Molecular Thermodynamic Model for Helix-Helix Docking and Protein Aggregation

Chau-Chyun Chen
Aspen Technology, Inc., Cambridge, MA 02141

Jonathan King
Dept. of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Daniel I. C. Wang
Dept. of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

The formation of aggregates, rather than correctly folded polypeptide chains, is a pressing problem in biotechnology that has been difficult to approach quantitatively. The competition between folding and aggregation has been carefully analyzed for bovine growth hormone (bGH) and can be attributed to incorrect helix-helix docking for this four-helix bundle protein. An extended molecular thermodynamic model reported here represents Gibbs energy changes associated with intramolecular and intermolecular helix-helix dockings occurring during protein folding and protein aggregation. The model incorporates (1) a semiempirical local composition Gibbs energy expression to account for the helix-helix hydrophobic interactions, which favor helix-helix docking and aggregation and (2) a Flory-Huggins-type Gibbs energy expression to describe the configurational entropy of the polypeptide backbone conformation, which favors disaggregation. For the folding and aggregation of bGH, the molecular thermodynamic model provides estimates for the Gibbs energies and thermodynamic stabilities of various conformations of bGH and qualitatively accounts for the competition between aggregation and productive folding. It also successfully describes the inhibition of aggregation found with peptides corresponding to bGH helical sequences and the effect of site-directed mutagenesis.

Introduction

Recombinant proteins produced at high levels during cell culture frequently accumulate as inclusion bodies, an inactive aggregated state of partially folded or misfolded polypeptide chains (Marston, 1986; Mitraki and King, 1989; Mitraki et al., 1991a; Cleland, 1993). Similar misfolded or aggregated states are involved in a number of human amyloid diseases, including Alzheimer's (Wetzel, 1992). The inclusion body state differs from precipitates of native proteins, which are easily resolubilized upon dilution or change of solvent (Mitraki et al., 1991a). When a desired protein does fail to fold *in vitro*, the general approach is to isolate the inclusion body form, solubilize the chains with strong denaturants, and empirically search for conditions in which the chains refold *in vitro*. Recoveries of correctly folded protein *in vitro* are often low, due to competing aggregation reactions related to the original intracellular problem.

A variety of experimental approaches indicate that aggregate formation represents relatively specific interactions among folding intermediates. The intermolecular forces involved in aggregate formation are noncovalent and are closely related to the interactions needed for correct folding (Mitraki and King, 1989; Mitraki et al., 1991a). The aggregated chains have been shown to retain secondary structure (Zettlmeissl et al., 1979). Using a nondenaturing detergent (lauryl maltoside) to specifically stabilize a labile intermediate to aggregation, Tandon and Horowitz (1986) successfully renatured rhodanese after guanidine hydrochloride denaturation. In the refolding of carbonic anhydrase, small numbers of polyethylene glycol molecules are bound to the critical folding intermediate preventing aggregation (Cleland and Wang, 1990a;

Correspondence concerning this article should be addressed to C.-C. Chen.

Cleland et al., 1992). Single amino acid substitutions in the tailspike of P22 increase off-pathway aggregation, while another class of suppressor mutations inhibits aggregation (Haase-Pettingel and King, 1988; Mitraki et al., 1991b, 1993).

The fundamental understanding of protein structure and function has developed out of studies of the solution state of proteins, and of crystals suitable for X-ray diffraction. The aggregated states, which are composed of less ordered forms of the polypeptides, have been difficult for protein biochemists to analyze or to model. These materials behave much more like organic polymers, with the aggregation reaction resembling phase transitions. The analytical methods and conceptual frameworks for approaching such materials are more developed in the field of chemical engineering than in protein biochemistry. Indeed, chemical engineers have been responsible for many of the recent advances in this area (Georgiou and DeBernardez-Clark, 1991; Himmel and Georgiou, 1993; Cleland, 1993). In the studies reported here we apply thermodynamic methods familiar in chemical engineering practice, but previously little employed in biochemistry, to the problem of protein aggregation.

There are a number of efforts to model protein aggregation. Fields et al. (1992) treated protein aggregation as a phase-separation problem. In contrast to the view just given that proteins aggregate in conformations intermediate between native and denatured states, they proposed that the denatured state aggregates. The model considers the aggregated state of mixture of amorphous copolymer plus solvent and that the driving forces are the hydrophobic interaction, which favors aggregation, and conformational and translational entropies, which favor disaggregation. A mean-field lattice statistical mechanics theory was proposed for the liquid-liquid phase equilibrium between the denatured and aggregated states of proteins in aqueous solutions. However, it is difficult to interpret results of the theory since the treatment of the aggregated state as amorphous polymer solution departs from experimental evidence.

Stigter and Dill (1993), using the model of Fields et al. (1992), accounted for the observed minimum in protein solubility as a function of urea concentration in water. Adding small urea concentrations decreases the protein solubility because the dominant action of urea is to convert native to denatured molecules that are prone to aggregation. At higher urea concentrations, further addition of urea leads to weakening of the attractions, and therefore to increased solubilities.

The preceding approaches are difficult to apply to particular proteins. We report here an extended molecular thermodynamic model to represent the Gibbs energy change associated with intramolecular and intermolecular helix-helix docking occurring during protein folding and protein aggregation. Molecular thermodynamics has been successfully used to derive semiempirical Gibbs energy expressions and to estimate the corresponding thermodynamic parameters for mixtures of electrolytes, nonelectrolytes, amino acids, antibiotics, micelles, and polymers (Prausnitz, 1979; Chen and Evans, 1986; Chen et al., 1989; Zhu et al., 1990; Chen, 1993). It takes advantage of a large body of experimental thermodynamic data developed by chemists and chemical engineers to study the phase behavior of both small molecules and large molecules (Renon and Prausnitz, 1968; Fredenslund et al., 1975). Recently, semiempirical Gibbs energy expressions have been successfully developed from molecular thermodynamics to represent the Gibbs energy of folding homopolypeptides from a random-coiled state to either the α-helical state or the β-sheet state (Chen et al., 1992), and the Gibbs energy of folding heteropolypeptides from random-coiled conformations into α-helical conformations (Zhu et al., 1992). The model yields a hydrophobicity scale for the twenty amino acid side chains, which compares favorably with established hydrophobicity scales. It accurately predicts thermodynamically favorable folding of a number of aqueous homopolypeptides from coil states into α-helices. The computed Gibbs energies correspond well with the experimental data on helicity of a number of natural polypeptides and synthetic polypeptides. Molecular thermodynamics provides a potentially practical framework to account for the intramolecular and intermolecular physical interactions that govern the phase behavior of protein systems.

