the density of the chain segments increases, the relative spatial positions of the segments remaining random, and then (II) reconfiguration of the chain in the condensed state so that solvophobic residues largely occupy the interior core of the molecule. These processes are discussed in turn below.

(I) Random Condensation. For this process, the free energy depends on the segment density, or radius of gyration. The statistical weight of chain configurations with radius of gyration between s and $s + ds$ is given by (Sanchez, 1979)

$$\Omega_I(s)\, ds = z^{n-1} P_0(s)\, \omega_{steric}(s)\, \omega_{contact}(s)\, ds \qquad (1)$$

z is the number of rotational isomeric states available to each bond pair, z^{n-1} being the total number of configurations of the molecule. (With negligible error for long chains, $n - 1$ is replaced by n in subsequent uses of this expression.) The internal rotational states are assumed to be of equal energy; specific interactions among neighboring residues along the chain such as those responsible for secondary structure are neglected. The distribution function

$$P_0(s) = \left(\frac{343}{15}\right)\left(\frac{14}{\pi\langle s^2\rangle_0}\right)^{1/2}\left(\frac{s^2}{\langle s^2\rangle_0}\right)^3 \exp\left(\frac{-7s^2}{2\langle s^2\rangle_0}\right) \qquad (2)$$

is due to Flory and Fisk (Flory & Fisk, 1966; Sanchez, 1979) and represents the fraction of configurations of long chains which are within the specified range of segment densities, and $\langle s^2\rangle_0$ is the mean squared radius of gyration of the randomly coiled molecule.

The factor $\omega_{steric}(s)$ accounts for the reduction in number of configurations due to the volume excluded to some chain segments by others and is given in the Flory approximation (Flory, 1953) by

$$\omega_{steric}(s) = \frac{\left(\frac{\rho_s}{n}\right)^n \left(\frac{n}{\rho_s}\right)!}{\left[\frac{n}{\rho_s} - n\right]!} \qquad (3)$$

where ρ_s is the local volume fraction of space occupied by chain segments of a molecule. The lattice treatment from which this expression derives requires partitioning of the chain into segments which are isodiametric (Flory, 1953, 1970). In compliance with this requirement, one chain segment here corresponds to approximately 1.4 amino acid residues. The amino acids in a globular protein can be represented as occupying cubic volumes whose edges range in length from 4.0 to 6.2 Å, with an average of 5.3 Å (Richards, 1977). The same interresidue separation would be predicted from atomic radial distribution functions around amino acid α-carbons; those functions have a broad maximum at approximately 5 Å (Crippen & Kuntz, 1978). A chain is thus required to be partitioned into segments whose center to center separation is approximately 5.3 Å. The separation between α-carbons is 3.8 Å; hence, the ratio of 1.4.

The statistical weight, $\omega_{contact}(s)$, in eq 1 is required to take into account the different free energies of contact of the different chain configurations. In some configurations, solvophobic residues that are not nearest neighbors along the chain will be adjacent to each other; these will be of lower free energy than configurations in which solvophobic residues are more extensively exposed to solvent. The statistical weight will depend on the number of these favorable contacts. This number may be calculated approximately through use of a spatial lattice, upon which the chains are considered to be configured so that neighboring monomers occupy contiguous lattice sites. The lattice sites may also contain solvent, provided $\rho_s \neq 1$. We consider a spherical lattice (Fisher, 1964; Dill & Flory, 1981) with $m = n/\rho_s$ sites of volume each equal to that of a chain segment. Each lattice site has q neighbors. The simple cubic lattice ($q = 6$) has previously been adopted for treatment of proteins (Crippen, 1974) and spherical micelles (Dill & Flory, 1981), but the character of the lattice is of little consequence here inasmuch as the subsequent predictions are relatively independent of q. To approximate the interfacial character of the molecule, the sphere is partitioned into a surface, or exterior, region (e) and an interior region (i). The fraction of sites that are at the surface is (Fisher, 1964; Dill & Flory, 1981)

$$f_e = \frac{m_e}{m} = \frac{3r^2 - 3r + 1}{r^3} \qquad (4)$$

where the radius, r, is given by

$$r = \left(\frac{3n}{4\pi\rho_s}\right)^{1/3} \qquad (5)$$

The fraction of interior sites is

$$f_i = \frac{m_i}{m} = 1 - f_e \qquad (6)$$

Let n_{xy} represent the number of residues of type x (x = h, p) situated in region y (y = i, e), and let

$$\Psi_{xy} = \frac{n_{xy}}{n_x} \qquad (7)$$

Conservation of residues of each type requires

$$\Psi_{xe} + \Psi_{xi} = 1 \qquad (8)$$

Conservation of volume in each region requires that the number of residues in region, y, n_y, is

$$n_y = n_{hy} + n_{py} \qquad (9)$$

For the purpose of computing the contact statistical weight,

chain segments are assumed to be distributed uniformly throughout the sphere; thus, eq 9 becomes

$$\Psi_{hy}\Phi_h + \Psi_{py}\Phi_p = f_y \ (y = i,e) \tag{10}$$

The following constraints then apply:

$$\Psi_{hi} \leq \min(1, f_i/\Phi_h) \qquad \Psi_{he} \geq \max(0, 1 - f_i/\Phi_h)$$
$$\Psi_{pi} \geq \max[0, (f_i - \Phi_h)/\Phi_p] \tag{11}$$
$$\Psi_{pe} \leq \min[1, (1 - f_i)/\Phi_p]$$

The statistical weight, $\omega_{contact}(s)$, will differ for different chain conformations and sequences. In order to compute this statistical weight, we make two approximations: (i) The first is the random copolymer approximation, according to which the number of favorable contacts is taken to be independent of sequence or is averaged over all viable sequences and thus depends only on the composition, i.e., on the fraction of residues which are solvophobic. Viable sequences are those that are capable of being configured to the specified density and residue distribution. (ii) We adopt the Bragg–Williams approximation (Hill, 1960) that the spatial distribution of residues is independent of conformation of the chain and depends only on the mean segment density. Inasmuch as the segment distribution may differ in the two regions i and e, the Bragg–Williams approximation is applied to each region independently. In accord with these assumptions, specific interactions are neglected; $\omega_{contact}(s)$ takes into account only the average interactions among spatial neighbors and solvent. The statistical mechanical approximations adopted herein are thus not valid for the description of a specific native globular structure and apply only to disordered states of the ensemble.

