Theory for the Aggregation of Proteins and Copolymers

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY · MAY 1992

Impact Factor: 2.78 · DOI: 10.1021/j100189a013

4 AUTHORS, INCLUDING:

Gregg B Fields
The Scripps Research Institute
237 PUBLICATIONS **8,793** CITATIONS
SEE PROFILE

Ken A Dill
Stony Brook University
389 PUBLICATIONS **27,769** CITATIONS
SEE PROFILE

Theory for the Aggregation of Proteins and Copolymers†

Gregg B. Fields,‡ Darwin O. V. Alonso, Dirk Stigter, and Ken A. Dill*

Department of Pharmaceutical Chemistry, University of California, 3333 California Street #102, San Francisco, California 94118 *(Received: September 26, 1991)*

We develop mean-field lattice statistical mechanics theory for the equilibrium between denatured and aggregated states of proteins and other random copolymers of hydrophobic and polar monomers in aqueous solution. We suppose that the aggregated state is a mixture of amorphous polymer plus solvent and that the driving forces are the hydrophobic interaction, which favors aggregation, and conformational and translational entropies, which favor disaggregation. The theory predicts that the phase diagram for thermal aggregation is an asymmetric closed loop, and for denaturants (guanidinium hydrochloride or urea) it is asymmetric with an upper consolute point. The theory predicts that a copolymer in a poor solvent will expand with increasing polymer concentration because of "screening" of the solvent interactions by the other chains; the chain ultimately reaches a theta-like state in the absence of solvent. The screening concentration depends strongly on the copolymer composition. We find two striking features of these copolymer phase diagrams. First, they are extraordinarily sensitive to the copolymer composition; a change of one amino acid can substantially change the aggregation behavior. Second, relative to homopolymers, copolymers should be stable against aggregation at concentrations that are higher by many orders of magnitude.

Protein precipitation is important in many processes, from cheese making in the food industry to disease states such as Alzheimer's disease and the growth of cataracts in eyes. Moreover, the aggregation of proteins is currently a major problem in biotechnology, often restricting the study or purification of bioengineered molecules. Aggregated proteins are often biologically inactive; proteins that are overexpressed in some cell types form "inclusion bodies".[1-4] Hence much effort has gone into developing strategies that avoid aggregation. Despite these efforts, such processes remain poorly understood. The present work aims to provide a simple physical model of protein aggregation.

Protein aggregation may be reversible or irreversible. We consider here only the equilibrium processes of aggregation and not the factors that control their rates. We do not consider the intermolecular covalent cross-linking processes that contribute to aggregation among disulfide-bonded proteins. We also do not consider protein crystallization. The processes of interest here have been referred to as "heat aggregation".[5] These appear to involve the association of denatured chains through hydrophobic interactions. For example, studies on the aggregation of bovine growth hormone by static and dynamic light scattering, size exclusion HPLC, fluorescence and second-derivative absorption spectroscopy, and circular dichroism;[6-10] of lysozyme by poly(acrylamide) gel electrophoresis,[11] absorption spectroscopy,[11] and dynamic light scattering;[12] of bovine carbonic anhydrase B by quasi-elastic light scattering;[13] and of P22 tailspike endorhamnosidase by poly(acrylamide) gel electrophoresis[14,15] support a model of intermolecular association driven by hydrophobic interactions between partially denatured protein chains.[4] Although there is not yet much evidence to show what conformations the individual protein molecules adopt in the aggregated state, we assume that they are simply highly entangled chains, as are found in highly concentrated bulk phases of amorphous polymer molecules. Our purpose here is to develop a simple model for the equilibrium among copolymer molecules denatured as individual molecules, and aggregated in the entangled state as a function of the temperature and solvent character and as a function of the copolymer composition and chain length.

Theory

It is well-known that homopolymers in a solution may separate into two phases containing a more dilute and a more concentrated polymer solution.[16] Here we treat the precipitation of denatured protein as a phase separation in a copolymer solution. Our aim is to determine how this phase separation depends on (i) the properties of the copolymer including chain length and hydrophobic composition, and on (ii) external variables such as temperature or denaturants in the solution. We use a mean-field lattice model described previously[17-19] for single denatured protein molecules. To describe the equilibrium between isolated and aggregated molecules, we add a term to account for the translational entropy following the treatment of Post and Zimm[20-23] based on the Flory-Huggins theory.[16] Provided that aggregation involves no native molecules, the native state need not be explicitly included in the calculation of the aggregation phase equilibrium. However, we believe that the theory presented here can be readily extended to treat situations in which native molecules are also involved in the aggregation equilibrium.

Each denatured protein molecule is represented as a copolymer on a lattice, composed of n "lattice monomers" (i.e., the number of amino acid residues divided by 1.4[17]). The number of protein molecules is N and the number of solvent molecules is n_s. Thus,

* Author to whom correspondence should be addressed.

† We dedicate this paper to Marshall Fixman on the occasion of his sixtieth birthday, with great appreciation for his profound contributions to physical chemistry.

‡ Present address: Department of Laboratory Medicine and Pathology and The Biomedical Engineering Center, University of Minnesota, Box 107, 420 Delaware Street S.E., Minneapolis, MN 55455.

(1) Marston, F. A. O. *Biochem. J.* **1986**, *240*, 1.
(2) Taylor, G.; Hoare, M.; Gray, D. R.; Marston, F. A. O. *Bio/Technology* **1986**, *4*, 553.
(3) Weir, M. P.; Sparks, J. *Biochem. J.* **1987**, *245*, 85.
(4) Mitraki, A.; King, J. *Bio/Technology* **1989**, *7*, 690.
(5) Jaenicke, R. *J. Polym. Sci. C* **1967**, *16*, 2143.
(6) Havel, H. A.; Kauffman, E. W.; Plaisted, S. M.; Brems, D. N. *Biochemistry* **1986**, *25*, 6533.
(7) Brems, D. N.; Plaisted, S. M.; Kauffman, E. W.; Havel, H. A. *Biochemistry* **1986**, *25*, 6539.
(8) Brems, D. N. *Biochemistry* **1988**, *27*, 4541.
(9) Brems, D. N.; Plaisted, S. M.; Havel, H. A.; Tomich, C.-S.-C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 3367.
(10) Brems, D. N.; Havel, H. A. *Proteins: Struct., Funct., Genet.* **1989**, *5*, 93.
(11) Wetzel, R.; Perry, L. J.; Mulkerrin, M. G.; Randall, M. In *Protein Design and the Development of New Therapeutics and Vaccines*; Hook, J. B., Poste, G., Eds.; Plenum: New York, 1990; pp 79.
(12) Wang, C.-C.; Cook, K. H.; Pecora, R. *Biophys. Chem.* **1980**, *11*, 439.
(13) Cleland, J. L.; Wang, D. I. C. *Biochemistry* **1990**, *29*, 11072.
(14) King, J. *Bio/Technology* **1986**, *4*, 297.
(15) Haase-Pettingell, C. A.; King, J. *J. Biol. Chem.* **1988**, *263*, 4977.
(16) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1953.
(17) Dill, K. A. *Biochemistry* **1985**, *24*, 1501.
(18) Dill, K. A.; Alonso, D. O. V.; Hutchinson, K. *Biochemistry* **1989**, *28*, 5439.
(19) Alonso, D. O. V.; Dill, K. A. *Biochemistry* **1991**, *30*, 5974.
(20) Post, C. B.; Zimm, B. H. *Biopolymers* **1979**, *18*, 1487.
(21) Post, C. B.; Zimm, B. H. *Biophys. J.* **1980**, *32*, 448.
(22) Post, C. B.; Zimm, B. H. *Biopolymers* **1982**, *21*, 2123.
(23) Post, C. B.; Zimm, B. H. *Biopolymers* **1982**, *21*, 2139.

