Thermal Stabilities of Globular Proteins†

Ken A. Dill,* Darwin O. V. Alonso, and Karen Hutchinson

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143

Received January 10, 1989; Revised Manuscript Received March 8, 1989

ABSTRACT: Statistical thermodynamic theory has recently been developed to account for the stabilities of globular proteins. Here we extend that work to predict the dependence on temperature. Folding is assumed to be driven by solvophobic interactions and opposed by the conformational entropy. The temperature dependence of the solvophobic interaction is taken from the transfer experiments on amino acids by Tanford and Nozaki and on model solutes by Gill and Wadsö. One long-standing puzzle has been why proteins denature upon heating, since the solvophobic force to fold strengthens with increasing temperature. This is resolved by the theory, which predicts two first-order phase transitions. "Cold denaturation" is driven principally by the weakening of the solvophobic interaction, but normal denaturation is driven principally by the gain of conformational entropy of the chain. Predictions of the thermodynamic state functions are in reasonable agreement with the calorimetric experiments of Privalov and Khechinashvili. Comparison of the theory with experiments suggests that there may be an additional enthalpic driving force toward folding which is not due to the solvophobic interaction.

A principal driving force for the folding of globular proteins is the solvophobic interaction (Kauzmann, 1959). This force alone, however, cannot account for the temperature dependence of protein folding (Ben-Naim, 1980; Creighton, 1985; Tanford, 1962) since proteins denature upon heating even though the solvophobic effect strengthens with temperature (Ben-Naim, 1980; Edsall & McKenzie, 1983; Gill & Wadsö, 1976; Gill et al., 1976). This raises the question: what other forces are important for folding?

We develop mean-field statistical thermodynamic theory for the temperature dependence of folding of globular proteins. This treatment shows the importance of two factors: (i) the conformational entropy of the chain and (ii) the heat capacity change of the solvophobic effect. A simple thermodynamic argument shows how the conformational entropy influences the thermal stability (Brandts, 1964). If folding is driven by a negative free energy change of clustering of the solvophobic residues into a globular structure (hϕ) and is opposed by a positive free energy change due to loss of conformational entropy upon folding (conf)

$$\Delta F_{\text{fold}} = \Delta F_{\text{h}\phi} + \Delta F_{\text{conf}} \tag{1}$$

then the temperature dependence of the total free energy of folding will be

$$\frac{\partial \Delta F_{\text{fold}}}{\partial T} = \frac{\partial \Delta F_{\text{h}\phi}}{\partial T} + \frac{\partial \Delta F_{\text{conf}}}{\partial T} \tag{2}$$

The strengthening of the solvophobic effect with temperature implies that the transfer of solvophobic residues from water to a solvophobic medium is characterized by $\partial \Delta F_{\text{h}\phi}/\partial T < 0$. The loss of a large conformational entropy upon folding ($\Delta S_{\text{conf}} \ll 0$) implies that $\partial \Delta F_{\text{conf}}/\partial T \gg 0$, and this conformational entropy contribution can dominate the solvophobic contribution, giving an overall decrease in the stability of the protein with increase in temperature ($\partial \Delta F_{\text{fold}}/\partial T > 0$).

This and recent related arguments (Baldwin, 1986; Becktel & Schellman, 1987) based on thermodynamics do not attempt a quantitative microscopic prediction of the temperature dependence of folding. We develop here a microscopic theory for thermal stability based on a recent statistical thermodynamic treatment of the balance of solvophobic and conformational forces in protein folding (Dill, 1985).

We note an important caveat in the use of the terms "hydrophobicity" and "solvophobicity" here. These terms have generally been applied to at least three situations: (i) transfer processes of nonpolar solutes between nonpolar media and aqueous solutions; (ii) the unusual temperature dependence whereby nonpolar molecules in a water solvent associate more strongly with increasing temperature; (iii) the ordering of water around nonpolar molecules. Some confusion arises since these three meanings are not the same. For example, recent arguments (Privalov & Gill, 1988) are persuasive that situation iii is not the principal driving force to expel nonpolar solutes from water at temperatures higher than 30 °C. Moreover, transfer processes (i) are quite general; they need not have the temperature dependence (ii) and may be driven by the full range of types of interaction, including van der Waals, electrostatics, hydrogen bonding, and others. Here we use the term solvophobicity to describe the experimental transfer process (i), for situations in which the temperature dependence (ii) is observed. We do not ascribe its microscopic origins to any particular specific type of interaction. We suppose only that these interactions arise from contact or near contact of nonpolar solutes (amino acid side chains) in aqueous media and that they are dominated by short-ranged forces.

Theory

We first briefly summarize the basic model. The free energy change upon folding of the protein is to be calculated as a function of temperature. Because only state functions are of interest, we are at liberty to choose fictitious pathways from the unfolded to the folded state. We construct the two pathways (Dill, 1985) depicted in Figure 1: (I) from the unfolded state to a state of "random condensation", wherein solvophobic and solvophilic residues are randomly distributed throughout a condensed globular structure; then (II), a process

† This work was supported by the NIH, the DARPA URI Program, and the Pew Scholars Program.

* Address correspondence to this author at the UCSF—Laurel Heights Campus, 3333 California St., Room 102, Box 1204, San Francisco, CA 94118.

0006-2960/89/0428-5439$01.50/0

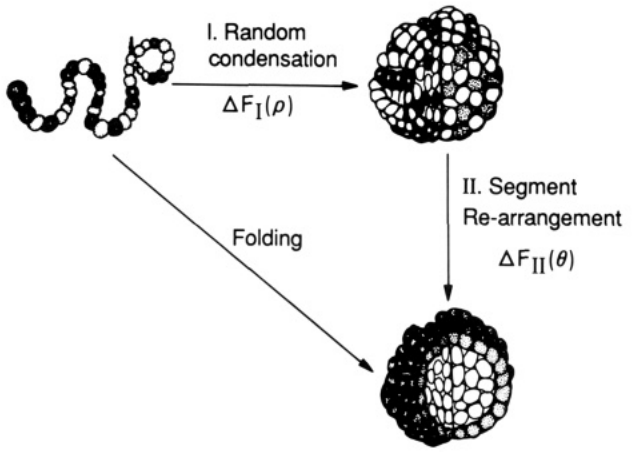

FIGURE 1: Folding model, $\Delta F_{fold} = \Delta F_I(\rho) + \Delta F_{II}(\theta)$.

in which residues in the condensed state are reconfigured so that solvophilic residues largely surround a core of solvophobic residues. Along path I, the thermodynamic functions depend only on "density" ρ, the mean volume fraction of space occupied by the chain segments. Along path II, they depend only on "ordering" θ, described below, which accounts for the segregation of solvophilic and solvophobic segments into inside and outside regions of the protein. ΔF_{fold} is the sum of these free energy changes over the two paths.

Path I: Random Condensation. Consider a protein with n residues, n_h of which are solvophobic and n_p of which are solvophilic. The fractional solvophobicity is given by $\Phi = n_h/n$. At all densities ρ the molecule is taken to be spherical with radius r, where ρ and r are related by

$$\rho = n/\left(\frac{4}{3}\pi r^3\right) \tag{3}$$

As ρ increases, there are fewer conformations available to the chain, largely because of the volume excluded by other segments, and this contributes to an entropy which disfavors collapse. However as ρ increases there are more contacts among the solvophobic residues, and this interaction favors collapse. The solvophobic and conformational contributions to the free energy difference per residue between the randomly condensed state ($\rho = 1$) and the unfolded state ($\rho = \rho^*$) are (see Appendix eq A27; Dill, 1985)

$$\frac{\Delta F_{h\phi\,I}}{nkT} = -\chi(T)\Phi^2(f_i + \sigma f_e)(1 - \rho^*) \tag{4}$$

$$\frac{\Delta F_{conf\,I}}{nkT} = \frac{-(1-\rho^*)}{\rho^*}\ln(1-\rho^*) + \frac{7}{2n}[\rho_0^{2/3} - (\rho_0/\rho^*)^{2/3}] - \frac{2}{n}\ln\rho^* \tag{5}$$

where kT is Boltzmann's constant multiplied by absolute temperature, $(1 - \sigma)$ is the fractional contact of a surface residue with the solvent, and $f_i(r)$ and $f_e(r)$ are the fractions of the spherical volume in the unfolded state which correspond to interior and exterior sites, respectively (see Appendix). The reference-state density is $\rho_0 = [19/(27n)]^{1/2}$ (Dill, 1985). The strength of the solvophobic interaction is characterized by $\chi(T)$, which is the negative logarithm of the solubility of isolated solvophobic amino acid side chains in water. It is described in detail in the Appendix. The change in free energy along path I is given by the sum of these two contributions:

$$\Delta F_I = \Delta F_{h\phi\,I} + \Delta F_{conf\,I} \tag{6}$$

The equilibrium density (ρ^*) of the unfolded state is calculated by finding the minimum in free energy by solving the equation

$$0 = \left.\frac{\partial(F_I/nkT)}{\partial\rho}\right|_{\rho^*} = -\Phi^2\chi(T)(f_i + \sigma f_e) - \frac{7}{3n}\frac{\rho_0^{2/3}}{\rho^{*5/3}} - \frac{1}{\rho^*}\left[\frac{\ln(1-\rho^*)}{\rho^*} + 1 - \frac{2}{n}\right] \tag{7}$$