Subject to the approximations above, the probability that a given site adjacent to a solvophobic segment is occupied by another solvophobic segment in the same region (i or e) is given by the volume fraction of such segments (Flory, 1953; Hill, 1960). Because there are $q - 2$ noncovalent neighbors and n_{hi} such segments, the total number of favorable contacts in the interior region of the molecule is

$$N_{ci} = \frac{(q-2)n_{hi}}{2}\left(\frac{n_{hi}}{m_i}\right) \tag{12}$$

where the factor of 2 is required to avoid double counting of contacts. Similarly for the surface region

$$N_{ce} = \frac{\sigma(q-2)n_{he}}{2}\left(\frac{n_{he}}{m_e}\right) \tag{13}$$

where σq represents the number of faces of a surface lattice site which are not adjacent to solvent. For the simple cubic lattice, $\sigma = {}^5/_6$. The curvature inherent in the present lattice will reduce this value somewhat. Richards has shown that the solvent-accessible surface area of globular proteins is approximately twofold greater than that of the equivalent sphere (Richards, 1977); therefore, for calculations herein we have used $\sigma = {}^2/_3$. Altering this value does not change the general conclusions drawn below; its principal effect is through a small destabilization of the folded state as σ is increased.

We assume each favorable contact has a free energy $-g$ relative to the solvated state. The value $-(q-2)g$ therefore represents the free energy of transfer at the temperature of interest of a solvophobic residue (in a chain) from the solvent to a medium consisting of pure solvophobic residues. For the treatment of water-soluble proteins, $-(q-2)g$ thus describes the transfer of hydrophobic residues from water to a hydrophobic environment; for the treatment of membrane proteins, $-(q-2)g$ characterizes the transfer of polar or charged residues from an apolar medium to a polar environment. The statistical weight due to such contacts is

$$\omega_{contact}(s) = \exp[g(N_{ci} + N_{ce})/(k_BT)] \tag{14}$$

where k_BT is Boltzmann's constant multiplied by absolute temperature. Through use of eq 7, 8, 10, and 12–14

$$\omega_{contact}(s) = \exp\left[\frac{n\epsilon\Phi_h^2\rho_s}{2}\left(\frac{\Psi_{hi}^2}{f_i} + \frac{\sigma(1-\Psi_{hi})^2}{f_e}\right)\right] \tag{15}$$

where $\epsilon = (q-2)g/(k_BT)$. The reference state is that of the random coil, in which all residues are nearly fully solvated and for which $\omega_{contact} \simeq 1$.

The configurational free energy of molecules of density ρ_s is

$$F_I(\rho_s,\Phi_h,n) = -k_BT \ln \Omega_I(\rho_s,\Phi_h,n) \tag{16}$$

By use of the relation (Sanchez, 1979)

$$\frac{s^2}{\langle s^2\rangle_0} = \left(\frac{\rho_0}{\rho_s}\right)^{2/3} \tag{17}$$

which defines ρ_0, and eq 1–3, 15, and 16, the free energy of random condensation ($\Psi_{hy} = f_y$; y = i, e) of the heteropolymer becomes

$$\frac{F_I(\rho_s) - F_I(\rho_0)}{nk_BT} = \frac{-\epsilon\Phi_h^2\rho_s}{2}(f_i + \sigma f_e) + \left(\frac{1-\rho_s}{\rho_s}\right)\ln(1-\rho_s) + 1 + \left(\frac{7}{2n}\right)\left[\left(\frac{\rho_0}{\rho_s}\right)^{2/3} - 1\right] \tag{18}$$

It is also useful to compute the free energy difference between the condensed state ($\rho_s = 1$) and the equilibrium state ($\rho_s = \rho_s^*$):

$$\frac{F_I(1) - F_I(\rho_s^*)}{nk_BT} = \frac{-\epsilon\Phi_h^2}{2}(1-\rho_s^*)(f_i + \sigma f_e) - \left(\frac{1-\rho_s^*}{\rho_s^*}\right)\times \ln(1-\rho_s^*) + \left(\frac{7}{2n}\right)\left[\rho_0^{2/3} - \left(\frac{\rho_0}{\rho_s^*}\right)^{2/3}\right] - \left(\frac{2}{n}\right)\ln\rho_s^* \tag{19}$$

The equilibrium density, ρ_s^*, is determined by the condition

$$\left.\frac{\partial F_I(\rho_s)}{\partial\rho_s}\right|_{\rho_s^*} = \frac{-\epsilon\Phi_h^2}{2}(f_i + \sigma f_e) - \frac{7\rho_0^{2/3}}{3n(\rho_s^*)^{5/3}} - \left(\frac{1}{\rho_s^*}\right)\left[\frac{\ln(1-\rho_s^*)}{\rho_s^*} + 1 - \frac{2}{n}\right] = 0 \tag{20}$$

where $\rho_0 \leq \rho_s^* \leq 1$.

For homopolymers ($\Phi_h = 1$), when surface effects are neglected ($f_i = 1, f_e = 0$), this treatment of condensation reduces to that of Sanchez (Sanchez, 1979; Sun et al., 1980). Inasmuch as the segment density changes sharply as a function of ϵ, or equivalently of temperature, a transition is predicted from the random coil to the condensed state. This solvent-induced, or "coil–globule", transition, is to be distinguished from the helix–coil transition that arises from interactions among near neighbors along the chain (Zimm & Bragg, 1959). The transition is predicted to be of second order provided that $\rho_0 = [19/(27n)]^{1/2}$, which specifies that the critical point occurs at $s^2 = \langle s^2\rangle_0$ as $n \to \infty$ (Sanchez, 1979). This condition specifies that, in the θ state ($\epsilon = 1$), long-chain molecules will adopt random coil conformations ($\rho_s = \rho_0$). A first-order transition is predicted by other values of ρ_0 (Post & Zimm, 1979; Sanchez, 1979). The order of the homopolymer collapse

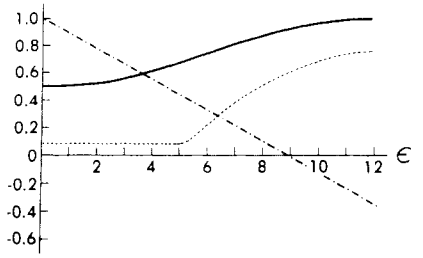

FIGURE 1: Predicted properties vs. ϵ, for $n = 100$ and $\Phi_h = 0.50$. (i) $\Delta F_{fold}/(nk_BT)$ (···), (ii) ρ_s (---), (iii) $\Psi_{hi}\Phi_h/f_i$, the fractional occupancy of interior sites by h residues (—).

transition is the subject of some dispute, the resolution of which will require a more rigorous approach than the mean-field treatment presented here [for additional discussion, see also de Gennes (1975), Stockmayer (1960), Lifschitz (1968), and Moore (1977)]. However, the purpose at hand is to calculate the free energy of folding, which is nearly independent of ρ_0 over a 30-fold range; thus, little error is incurred herein by the adoption of this expression for ρ_0.

