

Analytical model for studying how environmental factors influence protein conformational stability in solution

Jason K. Cheung and Prajakta S. Raverkar

Department of Chemical Engineering, The University of Texas at Austin, Austin, Texas 78712

Thomas M. Truskett

Department of Chemical Engineering and Institute for Theoretical Chemistry, The University of Texas at Austin, Austin, Texas 78712

(Received 26 September 2006; accepted 8 November 2006; published online 12 December 2006)

We introduce an analytical modeling strategy for probing the conformational stability of globular proteins in aqueous solution. In this approach, the intrinsic (i.e., infinite dilution) thermodynamic stability and coarse structural properties of the proteins, as well as the effective protein-protein interactions, derive from a heteropolymer collapse theory that incorporates predicted temperature- and pressure-dependent hydrophobic interactions. Protein concentration effects are estimated by integrating this information into a molecular thermodynamic model, which is an *ad hoc* generalization of the exact equilibrium theory of a one-dimensional binary mixture of square-well particles that interconvert through an isomerization (i.e., folding) reaction. The end result is an analytical multiscale modeling approach which, although still schematic, can predict that folded proteins exhibit a closed-loop region of stability in the pressure-temperature plane and that protein concentration has a nonmonotonic effect on protein stability, results consistent with qualitative trends observed in both experiments of protein solutions and simulations of coarse-grained protein models. © 2006 American Institute of Physics. [DOI: [10.1063/1.2403134](https://doi.org/10.1063/1.2403134)]

I. INTRODUCTION

Many proteins have been described as “marginally stable” because their native folds, responsible for biological activity, remain intact for only a narrow range of environmental conditions. Outside of these conditions, the proteins misfold or denature, often resulting in off-pathway aggregation,¹⁻³ which in turn decreases protein drug shelf life and/or reduces biological efficacy.⁴⁻⁶ The unfolding and aggregation of proteins have also been implicated in debilitating pathologies such as Alzheimer’s and Creutzfeldt-Jakob’s diseases.^{7,8} Consequently, understanding what destabilizes the native state of proteins in solution represents an important step in determining the appropriate processing conditions for pharmaceutical biotherapeutics as well as in providing a physical basis for studying protein-based diseases.

Protein stability in the type of concentrated solution environments relevant for most biological phenomena and pharmaceutical processes depends on both (i) the intrinsic (i.e., infinite dilution) thermodynamic stability of the protein molecules and (ii) the protein-protein interactions. As we explain below, hydrophobic interactions generally provide intrinsic stability to the native state, but they can also act to destabilize the native state through protein-protein interactions. Similar competitions emerge with other types of interactions (hydrogen bonding, electrostatics, etc.), but we focus on hydrophobic interactions in the present study because they represent a relatively generic and central driving force for protein folding and other assembly processes in aqueous solution,^{9,10} and thus serve as a natural starting point for our analysis.

We first consider the most basic effect of hydrophobic interactions on the folding of a single protein molecule (i.e., the case of infinite dilution). Kauzmann noted that changes in thermodynamic conditions that decrease the thermodynamic penalty for transferring nonpolar molecules from an oily phase into water similarly destabilize proteins, presumably by weakening the effective interactions that hold together the hydrophobic core of the native state.¹¹ Indeed, there is now ample evidence indicating that thermodynamic perturbations which weaken hydrophobic interactions, such as cooling to low temperature,¹²⁻¹⁴ applying high pressure,¹⁵⁻¹⁷ or introducing chaotropes,^{18,19} also tend to promote protein denaturation. In other words, hydrophobic interactions contribute substantially to the intrinsic native-state stability of individual protein molecules.

However, hydrophobic interactions also constitute a significant component of effective protein-protein interactions in solution,²⁰ and the strength of these attractions can depend sensitively on the conformational states of the participating proteins. For example, misfolded or denatured proteins may experience more favorable protein-protein interactions with one another because they desolvate a greater number of surface hydrophobic residues upon association than do their native-state counterparts. Thus, based solely on protein-protein interactions, one might expect hydrophobic interactions to destabilize the native state. The main point here is that since hydrophobic interactions impact both the intrinsic stability of a protein (generally favoring the native state) and the protein-protein attractions (potentially destabilizing the native state), a nontrivial competition can arise. Of course, properties specific to the protein system of interest will de-

termine which factor wins out under a given set of thermodynamic conditions. Unfortunately, there are few modeling strategies that currently allow one to make even qualitative predictions about this type of issue because it requires information about a wide variety of challenging physics, including hydrophobic interactions, protein folding, solvent-mediated protein-protein interactions, and the global thermodynamics of protein solutions.