The purpose of this work was to develop a molecular thermodynamic model for the physiologically important forms of protein aggregation, due to off pathway interactions among folding intermediates. The model represents a thermodynamic account of the stability of protein aggregate intermediates with α-helical structures. It does not at present represent the stability of intermediates in the formation of β-sheet proteins. Specifically, we reformulated the Gibbs energy model of Chen et al. (1992) and Zhu et al. (1992) to represent the changes associated with the nucleation and termination of multiple α-helical structures and with both intramolecular helix-helix docking and intermolecular helix-helix docking. These Gibbs energy expressions were derived to account for the helix-helix hydrophobic interactions, which favor aggregation, and the configurational entropies, which favor disaggregation.

Model Compound for Protein Aggregation Upon Refolding

One carefully analyzed example of competition between refolding and aggregation are the studies of bovine growth hormone (bGH) whose folding and aggregation pathways have been studied by Brems, Lehrman and coworkers (Brems, 1988; Brems, 1990; Lehrman et al., 1991). bGH is believed to be a member of the four-helix-bundle family of proteins. Brems and Havel (1989) showed that a peptide corresponding to one of the helices effectively suppressed the off pathway association reaction. The data and mechanisms established by Brems et al. made it possible for us to attempt to model the thermodynamic factors involved in protein aggregation of helical bundle proteins.

Based on the kinetic folding pathway, Brems (1988) suggested the following equilibrium folding and aggregation model for bGH. A similar model has been proposed by Cleland and Wang (1990b) for bovine carbonic anhydrase B (CAB) and by Wetzel (1992) for human interferon-γ and interleukin-1β (Wetzel and Chrunyk, 1993).

$$N \leftrightarrow I \leftrightarrow U$$

$$I \leftrightarrow I_{\text{assoc}}$$

$$I_{\text{assoc}} \rightarrow \text{precipitate}$$

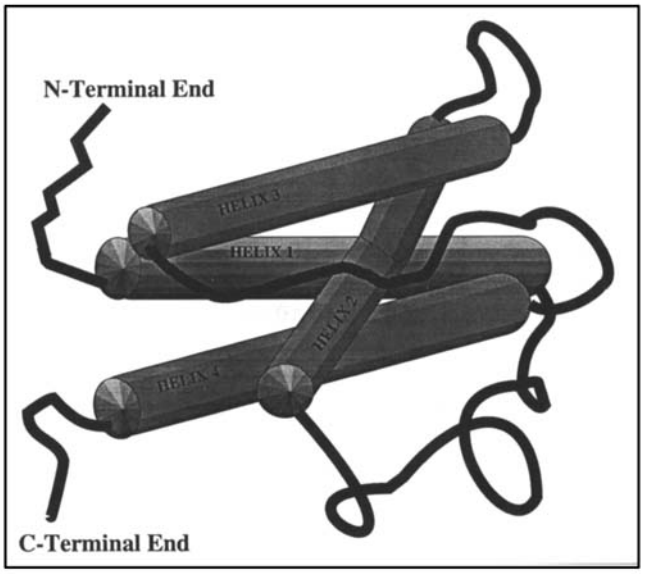

Figure 1. 3-D structure of native methionyl porcine somatotropin.

By Abdel-Meguid et al. (1987). Helix 1: residues 7–34; helix 2: residues 75–87; helix 3: residues 106–127; helix 4: residues 152–183.

Under native refolding conditions, a protein molecule in its unfolded state U forms the folding intermediate I, and then the partially folded intermediate I is completely refolded to the native conformation, N. While under aggregation conditions (such as high protein concentrations), intermediate I forms associated intermediate I_{assoc}, which is relatively insoluble and may lead to multimeric aggregate formation.

As summarized by Brems (1990), bGH consists of 191 amino acids and is a single-domain protein. Unfortunately, the three-dimensional structure of bGH has not been determined. However, the three-dimensional structure of porcine growth hormone (pGH), a protein closely related to bGH, has been solved and it is given in Figure 1. It is a four-helical bundle with the helices arranged in an antiparallel fashion (Abdel-Meguid et al., 1987). Several of these helices are amphipathic or dual-sided with hydrophobic and hydrophilic surfaces. From sequence homology and spectroscopic similarities, the architecture of the bovine species must be very similar (Brems, 1990).

Monomeric intermediate I has been described as a molten globule (Brems and Havel, 1989). It is compact like the native state, has a high content of α-helix, and has hydrophobic faces of the amphipathic helices that are relatively more exposed than in the native state. Additionally, the hydrophobic face of the third amphipathic helix is not completely buried as it is in the native state, and it readily interacts with a similar hydrophobic surface from another bGH molecule at higher protein concentrations. This intermolecular hydrophobic bonding between helices is critical to the stabilization of I_{assoc} (Brems et al., 1988).

Brems further showed that a peptide with sequence corresponding to the third helix of the growth hormone four-helix bundle could inhibit aggregation. The model that emerges from their studies is one in which the intermediate has an internal helix docking site exposed, and the free helix of one species docks against the helical docking site on another molecule forming the associated species. The helical peptides that are homologous for the unfolded sequence will dock against the aggregating site. The Brems model on bGH aggregation provides a framework for investigating aggregation of the family of α-helical bundle proteins.

Model for α-Helix Formation

We first examined the α-helical structure of bGH with the molecular thermodynamic model of Chen et al. (1992) and Zhu et al. (1992). Equation 1 gives the general formulation of the Gibbs energy of folding the residues from a random-coiled state into an α-helical state.

$$\Delta G^{c \to \alpha} = \Delta G^{lc, c \to \alpha} - T\Delta S^{\text{config}, c \to \alpha} \tag{1}$$

Here the local composition term, ΔG^{lc}, includes the enthalpy of α-helix formation and other entropic terms not included in the configurational entropy term, $T\Delta S^{\text{config}}$.

Zhu et al. (1992) presented the expression for the Gibbs energy of folding a polypeptide chain from the random-coiled conformation to an α-helical conformation J as follows:

$$\frac{\Delta G_J^{c \to \alpha}}{RT} = -\Sigma_i \frac{\Delta g_i^{c \to \alpha}}{RT} c_{iJ} \tag{2}$$

where

$$\frac{\Delta g_i^{c \to \alpha}}{RT} = -\frac{\Delta \tau_i}{Z} - \frac{\Delta s_i^{\text{config}, c \to \alpha}}{R}. \tag{3}$$

The specific expressions for $\Delta \tau_i$ and $\Delta s^{\text{config}, c \to \alpha}$ are to be derived in the next paragraphs; Z is the coordination number, c_{iJ} is the conformational index of residue i in α-helical conformation J. When summed over the amino acid sequence of a polypeptide chain, it accounts for the location and length of the helix. The value of c_{iJ} is 0 for residues in the coil state, and 1 for residues in the helical state.

Equation 3 was derived for a system consisting of a polypeptide chain and the nearest neighboring solvent molecules that constitute the hydration shell. The hydrated solvent molecules are packed out of the hydration shell when the helix is formed and the internal packing of the α-helix is void of water molecules.