(II) Chain Reconfiguration. Condensed heteropolymers of random organization may undergo further reduction in free energy by adopting configurations in which solvophobic segments largely reside in the molecular interior. For this process of reconfiguration, the gain in free energy due to additional favorable contacts may be computed by using eq 15 with $\rho_s = 1$. Entropy disfavors this ordered arrangement, however (Brandts, 1968; Volkenstein, 1977). The a priori probabilities for the various distributions of residues h and p in regions i and e are given approximately, as for the mixing of two solutes in two solvents, through use of binomial statistics:

$$\Omega_{II} = (f_i)^{n_{hi}+n_{pi}}(f_e)^{n_{he}+n_{pe}}\left(\frac{n_h!\ n_p!}{n_{hi}!\ n_{he}!\ n_{pi}!\ n_{pe}!}\right) \quad (21)$$

The use of this distribution is commensurate with classical approximations of chain statistics used previously herein. Thus, chain connectivity is taken into account in the following sense. The conditional probability that a site in a specific region is accessible to segment k, given that segment $k-1$ is located in a particular region, is approximated by the volume fraction of sites available in the region in which k is to be located. This approximation is obviously poor, however, if a significant fraction of the chain segments are located more than one layer away from the i/e interface, for then correlations should be important and the conditional probabilities should deviate significantly from the unconditional probabilities. This error should be small for the molecules of interest here: the error should increase with molecular size and will depend on the fraction of sites without access to the i/e interface, which is only 15% for molecules of 1200 amino acids, for example. Use of Stirling's approximation and eq 7, 8, and 10 lead to

$$\omega_{re} = \Omega_{II}^{1/n} = \left(\frac{f_e}{\Psi_{he}}\right)^{\Phi_h\Psi_{he}}\left(\frac{f_e}{\Psi_{pe}}\right)^{\Phi_p\Psi_{pe}}\left(\frac{f_i}{\Psi_{pi}}\right)^{\Phi_p\Psi_{pi}}\left(\frac{f_i}{\Psi_{hi}}\right)^{\Phi_h\Psi_{hi}} \quad (22)$$

In the random configurations, $\Psi_{xy} = f_y$ where y = i, e and x = h, p, and ω_{re} ($\Psi_{xy} = f_y$) = 1. If a system has "perfect" reconfigurational order, $\Psi_{hi} = 1$, $\Psi_{pe} = 1$, and $\Psi_{he} = 0$, $\Psi_{pi} = 0$, then $\omega_{re} = \phi_h^{\phi_h}\phi_p^{\phi_p}$. The total free energy of rearrangement is given by the Boltzmann relation, $F = -k_BT \ln(\omega_{re}\omega_{contact})$, and eq 15 and 22:

$$\frac{F_{II}(\Psi_{hi}^*) - F_{II}(f_i)}{nk_BT} = -\Phi_h\left\{\Psi_{he}^* \ln\left(\frac{f_e}{\Psi_{he}^*}\right) + \Psi_{hi}^* \ln\left(\frac{f_i}{\Psi_{hi}^*}\right)\right\} - \Phi_p\left[\Psi_{pe}^* \ln\left(\frac{f_e}{\Psi_{pe}^*}\right) + \Psi_{pi}^* \ln\left(\frac{f_i}{\Psi_{pi}^*}\right)\right] - \frac{\epsilon\Phi_h^2}{2}\left\{\frac{(\Psi_{hi}^*)^2 - f_i^2}{f_i} + \frac{\sigma[(1-\Psi_{hi}^*)^2 - f_e^2]}{f_e}\right\} \quad (23)$$

The equilibrium values of Ψ_{xy}, denoted by asterisks, are those which satisfy constraint eq 11 and the condition that

$$\left.\frac{\partial F_{II}}{\partial \Psi_{hi}}\right|_{\Psi_{hi}^*} = \ln\left(\frac{\Psi_{hi}^*\Psi_{pe}^*}{\Psi_{he}^*\Psi_{pi}^*}\right) - \epsilon\Phi_h^2\left(\frac{\Psi_{hi}^*}{f_i} - \frac{\sigma\Psi_{he}^*}{f_e}\right) = 0 \quad (24)$$

For $\epsilon \to \infty$

$$\Psi_{hi}^* = \frac{f_i - \delta}{\Phi_h} \qquad \Psi_{pi}^* = \frac{\delta}{\Phi_p}$$
$$\Psi_{he}^* = \frac{\Phi_h - f_i + \delta}{\Phi_h} \qquad \Psi_{pe}^* = \frac{\Phi_p - \delta}{\Phi_p} \quad (25)$$

where $\delta \ll 1$ is given through substitution of eq 25 into eq 24:

$$\delta = \left(\frac{f_i\Phi_p}{\Phi_h - f_i}\right)\exp\left\{-\epsilon\left[1 - \sigma\left(\frac{\Phi_h - f_i}{f_e}\right)\right]\right\} \quad (26)$$

The total free energy of folding is defined as the sum of the free energies given by eq 19 and 23:

$$\Delta F_{fold} = F_I(1) - F_I(\rho_s^*) + F_{II}(\Psi_{hi}^*) - F_{II}(f_i) \quad (27)$$

Predictions and Conclusions

The free energy of folding, given by eq 27, is a function of chain length, n, the fraction of residues which are hydrophobic, Φ_h, and ϵ, the free energy of association among solvophobic residues. Figure 1 shows the predicted dependence of ΔF_{fold} on ϵ. For small ϵ, the dominant contribution to the free energy is due to condensation (I); for $\epsilon \gtrsim 8$, the dominant contribution is due to reconfiguration (II). Taken together, the total free energy is nearly linear in ϵ over a relatively wide range (see Figure 1). The increase in ϵ along the abscissa of Figure 1 may be taken to represent the decrease in concentration of some denaturing agent such as guanidine hydrochloride (Gdn·HCl) or urea at fixed temperature. The predicted linear dependence of the free energy of folding on denaturant concentration, shown in Figure 1, has been observed experimentally (Green & Pace, 1974; Pace & Vanderburg, 1979; Schellman & Hawkes, 1980). Note that standard denaturants such as 8 M Gdn·HCl or 6 M urea are far from θ ($\epsilon = 1$) solvents (Creighton, 1979); they are approximately represented by $\epsilon = 7$. For the present purposes, it is of interest to consider a typical aqueous solvent, represented by a specific value of ϵ. In principle, for a given solvent and temperature, the value of ϵ can be obtained from free energy of transfer experiments. In practice, ϵ cannot be determined directly from solute transfer experiments (Nozaki & Tanford, 1971; Wolfenden, 1983), for the following implicit errors may not be negligible (Richards, 1974; Klapper, 1973; Karplus, 1980): (i) the free energy of transfer is assumed to be the same for isolated residues and those which are covalently linked in the chain, (ii) transfer experiments are between solvents in the liquid or gas phase, but the protein has the density of a solid, and (iii) transfer of a residue to an interfacial region such as a protein

interior is assumed to be identical with that of a bulk medium, yet free energies of solubilization can differ by a factor of 3 between interfacial and bulk solvents (Tanford, 1979; Lee, 1983).