Path II: Chain Reconfiguration. From the randomly condensed state, an even greater number of favorable solvophobic contacts will be made if the chain is free to reconfigure so that solvophilic residues are largely on the outside of the molecule surrounding a core of solvophobic residues. However, this process must compete with the decrease in entropy due to ordering, which arises from segregating the two classes of residues into the two spatial regions of the molecule. The variables characterizing this process are θ, which is the fraction of surface sites occupied by solvophobic residues, and x, which is the fraction of core sites occupied by solvophobic residues. They are constrained by the condition that

$$\theta f_e + x f_i = \Phi \tag{8}$$

The contact and conformational contributions to the free energy change of reconfiguration are (Appendix eq A39; Dill, 1985)

$$\frac{\Delta F_{h\phi\,II}}{nkT} = -\chi(T)[f_i(x^{*2} - \Phi^2) + \sigma f_e(\theta^{*2} - \Phi^2)] \tag{9}$$

$$\frac{\Delta F_{conf\,II}}{nkT} = f_i\left[x^*\ln\frac{x^*}{\Phi} + (1-x^*)\ln\frac{1-x^*}{1-\Phi}\right] + f_e\left[\theta^*\ln\frac{\theta^*}{\Phi} + (1-\theta^*)\ln\frac{1-\theta^*}{1-\Phi}\right] \tag{10}$$

and the change in free energy along path II is

$$\Delta F_{II} = \Delta F_{h\phi\,II} + \Delta F_{conf\,II} \tag{11}$$

The equilibrium values of the "order parameters", x^* and θ^*, are obtained by simultaneous solution of the constraint condition, eq 8, and

$$0 = \left.\frac{\partial(F_{II}/kT)}{\partial\theta}\right|_{\theta^*} = \ln\frac{x^*(1-\theta^*)}{\theta^*(1-x^*)} - 2\chi(T)(x^* - \sigma\theta^*) \tag{12}$$

The total free energy of folding is the sum of the free energies from the two pathways, eq 6 and 11:

$$\Delta F_{fold} = \Delta F_I + \Delta F_{II} \tag{13}$$

Temperature Dependence of the Solvophobic Interaction. The theory above depends on a quantity $\chi(T)$, which characterizes the strength of the elementary interaction of amino acids with each other and with the solvent. Its temperature dependence was not explicitly considered in the earlier treatment (Dill, 1985); here we use the data of Nozaki and Tanford (1971) and the equation of state of Gill and Wadsö (1976) to estimate $\chi(T)$.

It is observed that the transfer from the pure liquid or solid to water of a wide variety of solvophobic compounds is characterized by a large heat capacity change, which is relatively independent of temperature at least up to about 100 °C (Edsall & McKenzie, 1983; Frank & Evans, 1945; Gill et al., 1976; Privalov & Gill, 1988). Therefore

$$\chi(T) = \frac{1.4}{kT}[\Delta H^\circ + \Delta C_p(T - T_0) - T[\Delta S^\circ + \Delta C_p\ln(T/T_0)]] \tag{14}$$

where ΔH° and ΔS° are the enthalpy and entropy, respectively, of the exchange process at $T = T_0$. The factor of 1.4

accounts for the number of amino acids per lattice segment (Dill, 1985). We choose T_0 = 298 K. To characterize the temperature dependence of the elementary transfer process of the amino acids requires values for the constants, $\Delta H°$, $\Delta S°$, and ΔC_p, for typical solvophobic amino acids. The free energy of transfer of an average solvophobic side chain from the pure state to water, $kT\chi$(298 K), is taken to be 2.0 kcal mol^{-1}. Although this value differs depending on which of the many available solvophobicity scales is used, on which amino acids are considered solvophobic, and on the different possible solvent reference states, the differences among most of the choices are relatively small. We have arbitrarily adopted the set of Nozaki and Tanford [Table 13-1 of Tanford (1980); Nozaki & Tanford, 1971] of tryptophan, norleucine, phenylalanine, tyrosine, leucine, valine, methionine, and alanine as being solvophobic. [If isoleucine is substituted for norleucine, and cysteine is substituted for tyrosine, then this set is also the collection of the most hydrophobic amino acids observed in cyclohexane/water partitioning experiments (Radzicka & Wolfenden, 1988).] Their average free energy of transfer to water from ethanol or dioxane has been measured by Tanford and Nozaki to be 1.988 kcal mol^{-1}, hence the value chosen above. Gill et al. (Gill & Wadsö, 1976; Gill et al., 1976) have shown that the enthalpy of transfer at 298 K from the pure state to water is approximately 0 for a wide range of hydrocarbons, i.e., ΔH(298 K) = $\Delta H°$ = 0. We assume this value for the transfer of an average solvophobic side chain. Taken together, these values for the free energy and enthalpy specify that the average entropy of transfer of an amino acid side chain from the pure state to water is ΔS(298 K) = $\Delta S°$ = −6.7 cal K^{-1} mol^{-1}.

We estimate that the change in heat capacity upon transfer of the side chain of an average solvophobic amino acid is ΔC_p = 55 cal K^{-1} mol^{-1}. This value follows from two lines of evidence. First, Gill et al. (Gill & Wadsö, 1976; Gill et al., 1976) have shown a linear relationship between ΔC_p and the number of hydrogens in a wide range of hydrocarbons. If we consider as solvophobic the residues listed above or if we consider the residues with positive "hydropathy" from the scale of Kyte and Doolittle (1982), then the average number of hydrogens in solvophobic side chains is approximately 7, and the average number of carbons is 3. The equation of Gill and Wadso leads to ΔC_p = 55 cal K^{-1} mol^{-1} in that case. Second, heat capacities of transfer have been measured for some of the amino acids (Prasad & Ahluwalia, 1976; Spink & Wadsö, 1975), and they are observed to depend linearly on the number of carbons in the side chain. Of the amino acids for which measurements have been made, a reasonable choice for the "average" solvophobic residue is norvaline, inasmuch as its side chain of $-CH_2-CH_2-CH_3$ has seven hydrogens and three carbons. The heat capacity change between norvaline and glycine, and thus for the side chain, is approximately 49 cal K^{-1} mol^{-1} as measured by Prasad and Ahluwalia (1976). The measurements by Spink and Wadsö (1975) are in general agreement with those of Prasad and Ahluwalia.

By use of these values, Figure 2 shows predictions for the temperature dependence of transfer of the average solvophobic amino acid side chain from water to a pure medium of other solvophobic residues according to eq 14. It is evident from Figure 2 that there is a maximum strength of the solvophobic effect. If solvophobicity is measured by a quantity such as χ, the logarithm of a partition coefficient or a solubility, i.e., a free energy divided by kT, then the maximum solvophobicity will be observed to be near room temperature. This maximum is the point at which ΔH = 0. This is observed to be the case

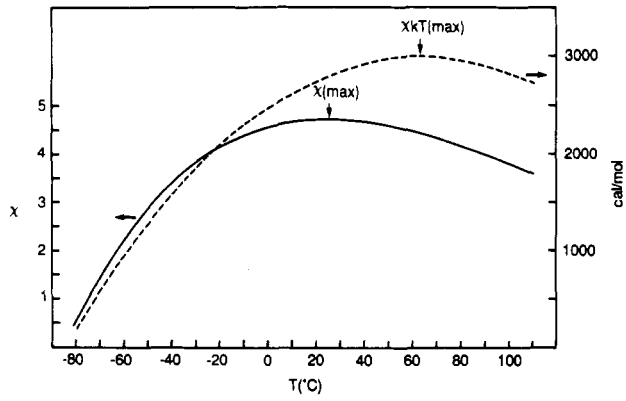

FIGURE 2: Calculated temperature dependence of the elementary solvophobic interaction. These curves are generated with $\Delta H°$ = 0, $\Delta S°$ = −6.7 eu, and ΔC_p = 55 cal/mol at 298 K in eq 14 (see text). The dashed curve is $kT\chi$, a free energy, and the solid line is χ, a free energy divided by kT, per chain segment. There are 1.4 amino acids per model chain segment (Dill, 1985); hence, the ordinate values must be divided by 1.4 to give transfer per amino acid.

for alkanes, benzene derivatives, and other hydrocarbons (Gill et al., 1976). On the other hand, if solvophobicity is measured simply as a free energy (not divided by kT), then the maximum solvophobicity will be observed to occur at approximately 60–80 °C; this is the point at which ΔS = 0. This is consistent with measurements on amino acids by Brandts (1964); similarly for a range of hydrocarbons the maxima are very broadly distributed from 60 to 200 °C (Privalov & Gill, 1988). Thus for the amino acids, as for simpler systems, the hydrophobic effect strengthens with increasing temperature up to a point at which, due to the large change in heat capacity, the hydrophobic effect weakens with further increase in temperature (Ben-Naim, 1980; Edsall & McKenzie, 1983). Although the side-chain transfer parameters adopted here pass a critical test of predicting the free energy maximum observed by Brandts, it would be of great value to have a better experimental basis for estimating these quantities. Although there are many measurements of free energies of amino acid transfer, there are few measurements of the temperature dependence. These data would be of interest both for polar and for nonpolar amino acids.

PREDICTIONS

(*i*) *Temperature Dependence of Folding.* We substitute eq 14, representing the temperature dependence of the elementary amino acid transfer process, into eq 4, 7, 9, and 12 in order to predict the free energy of folding of the protein as a function of temperature.