0022-3654/92/2096-3974$03.00/0

the total number of lattice sites occupied by the copolymer solution is

$$M = Nn + n_s \tag{1}$$

The chain is taken to be a random copolymer of two types of monomers: h (hydrophobic) and p (hydrophilic). The number of hydrophobic residues per chain is $n\Phi$; hence, the number of hydrophilic residues is $n(1 - \Phi)$.

We aim to calculate the free energy as a function of the copolymer concentration. We assume that the aggregated state is favored relative to the denatured state of isolated chains by advantageous association of the nonpolar monomers with each other in order to avoid contact with the aqueous solvent; the driving force is the contact free energy (E). The translational entropy of mixing (S_{mix}) favors the dispersion of polymer in the solvent relative to the aggregated state. In addition, the copolymer chain may adopt a different radius of gyration in the aggregated state relative to that in the dilute solution, since it will be in a solvent of a different character in the two different cases. Hence, there may also be a difference in conformational entropy (S_{conf}). Thus, the free energy of the solution will be described by

$$F = E - T(S_{mix} + S_{conf}) \tag{2}$$

First we compute the contact free energy per lattice site (E/MkT) as a function of the volume fraction of copolymer in the solution, $\psi = Nn/M$. We need a treatment that applies over the full range of copolymer concentration, from isolated single chains at low density to the high-density limit of pure copolymer. When the concentration of polymer is high, the Flory–Huggins theory,[16] generalized to account for copolymers of two types of monomers,[17,18] gives for the contact free energy term

$$E/MkT = -\chi\Phi^2\psi^2 \qquad \text{for } \psi \to 1 \tag{3}$$

where χ is the contact interaction parameter (here we take it to include both energetic and entropic contributions). Linear terms in ψ are left out here because they become unimportant constants in the chemical potential expressions later. On the other hand, when the copolymer is highly diluted, it follows from the collapse theory of isolated chains[17-19] that the contact free energy is

$$E/MkT = -\chi\Phi^2(f_i + \sigma f_e)\rho\psi \qquad \text{for } \psi \to 0 \tag{4}$$

where f_i is the fraction of residues in the copolymer interior, f_e is the fraction of residues at the copolymer exterior, σ is the fraction of the area of an exterior residue that makes contact with other residues (taken[17] to be equal to 2/3), and ρ is the chain density (i.e., the n chain segments are configured in a molecular sphere of n/ρ lattice sites; for the folded state, $\rho = 1$). The factor of ψ in this expression, multiplied by n, just sums the contact interaction over a single isolated chain in solution. The factors f_i and f_e account for "surface" effects of the isolated chain, i.e., the ability to segregate the two monomer types between the surface and interior; we assume the aggregated state is a macroscopic phase so large that surface effects are negligible.

Now to account for the range of copolymer concentrations between these two extremes, we suppose that the solution outside a particular chain is a mean-field of solvent plus segments of the other chains. It is shown in the Appendix that this leads to the following expression for the contact free energy for any ψ:

$$E/MkT = -\chi\Phi^2[\psi^2 + (f_i + \sigma f_e)\rho\psi(1 - \psi)^2] \tag{5}$$

This expression leads to eqs 3 and 4 above in the appropriate limits.

Next we consider the entropies of conformational change of the chains as a function of dilution. We follow Post and Zimm[22] in assuming that there are two separable degrees of freedom: the intramolecular conformational change of a chain, characterized here by changes in the density ρ, and the intermolecular "mixing" degree of freedom due to changes in the concentration ψ. Thus, as the copolymer concentration increases, the chains can expand or contract because the effective solvent character changes in the dilution process. The intramolecular conformational entropy is the sum of elastic and excluded volume terms[17,18,24]

$$\frac{S_{conf}}{Mk} = \psi\left[-\frac{7}{2n}\left(\frac{\rho_0}{\rho}\right)^{2/3} - \frac{2}{n}\ln\rho - \frac{(1-\rho)}{\rho}\ln(1-\rho)\right] \tag{6}$$

where ρ_0 is the density of the theta state, discussed below.

The intermolecular translational entropy is taken from the Flory–Huggins theory[16] for the mixing of a polymer with solvent:

$$\frac{S_{mix}}{Mk} = -\frac{\psi}{n}\ln\psi - (1-\psi)\ln(1-\psi) \tag{7}$$

The mixing entropy accounts for the number of ways to insert the polymers into fixed positions into the solution whereas the conformational entropy accounts for the change in compactness of each such polymer molecule at the given position. The Flory–Huggins mixing entropy is assumed independent of conformation, but is subject to the adequacy of the mean-field approximation that the chain monomers are uniformly distributed in the solution.

Substituting eqs 5, 6, and 7 into eq 2 gives the free energy of the protein solution:

$$\frac{F}{MkT} = -\chi\Phi^2[\psi^2 + (f_i + \sigma f_e)\rho\psi(1-\psi)^2] + \psi\left[\frac{7}{2n}\left(\frac{\rho_0}{\rho}\right)^{2/3} + \frac{2}{n}\ln\rho + \frac{(1-\rho)}{\rho}\ln(1-\rho)\right] + \frac{\psi}{n}\ln\psi + (1-\psi)\ln(1-\psi) \tag{8}$$

where, as in eq 6, we have omitted the terms proportional to ψ because they are thermodynamically irrelevant for our purposes.

Considerable simplification arises because of the appearance of χ and Φ only in the combination $\chi\Phi^2$ in all the thermodynamically relevant terms (see eq 8). Thus, the behavior of copolymers is the same as for homopolymers except for a rescaling of the χ parameter.

The coil density ρ depends on the copolymer concentration. To determine the equilibrium density, ρ^*, we minimize the free energy with respect to ρ; that is, ρ^* is evaluated from

$$(\partial F/\partial\rho)_{\Phi,n,N,n_s} = 0 \qquad \text{at } \rho = \rho^* \tag{9}$$

In order to evaluate ρ^*, we require an expression for the dependence of ρ_0, the density of the expanded reference state, on chain length, n. For the collapse of an isolated chain, a simple scaling argument[25] has led to the use of[17,24,26] $\rho_0 = (19/27n)^{1/2} = 0.839/n^{1/2}$. For protein collapse between native and denatured states, both of which states are relatively compact, the exact value of this expression has not been critical.[17] However, for the present problem, it is of value to consider this equation in more detail. We now compute a more accurate value of ρ_0 based on the persistence length of real polypeptides. Thus we find for myoglobin with $n = 110$ "lattice" residues[27] $\rho_0 = 0.012$ and, hence, for proteins with arbitrary chainlength $\rho_0 = 0.012(110/n)^{1/2} = 0.126/n^{1/2}$. In general, ρ^* differs from ρ_0 except in theta solvents or in polymer melts, where the chain configurations should become identical to those of ideal random flights. It is shown in Figure 1, however, that the present model does not lead to identity of ρ_0 and ρ^* in the bulk ($\psi = 1$). However, for common proteins, which range from $n = 100$–1000, the new choice, $\rho_0 = 0.126/n^{1/2}$, performs much better than the earlier choice of $\rho_0 = 0.839/n^{1/2}$.

Figure 2 shows another important difference between the two choices of reference state density, ρ_0. The collapse transition for isolated chains becomes much sharper when the polypeptide chain stiffness is modeled more accurately. This effect of "chain stiffness" in homopolymers has been previously pointed out by

(24) Dill, K. A.; Alonso, D. O. V. In *Protein Structure and Protein Engineering: Colloquium Mosbach*; Huber, T., Winnacker, E. L., Eds.; Springer-Verlag: Berlin, 1989; pp 51.
(25) Sanchez, I. C. *Macromolecules* **1979**, *12*, 980.
(26) Stigter, D.; Alonso, D. O. V.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 4176.
(27) Stigter, D.; Dill, K. A. *Biochemistry* **1990**, *29*, 1262.