As first proposed by Chen et al. (1992), the configurational entropy loss associated with the α-helix formation can be approximated as the difference between that of mixing solvents with a polymer chain of length n and that of mixing solvents with a polymer chain of length $n-1$.

$$\Delta s^{\text{config}, c \to \alpha} = (\Delta S^{c, FH}|_{n-1} - \Delta S^{c, FH}|_n). \tag{4}$$

The configurational entropy of mixing a disoriented polymer and the solvents to form a random-coiled polymer solution is well represented by the equation of Flory (1941, 1942) and Huggins (1941, 1942):

$$\Delta S_{\text{mix}}^{FH} = -R\left\{N_1 \ln\left(\frac{N_1}{N_1 + N_2 x}\right) + N_2 \ln\left(\frac{N_2 x}{N_1 + N_2 x}\right)\right\}. \quad (5)$$

Here, N_1 is the number of solvent molecules, N_2 is the number of polymer chain molecules of x segments long, and x is the number of the segments in a polymer chain. Chen et al. (1992) suggested a constant value, $-0.798\ R$, for the entropy loss per residue for α-helix formation, $\Delta s^{\text{config}, c \to \alpha}$.

Zhu et al. (1992) expressed the local composition term in terms of binary interaction energy parameters, τ_{ij}, of the nonrandom two-liquid theory of Renon and Prausnitz (1968).

$$\Delta\tau_i = 2(\tau_{W,R(i)} + \tau_{R(i),W}) - (\tau_{W,R(i)} + \tau_{R(i),W}) - \tau_{R(i+3),R(i)}$$

(for N-terminal end residues) (6)

$$\Delta\tau_i = 2(\tau_{W,R(i)} + \tau_{R(i),W}) - (\tau_{R(i-3),R(i)} + \tau_{R(i+3),R(i)})$$

(for central residues) (7)

$$\Delta\tau_i = 2(\tau_{W,R(i)} + \tau_{R(i),W}) - (\tau_{W,R(i)} + \tau_{R(i),W}) - \tau_{R(i-3),R(i)}$$

(for C-terminal end residues). (8)

The first terms in Eqs. 6 to 8 represent the intrinsic folding potential due to residue-solvent interactions, $\Delta\tau_i^*$:

$$\Delta\tau_i^* = 2(\tau_{W,R(i)} + \tau_{R(i),W}). \quad (9)$$

The second terms in Eqs. 6 and 8 represent the chain end effect. The last terms in Eqs. 6 to 8 represent the cooperative effect due to residue-residue interactions.

The model requires two binary interaction parameters, τ_{ij} and τ_{ji}, for each of the solvent-solvent interaction pairs, the solvent-residue interaction pairs, and the residue-residue interaction pairs. These binary interaction parameters have been estimated from the UNIFAC functional group contribution method with the electrostatic interactions neglected (Zhu et al., 1992). Here the interacting residues include both the backbone groups and the side-chain groups.

Zhu et al. (1992) demonstrated the utility of the Gibbs energy model in predicting the formation of α-helical secondary structure by examining the stability of α-helical conformations of a number of small natural polypeptides and synthetic polypeptides with single helical structure. The computed Gibbs energies correspond satisfactorily with the experimental data on helicity, a measure of the helical content of these polypeptides. However, the bGH protein molecule contains multiple helical structures and we need different Gibbs energy formulations to examine the locations and length of the helices.

Helix Nucleation Potential

The actual conformation of the bGH folding intermediate is not known. One route to its approximation is by "unfolding" the globular protein into its secondary structure elements and treating such species as models of the folding intermediates. A second approach, which we take below, is to compute the thermodynamically favored secondary helical structures formed after dilution of an unfolded molecule out of denaturant.

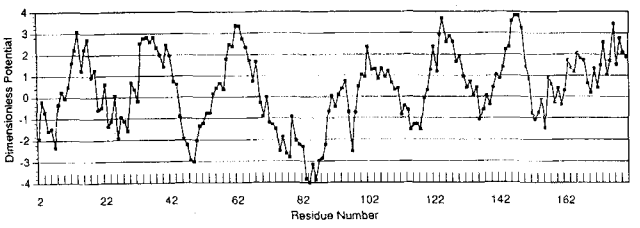

Figure 2. Computed bGH helix nucleation potential vs. amino acid residue number; nucleating segment length set to 10 residues.

Equation 2 can be used to evaluate a "helix nucleation potential" defined as the Gibbs energy change for a segment of the protein molecule in water to convert to the α-helical state from the random-coiled state. The nucleation potential will vary with the helix-starting residue and the helical chain length. Due to the chain end effect, short segments are not favored to form helices (Chen et al., 1992). The minimal chain length of a helix-nucleating segment, including both the N-terminal end residues and the C-terminal end residues, must be equal to or greater than four residues. Figure 2 gives the computed dimensionless helix nucleation potential for bGH with nucleating segment length fixed at ten residues and with nucleation-starting residues varied. To be a helix-nucleating segment, the helix nucleation potential must be negative. The more negative the helix nucleation potential is, the stronger the helix-forming tendency becomes. Once nucleated, the helix should propagate and grow in helical length until the helix is stopped by helix termination signals, including prolines and other helix termination constraints (Presta and Rose, 1988).

Although it is feasible to define and compute the helix nucleation potential, it is not feasible to predict exactly which residue should be the N-terminal residue of the helix. In this study, we arbitrarily set the N-terminal residue to be the first residue of a ten-residue helix with favorable helix nucleation potential. The computed results are not very sensitive to the length of the helices. The model predicts that, in a fully unfolded bGH molecule, there could be a number of potential helix-nucleating sites that would lead to helices of at least ten residues long. These major helix-nucleating sites are residues 1–8, 23–29, 45–54, 69–90, 112–119, and 152–156. Note that the helix nucleation potential alone is not sufficient to "predict" both locations and length of helices. It is necessary to define a helix termination potential as discussed in the next section. Furthermore, one must recognize that these helix-nucleating sites were solely suggested by the Gibbs energy analysis with other factors totally ignored.

Helix Termination Potential

To help determine helix-terminating locations, one could define a "helix termination potential" for any sequence of a polypeptide chain to be the Gibbs energy change of converting any three consecutive residues in a helical structure from the helical structure to the random-coil state. In other words, these three residues would now terminate the helix instead of further propagate the helix. This Gibbs energy change should

include the chain end effect of the preceding three residues ($i-3$ to $i-1$), the loss in residue-residue hydrophobic interactions associated with the three terminating residues (i to $i+2$), and the corresponding entropic gain in converting from the helical structure to the random-coil state. Equation 10 gives the helix termination potential at residue i.

$$\frac{\Delta G_i^{\alpha \to c}}{RT} = \sum_{i-3}^{i-1} \frac{\Delta \tau_i^*}{2Z} + \sum_{i}^{i+2} \frac{\Delta \tau_i}{Z} \Big|_{C-\text{terminal}} + \frac{3\Delta s_i^{\text{config}, c \to \alpha}}{R}. \quad (10)$$

Helix termination is favored if the helix termination potential is negative. Conversely, propagation of the helix should be favored if the helix termination potential is positive. Figure 3 shows the computed helix termination potential of the bGH polypeptide chain as a function of the residue location on the protein primary sequence. The results suggest a number of favored helix-terminating sites: residues 38–47, 65–71, 97–111, 127–134, 149–154, and 180–186. In addition, there is a less significant potential helix-terminating site: residues 19–21.