The predicted contributions of the contact interactions and conformational entropy components of stability of the protein are shown in Figure 3 as a function of temperature. The solvophobic contribution to ΔF_{fold} strengthens, that is, becomes more negative, with temperature over the full temperature range, in accord with the expectations expressed in the introduction of this work. It would seem puzzling that the solvophobic contribution to chain folding should continue to become more negative beyond about 70–80 °C, where the strength of the elementary solvophobic amino acid association process begins to weaken. However, this is a consequence of the density of the denatured state, which is also predicted to decrease above that temperature with a concomitant increase in the number of unfavorable solvent/solvophobic contacts. As a result the free energy of the unfolded state becomes more positive than if the unfolded state simply had fixed density. Thus the difference in free energies of the folded state relative

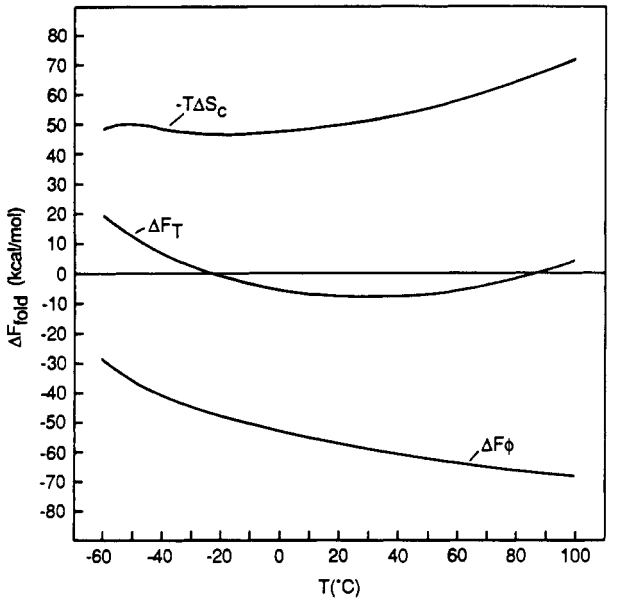

FIGURE 3: Calculated conformational entropy ($-T\Delta S_c$) and solvophobic (ΔF_ϕ) contributions to the free energy of folding (ΔF_T) for $n = 110$, $\Phi = 0.5$, and $\Delta H°$, $\Delta S°$, and ΔC_p as in Figure 2.

to the unfolded state continues to increase even beyond the temperature of maximum solvophobicity. It is evident from Figure 3 that the solvophobic contribution is predicted to have a strong temperature dependence at low temperatures but that this slope decreases with increasing temperature.

The contribution to the free energy of folding arising from the conformational entropy is positive and predicted to increase with temperature. There is little temperature dependence of this quantity at low temperatures, but the slope becomes larger at higher temperatures. The simplest expectation that the conformational entropy should be a constant independent of temperature (Brandts, 1964), and thus that the conformational free energy should be linear with temperature, is not supported by the theory. The reason is that, as noted above, the density of the unfolded state is predicted to increase in parallel with the strength of the solvophobic interaction (see Figure 4). Thus the density decreases rapidly with decreasing temperature at low temperatures, and the entropy of the unfolded state increases. The absolute difference in entropy between the folded and unfolded states increases, and the conformational free energy is higher than would be predicted at low temperatures by the linearity approximation.

The total free energy of folding is the sum of the contact and conformational free energies. It is clear from Figure 3 that at all temperatures the total free energy is a small difference of the two large contributions due to (i) solvent interactions and (ii) conformational entropy. A principal prediction of the theory is that the two transitions are first order (see Figure 4). This is the maximum possible cooperativity of a transition. This is consistent with the many observations that the folding of single-domain proteins is generally a two-state process: the folded and unfolded states are more stable than intermediates (Kim & Baldwin, 1982; Pace & Tanford, 1968; Privalov & Khechinashvili, 1974; Privalov, 1979).

The theory predicts two denaturation temperatures, points at which the total free energy equals 0. At those points, the equilibrium constant, the ratio of the number of native to denatured molecules, equals 1. At low temperatures, a "cold denaturation" is predicted. It arises because the temperature dependence at low temperatures is dominated by the solvophobic effect, the conformational contribution being nearly independent of temperature. Thus in that temperature range,

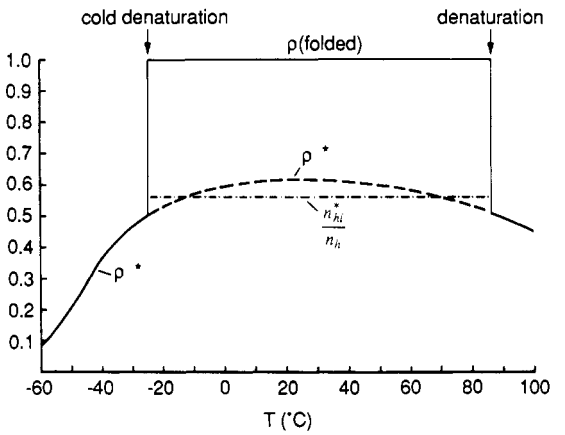

FIGURE 4: Predicted density vs temperature. Density of the state of lowest free energy (—); density of the unfolded reference state (– –); ordering (fraction of solvophobic residues in the interior, $n^*_{hi}/n_h = x^*f_i/\Phi$) (–·–) vs temperature. Calculations were performed with the same quantities as in Figure 3.

in accord with expectations expressed in the introduction, folding is favored by heating because the solvophobic effect strengthens with temperature, and cooling favors denaturation. Cold denaturation has been observed, or extrapolated from observations, for several proteins (Becktel & Schellman, 1987; Cho et al., 1982; Christensen, 1952; Nojima et al., 1977; Pace & Tanford, 1968; Pace, 1975; Privalov, 1979; Privalov et al., 1986).

On the other hand at high temperatures, heating is predicted to lead to unfolding. At these temperatures the conformational entropy dominates the temperature dependence, the slope of the solvophobic contribution being smaller. The free energy is reduced more by the gain in conformational freedom of the chain than by the gain in intrachain attraction of the nonpolar chain segments at high temperatures.

(*ii*) *Nature of the Unfolded State.* A principal conclusion from our calculations is shown in Figure 4: thermal denaturation leads to an unfolded state which is of relatively high density, but it is nevertheless highly solvated. The prediction shown, that the density of the unfolded state is about half that of the folded state, is in agreement with measurements of the volume of the unfolded molecules by viscometry (Privalov et al., 1986). It is worth noting that the density of the unfolded state upon thermal denaturation is not necessarily the same as that of the unfolded state due to solvent denaturation. The unfolded state is important insofar as it serves as the reference for free energy, the stability being the difference in free energies of folded and unfolded states. The density of the unfolded state makes two contributions to stability: (i) greater compactness reduces the conformational freedom of the chain, and (ii) greater compactness increases the incidence of solvophobic contacts in the ensemble of unfolded conformations. The density of the unfolded state is affected by external conditions such as temperature, in contrast to the density and order of the folded state, which are relatively independent of temperature (see Figure 4). Thus, while a principal prediction of the theory is that of a first-order transition between folded and unfolded states, and thus is consistent with a two-state model, nevertheless the two states, unfolded and folded, are of a very different nature. Whereas the unfolded "state" is an ensemble of microscopic conformations which changes with external conditions, the folded state ensemble does not, which is undoubtedly of value for biological function in changing environments.

The considerable variation of the density, and therefore the radius, of gyration of the unfolded molecule with temperature

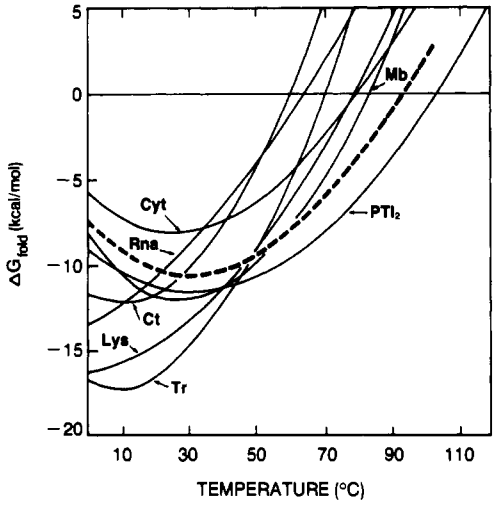

FIGURE 5: Comparison of theoretical (--) and experimental (—) (Privalov & Khechinashvili, 1974; Privalov, 1979) free energies of folding as a function of temperature. Mb, metmyoglobin; PTI_2, dimer of pancreatic trypsin inhibitor; Cyt, cytochrome *c*; Ct, α-chymotrypsin; Lys, lysozyme; Tr, β-trypsin; Rna, ribonuclease A.

is shown in Figure 4. With increasing temperature, the radius decreases, becoming smallest at about 20–30 °C, and then increases with further increase in temperature. This temperature dependence simply follows the similar temperature dependence of the interaction of water with the nonpolar residues. Because the strength of the hydrophobic interaction is maximum at this temperature, the radius of the unfolded molecule is smallest. Only at (experimentally unachievable) extremely low or high temperatures would the unfolded states of proteins resemble ideal (θ state) random coil conformations. The unfolded state density will also depend on the chain composition. For the particular case shown in Figure 4, the unfolded state will be quite compact relative to random coil molecules and will therefore have several hydrophobic contacts on average. The unfolded state will have considerably less exposure of nonpolar residues to water than would a random coil molecule. The unfolded states of proteins would be very poorly modeled as random coils; their conformational entropy and heat capacity are very different, their solvent exposure is very different, and their dependence on temperature and chain composition cannot be neglected as would be done if they were assumed to be random coils.

(*iii*) *Comparison with Experiments.* Figure 5 shows a comparison of the theoretical ΔF_{fold} with the differential scanning calorimetry experiments of Privalov et al. on a set of small single-domain globular proteins (Privalov & Khechinashvili, 1974; Privalov, 1979). The experiments shown are results evaluated at the pH of maximum stability; i.e., they are often near the isoelectric point where charge effects on stability are minimized. For the calculations shown in the figure, we have adopted the values indicated above for the thermodynamics of amino acid transfer ($\Delta H^\circ = 0$; $\Delta S^\circ = -6.7$ cal K^{-1} mol^{-1}; $\Delta C_p = 55$ cal K^{-1} mol^{-1}), and we have taken $\Phi = 0.5$ and $n = 110$ [corresponding to 154 residues, the mean value for this set of proteins. Note that each chain segment represents about 1.4 amino acids (Dill, 1985)]. The insignificant pV differences between Helmholtz and Gibbs free energies are neglected here. It is clear that there is considerable variability of the free energy of folding among proteins. Nevertheless, the theoretical prediction falls within this range and appears to predict well the typical temperature dependence of stability, including the normal and cold denaturation temperatures and temperature of maximum stability.