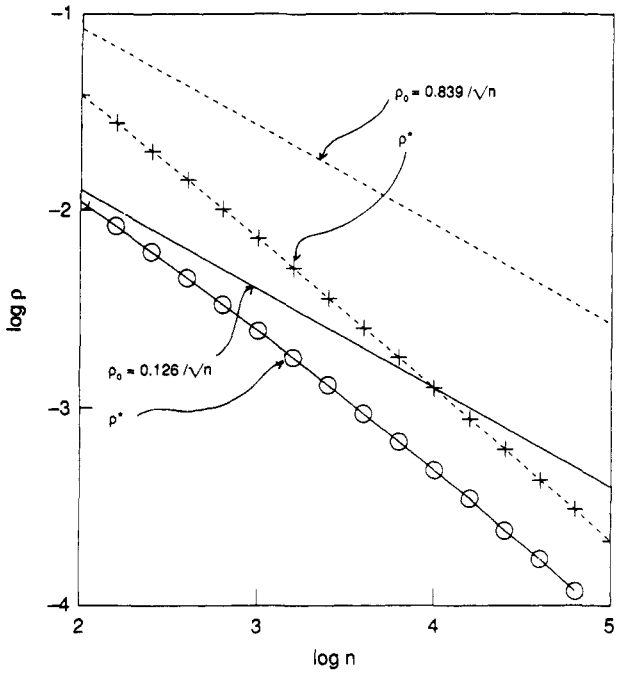

Figure 1. Coil density ρ versus n at $\psi = 1$. Reference state density ρ_0 and equilibrium density $\rho^*(\rho_0)$ for earlier choice of ρ_0(---, ×---×) and new choice of ρ_0(—, O—O).

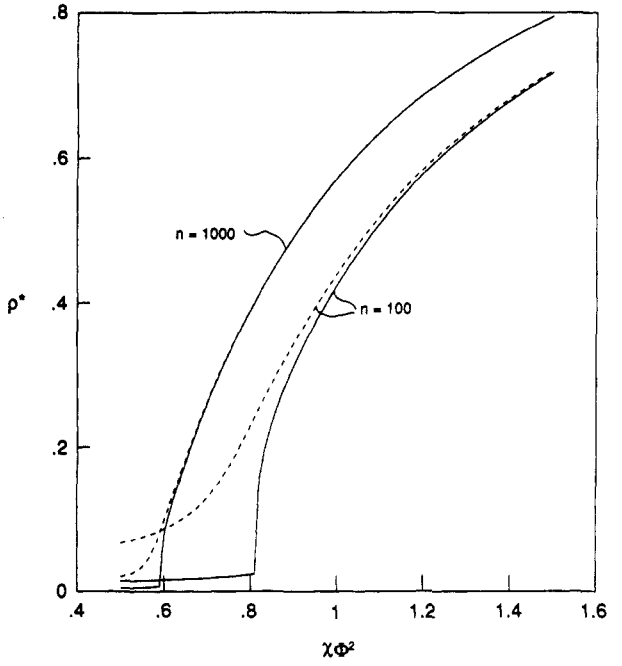

Figure 2. Equilibrium coil density ρ^* as a function of $\chi\Phi^2$ for $\psi = 0$, $n = 100$ or $n = 1000$ with reference state densities $\rho_0 = 0.839/n^{1/2}$ (---) and $\rho_0 = 0.126/n^{1/2}$ (—).

Post and Zimm.[20,21] As previously noted, however,[17] this effect is of diminishing importance in increasingly poor solvents; hence it is of less importance for some protein folding transitions, which take place between conformational states of relatively high chain density. The earlier expression $\rho_0 = 0.839/n^{1/2}$ gives a gradual transition of ρ^* with χ, consistent with the gradual change of ρ^* with temperature calculated for myoglobin.[28] On the other hand, the new, lower ρ_0 values give a sharp transition of ρ^* between compact and open coils for a χ value which decreases with increasing chain length. Such coil collapse is in accord with predictions by other theories, of Post and Zimm[20] and of Sanchez,[25] and with experiments reported for polystyrene in cyclohexane.[25,29]

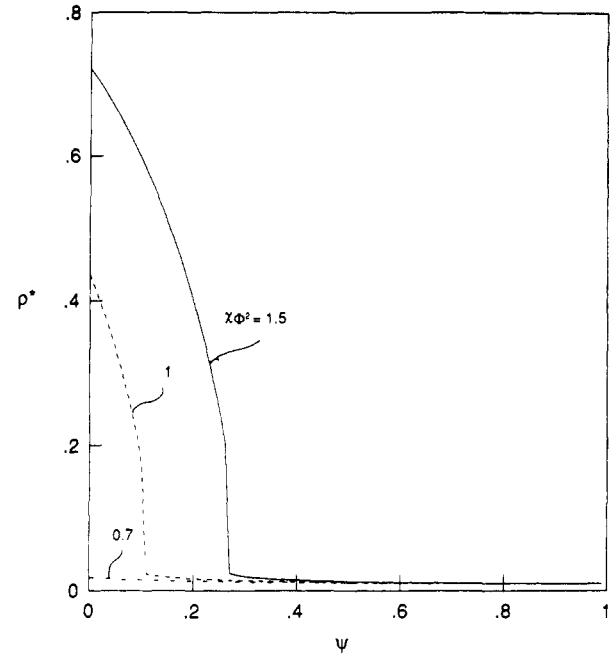

Figure 3. Equilibrium coil density ρ^* versus ψ for $n = 110$ with $\chi\Phi^2$ as indicated.

Thus $\rho_0 = 0.126/n^{1/2}$ is to be preferred and, unless indicated otherwise, further results in this paper are based on this ρ_0.

One of the main predictions of the theory is that increasing the concentration of copolymer in poor solvents causes the chains to expand because the interpenetrating neighboring chains "screen" out the unfavorable internal copolymer–solvent interactions. Figure 3 shows that ρ^* decreases as ψ, the copolymer concentration, increases, for various poor solvents and copolymer compositions. In particular, Figure 3 shows that in relatively good solvents, or when the fraction of solvent-averse monomers is small, the chains are expanded for all copolymer concentrations. However, as the solvent becomes poorer ($\chi\Phi^2$ increases), the chains are relatively compact at low copolymer concentrations, but they expand sharply with increasing copolymer concentration. There is an abrupt transition for chain expansion at a particular copolymer concentration, ψ^*. Higher polymer concentrations are required to achieve the expansion transition as the solvent becomes worse or the polymer becomes more solvophobic. It is shown below that this transition occurs in the two-phase region of ψ and, hence, is not experimentally accessible. Theory by Szleifer for homopolymers also predicts an increase of coil size with increasing ψ (see Figure 7 of ref 30). The present results suggest that high copolymer concentrations promote extended coils with close to random statistics and, consequently, with much intermingling and entanglement of neighbor chains. This might explain the difficulty of crystallizing many proteins from aqueous solution: entropy would favor their aggregation instead of crystallization.

Next we consider the phase equilibria predicted by the theory. Phase equilibria are readily described in terms of the chemical potentials. The chemical potential of the protein is obtained by taking the derivative of the free energy in eq 8 with respect to N:

$$\frac{\mu_{pr}}{kT} = \left[\frac{\partial(F/kT)}{\partial N}\right]_{\Phi,n,n_s} = -\chi\Phi^2 n[\psi + \rho^*(f_i + \sigma f_e)(1-\psi)^2] - \chi\phi^2 n(1-\psi)\psi a_1 + \log\psi + (n-1)\psi + \frac{7}{2}\frac{\rho_0^{2/3}}{\rho^{*2/3}} + 2\log\rho^* + n\left(\frac{1}{\rho^*} - 1\right)\log(1-\rho^*) + n(1-\psi)\psi a_2 \quad (10)$$

(28) Alonso, D. O. V.; Dill, K. A.; Stigter, D. *Biopolymers* **1991**, *31*, 1631.