In this work, our aim is to present a schematic approach based on analytical models to study the equilibrium conformational stability of folded proteins in concentrated solution environments. Specifically, our strategy treats proteins as “foldable” heteropolymers comprising hydrophobic and polar segments that qualitatively reflect the aqueous phase solubilities of the amino acids in the protein sequence. One of the central aspects of the modeling approach is that it accounts for, in an approximate manner, the energetic and entropic consequences associated with the formation of intra- and interprotein “contacts” between hydrophobic segments in the native and denatured states. Given that the strength of the hydrophobic interaction varies with thermodynamic conditions, so do the equilibrium probabilities associated with forming hydrophobic contacts and, consequently, finding proteins in their native or denatured conformations, respectively.

Our model is based on a recently introduced generalization²¹ of Dill and co-workers’ heteropolymer collapse (HPC) theory^{22,23} that incorporates an estimate of the effect that temperature and pressure have on the interactions between hydrophobic residues. The HPC theory is used in our framework to predict the intrinsic stability of the proteins and some average structural properties of their native and denatured conformations. Using the structural information predicted by this HPC theory, we determine state-dependent protein-protein interactions^{24,25} in solution. Protein concentration effects are then estimated by integrating this information and the intrinsic free energy of folding into a molecular thermodynamic model, which is an *ad hoc* generalization of the exact equilibrium theory²⁶ of a one-dimensional binary mixture of square-well particles that interconvert through an isomerization (in this case, folding) reaction.

One of the main predictions of this modeling strategy is that the effect of protein concentration on native-state stability depends strongly on the hydrophobic composition of the protein sequence. Proteins with low sequence hydrophobicity also have low intrinsic native-state stability, and thus have a high probability of being in the denatured state under conditions of infinite dilution. Moreover, since their interprotein attractions are weak, increasing the solution concentration of these proteins ultimately increases their native-state stability due to a self-crowding mechanism—the expanded denatured conformation gets “crowded”^{24,25,27–29} out of solution in favor of the compact native state. However, proteins with a higher sequence composition of hydrophobic residues show pronounced nonmonotonic behavior: increases in protein concentration destabilize the native state at low protein concentrations but restabilize the native state at high protein concentrations. The destabilization is due to the formation of very favorable protein-protein interactions involving the

denatured state, while the restabilization at high protein concentrations is again due to self-crowding. These trends qualitatively agree with previous simulation results of a coarse-grained protein model^{24,25} and appear consistent with the experimental behaviors of globular proteins in solution.^{30,31}

The outline of this article is as follows. First, we review some of the basic physics of the HPC theory that we employ for calculating the intrinsic properties of the proteins. We then discuss our method for accounting for protein-protein interactions and concentration effects. Finally, we analyze some of the predictions of our model and present concluding remarks.

II. INTRINSIC STABILITY OF A PROTEIN

We derive the intrinsic thermodynamics of folding and average structural properties of proteins in their native and denatured states using an equilibrium, two-state HPC theory.²¹ This particular approach is based on an HPC model originally introduced by Dill and co-workers^{22,23} and has been recently generalized²¹ so that it can approximately account for the effects of both temperature and pressure on hydrophobic interactions. Since the mathematical details of the HPC theory and its generalization are outlined in the original papers, we do not reproduce that information here. Rather, we discuss the main inputs and outputs of the model and some of the physical aspects that we feel will be helpful for understanding the results of the present work.

In this version of HPC theory, a protein is modeled as a random copolymer composed of hydrophobic and polar segments in an aqueous solution. The protein is treated at the single-molecule level (i.e., infinite dilution) in order to isolate the intrinsic stability of its native state, and thus interactions with other proteins in solution are not considered. The amino acid sequence of the protein is characterized by its number of effective copolymer segments N_s (related to the number of amino acid residues $N_r=1.4N_s$) (Ref. 23) and the fraction of those segments considered to be hydrophobic Φ (based on, e.g., an aqueous-phase solubility criterion^{23,32}). The folded (“native”) state is favored over the unfolded (“denatured”) state in this model under conditions for which the free energy benefit accompanying the net increase in the number of hydrophobic interactions in the folded conformation outweighs the entropic penalty associated with folding.