Due to the immense complexity of the protein folding process, one should not expect to establish the exact α-helical secondary structure of a protein molecule simply from a Gibbs energy analysis. However, the suggested helix-nucleating sites and helix-terminating sites would imply that a fully unfolded bGH molecule could partially fold into a protein intermediate with five α-helices made up of residues 1–37 (helix 1), 45–64 (helix $1\frac{1}{2}$), 69–96 (helix 2), 112–126 (helix 3), and 152–179 (helix 4). Figure 4 shows the suggested helical structure for the bGH folding intermediate prior to the intramolecular helix-helix docking. As mentioned before, the three-dimensional structure of bGH has not been determined. However, it is interesting that the suggested helical structure for the bGH folding intermediate is similar to the four-helical-bundle native structure reported by Abdel-Meguid et al. (1987) for methionyl porcine somatotropin (MPS), a genetically engineered variant of pGH. In MPS, the alanine residue at position one of pGH has been replaced with a methionine. The four helices of MPS are made up of residues 7–34, 75–87, 106–127, and 152–183. The region from 89 to 96 is also in an α-helical conformation. It can be thought of as an extension of the second helix. The region from 35 to 74 is predominantly loops and turns. One of these turns is a 1.5 turn helix (residues 53–58). The region from 128 to 151, on the other hand, has no well-defined secondary structure. Furthermore, residues 53 and 164 form one disulfide bridge, whereas residues 181 and 189 form the other (Abdel-Meguid et al., 1987).

Intramolecular Helix-Helix Docking

A naked helix may dock against an already formed helix as part of the folding process toward the native state. This intramolecular helix-helix docking must play a major role in the establishment of the tertiary structure of globular protein molecules.

In the following paragraphs, we formulate an expression to account for the Gibbs energy change associated with the intramolecular helix-helix docking. The Gibbs energy change represents a thermodynamic quantity for the protein molecule at the reference state of infinite dilution. We propose that the Gibbs energy of the intramolecular helix-helix docking be represented as the sum of the local composition contribution due to the helix-helix surface contact and the entropy contribution due to the loss of the configurational entropy of the nonhelical residues linking the helices. These nonhelical residues would lose their configurational entropy when the docking takes place and a tertiary structure is formed:

$$\Delta G^{\alpha+\alpha} = \Delta G^{lc, \alpha+\alpha} - T\Delta S^{\text{config}, \alpha+\alpha}. \quad (11)$$

The configurational entropy loss derived from the intramolecular helix-helix docking is proportional to m, the number of nonhelical residues that lose their configurational entropy when the helix-helix docking takes place. If the two

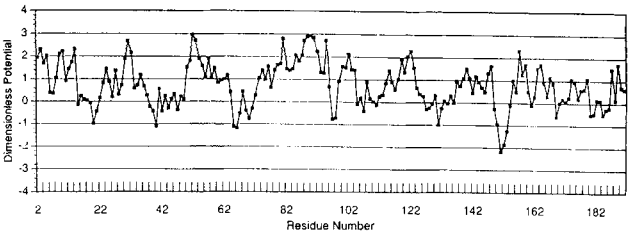

Figure 3. Computed bGH helix termination potential vs. amino acid residue number.

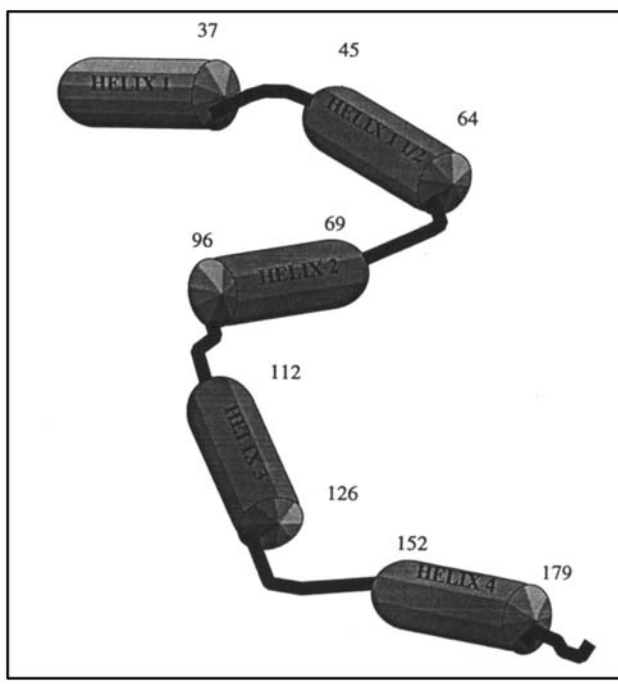

Figure 4. Model-predicted helical structure of bGH folding intermediate.

Helix 1: residues 1–37; helix $1\frac{1}{2}$: residues 45–64; helix 2: residues 69–96; helix 3: residues 112–126; helix 4: residues 152–179.

docking helices are adjacent to each other, then the configurational entropy loss is directly proportional to the number of nonhelical residues linking the two docking helices:

$$\Delta S^{\text{config},\alpha+\alpha} = m(\Delta S^{c,FH}|_{n-1} - \Delta S^{c,FH}|_{n}). \quad (12)$$

In other words, the closer together the two docking helices are, the smaller the configurational entropy loss will be. At $m=1$, $\Delta S^{\text{config},\alpha+\alpha}$ equals $-0.798\ R$ (Chen et al., 1992).

In order for the intramolecular helix-helix docking to be thermodynamically favorable, the Gibbs energy gain from the local composition term must overcome the entropy loss associated with the nonhelical residues linking the docking helices.

The local composition term for the intramolecular helix-helix docking, $\Delta G^{lc,\alpha+\alpha}$, is a function of the hydrophobicity of the two helix-docking surfaces. The surface hydrophobicity is determined by the compositions of the residues on the surface of the docking helices and the hydrophobicity of these individual residues. Here we account for the hydrophobicity with the intrinsic Gibbs energy of each docking residue and we compute the local contribution term per docking residue from the intrinsic Gibbs energy of the docking residue:

$$\frac{\Delta g_i^{lc,\alpha}}{RT} = -\frac{\Delta \tau_i^*}{Z}. \quad (13)$$

The intrinsic Gibbs energy for each residue, $\Delta\tau_i^*$, is given in Eq. 9.