(29) Swislow, G.; Sun, S.-T.; Nishio, I.; Tanaka, T. *Phys. Rev. Lett.* **1980**, *44*, 796.

(30) Szleifer, I. *J. Chem. Phys.* **1990**, *92*, 6940.

where

$$a_1 = 1 + \rho^*(f_i + \sigma f_c)(-2 + 2\psi) + \frac{d\rho^*}{d\psi}(f_i + \sigma f_c)(1 - \psi)^2 + \rho^*(1 - \psi)^2\frac{d\rho^*}{d\psi}(1 - \sigma)\left[-\left(\frac{4\pi}{3n}\right)^{1/3}\rho^{*-2/3} + 2\left(\frac{4\pi}{3n}\right)^{2/3}\rho^{*-1/3} - \frac{4\pi}{3n}\right] \quad (11)$$

$$a_2 = \frac{d\rho^*}{d\psi}\left[-\frac{7}{3n}\frac{\rho_0^{2/3}}{\rho^{*5/3}} - \frac{1}{\rho^*}\left[\frac{1}{\rho^*}\log(1 - \rho^*) + 1 - \frac{2}{n}\right]\right] \quad (12)$$

The chemical potential of the solvent is given by the derivative of F with respect to n_s:

$$\frac{\mu_s}{kT} = \left[\frac{\partial(F/kT)}{\partial n_s}\right]_{\Phi,n,N} = \chi\phi^2\psi^2 a_1 + \log(1 - \psi) + (1 - 1/n)\psi - \psi^2 a_2 \quad (13)$$

As a test of the computer program we have compared eqs 10 and 13 with $[(\Delta F/kT)/\Delta N]_{\Phi,n,n_s}$ and $[(\Delta F/kT)/\Delta n_s]_{\Phi,n,N}$ from eq 8 over the full range of ψ. Apart from a constant, the results always agreed to better than one part in 10^5.

To determine the phase equilibria, eqs 10 and 13 are solved as a function of ψ for given values of n, χ, and Φ. Phase boundaries are defined by the locus of pairs of values (ψ', ψ'') which simultaneously satisfy eqs 10 and 13:

$$\mu_{pr}(\psi') = \mu_{pr}(\psi'') \quad (14)$$

$$\mu_s(\psi') = \mu_s(\psi'') \quad (15)$$

Construction of the phase diagrams requires solution of the two nonlinear eqs 14 and 15 for ψ' and ψ''. There is no straightforward method for solving them. Our approach has been to inspect modified plots of μ_s and μ_{pr} versus ψ to locate regions of possible solutions, and then to use a double iteration method on a computer, using a rootsolver for ψ' nested in another rootsolver for ψ''. This gives numerical results that are accurate to better than 1 part in 10^5 for the values (ψ', ψ''). For very low polymer concentrations, the determination of ψ' and ψ'' can be decoupled as follows. In eq 13 we have $\mu_s \to 0$ for $\psi \to 0$. Therefore, for very small ψ', $<10^{-5}$, we may approximate eq 15 with $\mu_s(\psi'') = 0$ and solve this eq for ψ'' with the help of eq 13. For the resulting ψ'' we compute $\mu_{pr}(\psi'')$ from eq 10. According to eq 14, this value equals $\mu_{pr}(\psi')$, which result is then substituted into eq 10 to find ψ' by iteration.

Figure 4 shows the chemical potentials, μ_{pr} and μ_s versus ψ. These plots were modified by a vertical shift of the curves such that the first extremum was at $\mu_s = \mu_{pr} = 0$, followed by rescaling such that the second extremum was at $\mu_s = \mu_{pr} = 1$. Figure 5 shows the resulting curves for the parameter values $n = 110$, and $\chi\Phi^2 = 0.65$, with the rectangle giving the solution of eqs 14 and 15. We focus on the ψ values where $\mu = 0$ or 1, as indicated by the points A_p, A, B, and B_s. Inspection shows that eqs 14 and 15 can be satisfied simultaneously only for $\psi(A_p) < \psi' < \psi(A)$ and $\psi(B) < \psi'' < \psi(B_s)$. When the parameter $\psi\Phi^2$ is decreased, the points A and B move closer together until they merge at the critical point, for $\chi = \chi_c$ when $\psi' = \psi'' = \psi_c$ (compare Figures 5 and 6). As a further test of the computer program we have evaluated $F - (a + bN)$, where $a + bN$ is the straight line through $F(N')$ and $F(N'')$ with (N', N'') corresponding to (ψ', ψ''). Figure 7 shows that this line is indeed the common tangent to $F(N)$ as it should be.

In some of the calculations presented in the Results section, phase diagrams are constructed as $\psi\Phi^2$ versus copolymer concentration ψ. However, in modeling proteins it is generally of more practical value to have phase diagrams of either temperature versus copolymer concentration or of urea or guanidinium hydrochloride (GuHCl) denaturant concentration versus copolymer concentration. In order to compute either of these latter types of diagrams, we need to know the functional dependence of χ on temperature, or on the concentration of denaturing agent. For this purpose,

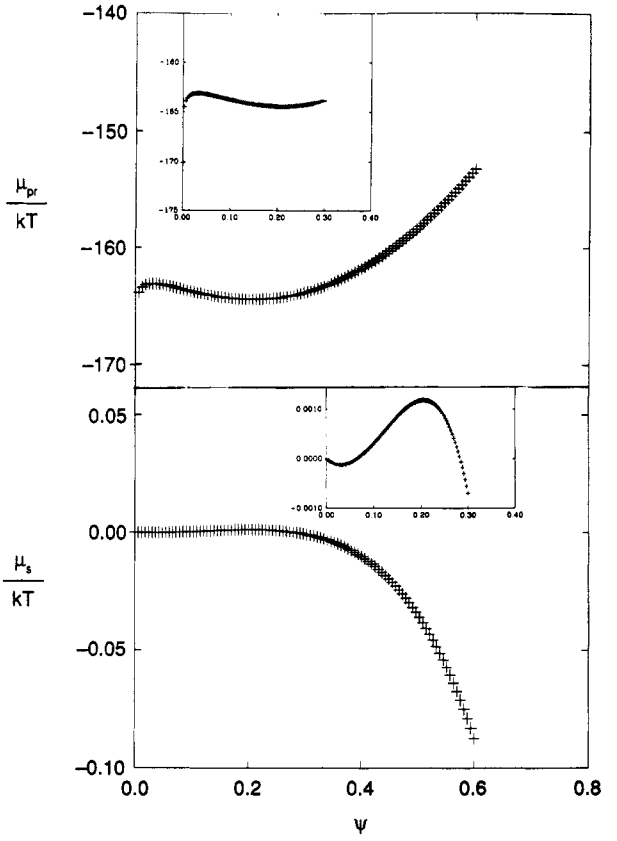

Figure 4. Chemical potentials of protein, μ_{pr}, and of solvent, μ_s, as a function of ψ.

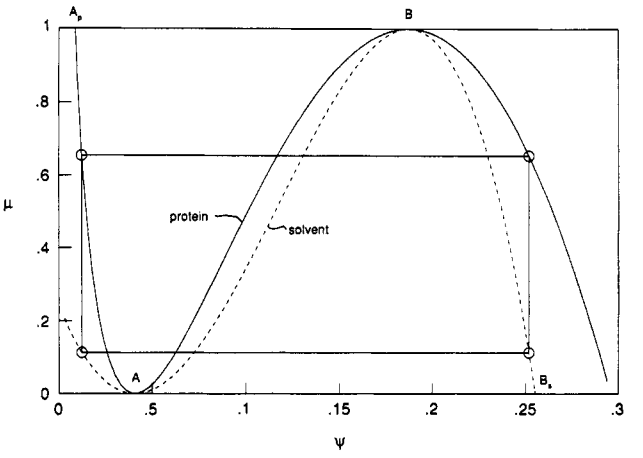

Figure 5. Shifted and rescaled chemical potentials of copolymer and of solvent versus ψ for $n = 110$ and $\chi\Phi^2 = 0.65$. Solution of eqs 14 and 15 given by rectangle.