Temperature T and pressure P play nontrivial roles in this HPC theory because they affect the magnitude of the solvent-mediated hydrophobic attraction between hydrophobic segments $\chi(T, P)$ (in units of $k_B T$, where k_B is the Boltzmann constant). The quantity $\chi(T, P)$ is estimated using a molecular thermodynamic model²¹ that accounts for the interfacial free energy change associated with bringing water into contact with the hydrophobic segments of the heteropolymer. Since both the model for $\chi(T, P)$ (Ref. 21) and the molecular thermodynamic description of water upon which the model is based³³ are described in detail elsewhere, we mention only a few of the key predictions below.

The inset of Fig. 1 shows $\chi(T, P)$ as a function of T at both $P=1$ bar and $P=2$ kbar.²¹ The maxima in $\chi(T, P)$ as a function of temperature reflect the characteristic solubility

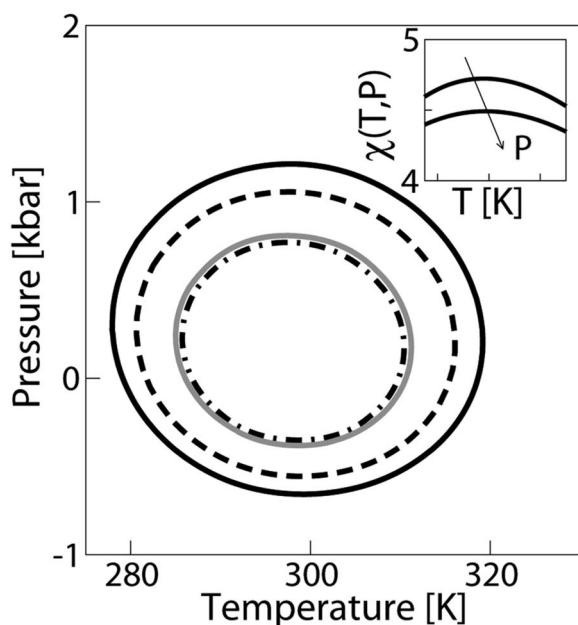


FIG. 1. Loci of midpoint folding transitions for a model protein of sequence hydrophobicity $\Phi=0.4$ and $N_r=154$ amino acid residues in the pressure-temperature plane for total protein concentrations $\rho_{\text{NN}}^3=0$ (solid black), 0.05 (dashed), 0.26 (dot-dashed), and 0.31 (solid gray). Areas enclosed by the curves indicate conditions that favor the native protein state. Inset: Strength of the effective attraction $\chi(T, P)$ between hydrophobic heteropolymer segments as a function of temperature (275–330 K) at $P=1$ bar and $P=2000$ bar (the arrow indicates increasing pressure).

minima that have been experimentally observed for nonpolar molecules in water.^{9,10,23} Capturing this trend in $\chi(T, P)$ allows HPC theory to predict both cold and warm denaturation transitions of proteins. Also note that elevating the pressure from $P=1$ bar and $P=2$ kbar reduces the strength of the attractive interactions between the hydrophobic polymer segments. As discussed previously,²¹ this decrease reflects the fact that high hydrostatic pressures are predicted to reduce the effective interfacial tension between water and a hydrophobic solute. This, in turn, leads to pressure unfolding of the heteropolymers in this HPC theory,²¹ a scenario similar to the one predicted originally by information theory calculations.^{15,34} Finally, we mention that the effective interfacial tension between water and a hydrophobic solute is also predicted to be reduced by large negative pressures (not shown),^{21,35} which in turn gives rise to negative pressure-induced unfolding of heteropolymers. Putting all of this together, one finds that the HPC theory employed here predicts a closed region of intrinsic thermodynamic stability for the native state in the pressure-temperature plane (see Fig. 1),²¹ which is qualitatively consistent with the available experimental data for small globular proteins.³⁶

The most important thermodynamic output of HPC theory for the present work is the intrinsic free energy change ΔG_f^0 associated with folding. The main structural outputs of the theory include the ratio of the radii of gyration of the native and denatured states, $R_N/R_D=\rho_s^*/1/3$ (where ρ_s^* is a reduced segment density in the denatured state),²³ and the fraction of solvent-exposed residues that are hydrophobic in the native state; i.e., Θ . It is assumed that the hydrophobic

composition of the solvent-exposed residues in the denatured state is identical to that of protein sequence itself; i.e., Φ .