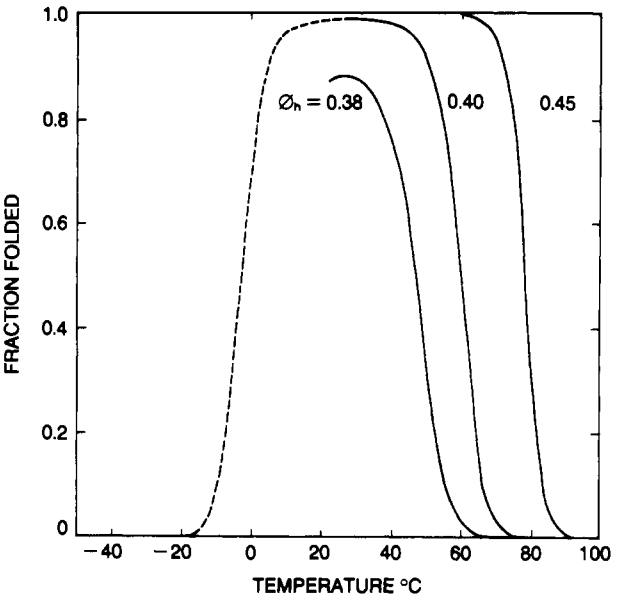

FIGURE 6: Calculated denaturation curves. Fraction of native protein, as a function of temperature for proteins for $n = 100$ and $\Phi = 0.38$, 0.40, and 0.45. The dashed line for $\Phi = 0.40$ shows the cold denaturation.

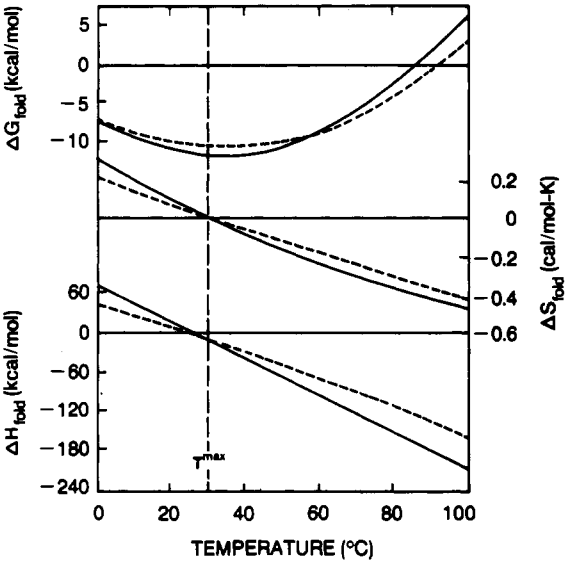

FIGURE 7: Entropy and enthalpy of protein folding. Calculated (--) and experimental (—) (Privalov & Khechinashvili, 1974) entropy, enthalpy, and free energy of folding for metmyoglobin.

Figure 6 shows a common alternative representation of the folding equilibrium, the fraction of native molecules as a function of temperature. This denaturation profile contains the same information but in a different form that that of the dependence of free energy on temperature. It is computed through use of

$$f_{native} = \frac{\exp(-\Delta F_{fold}/kT)}{1 + \exp(-\Delta F_{fold}/kT)} \qquad (15)$$

A more stringent test of the theory is the enthalpy and entropy of folding, as functions of temperature. Predictions are compared in Figure 7 with experiments of Privalov and Khechinashvili on myoglobin (Privalov & Khechinashvili, 1974; Privalov, 1979). The theoretical enthalpy is simply taken as the enthalpic contribution to the solvophobic interaction. The theoretical entropy is calculated as the sum of solvophobic and conformational contributions. Whereas the temperature dependences of the enthalpy and entropy are qualitatively

Table I: Comparison of Theoretical and Experimental (Privalov & Khechinashvili, 1974) $\Delta C_{p\,\text{fold}}$'s

protein	n residues	Φ^a	$\Delta C_{p\,\text{exp fold}}$ (cal g^{-1} K^{-1})	$\Delta C_p/n_{\text{exp fold}}$ [cal (mol of residue)$^{-1}$ K^{-1}]	$\Delta C_p/n_{\text{theory fold}}$ (cal/mol of residue)	S-S bonds	heme group
Mb	153	0.43	−0.15	−16.9	−14.3	0	yes
α-CT	241	0.39	−0.12	−12.6	−18.7	5	no
Lys	129	0.36	−0.11	−12.3	−15.1	4	no
cyt *c*	104	0.31	−0.13	−14.5	−13.4	0	yes
RNase	124	0.32	−0.09	−9.9	−14.7	4	no

[a] Φ was calculated by counting the nonpolar amino acids (A, I, L, V, M, F, W, and Y) in a given protein and dividing by n.

predicted by the theory, the absolute errors are seen to be significantly greater at high temperatures than those of the free energy, but they compensate.

A principal thermal property of protein folding is $\Delta C_{p\,\text{fold}}$, which can be calculated from theory by $\partial \Delta H(T)_{\text{fold}}/\partial T$. Comparisons of theoretical and experimental values are shown in Table I; the agreement is good. It should be noted that there is considerable variation in experimental measurements of $\Delta C_{p\,\text{fold}}$. Experimental values of $\Delta C_{p\,\text{fold}}$ for ribonuclease A range from −9.9 to −24 cal (mol of residue)$^{-1}$ K^{-1} (Edelhoch & Osborne, 1976; Freire & Biltonen, 1978), predominantly clustering around −16 cal (mol of residue)$^{-1}$ K^{-1}. This is to be compared with the theoretical value of −14.7 cal (mol of residue)$^{-1}$ K^{-1}. The theory predicts that $\Delta C_{p\,\text{fold}}$ is relatively independent of temperature: the variation in the theoretical value of $\Delta C_{p\,\text{fold}}$ over the temperature range 0–80 °C is about 7%, which is in agreement with estimates of the experimental variation of 5–10% (Becktel & Schellman, 1987). Above 80 °C there is a variation of about 15% for proteins with $\Phi > 0.45$.

These predictions call into question a common thermodynamic assumption made for the purpose of microscopic interpretation of the balance of forces in protein folding. It is common to divide $\Delta C_{p\,\text{fold}}$ by n, the number of residues in the chain, and to assume that this heat capacity contribution per residue will be independent of chain length. The theory shows, however, that $\Delta C_{p\,\text{fold}}/n$ will be strongly dependent on both n and Φ (Figure 8). This is most readily understood by considering how the folding process changes if the number of solvophobic residues is increased while the chain length is held constant (i.e., increasing Φ for fixed n); see either the left-hand or right-hand column of Figure 8. When there are few h residues, the unfolded state has low density (bottom-row panels in Figure 8), and in the folded state most of the h residues are in the interior core of the protein (θ^* is small; third-row panels in Figure 8). Added h residues contribute principally to further filling the interior of the core of the protein (x^* increases, θ^* remains small). The principal factor affecting the thermodynamic changes of $\Delta H_{\text{fold}}/n$, $\Delta C_{p\,\text{fold}}/n$, and $\Delta F_{h\phi\,\text{fold}}/n$ is the change in fractional exposure of solvophobic groups upon folding Δa, given per residue by

$$\Delta a = -\Phi^2(f_i + \sigma f_e)(1 - \rho^*) - f_i(x^{*2} - \Phi^2) - \sigma f_e(\theta^{*2} - \Phi^2) \quad (16)$$

(Changes in exposure are defined to be negative when h residues become increasingly buried or sequestered from solvent contact, in the folded state relative to the unfolded state.) This change in exposure and the individual pathway contributions to it are shown in the second-row panels in Figure 8. The change in heat capacity of folding, $\Delta C_{p\,\text{fold}}/n$, is approximately equal to the change in exposure multiplied by the elementary amino acid transfer process at 25 °C and is shown in the first-row panels of Figure 8. Likewise, $\Delta H_{\text{fold}}/n$ and $\Delta F_{h\phi\,\text{fold}}/n$ are also calculated by multiplying the appropriate elementary value by the change in exposure. For small Φ, added h residues

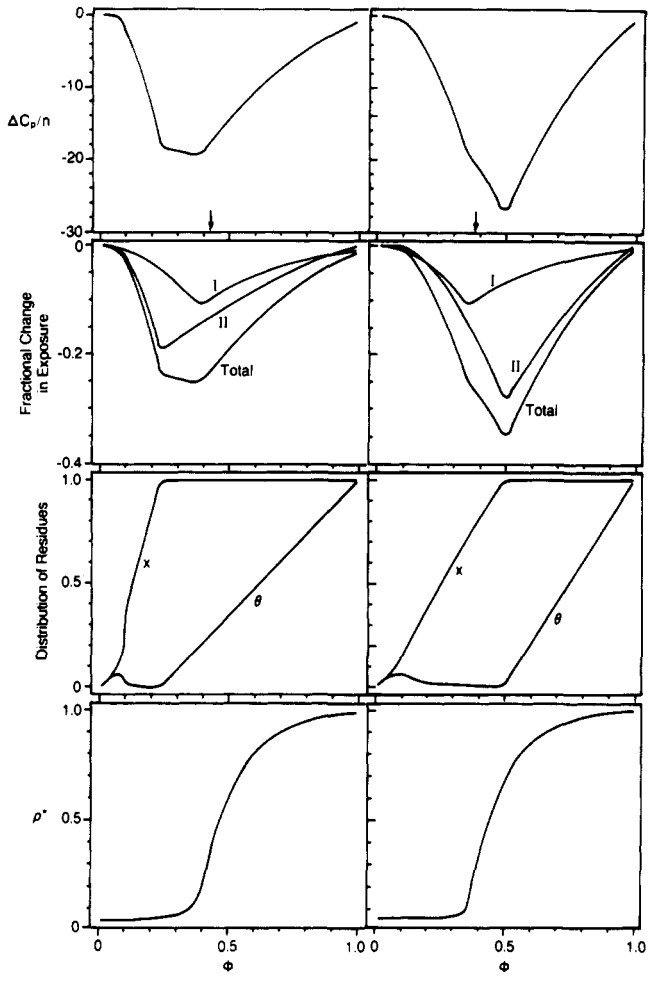

FIGURE 8: Composition dependence of $\Delta C_{p\,\text{fold}}/n$, Δa (see eq 16), x^*, and θ^* (see eq 12); $x^* = n^*_{\text{hi}}/n_i$ = equilibrium fraction of core filled by h residues. $\theta^* = n^*_{\text{he}}/n_e$ = fraction of surface filled by h residues, and ρ^* (density of the unfolded state, see eq 7). Left column, n = 75; right column, n = 450.

principally contribute to added hh contacts in the folded core; hence, there is less h exposure to solvent upon folding, and $\Delta C_{p\,\text{fold}}/n$ becomes more negative. In short, each added h residue helps stabilize the molecule because it has reduced exposure to solvent upon folding (i.e., the number of hh contacts increases upon folding); it is this change in burial of h residues upon folding which contributes to ΔH_{fold}.