For the predictions below, we therefore consider ϵ to be a semiempirical parameter and adopt the value $\epsilon = 10$ as plausible on the following grounds. In accord with the conventions used above, the globular state will be more stable than the unfolded state when $\Delta F_{fold} < 0$. For $n = 100$, and $\phi_h = 0.5$, the globular state is predicted to be stable if $\epsilon > 9$ (see Figure 1); $\epsilon = 10$ corresponds to a net stability of approximately 9 kcal/mol, which is in the range observed for typical globular proteins by differential scanning calorimetry experiments (Tanford, 1968; Privalov, 1979; Pace, 1975). The value of $\epsilon = 10$ corresponds to the transfer of butane or pentane from water to the interior of a micelle, the transfer of toluene from water to the pure liquid, or the transfer of propane or butane from water to CCl_4, at 25 °C (Tanford, 1980). It is, however, approximately 30% greater than the free energy of transfer of tryptophan from water to methanol (Nozaki & Tanford, 1971) and is about 40% larger than that observed from residue distributions in proteins for the transfer of an average solvophobic residue to a medium of identical residues (R. Jernigan, personal communication) when the relative sizes of lattice segments and residues are taken into account. In the present model, the requirement that ϵ exceed the measured free energy of transfer by 40% can be attributed largely to two factors: (i) other intramolecular interactions are neglected here, and cooperative hydrogen-bonded units, for example, tend to favor the folded state (Schellman, 1955; Kauzmann, 1959), and (ii) the van der Waals–Flory approximation for excluded volume leads to overestimation of the magnitude of the entropy of random condensation by an estimated 10–20% (Gordon et al., 1976; Flory, 1982).

Two variables characterize the degree of folding: ρ_s^*, the equilibrium chain segment density of the unfolded state, which specifies the state of the system along pathway I, and Ψ_{hi}^*, the degree of distributional order in the condensed state, which specifies the state along pathway II. Both quantities increase with ϵ (see Figure 1). For $\epsilon = 10$, $n = 100$, and $\Phi_h = 0.5$, taken here to be representative of a typical small protein, we make two observations (see Figure 1). First, the equilibrium segment density of the unfolded molecule is predicted to be significantly higher than that of the random coil state. (We adopt the term "unfolded" to refer to the state $\rho_s = \rho_s^*$ and "random coil" to refer to the θ state, $\rho_s = \rho_0$.) A density difference in this direction is observed by hydrodynamic measurements (Tanford, 1968; Brandts, 1968; Tanford & Aune, 1970). Even at this high density, however, the unfolded chain will be highly solvated. Second, the globular state, more stable than the unfolded state under these conditions, is characterized by a distribution in which solvophilic residues largely surround a core of solvophobic residues, in agreement with classical observations (Kauzmann, 1959; Fisher, 1964; Tanford, 1968; Nemethy et al., 1981). However, it is noteworthy that this separation of residues is predicted to be incomplete; a small percentage of the residues should be in their nonpreferred regions. This is a consequence of the reconfiguration entropy, which drives the system toward distributional disorder. Under the conditions specified above, for example, solvophobic residues outnumber interior sites; approximately 40% of the solvophobic residues are predicted to be at the surface, and it is interesting that approximately two solvophilic residues are predicted to be buried in the solvophobic core. In general, it is more stabilizing to add a hydrophobic residue at the protein surface than to add a polar residue to the protein interior; in both cases, the added chain segment contributes destabilizing entropy but only in the former case is the free energy reduced due to hydrophobic contacts.

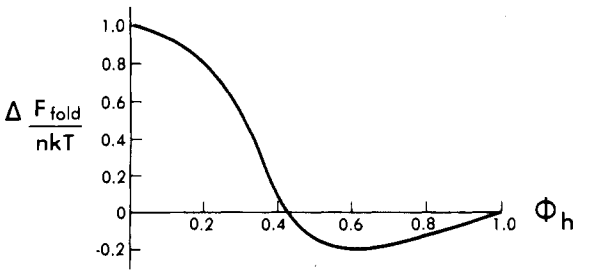

FIGURE 2: Free energy of folding vs. fraction of residues which are solvophobic, for $n = 100$ and $\epsilon = 10$. For these conditions, the globular state is stable if more than 42% of the residues are solvophobic; otherwise, the unfolded state is preferred.

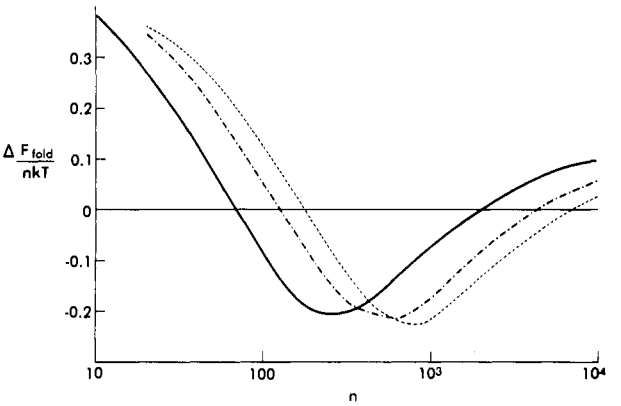

FIGURE 3: Free energy of folding vs. chain length for a spherical molecule (—), or for molecules with "domains": two spheres (- - -) and three spheres (- - -); $\Phi_h = 0.45$ and $\epsilon = 10$. A spherical molecule with fewer than 70 or more than 2000 residues is predicted to be unfolded; chains of intermediate length are predicted to be stable. With increasing chain length, domains are predicted to be more stable than a spherical shape.

The stability of the globular state depends on the composition and length of the chain. For a given chain length, theory predicts that there is an optimal fractional hydrophobicity, Φ_h, for maximum stability (see Figures 2 and 4). The unfolded state is preferred if too few of the residues are solvophobic, for there is too little driving force to overcome the configurational entropy. Stability decreases relative to the optimal value as $\Phi_h \rightarrow 1$ because solvophobic residues in excess of the number required to fill the core must be distributed at the molecular surface in contact with solvent. These predictions are supported by the observations that most soluble globular proteins have fractional hydrophobicities in the range 35–80%, depending on the criteria by which residues are partitioned into solvophobic and solvophilic classes (Kauzmann, 1959; Tanford, 1968; Lee & Richards, 1971).