Analysis of the pattern of residue-to-residue contacts at the interface of fifty helix-to-helix packings observed in ten proteins of known structure supports a model for helix-to-helix packing in which the ridges and grooves on the helix surface intercalate (Chothia et al., 1981). The most common way of packing α-helices, as shown in Figure 5, is by fitting the ridges formed by a row of residues separated in sequence by four in one helix into the same type of grooves in the other helix (Branden and Tooze, 1991). These class 4–4 packings involve, on each helix, two adjacent $\pm 4n$ rows made up of residues with the following sequence relationship: i, $i+3$, $i+4$, $i+7$, $i+8$, and $i+11$ (Chothia et al., 1981). In this article, we consider only the helix-helix docking with class 4–4 packings.

For the purpose of quantifying the local composition term associated with helix-helix docking, we characterize the helical surface hydrophobicity of a docking helix at residue i with the absolute value of the summation of the intrinsic Gibbs energy of the residues located at positions i, $i+3$, $i+4$, $i+7$, $i+8$, and $i+11$. In other words, we compute $\Delta G_i^{lc,\alpha}$ by summing up $\Delta g^{lc,\alpha}$ of the residues located at positions i, $i+3$, $i+4$, $i+7$, $i+8$, and $i+11$. In doing so, we ignored the cooperative term because the term is relatively minor in comparison to the intrinsic term expressed in Eq. 13. $\Delta G_{ij}^{lc,\alpha+\alpha}$ is then the sum of $\Delta G^{lc,\alpha}$ of the two docking helices at residues i and j, respectively:

$$\Delta G_{ij}^{lc,\alpha+\alpha} = \Delta G_i^{lc,\alpha} + \Delta G_j^{lc,\alpha}. \quad (14)$$

If the entropic contribution is constant, then the helical surface position with the strongest surface hydrophobicity should represent the most favorable helix-helix docking site. Figure 6 shows the computed surface hydrophobicity profile of the bGH folding intermediate. The profile was computed as if the entire bGH chain forms a single helix. The profile has a jagged appearance because the values go up and down as we move around the amphipathic helix from the hy-

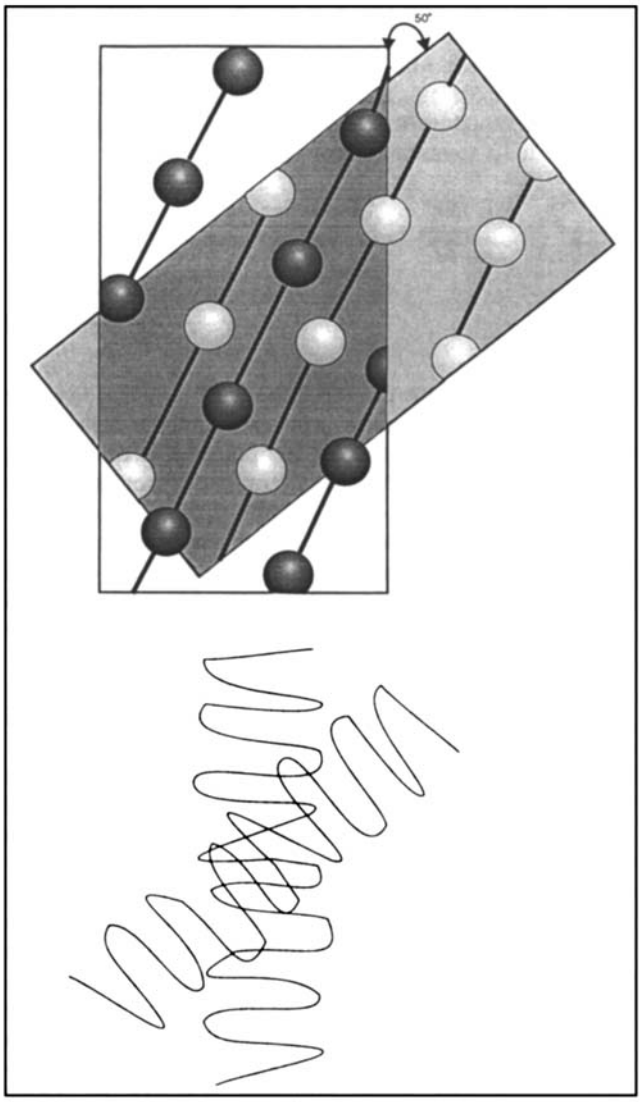

Figure 5. Class 4-4 helix-helix packing. (Branden and Tooze, 1991.)

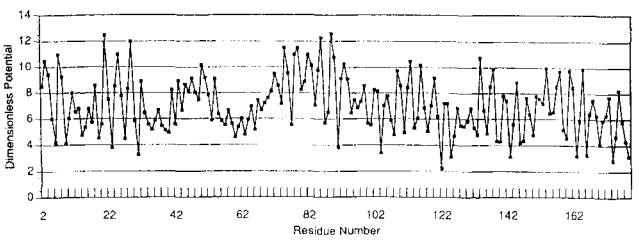

Figure 6. Computed bGH helical surface hydrophobicity vs. amino acid residue number.

Figure 7. Model-predicted helical structure of bGH folding intermediate after intramolecular docking.

drophobic side to the hydrophilic side. The figure suggests that both helices 1 (residues 1–37) and 2 (residues 69–96) have a number of strong hydrophobic surfaces with the absolute value of $\Delta G_i^{lc,\alpha}/RT$ in the range of 10 to 12. The resulting $\Delta G_{ij}^{lc,\alpha+\alpha}/RT$ for docking helices 1 and 2 would be in the range of -20 to -24. The helices 3 (residues 112–126) and 4 (residues 152–179) do not have such strong hydrophobic surfaces. The absolute values of their $\Delta G_i^{lc,\alpha}/RT$ are in the range of 8 to 10. The existence of helix $1\frac{1}{2}$ (residues 45–64) seems to have a favorable effect on the intramolecular docking because it would reduce the number of nonhelical residues between helices 1 and 2 and the associated entropy loss when the intramolecular docking takes place. Therefore, it seems likely that helices 1 and 2 have a stronger tendency to dock and form a "molten globular molecule," while helices 3 and 4 tend to hang loose. In other words, the model computations are consistent with the Brems model that, in the bGH folding intermediate, helices 1 and 2 dock intramolecularly, while helices 3 and 4 do not. Figure 7 depicts such a folding intermediate.

Intermolecular Helix-Helix Docking

A free peptide helix may dock against a helix of another peptide molecule to inhibit the folding of the peptide molecule. It takes place through the naked helix to form helix docking against the already formed helices of another species. Such an intermolecular helix-helix docking step may result in protein association and aggregation.

The intermolecular helix-helix docking can be conveniently treated as a chemical equilibrium relationship between the naked helices in the aqueous phase and the intermolecularly docked helices. In doing so, we must recognize that other factors, such as kinetics or diffusion, may play important roles in the docking process.