This process changes character, however, when Φ exceeds a critical value, due to two physical factors: (i) change in ρ^* due to addition of an h residue and (ii) the limited size of the core of the protein. First, beyond a certain value of Φ, each added h residue causes the density of the unfolded state to increase sharply (see the fourth-row panels of Figure 8), reducing the average exposure in the unfolded state, and this dominates path I. In short, an added h residue will reduce exposure in the unfolded state more than in the folded state, with an overall destabilizing effect on ΔH_{fold}. For short chains,

this physical process dominates the decrease of hydrophobic stability upon addition of an h residue at high Φ. Second, beyond a (generally different) critical value of Φ, the core has become saturated with h residues, and subsequent h residues must be added to the surface of the protein (see third-row panels, Figure 8). This physical process dominates path II and also leads to lesser relative burial (in the folded state relative to the unfolded state) upon addition of an h residue than if the core were not saturated.

Hence, $\Delta C_{p\,\mathrm{fold}}/n$ is not independent of Φ, nor is it independent of n. The dependence on n arises from the increase in volume-to-surface ratio of the protein with increase in n; a greater percentage of h residues can be accommodated in the cores of proteins with large n.

The theory predicts that for a series of proteins differing in composition, the curves of ΔH_{fold} vs T will have different slopes and will all intersect at a point, T = 25 °C, ΔH_{fold} = 0. This theoretical prediction is an immediate consequence of the assumption that the enthalpy of transfer of the individual residues of both types (h and p) is zero at that temperature, i.e., $\Delta H(298)$ = 0. Spectrophotometric van't Hoff experiments on chymotrypsinogen and two of its chemical derivatives show such an intersection point at T = 10 °C, ΔH_{fold} = 0 (Shiao et al., 1971), in reasonable agreement with the theory. Similar experiments by Shortle et al. on staphylococcal nuclease mutants also show that if there is any convergence of $\Delta H_{\mathrm{fold}}(T)$ for different mutants, then it occurs below 20 °C and small ΔH (Shortle et al., 1988). However, the calorimetric experiments of Privalov et al. (Privalov & Khechinashvili, 1974; Privalov, 1979) show a very different intersection point, at T = 110 °C, ΔH_{fold} = −1.2 kcal/mol, for the series of proteins shown in Figure 5 (Baldwin, 1986; Privalov & Gill, 1988). Except for myoglobin, the Privalov data imply that there is an extra enthalpic driving force toward folding at 25 °C, in addition to the solvophobic force. This driving force must involve either solvophilic residues or all residues but cannot readily be accounted for by nonpolar residues alone. It is not yet clear why there is such a large discrepancy between these two experiments, both of which derive the enthalpies from experiments of changing pH. One principal difference between them is the use of ionization-matched buffers in the Privalov experiment to offset small enthalpies of ionization of the protein. Questions regarding the nature of this interaction have motivated our development here of a lattice treatment of sufficient generality to account for the differences in interactions among h, p, and solvent (see Appendix). However, we will not further explore those refinements in the present work nor attempt to estimate the several constants required. The next level of refinement required, beyond the model presented here based on three parameters, $\Delta H°$, $\Delta S°$, and ΔC_p, would require nine such parameters, the other six being for the polar/solvophobic and the polar/solvent interactions. We do not believe such an attempt would yet be justified inasmuch as too little experimental data are currently available on the temperature dependence of transfer processes of polar residues.

Figures 9 and 10 show the effects of changing composition, chain length, $\Delta H°$, $\Delta S°$, and ΔC_p on the temperature dependence of stability. One principal conclusion is that increased chain length (at fixed Φ) is predicted to increase the stability at all experimentally accessible temperatures. However, whereas the present model treats only spherical proteins, the dependence of stability on chain length will be substantially affected if the protein has the freedom to change its shape (Dill, 1985). For example, with increasing molecular weight, proteins tend to be configured into multiple domains; the free

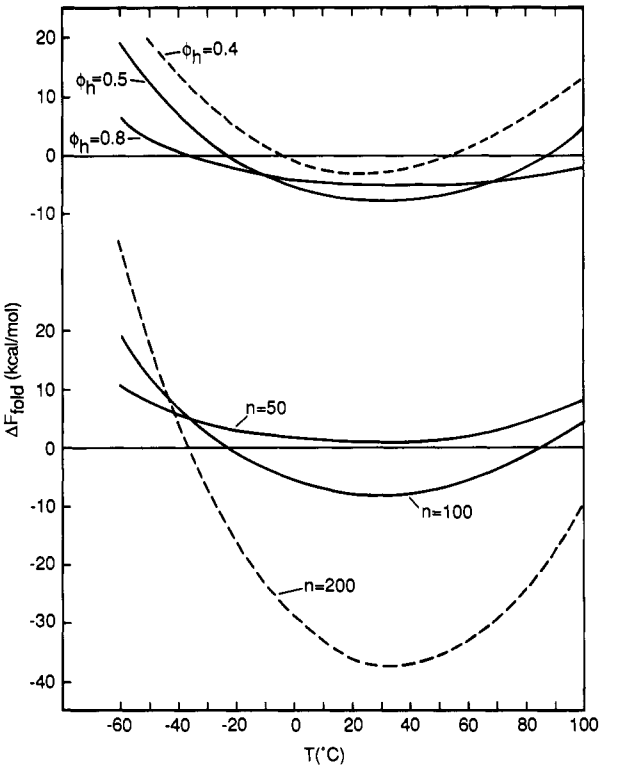

FIGURE 9: Dependence of free energy of folding on n and Φ. The dependence on n is calculated for Φ = 0.5, and the dependence on Φ is calculated for n = 100.

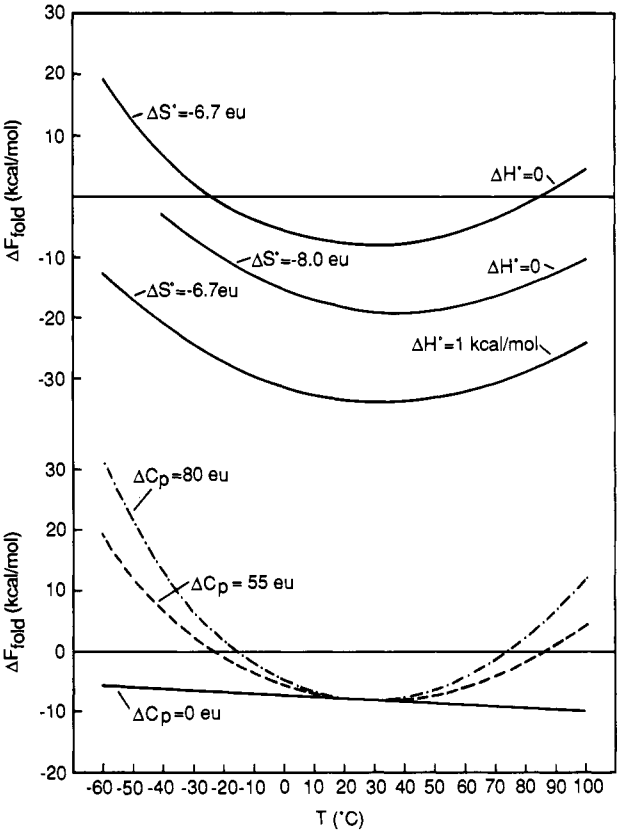

FIGURE 10: Dependence of free energy of folding on the thermodynamic quantities $\Delta S°$, $\Delta H°$, and ΔC_p for the elementary amino acid transfer process, for n = 100 and Φ = 0.5.

energy of folding is significantly dependent upon the size, number, and shape of domains (Dill, 1985). Currently available experiments cover too narrow a range of chain lengths to test this predicted chain-length dependence of stability. A

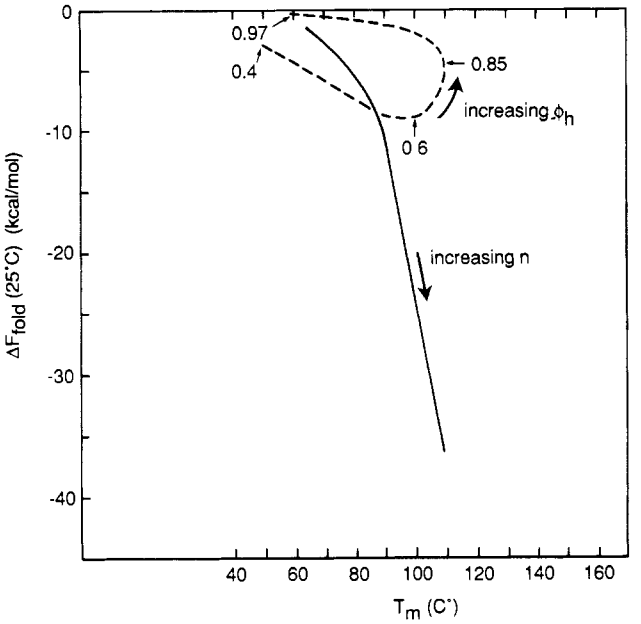

FIGURE 11: Relationship between denaturation temperature and thermodynamic stability (at 25 °C) as a function of n (—) and Φ (--). n dependence was calculated with $\Phi = 0.5$, and Φ dependence was calculated with $n = 100$.

second conclusion is that increasing the heat capacity increases the curvature of free energy with temperature. It is interesting that if ΔC_p were zero, then, due to the large negative $\Delta S°$ from the solvophobic effect, proteins would not denature at higher temperatures. Note that small changes in any of these quantities lead to significant changes in most of the parameters characterizing stability, particularly the denaturation temperatures and free energies of folding. Relatively minor changes in the amino acid composition can lead to significantly more or less stable proteins than the native molecule. This implies that the requirements may not be very stringent for production of proteins which can function at very high (or low) temperatures.