The theory predicts that there is also an optimum chain length for maximum stability (see Figures 3 and 4). Short-chain molecules ($n < 70$, for $\Phi_h = 0.45$) should not fold. If condensed, these molecules would have little interior volume; the small free energy gained through favorable contacts of the surface residues would not be sufficient to overcome the entropy of folding. This prediction is supported by the observation that few short peptides occur in single stable conformations (Wetlaufer, 1981). Exceptions occur, however, when disulfide bonds cross-link the chain or when extensive secondary structure is prevalent, for the conformational entropy is thereby reduced (Holladay & Puett, 1976). Note that the

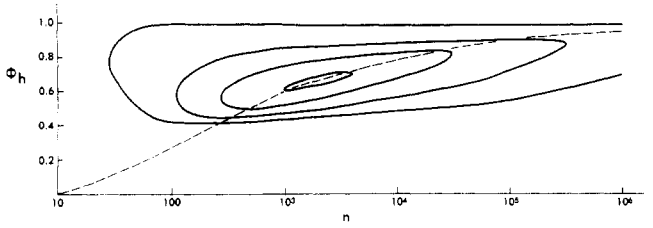

FIGURE 4: Contour plot of $\Delta F_{fold}/(nk_BT)$ vs. n and Φ_h, summarizing results from Figures 2 and 3. Reduced free energy, from the outside to the inside, contours represent 0, −0.2, −0.4, and −0.6; greatest stabilities of folded molecules toward the center, unfolded beyond the outermost contour. For these calculations, $\epsilon = 10$, and the molecule is assumed to be spherical. The dashed line shows the dependence of f_i on n. The fact that the minimum free energy region coincides with the dashed line predicts that, for long enough chains, the greatest stability of the folded state occurs when $\Phi_h \simeq f_i$; i.e., it is advantageous for a spherical protein to be configured so that its surface/volume ratio is not the minimum possible but is equal to the ratio of solvophilic/solvophobic residues.

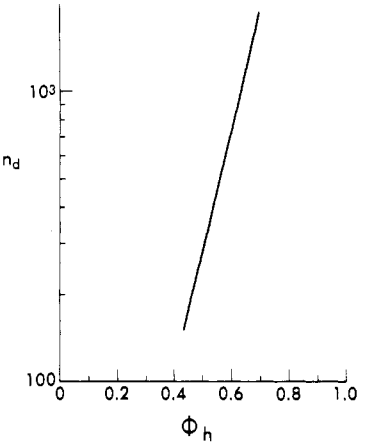

FIGURE 5: Domain size, n_d, is predicted to depend on the fraction of residues which are solvophobic, for $n = 100$ and $\epsilon = 10$.

critical chain length, corresponding to zero free energy of folding, and above which folding is predicted to be favored, should depend on Φ_h and on specific interactions; thus, the value of $n = 70$ cited above does not apply to an specific protein and should be considered only illustrative. Very long chains should be less stable than those of intermediate length, for given $\Phi_h < f_i$, since solvophilic residues would be required to be buried in the interior. Indeed, few globular proteins consist of more than a few thousand residues. The observation that stability is nearly independent of molecular weight (Privalov, 1979) applies to a range of chain lengths too narrow to test the present hypothesis.

A protein will adopt a geometric structure for which its free energy is a minimum. That structure will not necessarily have a minimum surface/volume ratio. The theory predicts that greatest stability can be achieved for those surface/volume ratios, as established by the chain length for fixed geometry, which are approximately equal to the ratio of the number of solvophilic/solvophobic residues, provided chains are longer than $n = 500$ (see Figure 4). For shorter chains, it is advantageous to have enough solvophobic residues to condense, even if many of them must be at the surface [see Figure 4; in the region $70 < n < 500$, the minimum free energies occur for Φ_h greater than the value required to just fill the core (given by the dashed line)]. This is in accord with the observations that many of the atoms at the surfaces of small proteins ($n < 200$) are solvophobic (Richards, 1977; Lee & Richards, 1971; Janin, 1979; Shrake & Rupley, 1973); the surfaces of small micelles are similarly highly solvophobic (Dill, 1984a; Dill et al., 1984b). The theory predicts that the fraction of solvophobic residues which are buried increases with molecular weight, in agreement with observation (Chothia, 1976). The fractional solvophobicity thus determines the optimal chain length, provided the molecule is spherical. However, a molecule may also alter its geometry or may form globular domains (Wetlaufer, 1973) in order to maximize its stability. Because of this freedom to deviate from a spherical shape to maximize stability, chain length need not be correlated with fractional hydrophobicity in real proteins (Fisher, 1964; Kyte & Doolittle, 1982).

The present theory provides an explanation for the existence of domains within proteins. For the predictions described above, we have assumed that the globular molecule is a single sphere. For comparison, consider the folding of a molecule comprised of n/n_d nonoverlapping spheres, where n_d is the number of residues per spherical domain. The free energy of reconfiguration, process II, is then computed subject to substitution of n_d for n in eq 5. The free energies of folding for n/n_d domains and $(n/n_d) - 1$ domains are then compared; the value of n for which they are equal is taken to establish n_d (see Figure 3). Free energies of folding may be readily calculated for other geometric shapes. It is predicted that for short chains, a single domain is preferred (see Figure 3). For longer molecules of the same fractional hydrophobicity, more domains are preferred. Domain size is predicted to be nearly independent of chain length, in accord with observation (Rose, 1979). Domain size is predicted to increase rapidly with Φ_h (see Figure 5); correspondingly, the number of domains per molecule should decrease as Φ_h increases. In contrast to the view that domains are required for kinetic reasons (Wetlaufer, 1973, 1980, 1981), the principal conclusion here is that, under the circumstances outlined above, structures comprised of domains are thermodynamically more stable than spherical structures.

The present treatment provides a basis for computing the relative molecular contributions to the conformational entropy of folding. The conformational entropy of folding from the random coil conformations to the ensemble of the low free energy, globular states of perfect reconfigurational order ($\Psi_{hi} = \Psi_{pe} = 1$, $\Psi_{he} = \Psi_{pi} = 0$) is predicted to be approximately

$$\Delta S_{fold} = nk_B \ln\left(\frac{\omega_{re}}{a}\right) \qquad (28)$$

For $\Phi_h = 0.5$, $\Delta S_{fold}/n = -3.0$ cal K^{-1} (mol of residues)$^{-1}$, for example. Of greater interest is the entropy of folding from the unfolded state ($\rho_s = \rho_s^*$) to the equilibrium folded state; in this case, the chain conformational entropy is computed from eq 19 and 23 exclusive of the contact free energy terms. For a molecule of length corresponding to the 58 residues of bovine pancreatic trypsin inhibitor (BPTI), the entropy of folding is predicted to range from −1.0 to −1.4 cal K^{-1} (mol of residues)$^{-1}$ for $\Phi_h = 0.5$–0.35, respectively. These values represent a lower bound on the magnitude of the entropy of folding, since the folded "state" of the theory does not represent a single native conformation; it represents an ensemble of configurations of constant Ψ_{xy}. Thus, these values are only illustrative and do not apply to a specific protein. Nevertheless, for comparison, the conformational entropy of folding of BPTI observed by differential scanning calorimetry experiments is −3.2 cal K^{-1} (mol of residues)$^{-1}$ (Privalov, 1979). BPTI has less secondary structure, and is thus more relevant for comparison, than other small globular proteins for which such measurements are available [−4.2 cal K^{-1} (mol of residues)$^{-1}$ is the average for one set of five small proteins (Privalov,

1979)]. Interpretation of the experiments depends on the assumption that no hydrogen bonding or secondary structure contributes to the entropy at the boiling temperature of the solution. The contributions to the entropy of folding from secondary structure and side chain constraints, not taken into account in the present theory, are also likely to be important in the folding of real proteins.