It takes two naked helices, I, to dock against each other and form a dimer, II, a first step in the development of I_{assoc}.

$$2I_{\text{aq}} \rightleftharpoons II_{\text{dock}}. \tag{15}$$

The driving force for the intermolecular helix-helix docking, or the intermolecular helix-helix docking potential, can be defined as the difference between the chemical potentials of the docked helices and the chemical potential of the free helices at their standard state, that is, the infinite dilution state:

$$\Delta G^0_{II,\text{docking}} = 2\left(\mu^0_{I,\text{dock}} - \mu^0_{I,\text{aq}}\right). \tag{16}$$

Similar to the intramolecular helix-helix docking, the driving force for the intermolecular helix-helix docking is derived from the hydrophobic interaction between the docking helical surfaces, that is, $\Delta G_{ij}^{lc,\alpha+\alpha}$ of Eq. 14. The entropy loss due to intermolecular docking is concentration dependent. In this article, we approximate the entropy loss as the difference between the entropy of mixing 2 moles of monomer with 55.51 moles of water and the entropy of mixing 1 mole of dimers with 55.51 moles of water, or $3.66R$:

$$\Delta G^0_{II,\text{docking}} = \Delta G_{ij}^{lc,\alpha+\alpha} - T*(-3.66R). \tag{17}$$

The chemical equilibrium relationship between the naked helices and the docked helices can then be expressed as Eq. 18:

$$K_{II,\text{docking}} = \frac{a_{II}}{a_I^2} = \exp\left(\frac{-\Delta G^0_{II,\text{docking}}}{RT}\right). \tag{18}$$

It has been proposed that the intermolecular docking takes place at helix 3 to form the bGH dimer and subsequently the aggregate (Brems and Havel, 1989). Assuming a helix 3–helix 3 intermolecular docking with docking sites being residue 113, the strongest hydrophobic site on helix 3, $\Delta G_{ij}^{lc,\alpha+\alpha}/RT$ would be -21, $\Delta G^0_{II,\text{docking}}/RT$ would be -17.3, $K_{II,\text{docking}}$ would be 3.4×10^7, or 6×10^5 M^{-1}. Lehrman et al. (1991) reported an experimental value of 1.6×10^5 M^{-1}.

Activity of a helix is the product of the helical concentration and the unsymmetric activity coefficient, γ^*, of the helix:

$$a_I = x_I\gamma_I^*. \tag{19}$$

The unsymmetric activity coefficient of the helix is a function of the concentration and composition of the helix, the solvent, the cosolvents, and any denaturing agents. Our molecular thermodynamic formation can be extended to investigate the effects of helper helices, cosolvents, and the solvent compositions upon the activity coefficients of the helices. Typically, without the presence of cosolvents and denaturing agents, one would assume that the unsymmetric activity coef-

ficient of the helix would remain close to unity due to the low helix concentration. In this article, we approximate the activities of the helices with the concentrations of the helices:

$$K_{II,\text{docking}} = \frac{a_{II}}{a_I^2} = \frac{x_{II}}{x_I^2}. \tag{20}$$

From Eq. 20, one could compute the concentrations of each helical species.

While the entropy loss associated with the nonhelical residues between two intramolecularly docked helices is the barrier for the intramolecular docking, the low concentration of the docking helices is the barrier for the intermolecular helix-helix docking. For the bGH system, the bGH folding intermediate corresponds to the bGH molecule with naked helices. After intermolecular helix 3–helix 3 docking, the bGH folding intermediate forms dimers. Figure 8 gives computed yields of the bGH monomer and dimer vs. total bGH concentrations with $K_{II,\text{docking}}$ fixed at 4×10^5 M^{-1}. At total bGH concentrations greater than about 0.01 mg/mL, the model predicts that bGH is present mostly as dimers (or aggregates).

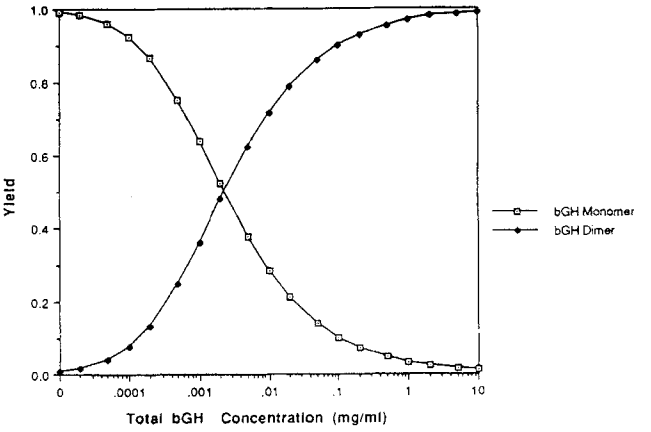

Figure 8. Computed fractional yields of bGH species vs. total bGH concentrations.

$K_{\text{II,docking}}$ fixed at 4×10^5 M^{-1}.

Helical Fragments

The fragment consisting of residues 96–133 of bGH contains helix 3 of the bGH folding intermediate (Brems et al., 1987). The presence of this helical fragment in tenfold excess effectively prevents aggregation of bGH (Brems, 1990). Our helix nucleation/termination potential computations on the fragment indicate that the fragment does form helix 3. We believe helix 3 on the fragment prevents aggregation by docking with helix 3 of the bGH folding intermediate and blocking it from docking with another bGH molecule.

In addition to being in equilibrium with its dimer, *II*, when the helical fragment consisting of residues 96–133 is present in solution, the bGH folding intermediate is also in equilibrium with what we called blocked bGH. Blocked bGH, *IJ*, is formed when the helical fragment, *J*, docks with bGH at helix 3. Also, the helical fragment is in equilibrium with its dimer, *JJ*:

$$I_{\text{aq}} + J_{\text{aq}} \rightleftharpoons IJ_{\text{dock}} \tag{21}$$

$$2J_{\text{aq}} \rightleftharpoons JJ_{\text{dock}} \tag{22}$$

$$K_{IJ,\text{docking}} = \frac{a_{IJ}}{a_I a_J} = \frac{x_{IJ}}{x_I x_J} \tag{23}$$

$$K_{JJ,\text{docking}} = \frac{a_{JJ}}{a_J^2} = \frac{x_{JJ}}{x_J^2}. \tag{24}$$

We can solve for the equilibrium constants, K_{II}, K_{IJ}, and K_{JJ}, and the concentrations of each species if the ΔG's are known. Here we assume the three equilibrium constants are equal since docking takes place between the same helical surfaces in each case (those which make up helix 3). Figure 9 shows computed yields of various bGH species vs. total bGH concentrations when the helical fragment 96–133 is present at ten times the total bGH concentration. The major species present is predicted to be blocked bGH, except for at low total bGH concentrations, where the monomer becomes the significant species. Compared to the case without the helical fragment (Figure 8), the bGH dimer concentration is much lower in Figure 9. This indicates that the fragment 96–133 would be effective in preventing aggregation as shown by Brems (1990).