It is widely assumed that any change of structure which leads to increased denaturation temperature should also lead to increased stability (more negative ΔF_{fold} at some specified temperature other than that of denaturation). Thermodynamics however does not specify any particular relationship between the denaturation temperature and the free energy of folding (at a given temperature). Becktel and Schellman (1987) have explored the consequences of small perturbations which are assumed to offset the free energy by an additive constant. The limitations of this assumption, and more generally the microscopic origins of this dependence, can only be explored through statistical mechanical theory. Figure 11 shows the prediction of the present model that in general when chain length or composition are changed so as to increase the stability, i.e., to cause the free energy of folding to become more negative, then it is accompanied by an increase in the denaturation temperature (and a decrease in the cold denaturation temperature), confirming the common expectation and the Becktel and Schellman assumption. However, there are certain circumstances in which denaturation temperature is predicted to increase when the stability at 25 °C decreases, for example, with increasing percentage of solvophobic residues in excess of 60%.

ACKNOWLEDGMENTS

We thank Tom Alber, Wayne Bolen, Stan Gill, Peter Privalov, and Marcelo Santoro for helpful discussions.

APPENDIX

In this appendix, we use lattice methods to derive the theory for folding which leads to eq 4, 5, 9, and 10 in the text. This derivation differs from previous work (Dill, 1985) in the following respects. (i) In the present work we consider all the possible interactions of polar with solvophobic residues (χ_{hp}), of solvent with solvophobic residues (χ_{hs}), and of solvent with polar residues (χ_{ps}). In previous work (Dill, 1985) we considered the limiting approximation in which solvent and polar residues are of identical chemical character ($\chi_{hp} \approx \chi_{hs}$ and $\chi_{ps} \approx 0$). (ii) We develop the theory for pathway II in an alternative chemical potential formulation, which is of more general advantage for some applications. (iii) We simplify the notation and correct some insignificant errors of the earlier work.

The protein is assumed spherical with radius r (in units of amino acid diameters) and has a fraction

$$f_i = \left(\frac{r-1}{r}\right)^3 \tag{A1}$$

of the sites in the protein core and a fraction

$$f_e = 1 - f_i \tag{A2}$$

at the surface.

For each of the two pathways, condensation (I) and reconfiguration (II), the free energy is given by

$$F = E_{contact} - TS_{conf} \tag{A3}$$

where $E_{contact}$ ($=F_{contact} = F_{h\phi}$) is the solvophobic free energy contribution of residues and solvent in contact with their spatial neighbors. This free energy is the sum over all neighbor contacts:

$$E_{contact} = \sum_{i,e,o,b}^{regions} \sum_{l=j}^{components} \sum_{j} w_{lj} m_{lj} \tag{A4}$$

where w_{lj} is the reversible work of bringing units of type l and j from infinite separation to contact and m_{lj} is the number of contacts between l and j species; regions are defined below. S_{conf} is the conformational entropy of the chain and is calculated with the Boltzmann equation, $S_{conf} = k \ln \omega_{conf}$.

(I) Random Condensation

For the purpose of computing the numbers of contacts required to evaluate eq A4, the molecule is considered embedded in a spherical volume containing an interior core (i) and exterior surface shell (e), containing a first solvation layer (o), immediately outside the surface shell, and surrounded by bulk solvent (b). Except at the highest density, along path I all these regions may be solvated; hence, eq A4 may be expressed explicitly as

$$\begin{aligned} E_{1\,contact} = E_I = {} & w_{hh}m_{hhi} + w_{pp}m_{ppi} + w_{ss}m_{ssi} + w_{hp}m_{hpi} + \\ & w_{hs}m_{hsi} + w_{ps}m_{psi} + w_{hh}m_{hhe} + w_{pp}m_{ppe} + w_{ss}m_{sse} + \\ & w_{hp}m_{hpe} + w_{hs}m_{hse} + w_{ps}m_{pse} + w_{hs}m_{hseo} + w_{ps}m_{pseo} + \\ & w_{ss}m_{sseo} + w_{ss}m_{sso} + w_{ss}m_{ssb} \end{aligned} \tag{A5}$$

where s, h, and p represent solvent, solvophobic, and solvophilic residues, respectively, and eo indicates contacts which occur across the interface between the protein surface and the first solvation layer.

Conversion from numbers of contacts to numbers of residues is performed according to standard lattice counting methods (Hill, 1960):

$$qn_{x\alpha} = 2m_{xx\alpha} + m_{xy\alpha} + m_{xz\alpha} \tag{A6}$$

where $n_{x\alpha}$ is defined as the number of units of species x in region α, $m_{xy\alpha}$ is the number of xy contacts in region α (x, y,

z = h, s, p), and q is the lattice coordination number, the number of neighbors of a residue or solvent molecule, e.g.

$$qn_{hi} = 2m_{hhi} + m_{hpi} + m_{hsi} \qquad qn_{sb} = 2m_{ssb}$$

Using the Bragg–Williams approximation, we have

$$m_{xy\alpha} = q\frac{n_{x\alpha}n_{y\alpha}}{n_\alpha} \tag{A7}$$

e.g.

$$m_{hpi} = q\frac{n_{hi}n_{pi}}{n_i}$$

Any residue at the surface of the protein has a fraction $(1 - \sigma)$ of its contact area exposed to solvent; that is, it makes $(1 - \sigma)q$ contacts with the solvent outside the sphere occupied by the protein and σq contacts within the sphere; thus

$$\begin{aligned} m_{hpe} &= \sigma q\frac{n_{he}n_{pe}}{n_e} \qquad m_{hseo} = (1-\sigma)qn_{he} \\ m_{pseo} &= (1-\sigma)qn_{pe} \qquad m_{sso} = qn_{so}/2 \\ m_{sseo} &= (1-\sigma)qn_{se} \end{aligned} \tag{A8}$$

For the present calculations we have used $\sigma = {}^2/_3$ (Dill, 1985).

The contact free energies can be expressed more conveniently in terms of dimensionless exchange free energies:

$$\chi_{xy} = \frac{q}{kT}\left(w_{xy} - \frac{w_{xx} + w_{yy}}{2}\right) \tag{A9}$$

where x and y represent h, p, or s to define χ_{hp}, χ_{hs}, and χ_{ps}. With the substitutions of eq A6–A9 and division by kT, the contact free energy (eq A5) becomes

$$\begin{aligned} \frac{E_I}{kT} = {} & \frac{q}{2kT}(w_{hh}n_{hi} + w_{pp}n_{pi} + w_{ss}n_{si}) + \frac{n_{hi}n_{pi}}{n_i}\chi_{hp} + \\ & \frac{n_{hi}n_{si}}{n_i}\chi_{hs} + \frac{n_{pi}n_{si}}{n_i}\chi_{ps} + \frac{q\sigma}{2kT}(w_{hh}n_{he} + w_{pp}n_{pe} + w_{ss}n_{se}) + \\ & \sigma\left(\frac{n_{he}n_{pe}}{n_e}\chi_{hp} + \frac{n_{he}n_{se}}{n_e}\chi_{hs} + \frac{n_{pe}n_{se}}{n_e}\chi_{ps}\right) + \\ & \frac{q(1-\sigma)}{kT}(w_{hs}n_{he} + w_{ps}n_{pe} + w_{ss}n_{se}) + \frac{q\sigma w_{ss}}{2kT}n_{so} + \frac{qw_{ss}}{2kT}n_{sb} \end{aligned} \tag{A10}$$

The constraints of the system are

$$n_{hi} + n_{he} = n_h \tag{A11a}$$

$$n_{pi} + n_{pe} = n_p \tag{A11b}$$

$$n_{si} + n_{se} + n_{so} + n_{sb} = n_s \tag{A11c}$$

$$n_{he} + n_{pe} + n_{se} = n_e \tag{A11d}$$

$$n_{hi} + n_{pi} + n_{si} = n_i \tag{A11e}$$

$$n_e = n_{so} \tag{A11f}$$

which may be used to simplify eq A10:

$$\begin{aligned} \frac{E_I}{kT} = {} & \frac{n_{hi}n_{pi}}{n_i}\chi_{hp} + \frac{n_{hi}n_{si}}{n_i}\chi_{hs} + \frac{n_{pi}n_{si}}{n_i}\chi_{ps} + \\ & \sigma\left(\frac{n_{he}n_{pe}}{n_e}\chi_{hp} + \frac{n_{he}n_{se}}{n_e}\chi_{hs} + \frac{n_{pe}n_{se}}{n_e}\chi_{ps}\right) + (1-\sigma)(n_{he}\chi_{hs} + \\ & n_{pe}\chi_{ps}) + \frac{q}{2kT}(w_{hh}n_h + w_{pp}n_p + w_{ss}n_s) \end{aligned} \tag{A12}$$

The last term in parentheses is a constant for a given molecule.