Reversibility of Protein Folding

That a protein can fold has been widely considered paradoxical. On the one hand, classical experiments support the "thermodynamic hypothesis" that some proteins seek and find the conformation of minimum free energy (Anfinsen, 1973). On the other hand, the number of accessible states has been considered to be too large, and phase space too vast, for this search to succeed in a reasonable time (Levinthal, 1968; Wetlaufer, 1981, 1973, 1980; Anfinsen, 1973; Wetlaufer & Ristow, 1973; Anfinsen & Scheraga, 1975; Sternberg & Thornton, 1978; Cantor & Schimmel, 1980). A chain molecule has $N_{rc} = z^{n-1}$ random coil configurations. If the configurations were sampled through random search, the time required to find a specific one would be proportional to $N_{rc}\nu^{-1}$, where ν is the trial frequency. For $n = 100$, this search time would be measured in ages of the universe, inasmuch as z has been taken to be 2 (Anfinsen & Scheraga, 1975), 3 (Cantor & Schimmel, 1980), 4 (Anfinsen, 1973), 5 (Karplus & Weaver, 1976), 9 (Wetlaufer, 1973), or 10 (Sternberg & Thornton, 1978). These estimates of the number of accessible states have motivated "kinetic" theories of folding: "since it is doubtful that excluded volume could reduce the folding time to the right time range, then there must be some initial event [or well-defined sequence of events (Levinthal, 1968)] in the folding process or pathway which directs the folding (Wetlaufer, 1973)." The kinetic view thus holds that certain critical events must be essentially irreversible, characterized by free energy barriers too high to permit the search of many other conformations on the biological time scale. It follows from this hypothesis that the global free energy minimum is not necessarily accessible. Hence, there is a paradox: if the molecule finds the global free energy minimum, how does it succeed so rapidly? According to Chothia, "The central problem at all levels of biological structure is to understand how the intrinsic entropy of its various substances is overcome to form particular stable structures in a finite time..." (Chothia, 1980).

The resolution to the paradox lies in the fact that the above approach overestimates the folding time by many tens of orders of magnitude. There are far fewer accessible globular states than random coil states. Excluded volume is not negligible; it is of overwhelming importance. The factor by which the number of random coil configurations is diminished due to excluded volume in the globular state is approximately $\omega_{steric}(s) = \exp(-n)$, for $\rho_s = 1$ (see eq 3). Thus, only an exceedingly small fraction of phase space is accessible; 10^{-44} of the states are accessible for $n = 100$. The a priori conformational freedom per chain segment, z, is diminished by a factor $a = e = 2.718$ according to the Flory approximation used here. In better approximations, this factor depends somewhat on the lattice coordination number; in the Huggins approximation, for example, $a = 2.25$ for the simple cubic lattice (Kasteleyn, 1963; Flory, 1982; Gordon et al., 1976).

Thus, the time required to fold cannot be identified with that required for random search of all possible conformations. The assumption that the energies of all the rotational isomers are equal, or nearly so, only applies to random coil configurations. It does not apply to globular states for which the overwhelming majority of conformations are prohibited by steric constraints and for which accessible conformations differ greatly in free energy. The accessible states will not be searched randomly, as if phase space were a flat landscape, for the free energies will direct the folding.

The number of accessible globular conformations depends critically on z, whose value is estimated in the next few paragraphs. It is defined by $z = z_{rc}/z_g$ where z_{rc} represents the accessibility of phase space of a bond pair in the random coil state and is given by the partition function

$$z_{rc} = \int_0^{2\pi}\int_0^{2\pi} e^{-E(\Phi,\Psi)}\, d\Phi d\Psi \qquad (29)$$

and z_g represents the corresponding quantity for a "single" conformation in the globular state. Φ and Ψ angles are defined in the usual manner for polypeptides (Flory, 1969) and should not be confused with the subscripted quantities presented earlier. This definition for z differs from that used for the treatment of the helix–coil transition. Many semiempirical and quantum mechanical potentials have been used to calculate the internal energy, E, as a function of bond angles Φ and Ψ (Weiner et al., 1984; Zimmerman et al., 1977; Pullman & Pullman, 1974; Brant et al., 1967). From them, we can obtain z_{rc}. No simulations have yet been performed over the full range of Φ and Ψ for dipeptides in water, however. We adopt the value of Brant et al. (1967) of $z_{rc} = 4118$ deg^2 (Flory, 1971) for L-alanine as being most representative of a dimer in water, inasmuch as that potential energy function correctly predicts the characteristic ratio of polyalanine in dilute solution. More recent semiempirical potentials are less appropriate for our purposes, for they are parameterized for the gas phase in which conformation space is more restricted than it would be for aqueous solutions, since electrostatic interactions and intramolecular hydrogen bonds should be stronger in the former. For example, using the potential of Weiner et al. (1984), we find that the configuration integral is approximately 1125 deg^2 for alanine and 1610 deg^2 for glycine. Note that this integral evaluated by using the gas-phase potential is smaller by a factor of nearly 3.7 than that of the solution potential. Thus, inasmuch as the configuration integral is highly sensitive to the potential parameters, which are not yet well-established for dipeptides in aqueous solution, the value of z_{rc} adopted herein should only be considered to be an approximate upper bound. Although z_{rc} should vary among amino acids, being larger for glycine and smaller for isoleucine and threonine, the value for alanine should not seriously misrepresent an average amino acid.