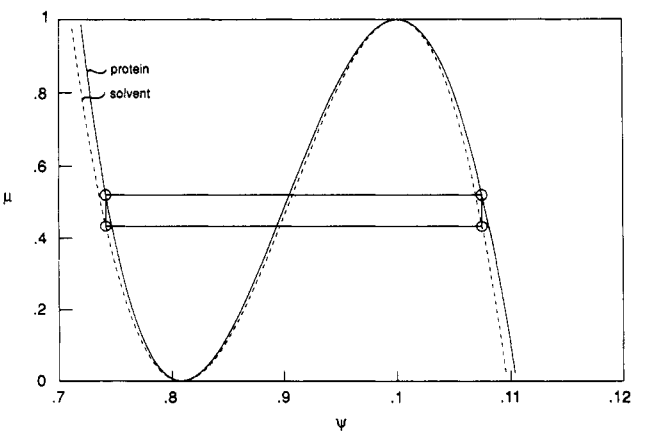

Figure 6. Same as Figure 5 for $\chi\Phi^2 = 0.614$.

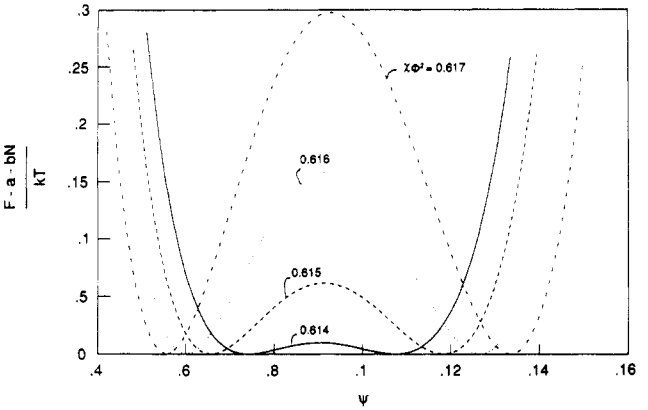

Figure 7. Nonlinear part of solution free energy, $(F - a - bN)/kT$, versus ψ for $n = 110$, $n_s = 10^5$, and $\chi\Phi^2$ as indicated.

we use previously derived relationships. For the thermal phase diagram, we use the expression[18]

$$\chi = \frac{1.4}{kT}[\Delta H^\circ + \Delta C_p(T - T_0) - T[\Delta S^\circ + \Delta C_p \ln (T/T_0)]] \quad (16)$$

where $T_0 = 298$ K, $\Delta S^\circ = -6.7$ cal K^{-1} mol^{-1}, $\Delta H^\circ = 0$, and $\Delta C_p = 55$ cal K^{-1} mol^{-1}. This expression is obtained from parametrization of the partition coefficients of nonpolar amino acids between water and ethanol or *p*-dioxane measured by Nozaki and Tanford,[31] and of small molecule nonpolar solutes measured by Gill and Wadsö.[32] For the urea phase diagrams, we use the expression[19]

$$\chi = \frac{1.4}{kT}[g_{water} + g_1[\text{urea}]] \quad (17)$$

where $T = 298$ K, $g_{water} = 2000$ cal mol^{-1}, and $g_1 = -36.8$ cal M^{-1} mol^{-1}. This was derived by a similar parametrization of the transfer of nonpolar amino acids between oil and aqueous urea solutions measured by Nozaki and Tanford.[31,33] For the GuHCl phase diagrams below, we use the expression[19]

$$\chi = \frac{1.4}{kT}[g_{water} + g_1[\text{GuHCl}] + g_2[\text{GuHCl}]^2] \quad (18)$$

where $T = 298$ K, $g_1 = -111.0$ cal M^{-1} mol^{-1}, and $g_2 = 5.76$ cal M^{-2} mol^{-1}, also parametrized from the data of Nozaki and Tanford.[31,34]

Results and Discussion

Predicted phase diagrams ($\chi\Phi^2$ versus ψ) for the stable phases are shown in Figure 8 for copolymers of different chain lengths ($n = 110$ and $n = 1000$). For comparison, Figure 8 shows also the Flory–Huggins result for homopolymers with $n = 110$. These phase diagrams all show upper critical solution behavior and have an asymmetry characteristic of polymers in solvents of small molecules; i.e., the critical point occurs at a relatively small volume fraction of polymer. We find, as is well-known from the Post and Zimm[22] and Flory–Huggins[35] theories, that the two-phase region narrows and peaks at lower polymer concentrations as the chain length increases, for fixed composition. Compared with the Flory–Huggins results, the present theory disfavors aggregation, corresponding to a change of about 2% in $\chi\Phi^2$ in the example shown here. The difference is attributable to our accounting for the freedom for chain conformational change upon dilution, neglected in the Flory–Huggins theory. Figure 8 also shows that aggregation is favored (i.e., the two-phase region enlarges) when

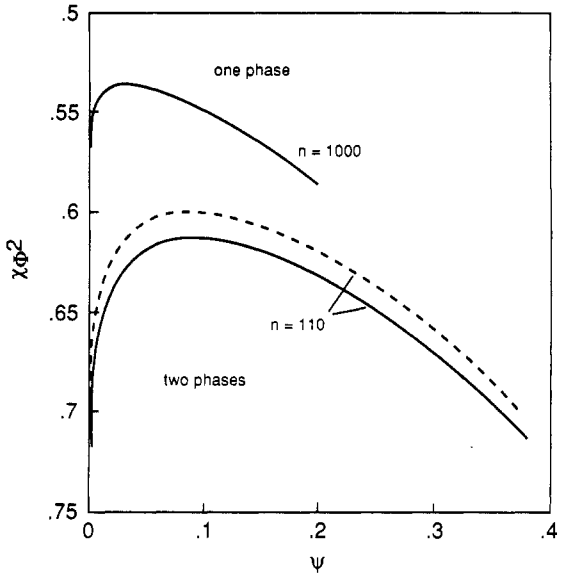

Figure 8. $\chi\Phi^2$ vs ψ phase diagrams of present theory (—) for $n = 110$ and $n = 1000$ compared with Flory–Huggins theory (- - -) for $n = 110$.

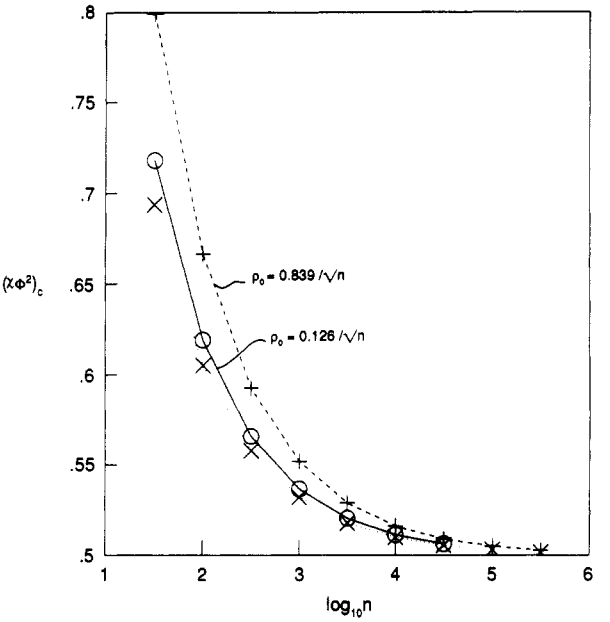

Figure 9. Critical values of $\chi\Phi^2$ for phase separation as a function of n for present theory with old (+) and new (○) reference state coil density ρ_0 compared with theory of Flory–Huggins (×).

the strength of the nonpolar interaction, χ, is increased or when the nonpolar fraction, Φ, of the chain is increased. Because of the Φ^2 dependence, the phase boundaries can undergo remarkably large shifts with very small changes in copolymer composition. The practical consequence of this sensitivity is that the change of a single amino acid in a protein could cause a protein to precipitate, or could cause an insoluble protein to become soluble. The present theory predicts about the same width of the coexistence region as the Flory–Huggins theory.