Since the residues have a random spatial distribution along path I

$$\frac{n_{he}}{n_h} = \frac{n_{pe}}{n_p} = f_e \qquad \frac{n_{hi}}{n_h} = \frac{n_{pi}}{n_p} = f_i \tag{A13}$$

We divide eq A12 by n and substitute the definitions of f_i, f_e, ρ, and Φ to obtain the following expression for the contact free energy along path I (see Dill, 1985):

$$\begin{aligned} \frac{E_I}{nkT} = {} & (f_i + \sigma f_e)\rho[\Phi(1-\Phi)\chi_{hp} - \\ & \Phi\chi_{hs} - (1-\Phi)\chi_{ps}] + \text{constant} \end{aligned} \tag{A14}$$

The conformational entropy for condensation is the sum of an excluded volume (osmotic pressure) contribution

$$\frac{S_I^{ev}}{nk} = \frac{-(1-\rho)}{\rho}\ln(1-\rho) - 1 \tag{A15}$$

and an elastic term (Dill, 1985; Dill & Alonso, 1988)

$$\frac{S_I^{elas}}{nk} = \frac{-7}{2n}\left(\frac{\rho_0}{\rho}\right)^{2/3} - \frac{2}{n}\ln(\rho/\rho_0) - \ln c \tag{A16}$$

where c is a constant.

The total free energy for path I is obtained through use of eq A3, A14, A15, and A16:

$$\begin{aligned} \frac{F_I}{nkT} = {} & (f_i + \sigma f_e)\rho[\Phi(1-\Phi)\chi_{hp} - \Phi\chi_{hs} - (1-\Phi)\chi_{ps}] + \\ & \frac{1-\rho}{\rho}\ln(1-\rho) + \frac{7}{2n}\left(\frac{\rho_0}{\rho}\right)^{2/3} + \frac{2}{n}\ln(\rho/\rho_0) + \text{constant} \end{aligned} \tag{A17}$$

Equation A17 is identical with the earlier theory (Dill, 1985) when $\chi_{hp} \approx \chi_{hs} \equiv \chi$ and $\chi_{ps} \approx 0$; whereupon, we have

$$\begin{aligned} \frac{F_I}{nkT} = {} & -(f_i + \sigma f_e)\rho\chi\Phi^2 + \frac{1-\rho}{\rho}\ln(1-\rho) + \frac{7}{2n}\left(\frac{\rho_0}{\rho}\right)^{2/3} + \\ & \frac{2}{n}\ln(\rho/\rho_0) + \text{constant} \end{aligned} \tag{A18}$$

In the earlier work the strength of the solvophobic interaction was represented by the quantity ϵ; it is related to χ by $\chi = \epsilon/2$.

The Unfolded State. The unfolded state is that of minimum free energy along path I, in which the solvophobic driving force to condense and the conformational entropy opposing condensation are balanced:

$$\left.\frac{\partial(F_I/nkT)}{\partial\rho}\right|_{\rho=\rho^*} = 0 \tag{A19}$$

Hence, ρ^* identifies the density of the molecule in the equilibrium unfolded state. The contact contribution to this derivative is

$$\begin{aligned} \frac{\partial(E_I/nkT)}{\partial\rho} = {} & [\Phi(1-\Phi)\chi_{hp} - \Phi\chi_{hs} - (1-\Phi)\chi_{ps}] \times \\ & \left[f_i + \sigma f_e + \rho\left(\frac{\partial f_i}{\partial\rho} + \sigma\frac{\partial f_e}{\partial\rho}\right)\right] \end{aligned} \tag{A20}$$

$$\begin{aligned} = {} & [\Phi(1-\Phi)\chi_{hp} - \Phi\chi_{hs} - (1-\Phi)\chi_{ps}] \times \\ & \left[f_i + \sigma f_e + \rho\frac{\partial f_i}{\partial\rho}(1-\sigma)\right] \end{aligned} \tag{A21}$$

where

$$\frac{\partial f_i}{\partial\rho} = -\frac{4\pi(r-1)^2}{3n} \qquad \frac{\partial f_e}{\partial\rho} = -\frac{\partial f_i}{\partial\rho} \tag{A22}$$

which follow from eq 3, A1, and A2. The $\rho(\partial f_i/\partial\rho)(1-\sigma)$ term is much smaller than $f_i + \sigma f_e$; hence, we may use the approximation

$$\frac{\partial(E_I/nkT)}{\partial\rho} = [\Phi(1-\Phi)\chi_{hp} - \Phi\chi_{hs} - (1-\Phi)\chi_{ps}](f_i + \sigma f_e) \quad \text{(A23)}$$

The equilibrium condition, eq A19, is determined from eq A23 and the derivative of eq A15 and A16:

$$0 = \frac{\partial(F_I/nkT)}{\partial\rho} = [\Phi(1-\Phi)\chi_{hp} - \Phi\chi_{hs} - (1-\Phi)\chi_{ps}] \times (f_i + \sigma f_e) - \frac{7}{3n}\frac{{\rho_0}^{2/3}}{{\rho^*}^{5/3}} - \frac{1}{\rho^*}\left[\frac{\ln(1-\rho^*)}{\rho^*} + 1 - \frac{2}{n}\right] \quad \text{(A24)}$$

This leads to eq 7 when $\chi_{hp} = \chi_{hs}$ and $\chi_{ps} = 0$.

Change in Free Energy upon Condensation: Pathway I. The change in free energy for the collapse from the unfolded state, $\rho = \rho^*$, to a randomly ordered compact state, $\rho \equiv 1$, is given with eq A17:

$$\frac{\Delta F_I}{nkT} = \frac{F_I(1) - F_I(\rho^*)}{nkT} = [\Phi(1-\Phi)\chi_{hp} - \Phi\chi_{hs} - (1-\Phi)\chi_{ps}](f_i(1) + \sigma f_e(1) - \rho^*[f_i(\rho^*) + \sigma f_e(\rho^*)]) - \frac{1-\rho^*}{\rho^*}\ln(1-\rho^*) + \frac{7}{2n}[{\rho_0}^{2/3} - (\rho_0/\rho^*)^{2/3}] - \frac{2}{n}\ln\rho^* \quad \text{(A25)}$$

In this case, as with the equilibrium condition, the dependencies of f_e and f_i on ρ are small, contributing less than a few percent to the overall free energy; hence, we use the approximation

$$\frac{\Delta F_I}{nkT} = [\Phi(1-\Phi)\chi_{hp} - \Phi\chi_{hs} - (1-\Phi)\chi_{ps}](f_i + \sigma f_e)(1-\rho^*) - \frac{1-\rho^*}{\rho^*}\ln(1-\rho^*) + \frac{7}{2n}[{\rho_0}^{2/3} - (\rho_0/\rho^*)^{2/3}] - \frac{2}{n}\ln\rho^* \quad \text{(A26)}$$

In the limit in which $\chi \equiv \chi_{hp} \approx \chi_{hs}$ and $\chi_{ps} \approx 0$, then this reduces to

$$\frac{\Delta F_I}{nkT} = -\chi\Phi^2(f_i + \sigma f_e)(1-\rho^*) - \frac{1-\rho^*}{\rho^*}\ln(1-\rho^*) + \frac{7}{2n}[{\rho_0}^{2/3} - (\rho_0/\rho^*)^{2/3}] - \frac{2}{n}\ln\rho^* \quad \text{(A27)}$$

eq 4 and 5 in the text.

(II) Reconfiguration

In this process the molecule is at maximum density ($\rho = 1$) and the free energy depends on the distribution of the h and p residues between the internal core and the external surface of the protein. For $\rho = 1$, eq A4 gives

$$E_{II} = w_{hp}m_{hpi} + w_{hh}m_{hhi} + w_{pp}m_{ppi} + w_{hp}m_{hpe} + w_{hh}m_{hhe} + w_{pp}m_{ppe} + w_{hs}m_{hseo} + w_{ps}m_{pseo} \quad \text{(A28)}$$

In this case all the solvent interactions occur only at the surface of the globular protein.

As before, we use lattice counting methods, the Bragg–Williams approximation, and the substitution of χ's to give

$$\frac{E_{II}}{kT} = \chi_{hp}\frac{n_{hi}n_{pi}}{n_i} + \frac{qw_{hh}}{2kT}n_{hi} + \frac{qw_{pp}}{2kT}n_{pi} + \sigma\chi_{hp}\frac{n_{he}n_{pe}}{n_e} + \frac{\sigma qw_{hh}}{2kT}n_{he} + \frac{\sigma qw_{pp}}{2kT}n_{pe} + (1-\sigma)\frac{qw_{hs}}{kT}n_{he} + (1-\sigma)\frac{qw_{ps}}{kT}n_{pe} \quad \text{(A29)}$$

External solvent interactions need not be accounted for explicitly in eq A29 since they do not change as a function of the path variable.

The reconfigurational entropy describes a "mixing" process of h- and p-type "solutes" in internal (i) and external (e) regions ("solvents"). The number of configurations of the system (ω_{II}) is

$$\omega_{II} = \frac{n_i!}{n_{hi}!n_{pi}!}\frac{n_e!}{n_{he}!n_{pe}!} \quad \text{(A30)}$$

Subject to Stirling's approximation and the constraints of eq A11, the reconfigurational entropy becomes

$$S_{II}/k = \ln\omega_{II} = n_{hi}\ln\frac{n_i}{n_{hi}} + n_{pi}\ln\frac{n_i}{n_{pi}} + n_{he}\ln\frac{n_e}{n_{he}} + n_{pe}\ln\frac{n_e}{n_{pe}} \quad \text{(A31)}$$

The sum of the contact and conformational contributions, eq A29 and A31, respectively, gives the free energy at any point along path II:

$$\frac{F_{II}}{kT} = \chi_{hp}\frac{n_{hi}n_{pi}}{n_i} + \frac{q}{2}\frac{w_{hh}}{kT}n_{hi} + \frac{q}{2}\frac{w_{pp}}{kT}n_{pi} + \sigma\chi_{hp}\frac{n_{he}n_{pe}}{n_e} + \sigma\frac{q}{2}\frac{w_{hh}}{kT}n_{he} + \sigma\frac{q}{2}\frac{w_{pp}}{kT}n_{pe} + (1-\sigma)q\frac{w_{hs}}{kT}n_{he} + (1-\sigma)\times q\frac{w_{ps}}{kT}n_{pe} + n_{hi}\ln\frac{n_{hi}}{n_i} + n_{pi}\ln\frac{n_{pi}}{n_i} + n_{he}\ln\frac{n_{he}}{n_e} + n_{pe}\ln\frac{n_{pe}}{n_e} \quad \text{(A32)}$$