We adopt the value $z_g = 1600$ deg^2 on the grounds that a conformation of a dipeptide in a globular protein should be topologically "the same" as another dimer of the same constitution if it deviates by no more than ±20 deg ($40^2 = 1600$). Since z_g cannot be reliably calculated at present from a configuration integral such as that of eq 29, this estimate, which is crude at best, has the following basis. Typical root mean square variabilities of Φ and Ψ angles in molecular dynamic stimulations of BPTI are approximately ±15 deg and range from ±10 to ±30 deg (van Gunsteren & Karplus, 1982). This should underestimate $z_g^{1/2}$ in that it represents only high-frequency motions, typical maximum variations over 25 ps are a factor of 6 greater than this (van Gunsteren & Karplus, 1982), and molecular dynamics fluctuations are generally about half those measured in protein crystals (Karplus & McCammon, 1983). Estimates of the changes of Φ and Ψ angles which occur upon refinement of a given X-ray crystallographic structure of a protein range from ±10 to ±30 deg

or more (Ramachandran & Sasisekharan, 1968; Wu & Kabat, 1973; R. M. Stroud, personal communication). Furthermore, secondary structures are identifiable for deviations of Φ and Ψ angles of ± 40 deg from the mean (Chou & Fasman, 1978). Therefore, a plausible criterion is that the conformation of a dimer in the globular state is topologically the same as that of another if Φ and Ψ angles agree to within ± 20 deg or are within approximately the same 1.2% of the total conformation space. Such variations must be assumed to be uncorrelated with others along the protein backbone; even small errors, if correlated along the chain, would give rise to vastly different conformations (Burgess & Scheraga, 1975). Therefore, $z \leq 3.8$ $[(4118/1600)^{1.4}]$; the exponent is required to take into account the relative sizes of lattice segments and amino acids. The value of z is further reduced if a protein has specific secondary structure. Those regions of the chain with strong propensity to adopt helical conformations will be represented by a value of z approaching 1; this characteristic of some chains is in part responsible for their helix–coil transitions.

We conclude that an upper bound on the number of conformations available to a molecule in the globular state is $(z/a)^n = (1.7)^n$ (for $z = 3.8$ and $a = 2.25$). The number of conformations of relatively low free energy is significantly smaller than this. For example, the number of conformations that have the equilibrium distribution of solvophobic and solvophilic residues between interior and exterior sites (Ψ_{xy}^*) is $(z\omega_{re}/a)^n = (1.4)^n$ (for $n = 100$, $\epsilon = 10$, and $\Phi_h = 0.5$). Inasmuch as the value of z can only be crudely estimated at present, little significance should be attached to this number, per se. The important point is that, for molecules which have viable globular structures, the number of accessible states is an exceedingly small fraction of the number of conformations accessible in the random coil state.

A principal conclusion of this work is that irreversibility is not required to account for the folding of at least some proteins. Through a biased reversible search, a protein could readily fold to conformations at or near the global free energy minimum in a time commensurate with biological function. This is not to be construed as an argument that there are not kinetic barriers to protein folding, since any physical process that occurs in finite time must be limited by some kinetic constraint. The present model does not address the kinetics or mechanism of folding. Inasmuch as it provides an upper bound on the number of accessible states, however, it implies that the primary kinetic barriers to folding are not those which might be imposed to avoid the plethora of random coil configurations; rather, the principal barriers are likely to correspond to reconfiguration of the chain through the relatively small number of low free energy condensed states, for which the topology is tortuous. This view is consistent with the evidence that only the fastest folding events (in the submillisecond range) have rates dependent on solvent viscosity; slower folding events are observed to be viscosity independent (Tsong, 1982; Tsong & Baldwin, 1978; Baldwin, 1980).

The validity of these conclusions extends beyond the limitations imposed by the approximations of the present treatment, reiterated here. Conformations refer to different backbone topologies, exclusive of side chains, averaged over vibrational degrees of freedom of a given molecule and over sequences of different molecules. We have assumed the molecule forms a sphere; geometric degrees of freedom are thus not taken into account. The most fundamental assumption of this work is that the chain distribution function is factorable into independent terms. Factorability of the distribution function implies the additivity of contributions to the free energy: (i) due to nearest-neighbor rotational isomeric states, (ii) from the excluded volume constraints, and (iii) from solvent interactions. It is further implied that the state of the system which minimizes the total free energy is that for which each contribution is independently minimized. To the extent that this approximation holds, steric forces reduce the likelihood of all conformations, and solvent forces induce the relocation of all residues, independently of the orientational disposition of a bond relative to its bonded neighbors along the chain.

The separability principle, and the approximation that interactions are nearly independent, has three bases. First, the predictions of the theory are in general accord with experiments. Second, the premise of intramolecular and intermolecular separability has led to widely successful predictions of properties of polymers in amorphous condensed phases (Flory, 1956, 1977, 1979; de Gennes, 1979). The underlying principle is that the relative orientation of bonds connecting adjacent monomers along a chain should be virtually unaffected by even strong interactions among other spatial neighboring molecules or submolecules, provided those neighbors are randomly arrayed and interact nonspecifically, for then those interactions will approximately cancel (Flory, 1979; de Gennes, 1979). That premise is less suitable for native proteins than for amorphous polymers, but it provides a satisfactory basis for the present approach, in which conformations are taken to be averaged over sequences. Third, the configurations of model dipeptides are useful predictors of secondary structures within globular proteins (Anfinsen & Scheraga, 1975; Chou & Fasman, 1978; Nemethy & Scheraga, 1977; Scheraga, 1980; Garnier et al., 1978). The successes, and failures (Kabach & Sander, 1984), of such predictions should be attributed to the degree to which interactions are independent, rather than to the degree to which "short-range forces dominate" and determine folding, as one view has held (Anfinsen & Scheraga, 1975; Scheraga, 1980, 1983). Nearest-neighbor intramolecular forces dominate only in the θ state ($\epsilon = 1$); in the globular state ($\epsilon = 10$), solvent forces are exceptionally strong. Indeed, without solvent forces, chain molecules would have no globular state.

ACKNOWLEDGMENTS

I thank Robert Baldwin, Paul Flory, Peter Kollman, Irwin Kuntz, Isaac Sanchez, Robert Stroud, and Bruno Zimm for helpful comments and Scott Weiner for making his raw data available to me.

REFERENCES

Anfinsen, C. B. (1973) *Science (Washington, D.C.) 181*, 223.
Anfinsen, C. B., & Scheraga, H. A. (1975) *Adv. Protein Chem. 29*, 205.
Baldwin, R. L. (1980) in *Protein Folding* (Jaenicke, R., Ed.) p 369, Elsevier, Amsterdam.
Brant, D. A., Miller, W. G., & Flory, P. J. (1967) *J. Mol. Biol. 23*, 47.
Brandts, J. F. (1968) in *Structure and Stability of Biological Macromolecules* (Timasheff, S. N., & Fasman, G. D., Eds.) p 213, Marcel Dekker, New York.
Brown, K. G., et al. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1467.
Burgess, & Scheraga, H. A. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1221.
Cantor, C. R., & Schimmel, P. R. (1980) *Biophysical Chemistry*, p 296, 299, W. H. Freeman, San Francisco, CA.
Chothia, C. (1975) *Nature (London) 254*, 304.
Chothia, C. (1976) *J. Mol. Biol. 105*, 1.