We now consider how the upper critical solution points vary with chain length. In Figure 9 we compare $(\chi\Phi^2)_c$ versus log n for three models: the present theory with the two choices of the reference state density, and the Flory–Huggins theory. All the models approach the classical Flory–Huggins limit, $\chi_c = 0.5$ for homopolymers ($\Phi = 1$) in the limit of large n. Figure 10 shows the critical polymer concentration, log ψ_c, versus log n for the three models. The slope of this curve from the present theory is nearly the same as the value of about 0.5 of the Flory–Huggins theory. Both Figures 9 and 10 show that the critical behavior with the new ρ_0 is closer to Flory–Huggins than with the earlier expression for ρ_0. Also plotted in Figure 10 are the results of Szleifer[30] (taken

(31) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1971**, *246*, 2211.
(32) Gill, S. J.; Wadsö, I. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2955.
(33) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1963**, *238*, 4074.
(34) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1970**, *245*, 1648.
(35) Shultz, A. R.; Flory, P. J. *J. Am. Chem. Soc.* **1952**, *74*, 4760.

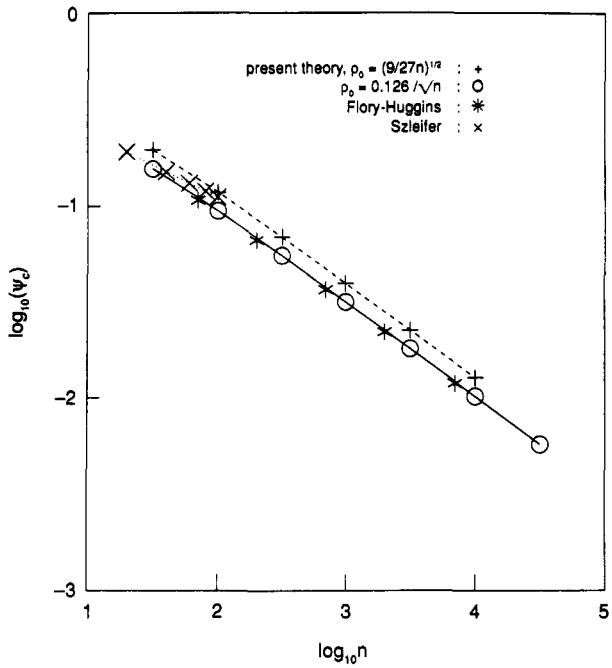

Figure 10. Copolymer concentration at critical phase separation, ψ_c, of present theory for the two ρ_0 values compared with results of Flory–Huggins[16] and Szleifer.[30]

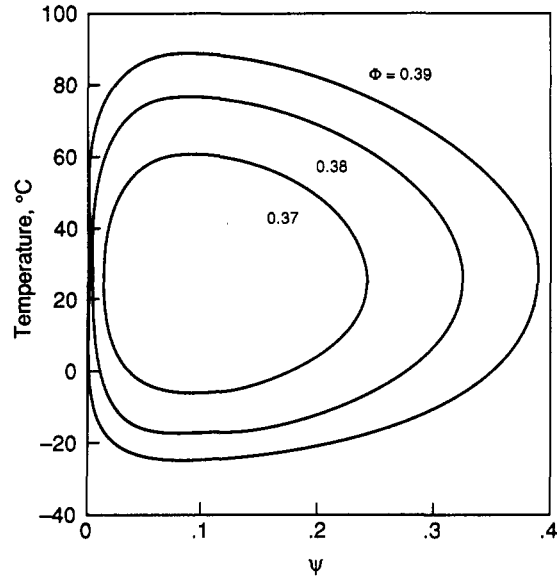

Figure 11. Phase diagram of temperature (T) versus ψ for n = 110 and Φ = 0.37, 0.38, or 0.39.

from Figure 3 of his paper) which indicate a slope close to the experimental value 0.38.

Figure 11 shows the predictions for thermal phase diagrams, for which we have used eq 16 for the temperature dependence of χ. The theory predicts a closed-loop phase diagram with both upper and lower consolute temperatures. The upper phase boundary indicates that heating a two-phase solution leads to mixing. This arises because heating favors the gain of translational entropy due to the dispersion of polymer. At these high temperatures, the temperature dependence of this translational entropy component is more important than the strengthened hydrophobic interaction that also results from heating. The lower phase boundary similarly arises from two competing effects. On the one hand, cooling (at low temperatures) leads to a loss of the translational entropy; this component should favor aggregation. However, more important at these low temperatures is the weakening of the hydrophobic interaction upon cooling; this favors dispersion of the polymer in the solvent. Hence, in principle, proteins could be made soluble by cooling below this temperature.

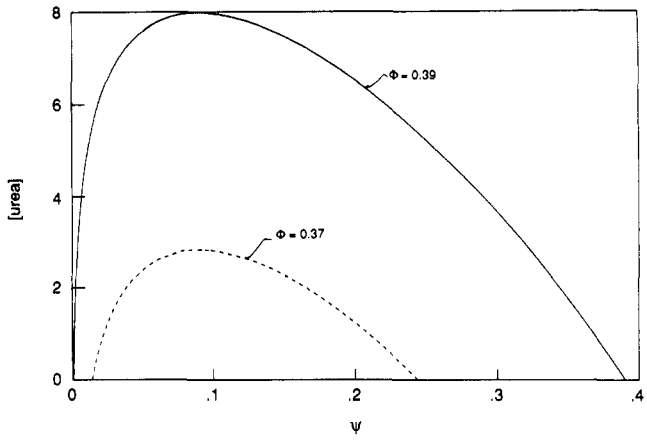

Figure 12. Phase diagram of [urea] versus ψ for n = 110 and Φ = 0.39 or 0.37.

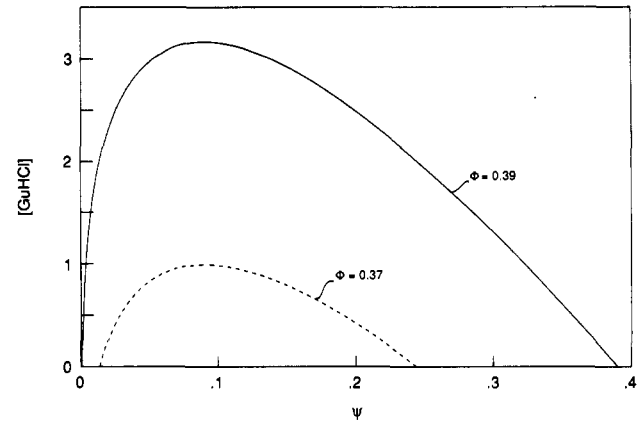

Figure 13. Phase diagram of [GuHCl] versus ψ for n = 110 and Φ = 0.39 or 0.37.