Brems (1990) also showed that the fragment consisting of residues 96–112 of bGH was not effective in preventing aggregation. Our computations on helix nucleation and termination potentials suggest that the fragment 96–112 does not form a helix and, as a result, it would not dock intermolecularly with bGH to form blocked bGH.

Site-Directed Mutagenesis

Human growth hormone (hGH), which is unable to aggregate via the bGH aggregation process, differs from the bovine primary sequence at eight positions within the bGH region with amino acid residues 109–133. Not long ago, Lehrman et al. (1991) characterized the folding of a bGH analogue (8H-bGH) that contains the hGH sequence between amino acid

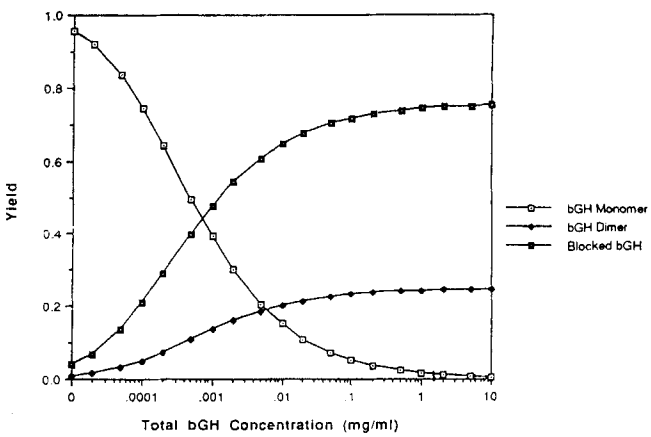

Figure 9. Computed fractional yields of bGH species vs. total bGH concentrations with helical fragment.

96 – 133, $K_{\text{II,docking}}$ fixed at 4×10^5 M^{-1}.

residues 109–133 at low and high concentrations. The equilibrium folding characteristics of bGH and 8H-bGH are similar when monitored at low protein concentrations (≤ 2 μM). Also, the structures of the bGH and 8H-bGH monomeric folding intermediates are similar. In contrast, partial denaturation of 8H-bGH at higher concentrations (> 2 μM) leads to significantly less aggregation than is observed for bGH. The equilibrium constants for the formation of soluble bGH and 8H-bGH aggregates differ by one order of magnitude (1.6×10^5 M^{-1} and 1.4×10^4 M^{-1}, respectively).

Given the primary sequence of 8H-bGH, we could compute $\Delta G_i^{lc,\alpha}/RT$ for 8H-bGH at residue site 113. The $\Delta G_i^{lc,\alpha}/RT$ for 8H-bGH (-9.2) is found to be about 1.3 higher than the $\Delta G_i^{lc,\alpha}/RT$ for bGH (-10.5). The resulting difference in $\Delta G_{ij}^{lc,\alpha+\alpha}/RT$, 2.6 corresponds well to the one order of magnitude difference in the equilibrium constants for the formation of soluble bGH and 8H-bGH aggregates. The model correctly suggests that the site-directed mutagenesis weakens the hydrophobic interactions between the intermolecular helix-helix docking sites.

Conclusions

Based on molecular thermodynamics, we derived Gibbs energy expressions for the nucleation and termination of helical structures in protein folding intermediates. We also developed Gibbs energy expressions for the intramolecular helix-helix docking and the intermolecular helix-helix docking. These expressions help facilitate investigations of the thermodynamic stabilities of protein intermediates identified in the protein folding and aggregation pathway. The thermodynamic stability represents one of several major driving forces in the protein folding and aggregation pathways. Other factors include folding kinetics and diffusion.

Applying the molecular thermodynamic model to bGH, we obtained computed results that show: (1) favorable conversion of bGH molecules from the random coil state to a helical structure similar to the native structure of methionyl porcine somatotropin with four-helical bundles, (2) favorable intramolecular docking of helix 1 and helix 2, (3) favorable intermolecular docking of helix 3 from one bGH molecule and helix 3 from another bGH molecule, (4) a bGH fragment of helix 3 inhibits the intermolecular helix-helix docking by competing for the intermolecular docking site, and (5) site-directed mutagenesis could alter the hydrophobic interactions between the intermolecular helix-helix docking sites.

The entropic contribution to the calculation was influenced by the presence of relatively long loops between the bGH helices, one of the factors favoring the intermolecular interaction leading to aggregation. The competition between folding and aggregation for a four-helix bundle with tight turns rather than long loops would be different. We have not found experimental data on the folding vs. aggregation of such proteins.

These studies are being continued to include (1) the effect of small molecules, including denaturing agents, upon the stabilities of various protein conformations, (2) the effects of folding reaction kinetics in the folding pathway and the aggregation pathway, and (3) the derivation of molecular thermodynamic models for other secondary structures, such as β sheets.

Acknowledgments

This work was supported by the National Science Foundation under the Engineering Research Center Initiative to the Biotechnology Process Engineering Center (Cooperative Agreement CDR-8-03014). The authors wish to thank Todd Riccobene for help in the computations.

Notation

R = gas constant
s = entropy, per residue basis
T = temperature, K
x = liquid-phase mole fraction
μ = chemical potential

Superscripts and subscripts

c = random coil
$\alpha \rightarrow c = \alpha$ helical to random coil
aq = aqueous phase
mix = mixing
R = residue
W = water

Literature Cited

Abdel-Meguid, S. S., H.-S. Shieh, W. W. Smith, H. E. Dayringer, B. N. Violand, and L. A. Bentle, "Three-Dimensional Structure of a Genetically Engineered Variant of Porcine Growth Hormone," *Proc. Nat. Acad. Sci. U.S.A.*, **84**, 6434 (1987).

Branden, C., and J. Tooze, *Introduction to Protein Structure*, Garland, New York and London, p. 36 (1991).

Brems, D. N., S. M. Plaisted, E. W. Kauffman, M. Lund, and S. R. Lehrman, "Helical Formation in Isolated Fragments of Bovine Growth Hormone," *Biochemistry*, **26**, 7774 (1987).

Brems, D. N., "Solubility of Different Folding Conformers of Bovine Growth Hormone," *Biochemistry*, **27**, 4541 (1988).

Brems, D. N., S. M. Plaisted, H. A. Havel, and C.-S. C. Tomich, "Stabilization of an Associated Folding Intermediate of Bovine Growth Hormone by Site-Directed Mutagenesis," *Proc. Nat. Acad. Sci. U.S.A.*, **85**, 3367 (1988).

Brems, D. N., "Folding of Bovine Growth Hormone," *Protein Folding: Deciphering the Second Half of the Genetic Code*, Amer. Assoc. for Adv. of Sci., p. 129 (1990).