In terms of the dimensionless variables, $\theta = n_{he}/n_e$ and $x = n_{hi}/n_i$, the free energy per residue is

$$\frac{F_{II}}{nkT} = \chi_{hp}x(1-x)f_i + xf_i\frac{q}{2}\frac{w_{hh}}{kT} + (1-x)f_i\frac{q}{2}\frac{w_{pp}}{kT} + \chi_{hp}\sigma\theta(1-\theta)f_e + \theta f_e\sigma\frac{q}{2}\frac{w_{hh}}{kT} + (1-\theta)f_e\sigma\frac{q}{2}\frac{w_{pp}}{kT} + \theta f_e(1-\sigma)q\frac{w_{hs}}{kT} + (1-\theta)f_e(1-\sigma)q\frac{w_{ps}}{kT} + f_ix\ln x + f_i(1-x)\ln(1-x) + f_e\theta\ln\theta + f_e(1-\theta)\ln(1-\theta) \quad \text{(A33)}$$

The Folded State. We can determine the equilibrium condensed state (the "native" state in the model) either by utilizing relationships among chemical potentials for h and p in the two "phases", i and e, or by finding the minimum in total free energy. The latter approach was used in earlier work; we develop the chemical potential approach here. The constraints in eq A11 can be expressed as

$$dn_{hi} = -dn_{he} = dn_{pe} = -dn_{pi}$$

Hence, the state of equilibrium is given by

$$\mu_{hi} - \mu_{he} = \mu_{pi} - \mu_{pe} \quad \text{(A34)}$$

These chemical potentials are obtained from the appropriate derivatives of the free energy given in eq A32:

$$\frac{\mu_{hi}}{kT} = \left(\frac{\partial F/kT}{\partial n_{hi}}\right)_{T,n_{he},n_{pe},n_{pi}} = \ln x + \chi_{hp}(1-x)^2 + \frac{qw_{hh}}{2kT}$$

$$\frac{\mu_{he}}{kT} = \left(\frac{\partial F/kT}{\partial n_{he}}\right)_{T,n_{hi},n_{pi},n_{pe}} = \ln\theta + \sigma\chi_{hp}(1-\theta)^2 + \sigma q\frac{w_{hh}}{2kT} + (1-\sigma)\frac{qw_{hs}}{kT}$$

$$\frac{\mu_{pi}}{kT} = \left(\frac{\partial F/kT}{\partial n_{pi}}\right)_{T,n_{hi},n_{pe},n_{he}} = \ln(1-x) + \chi_{hp}x^2 + \frac{qw_{pp}}{2kT}$$

$$\frac{\mu_{pe}}{kT} = \left(\frac{\partial F/kT}{\partial n_{pe}}\right)_{T,n_{hi},n_{pi},n_{he}} = \ln(1-\theta) + \sigma\chi_{hp}\theta^2 + \sigma q\frac{w_{pp}}{2kT} + (1-\sigma)\frac{qw_{ps}}{kT} \quad \text{(A35)}$$

Substitution of these chemical potentials into the equilibrium expression, eq A34, permits prediction of the equilibrium state of "ordering" of the residues, θ^* and x^*. Thus the folded state is described by the values θ^* and x^* that satisfy

$$0 = \chi_{hp}[(1 - 2x^*) - \sigma(1 - 2\theta^*)] + (1 - \sigma)(\chi_{ps} - \chi_{hs}) + \ln \frac{x^*(1 - \theta^*)}{\theta^*(1 - x^*)} \quad \text{(A36)}$$

and satisfy the constraint equation (eq 8) in the main text. In the approximation adopted earlier that $\chi_{hp} \approx \chi_{hs} \equiv \chi$ and $\chi_{ps} \approx 0$, the state of equilibrium is predicted by

$$0 = \ln \frac{x^*(1 - \theta^*)}{\theta^*(1 - x^*)} - 2\chi(x^* - \sigma\theta^*) \quad \text{(A37)}$$

A small error appeared in the earlier treatment (Dill, 1985); eq 24 of Dill (1985) should read

$$0 = \ln \frac{\Psi^*_{hi}\Psi^*_{pe}}{\Psi^*_{he}\Psi^*_{pi}} - \Phi\epsilon\left(\frac{\Psi^*_{hi}}{f_i} - \frac{\sigma\Psi^*_{he}}{f_e}\right)$$

(which is identical with eq 12 and A37) rather than

$$0 = \ln \frac{\Psi^*_{hi}\Psi^*_{pe}}{\Psi^*_{he}\Psi^*_{pi}} - \Phi^2\epsilon\left(\frac{\Psi^*_{hi}}{f_i} - \frac{\sigma\Psi^*_{he}}{f_e}\right)$$

This error has a negligible effect on the predictions given earlier.

Change in Free Energy upon Rearrangement: Pathway II. The difference in free energy from the randomly condensed state ($\theta = x = \Phi$) to the folded state ($\theta = \theta^*$, $x = x^*$) is calculated with eq A33:

$$\frac{\Delta F_{II}}{nkT} = \frac{F_{II}(\theta^*,x^*) - F_{II}(\Phi,\Phi)}{nkT} = \chi_{hp}(f_i[x^*(1 - x^*) - \Phi(1 - \Phi)] + f_e\sigma[\theta^*(1 - \theta^*) - \Phi(1 - \Phi)]) + (\theta^* - \Phi)f_e(1 - \sigma)(\chi_{hs} - \chi_{ps}) + f_i\left[x^* \ln \frac{x^*}{\Phi} + (1 - x^*) \ln \frac{1 - x^*}{1 - \Phi}\right] + f_e\left[\theta^* \ln \frac{\theta^*}{\Phi} + (1 - \theta^*) \ln \frac{1 - \theta^*}{1 - \Phi}\right] \quad \text{(A38)}$$

In the limit in which $\chi_{hp} \approx \chi_{hs} \equiv \chi$ and $\chi_{ps} \approx 0$, this reduces to eq 9 and 10 in the main text:

$$\frac{\Delta F_{II}}{nkT} = -\chi[f_i(x^{*2} - \Phi^2) + \sigma f_e(\theta^{*2} - \Phi^2)] + f_i\left[x^* \ln \frac{x^*}{\Phi} + (1 - x^*) \ln \frac{1 - x^*}{1 - \Phi}\right] + f_e\left[\theta^* \ln \frac{\theta^*}{\Phi} + (1 - \theta^*) \ln \frac{1 - \theta^*}{1 - \Phi}\right] \quad \text{(A39)}$$

References

Baldwin, R. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8069–8072.

Becktel, W. J., & Schellman, J. A. (1987) *Biopolymers 26*, 1859–1877.

Ben-Naim, A. (1980) *Hydrophobic Interactions*, Plenum Press, New York.

Brandts, J. F. (1964) *J. Am. Chem. Soc. 86*, 4302–4314.

Cho, K. C., Poon, H. J., & Chey, C. L. (1982) *Biochim. Biophys. Acta 701*, 206–215.

Christensen, J. K. (1952) *C. R. Trav. Lab. Carlsberg, Ser. Chim. 28*, 37–169.

Creighton, T. E. (1985) *J. Phys. Chem. 89*, 2452–2459.

Dill, K. A. (1985) *Biochemistry 24*, 1501–1509.

Dill, K. A., & Alonso, D. O. V. (1988) in *Colloquium-Mosbach der Gesellschaft für Biologische Chemie* (Huber, T., & Winnacker, E. L., Eds) *Protein Structure and Protein Engineering*, Vol. 39, pp 51–58, Springer-Verlag, Berlin.

Edelhoch, H., & Osborne, J. C. (1976) *Adv. Protein Chem. 30*, 183–250.

Edsall, J. T., & McKenzie, H. A. (1983) *Adv. Biophys. 16*, 53–183.

Frank, H. S., & Evans, M. W. (1945) *J. Chem. Phys. 13*, 507–532.

Freire, E., & Biltonen, R. L. (1978) *Biopolymers 17*, 463–479.

Gill, S. J., & Wadsö, I. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 2955–2958.

Gill, S. J., Nichols, N. F., & Wadsö, I. (1976) *J. Chem. Thermodyn. 8*, 445–452.

Hill, T. L. (1960) *An Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, MA.

Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1–63.

Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Nojima, H., Ikai, A., Oshima, T., & Noda, H. (1977) *J. Mol. Biol. 116*, 429–442.

Nozaki, Y., & Tanford, C. (1971) *J. Biol. Chem. 246*, 2211–2217.

Pace, C. N. (1975) *CRC Crit. Rev. Biochem. 3*, 1–43.

Pace, C. N., & Tanford, C. (1968) *Biochemistry 7*, 198–208.

Prasad, K. P., & Ahluwalia, J. C. (1976) *J. Solution Chem. 5*, 491–507.

Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167–241.

Privalov, P. L., & Khechinashvili, N. N. (1974) *J. Mol. Biol. 86*, 665–684.

Privalov, P. L., & Gill, S. J. (1988) *Adv. Protein Chem. 39*, 191–234.

Privalov, P. L., Griko, Y. V., Venyaminov, S. Y., & Kutyshenko, V. P. (1986) *J. Mol. Biol. 190*, 487–498.

Radzicka, A., & Wolfenden, R. (1988) *Biochemistry 27*, 1664–1670.

Shiao, D. F., Lumry, R., & Fahey, J. (1971) *J. Am. Chem. Soc. 93*, 2024–2035.

Shortle, D., Meeker, A. K., & Freire, E. (1988) *Biochemistry 27*, 4761–4768.

Spink, C. H., & Wadsö, I. (1975) *J. Chem. Thermodyn. 7*, 561–572.

Tanford, C. (1962) *J. Am. Chem. Soc. 84*, 4240–4247.

Tanford, C. (1980) *The Hydrophobic Effect*, Wiley, New York.