Note that this may often be impractical, however, since this temperature is predicted to be below the freezing point of water. Figure 11 also shows the prediction that changing the nonpolar composition by only a few percent (a few monomers) can change the aggregation behavior of a protein dramatically. Thermal phase diagrams have been obtained for several proteins, including lysozyme,[36,37] arachin,[38] and γ-crystallin.[39,40]

Figures 12 and 13 show that copolymer precipitation can also be controlled by denaturants. These properties are of interest because denaturants are used to solubilize proteins in inclusion bodies.[1,3,4] For both urea (Figure 12) and GuHCl (Figure 13), the theory predicts upper critical solution behavior. Above a certain concentration of denaturant, the hydrophobic interaction is weak relative to the translational entropy, so the molecules disaggregate. At lower denaturant concentrations, aggregation into a two-phase solution is driven by the relatively strong nonpolar interactions. These phase diagrams indicate that aggregated protein could be dissolved by first increasing the denaturant concentration, then diluting the protein, and then removing the denaturant, provided there are no kinetic barriers. As for thermal aggregation, the denaturant-mediated disaggregation is also predicted to be very sensitive to amino acid composition. For example Figure 12 shows that replacing only two or three polar amino acids by nonpolar ones can lead to a need for 8 M urea to disaggregate rather than 3 M urea.

(36) Ishimoto, C.; Tanaka, T. *Phys. Rev. Lett.* **1977**, *39*, 474.
(37) Taratuta, V. G.; Holschbach, A.; Thurston, G. M.; Blankschtein, D.; Benedek, G. B. *J. Phys. Chem.* **1990**, *94*, 2140.
(38) Tombs, M. P.; Newsom, B. G.; Wilding, P. *Int. J. Peptide Protein Res.* **1974**, *6*, 253.
(39) Thomson, J. A.; Schurtenberger, P.; Thurston, G. M.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7079.
(40) Broide, M. L.; Berland, C. R.; Pande, J.; Ogun, O. O.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5660.

Consistent with the theory, extreme sensitivity of GuHCl-induced aggregation to nonpolar content has been observed for bovine growth hormone (bGH).[9,41] For that molecule, lysine-112 of native bGH was changed to leucine by single-site mutagenesis. At a given GuHCl concentration, the more hydrophobic mutant bGH precipitated at much lower protein concentrations than the wild type bGH. In a separate experiment, eight residues were mutated to give a slightly more hydrophilic bGH. It was observed to precipitate at higher protein concentration than the wild type bGH.[41] We note that the present mean-field model is not suitable for predicting single-site changes. Moreover, for such extreme sensitivities, the accurate establishment of the value for nonpolar composition, Φ, for use in this model requires careful choice of a hydrophobicity scale.

Although our focus here is on copolymer aggregation processes, it is of interest to compare these results with those predicted for homopolymer aggregation. In the homopolymer limit ($\Phi = 1$), our present model predictions can be compared with those of the Post and Zimm model.[20-23] One main respect in which our model differs from that of Post and Zimm is that we retain the full excluded volume term in the Flory approximation, rather than approximating it by expansion around the theta state. Thus, the present model is more suitable for treating processes of complete collapse as occur in proteins; this was not of importance in the Post and Zimm work. Nevertheless, our predictions in the homopolymer limit are in reasonable agreement with those of Post and Zimm for the homopolymer situations that they modeled.

The low concentration limit of the two-phase region in the phase diagram is referred to as the solubility limit of the polymer. The present theory predicts that such solubility limits of copolymers, on the one hand, and homoplymers, on the other hand, should be extremely different, since the relevant interaction parameter is $\chi\Phi^2$. Homopolymers should be exceedingly insoluble relative to copolymers [for a given type of solvent-aversive (h) monomer, if the comonomer (p) is not solvent averse]. The relatively high solubilities of copolymers is undoubtedly important for proteins which, of necessity, must be soluble within biological cells, wherein the protein concentrations are high. Computations of the solubility limits of heteropolymers and homopolymers of chain length $n = 110$ show that molecules that have the nonpolar compositions of typical proteins ($\Phi = 0.40$) are nearly 50 orders of magnitude more soluble than an equivalent homopolymer chain of pure nonpolar amino acids. For example, polyleucine should be extremely insoluble. Under conditions where the folded state of a protein is stable, its solubility is greater than that of the denatured state.

Conclusions

We previously developed a mean-field theory of protein folding equilibria, based on a treatment of the stabilities of isolated chains[19,20] that assumes protein folding is driven by hydrophobic interactions and opposed by chain conformational entropies. This model has been compared with experimental measurements of stabilities as a function of temperature,[20] urea and GuHCl concentrations,[21] and pH and ionic strength.[28,29] In the present work we have extended the theory to treat translational entropy that favors dispersion of the denatured protein chains in solution, and the hydrophobic interaction that favors chain–chain association. The theory predicts that at sufficiently high protein concentration precipitation occurs. The lowest protein concentration at which this occurs is the solubility limit. The theory predicts an upper consolute temperature as in simple binary solutions, but also predicts that for thermal aggregation, there should be a lower consolute temperature as well, i.e., a temperature *below which* proteins will be soluble. The theory predicts that the solubility limit, the upper and lower consolute points, and the other points on the phase boundaries are extremely sensitive to the length and nonpolar composition of the chain. The theory indicates that heteropolymers such as proteins should be extraordinarily more soluble in water than homopolymers of the corresponding nonpolar amino acids; the solubility limit for proteins is predicted to be higher by tens of orders of magnitude than for homopolymers.

Acknowledgment. We thank Linda DeYoung for many helpful comments and NIH and the DARPA University Research Initiative for financial support.

Appendix

In this Appendix, we derive an expression for the contact free energy for copolymer solutions over the complete range of polymer concentrations, based on a mean-field approximation. Equation 3 gives the Flory–Huggins expression for the contact interactions in a copolymer solution, assuming a uniform distribution of the solvent and two monomer types. To account for the behavior at low concentrations, we first single out one polymer molecule, configured in a sphere with n/ρ lattice sites. We assume that this sphere, apart from having the one chain, is permeated by the average solution outside the chain, that is, by a solution with polymer volume fraction

$$\psi_{\text{bulk}} = \frac{(N-1)n}{(N-1)n + n_s} \tag{A1}$$

or, in terms of the total solution concentration $\psi = Nn/(Nn + n_s)$,

$$\psi_{\text{bulk}} = \psi - \frac{\psi(1-\psi)}{N} - O\left(\frac{1}{N^2}\right) \quad \text{for } n \ll Nn + n_s \tag{A2}$$

Inside the sphere, $n + (n/\rho - n)\psi_{\text{bulk}}$ sites are occupied by polymer segments. Therefore, the polymer volume fraction in the sphere with n/ρ sites is

$$\psi_{\text{sphere}} = \rho + (1-\rho)\psi_{\text{bulk}} \tag{A3}$$

Thus, we divide the solution into two homogeneous parts, the sphere of n/ρ sites with volume fraction ψ_{sphere} of polymer, and the remaining solution outside the sphere, with $Nn + n_s - n/\rho$ sites and volume fraction ψ_{bulk} of polymer. When we apply the Flory–Huggins theory to both parts of the solution independently, the total contact energy for the solution, E_1, is (compare eq 3)

$$\frac{E_1}{kT} = -\chi\Phi^2\left[\frac{n}{\rho}{\psi_{\text{sphere}}}^2 + \left(Nn + n_s - \frac{n}{\rho}\right){\psi_{\text{bulk}}}^2\right] \tag{A4}$$

or, with eqs A2 and A3,

$$\frac{E_1}{kT} = -\chi\Phi^2[(Nn + n_s)\psi^2 + n\rho(1-\psi)^2] + O\left(\frac{1}{N}\right) \tag{A5}$$