Chothia, C. (1980) in *Protein Folding* (Jaenicke, R., Ed.) p 583, Elsevier, Amsterdam.
Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251.
Connolly, M. L. (1981) Ph.D. Thesis, University of California, Berkeley, CA.
Creighton, T. E. (1979) *J. Mol. Biol. 129*, 235.
Crippen, G. M. (1974) *J. Theor. Biol. 45*, 327.
Crippen, G. M., & Kuntz, I. D. (1978) *Int. J. Pept. Protein Res. 12*, 47.
deGennes, P. G. (1975) *J. Phys. (Paris) 36*, L-55.
deGennes, P. G. (1979) *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY.
Dill, K. A. (1984) in *Surfactants in Solution* (Mittal, K. L., & Lindman, B., Eds.) Vol. 1, p 307, Plenum Press, New York.
Dill, K. A., & Flory, P. J. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 676.
Dill, K. A., Koppel, D. E., Cantor, R. S., Dill, J. D., Bendedouch, D., & Chen, S. H. (1984) *Nature (London) 309*, 42.
Edsall, J. T. (1968) in *Structural Chemistry and Molecular Biology*, (Rich, A., & Davidson, N., Eds.) p 88, W. H. Freeman, San Francisco, CA.
Edsall, J. T., & McKenzie, H. A. (1983) *Adv. Biophys. 16*, 53.
Fisher, H. (1964) *Proc. Natl. Acad. Sci. U.S.A. 51*, 1285.
Flory, P. J. (1953) *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY.
Flory, P. J. (1956) *Proc. R. Soc. London, Ser. A 234*, 60.
Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, Wiley, New York.
Flory, P. J. (1970) *Discuss. Faraday Soc. 49*, 7.
Flory, P. J. (1971) *Pure Appl. Chem. 26*, 309.
Flory, P. J. (1977) *Ber. Bunsenges. Phys. Chem. 81*, 885.
Flory, P. J. (1979) *Faraday Discuss. Chem. Soc. 68*, 14.
Flory, P. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4510.
Flory, P. J. & Fisk, S. (1966) *J. Chem. Phys. 44*, 2243.
Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97.
Gavish, B., et al. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 750.
Gordon, M., Kapadia, P., & Malakin, A. (1976) *J. Phys. A: Math. Gen. 9*, 751.
Greene, R. F., & Pace, C. N. (1974) *J. Biol. Chem. 249*, 5388.
Hill, T. L. (1960) *Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, MA.
Holladay, L. A. & Puett, D. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 1199.
Janin, J. (1979) *Nature (London) 277*, 491.
Kabsch, W., & Sander, C. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1075.
Karplus, M. (1980) *Biophys. J. 32*, 45.
Karplus, M., & Weaver, D. L. (1976) *Nature (London) 260*, 404.
Karplus, M., & McCammon, J. A. (1983) *Annu. Rev. Biochem. 52*, 263.
Kasteleyn, P. W. (1963) *Physica (Amsterdam) 29*, 1239.
Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1.
Klapper, M. H. (1971) *Biochim. Biophys. Acta 229*, 557.
Klapper, M. H. (1973) *Prog. Bioorg. Chem. 2*, 55.
Kuntz, I. D., & Kauzmann, W. (1974) *Adv. Protein Chem. 28*, 239.
Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105.
Lee, B. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 622.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379.
Levinthal, C. (1968) *J. Chim. Phys. 65*, 44.
Lifschitz, I. M. (1968) *Zh. Eksp. Teor. Fiz. 55*, 2408.
Moore, M. A. (1977) *J. Phys. A: Math. Gen. 10*, 305.
Nemethy, G., & Scheraga, H. A. (1977) *Q. Rev. Biophys. 10*, 239.
Nemethy, G., Peer, W. J., & Scheraga, H. A. (1981) *Annu. Rev. Biophys. Bioeng. 10*, 459.
Nozaki, Y., & Tanford, C. (1971) *J. Biol. Chem. 246*, 2211.
Pace, N. (1975) *CRC Crit. Rev. Biochem. 3*, 1.
Pace, N. & Vanderburg, K. E. (1979) *Biochemistry 18*, 288.
Post, C. B., & Zimm, B. H. (1979) *Biopolymers 18*, 1487.
Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167.
Pullman, B., & Pullman, A. (1974) *Adv. Protein Chem. 28*, 347.
Ramachandran, G. M., & Sasisekharan, V. (1968) *Adv. Protein Chem. 23*, 283.
Richards, F. M. (1974) *J. Mol. Biol. 82*, 1.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151.
Rose, G. (1979) *J. Mol. Biol. 134*, 447.
Sanchez, I. C. (1979) *Macromolecules 12*, 980.
Schellman, J. A. (1955) *C. R. Trav. Lab. Carlsberg Ser. Chim. 29*, 230.
Schellman, J. A., & Hawkes, R. B. (1980) in *Protein Folding* (Jaenicke, R., Ed.) p 331, Elsevier, Amsterdam.
Scheraga, H. A. (1980) in *Protein Folding* (Jaenicke, R., Ed.) p 261, Elsevier, Amsterdam.
Scheraga, H. A. (1983) *Biopolymers 22*, 1.
Schultz, G. E., & Schirmer, R. H. (1979) *Principles of Protein Structure*, Springer-Verlag, New York.
Shrake, A., & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351.
Sternberg, M. J. E., & Thornton, J. M. (1978) *Nature (London) 271*, 15.
Stockmayer, W. H. (1960) *Makromol. Chem. 35*, 54.
Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236.
Sun, S. T., et al. (1980) *J. Chem. Phys. 73*, 5971.
Tanford, C. (1962) *J. Am. Chem. Soc. 84*, 4240.
Tanford, C. (1968) *Adv. Protein Chem. 23*, 121.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1.
Tanford, C. (1979) *Proc. Natl. Acad. Sci. 76*, 4175.
Tanford, C. (1980) *The Hydrophobic Effect*, 2nd ed., Wiley, New York.
Tanford, C., & Aune, K. C. (1970) *Biochemistry 9*, 206.
Tsong, T. Y. (1982) *Biochemistry 21*, 1493.
Tsong, T. Y., & Baldwin, R. L. (1978) *Biopolymers 17*, 1669.
van Gunsteren, W. F., & Karplus, M. (1982) *Macromolecules 15*, 1528.
Volkenstein, M. V. (1977) *Molecular Biophysics*, Chapter 4, Academic Press, New York.
Weiner, S. J., et al. (1984) *J. Am. Chem. Soc. 106*, 765.
Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 697.
Wetlaufer, D. B. (1980) in *Protein Folding* (Jaenicke, R., Ed.) p 323, Elsevier, Amsterdam.
Wetlaufer, D. B. (1981) *Adv. Protein Chem. 34*, 61.
Wetlaufer, D. B., & Ristow, S. (1973) *Annu. Rev. Biochem. 42*, 135.
Wolfenden, R. (1983) *Science (Washington, D.C.) 222*, 1087.
Wu, T. T., & Kabat, E. A. (1973) *J. Mol. Biol. 75*, 13.
Zimm, B. H., & Bragg, J. K. (1959) *J. Chem. Phys. 31*, 526.
Zimmerman, S. S., et al. (1977) *Macromolecules 10*, 1.