Brems, D. N., and H. A. Havel, "Folding of Bovine Growth Hormone is Consistent with the Molten Globule Hypothesis," *Proteins: Struct. Funct. Genet.*, **5**, 93 (1989).

Chen, C.-C., and L. B. Evans, "A Local Composition Model for the Excess Gibbs Energy of Aqueous Electrolyte Systems," *AIChE J.*, **32**, 444 (1986).

Chen, C.-C, Y. Zhu, and L. B. Evans, "Phase Partitioning of Biomolecules: Solubilities of Amino Acids," *Biotechnol. Prog.*, **5**, 111 (1989).

Chen, C.-C., Y. Zhu, J. King, and L. B. Evans, "A Molecular Thermodynamic Approach to Predict the Secondary Structure of Homopolypeptides in Aqueous Solution," *Biopolymers*, **32**, 1375 (1992).

Chen, C.-C., "A Segment-Based Local Composition Model for the Gibbs Energy of Polymer Solutions," *Fluid Phase Equilibria*, **83**, 301 (1993).

Chothia, C., M. Levitt, and D. Richardson, "Helix to Helix Packing in Proteins," *J. Mol. Biol.*, **145**, 215 (1981).

Cleland, J. L., and D. I. C. Wang, "Cosolvent Assisted Protein Refolding," *Bio/Technol.*, **8**, 1274 (1990a).

Cleland, J. L., and D. I. C. Wang, "Refolding and Aggregation of Bovine Carbonic Anhydrase B: Quasi-Elastic Light Scattering Analysis," *Biochemistry*, **29**, 11072 (1990b).

Cleland, J. L., C. Hedgepeth, and D. I. C. Wang, "Polyethylene Glycol Enhanced Refolding of Bovine Carbonic Anhydrase B," *J. Biol. Chem.*, **267**, 13327 (1992).

Cleland, J. L., "Impact of Protein Folding on Biotechnology," *Protein Folding: In Vivo and In Vitro*, Am. Chem. Soc. Symposium Series 526, Amer. Chem. Soc., p. 1 (1993).

Fields, G. B., D. O. V. Alonso, D. Stigter, and K. A. Dill, "Theory for the Aggregation of Proteins and Copolymers," *J. Phys. Chem.*, **96**, 3974 (1992).

Flory, P. J., "Thermodynamics of High Polymer Solutions," *J. Chem. Phys.*, **9**, 660 (1941).

Flory, P. J., "Thermodynamics of High Polymer Solutions," *J. Chem. Phys.*, **10**, 51 (1942).

Fredenslund, A., R. L. Jones, and J. M. Prausnitz, "Group Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures," *AIChE J.*, **21**, 1086 (1975).

Georgiou, G., and E. DeBernardez-Clark, *Protein Refolding*, Amer. Chem. Soc. Symp. Ser. 470 (1991).

Haase-Pettingell, C., and J. King, "Formation of Aggregates from a Thermolabile *in Vivo* Folding Intermediate in P22 Tailspike Maturation," *J. Biol. Chem.*, **263**, 4977 (1988).

Himmel, M. E., and G. Georgiou, *BioCatalyst Design for Stability and Specificity*, Amer. Chem. Soc. Symp. Ser. 516 (1993).

Huggins, M. L., "Solutions of Long Chain Compounds," *J. Chem. Phys.*, **9**, 440 (1941).

Huggins, M. L., "Some Properties of Solutions of Long-Chain Compounds," *J. Phys. Chem.*, **46**, 151 (1942).

Lehrman, S. R., J. L. Tuls, H. A. Havel, R. J. Haskell, D. J. Putnam, and C.-S. C. Tomich, "Site-Directed Mutagenesis to Probe Protein Folding: Evidence That the Formation and Aggregation of a Bovine Growth Hormone Folding Intermediate Are Dissociable Processes," *Biochemistry*, **30**, 5777 (1991).

Marston, F. A. O., "The Purification of Eukaryotic Polypeptides Synthesized in *Escherichia coli*," *Biochem. J.*, **240**, 1 (1986).

Mitraki, A., and J. King, "Protein Folding Intermediates and Inclusion Body Formation," *Bio/Technol.*, 7, 690 (1989).

Mitraki, A., C. Haase-Pettingell, and J. King, "Mechanisms of Inclusion Body Formation," *Protein Refolding*, Amer. Chem. Soc. Symp. Ser. 470, p. 35 (1991a).

Mitraki, A., B. Fane, C. A. Haase-Pettingell, J. Sturtevant, and J. King, "Global Suppression of Protein Folding Defects and Inclusion Body Formation," *Science*, **253**, 54 (1991b).

Mitraki, A., M. Danner, J. King, and R. Seckler, "Temperature-Sensitive Mutations and Second-Site Suppressor Substitutions after Folding of the P22 Tailspike Protein *in Vitro*," *J. Biol. Chem.*, **268**, 20071 (1993).

Prausnitz, J. M., "Molecular Thermodynamics for Chemical Process Design," *Science*, **205**, 759 (1979).

Presta, L. G., and G. D. Rose, "Helix Signals in Proteins," *Science*, **240**, 1632 (1988).

Renon, H., and J. M. Prausnitz, "Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures," *AIChE J.*, **14**, 135 (1968).

Stigter, D., and K. Dill, "Theory for Protein Solubilities," *Fluid Phase Equilibria*, **82**, 237 (1993).

Tandon, S., and P. Horowitz, "Detergent-Assisted Refolding of Guanidinium Chloride-Denatured Rhodanese," *J. Biol. Chem.*, **261**, 15615 (1986).

Wetzel, R., *Stability of Protein Pharmaceuticals: In Vivo Pathways of Degradation and Strategies for Protein Stabilization*, T. J. Ahern and M. C. Manning, eds., Plenum, New York, p. 43 (1992).

Wetzel, R., and B. A. Chrunyk, "Mutational Effects on Inclusion Body Formation," in *Biocatalyst Design for Stability and Specificity*, Amer. Chem. Soc. Symp. Ser. *516*, p. 116 (1993).

Zettlmeissl, G., R. Rudolph, and R. Jaenicke, "Reconstitution of Lactic Dehydrogenase. Noncovalent Aggregation vs. Reactivation: 1. Physical Properties and Kinetics of Aggregation," *Biochemistry*, **18**, 5567 (1979).

Zhu, Y., L. B. Evans, and C.-C. Chen, "Representation of Phase Equilibrium Behavior of Antibiotics," *Biotech. Prog.*, **6**, 266 (1990).

Zhu, Y., C.-C. Chen, J. King, and L. B. Evans, "Molecular Thermodynamic Model to Predict the α-Helical Secondary Structure of Polypeptide Chains in Solution," *Biochemistry*, **31**, 10591 (1992).

Manuscript received July 15, 1994, and revision received Oct. 18, 1994.