Equation A5 implies that the single sphere has no contacts with the solution around it, i.e., chain contacts in the "test sphere" are described by ${\psi_{\text{sphere}}}^2$ term, and polymer–polymer contacts external to the test sphere are described by the ${\psi_{\text{bulk}}}^2$ term in eq A4. Hence, it remains to correct E_1 for the contact energy E_1^{eo} between the exterior layer of the sphere and the first solution layer around it, designated e and o, respectively. As in eq 4, we assume that in layer e, which has $n_e = f_e n/\rho$ sites, a fraction $1 - \sigma$ of the q contacts per site is made with layer o. In layer e the number of hydrophobic (h), polar (p), and solvent (s) occupied sites are respectively

$$n_h = n_e\Phi\psi_{\text{sphere}}, \quad n_p = n_e(1-\Phi)\psi_{\text{sphere}}, \quad n_s = n_e(1-\psi_{\text{sphere}}) \tag{A6}$$

The fractions of h, p, and s sites in the layer o are respectively

$$H = \Phi\psi_{\text{bulk}}, \quad P = (1-\Phi)\psi_{\text{bulk}}, \quad S = 1 - \psi_{\text{bulk}} \tag{A7}$$

The contact energy between two neighbor sites x and y is w_{xy} where x and y are h, p, or s. In keeping with the Flory–Huggins theory we assume random distributions of h, p, and s in each layer. When each site has q neighbors, the contact energy between layers e and o is

(41) Lehrman, S. R.; Tuls, J. L.; Havel, H. A.; Haskell, R. J.; Putnam, S. C. *Biochemistry* **1991**, *30*, 5777.
(42) Tanford, C. *The Hydrophobic Effect*; Wiley-Interscience: New York, 1980.
(43) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664.

$$E_1^{eo} = (1-\sigma)q[w_{hh}n_hH + w_{pp}n_pP + w_{ss}n_sS + w_{hp}(n_hP + n_pH) + w_{hs}(n_hS + n_sH) + w_{ps}(n_pS + n_sP)] \quad (A8)$$

Treating interactions with polar and solvent sites alike, the interaction parameter is

$$\chi = \frac{q}{kT}\left(w_{hp} - \frac{w_{hh} + w_{pp}}{2}\right) = \frac{q}{kT}\left(w_{hs} - \frac{w_{hh} + w_{ss}}{2}\right) \quad (A9)$$

and

$$w_{ps} - \frac{w_{pp} + w_{ss}}{2} = 0 \quad (A10)$$

With eqs A9 and A10 reduction of eq A8 yields

$$\frac{E_1^{eo}}{kT} = (1-\sigma)n_e\chi\left[\frac{n_h}{n_e}(P+S) + \left(\frac{n_p}{n_e} + \frac{n_s}{n_e}\right)H\right] + \frac{(1-\sigma)q}{2kT}n_e\left[w_{hh}\left(\frac{n_h}{n_e} + H\right) + w_{pp}\left(\frac{n_p}{n_e} + P\right) + w_{ss}\left(\frac{n_s}{n_e} + S\right)\right] \quad (A11)$$

and with eqs A6 and A7 one obtains

$$\frac{E_1^{eo}}{kT} = (1-\sigma)n_e\chi\Phi(\psi_{sphere} + \psi_{bulk} - 2\Phi\psi_{sphere}\psi_{bulk}) + (1-\sigma)\frac{qn_e}{2kT}[(\psi_{sphere} + \psi_{bulk})(\Phi w_{hh} + (1-\Phi)w_{pp}) + (2 - \psi_{sphere} - \psi_{bulk})w_{ss}] \quad (A12)$$

We add E_1^{eo} to E_1 in eqs to account for the $(1-\sigma)n_eq$ sphere/solution contacts. Since the total number of contacts between lattice sites is invariant, we must now subtract from E_1 the amount E_2^{eo} which is the contact energy of $(1-\sigma)n_e$ sphere sites among themselves and of $(1-\sigma)n_e$ bulk solution sites among themselves. Methods similar to those used in deriving eq A12 now yield

$$\frac{E_2^{eo}}{kT} = (1-\sigma)n_e\chi\Phi(\psi_{sphere} + \psi_{bulk} - \Phi\psi_{sphere}^2 - \Phi\psi_{bulk}^2) + (1-\sigma)\frac{qn_e}{2kT}[(\psi_{sphere} + \psi_{bulk})(\Phi w_{hh} + (1-\Phi)w_{pp}) + (2 - \psi_{sphere} - \psi_{bulk})w_{ss}] \quad (A13)$$

The correction of eq A5 for the surface contacts per polymer sphere is now the difference of eqs A12 and A13

$$\frac{E^{eo}}{kT} = \frac{E_1^{eo}}{kT} - \frac{E_2^{eo}}{kT} = (1-\sigma)n_e\chi\Phi^2(\psi_{sphere} - \psi_{bulk})^2 \quad (A14)$$

or, with eqs A2 and A3,

$$\frac{E^{eo}}{kT} = (1-\sigma)f_en\rho\chi\Phi^2(1-\psi)^2 + O\left(\frac{1}{N}\right) \quad (A15)$$

Equations A5 and A15 now yield for the corrected contact energy of the two-part solution (copolymer sphere and remainder)

$$\frac{E_1 + E^{eo}}{kT} = -\chi\Phi^2[(Nn + n_s)\psi^2 + (f_i + \sigma f_e)\rho n(1-\psi)^2] + O\left(\frac{1}{N}\right) \quad (A16)$$

In eq A16 the first term in the brackets represents the Flory-Huggins result for the homogeneous polymer solution, eq 3. Therefore, the second term accounts for the extra intramolecular contact energy of the single polymer sphere. Since there is such a contribution for each polymer molecule the corrected contact energy for the copolymer solution becomes

$$\frac{E}{kT} = -\chi\Phi^2[(Nn + n_s)\psi^2 + (f_i + \sigma f_e)\rho Nn(1-\psi)^2] + O(N^0) \quad (A17)$$

Dropping terms of order N^0, dividing by $M = Nn + n_s$, and using $\psi = Nn/(Nn + n_s)$ converts eq A17 into eq 5 of the main text.

We note that the derivation of the Flory-Huggins expression in eq 3 yields terms proportional to ψ, similar to those in eq A12. Such terms have been suppressed because they have no thermodynamic significance.

Brownian Dynamics Simulations of an Order-Disorder Transition in Sheared Sterically Stabilized Colloidal Suspensions

Angeliki Artemis Rigos*

Department of Chemistry, Merrimack College, North Andover, Massachusetts 01845

and Gerald Wilemski*

Physical Sciences Inc., 20 New England Business Center, Andover, Massachusetts 01810

(Received: September 24, 1991)

The shear thinning behavior of a sterically stabilized nonaqueous colloidal suspension was investigated using nonequilibrium Brownian dynamics simulations of systems with 108 and 256 particles. At a volume fraction of 0.4, the suspension is thixotropic: it has a reversible shear thinning transition from a disordered state to an ordered, lamellar state with triangularly packed strings of particles. The time scale for the transition is set by the free particle diffusion constant. For the smaller system, the transition occurs gradually with increasing shear rate. For the larger system, the transition is sharp and discontinuous shear thinning is found.

Introduction

Colloidal suspensions are attractive candidates to study for unusual shear rate dependent phenomena. Although they are simpler in structure and dynamics than polymer solutions and melts, they exhibit much of the same rheological behavior. Moreover, structural transitions occur at moderate shear rates due to the large particle sizes and low Reynolds number flows involved. This is in contrast to molecular liquids where extremely large shear rates, well outside the experimentally accessible range, would be required to observe comparable effects found in computer simulations using nonequilibrium molecular dynamics (NEMD).[1-5]

(1) Erpenbeck, J. J. *Phys. Rev. Lett.* **1984**, *52*, 1333.
(2) Woodcock, L. V. *Phys. Rev. Lett.* **1985**, *54*, 1513; *Chem. Phys. Lett.* **1984**, *111*, 455.

0022-3654/92/2096-3981$03.00/0