For sequence parameters typical of those observed in small to medium-size globular proteins,³⁷ HPC theory makes some other basic predictions. Specifically, for a given chain length, a protein with lower sequence hydrophobicity (lower Φ) exhibits decreased intrinsic native-state stability (higher ΔG_f^0), a smaller fraction of solvent-exposed residues that are hydrophobic in the native state (lower Θ), and a more expanded denatured state relative to the native conformation (smaller ρ_s^*). As we discuss in the next section, we use this structural data to estimate effective protein-protein interactions in solution,²⁴ which in turn enables us to study concentration effects on protein stability.

III. CONCENTRATION EFFECTS ON PROTEIN STABILITY

In the previous section, we described how HPC theory can be used to model the effects that temperature, pressure, and protein sequence properties (N_r , Φ) have on the intrinsic free energy difference between, and some average structural properties pertaining to, the folded and unfolded states. We now outline how this information can be used together with a simple molecular thermodynamic model to predict the qualitative effects of protein concentration on native-state stability. Our specific strategy here is to (i) use the structural properties from HPC theory to estimate effective “square-well” protein-protein interaction parameters appropriate for treating the proteins as a binary mixture of interconverting (i.e., folding/unfolding) particles with intrinsic free energy difference ΔG_f^0 and to (ii) analyze the equilibrium behavior of this system using a thermodynamic model based on an *ad hoc* generalization of the exact one-dimensional theory for square-well particle mixtures.²⁶

The square-well potential for a binary system has the following mathematical form:

$$\begin{aligned}
 V_{ij}(r) &= \infty, & r \leq \sigma_{ij}^{\text{eff}}, \\
 V_{ij}(r) &= -u_{ij}, & \sigma_{ij}^{\text{eff}} < r \leq \lambda \sigma_{ij}^{\text{eff}}, \\
 V_{ij}(r) &= 0, & r > \lambda \sigma_{ij}^{\text{eff}}.
 \end{aligned}
 \tag{1}$$

In the above relation, r is the center-to-center distance separating interacting proteins of states i and j where $ij \in (\text{NN}, \text{ND}, \text{DD})$, σ_{ij}^{eff} is the effective interprotein diameter, u_{ij} is the magnitude of the effective “contact” attraction, and $(\lambda-1)\sigma_{ij}^{\text{eff}}$ is the range of that effective attraction.

In order to determine the (generally state-dependent) values of the parameters σ_{ij}^{eff} , u_{ij} , and λ in this particle-like description of heteropolymer globules, we first consider a similarly shaped continuous potential³⁸ $V_{ij}^c(r)$ that has previously been used as a coarse-grained model for the effective interactions of globular proteins, and is known to qualitatively capture many aspects of protein solution thermodynamics and phase behavior:^{24,25,38,39}

$$V_{ij}^c(r) = \infty, \quad r < \sigma_{ij}, \quad (2)$$

$$V_{ij}^c(r) = \frac{\varepsilon_{ij}}{625} \left\{ \left[\left(\frac{r}{\sigma_{ij}} \right)^2 - 1 \right]^{-6} - 50 \left[\left(\frac{r}{\sigma_{ij}} \right)^2 - 1 \right]^{-3} \right\}, \quad r \geq \sigma_{ij}.$$

In earlier studies,^{24,25} means for estimating parameters σ_{ij} and ε_{ij} of Eq. (2) from HPC theory were introduced. Specifically, for the effective interprotein diameters, the following relationships were used, $\sigma_{DD}/\sigma_{NN} \approx R_D/R_N = \rho_s^{*-1/3}$ and $\sigma_{ND}/\sigma_{NN} \approx (1+R_D/R_N)/2 = (1+\rho_s^{*-1/3})/2$. Furthermore, simple mean-field arguments were employed to derive analytical expressions for the magnitudes of the protein-protein “contact” attractions ε_{ij} .²⁴ Consistent with the level of modeling in HPC theory, the derivation assumed that the primary driving force for protein attractions is the favorable free energy change associated with desolvating exposed hydrophobic residues on the surfaces of the participating proteins when they associate reversibly. The qualitative validity of this assumption is supported by statistical analysis of protein-protein interfaces, which reveal a higher fraction of hydrophobic residues at their binding sites,²⁰ and also by the strong role that hydrophobic interactions are known to play in protein aggregation processes.⁴

The strength of the interprotein hydrophobic interactions in this picture depends on the magnitude of the attraction between hydrophobic polymer segments $\chi(T, P)$, the respective fractions of the solvent-exposed residues on the surfaces of the participating proteins that are hydrophobic (Θ or Φ), the number of polymer segments N_s in the protein sequence, and, for denatured proteins, the reduced polymer segment density ρ_s^* . The expressions for the magnitudes of the native-denatured ε_{ND} , denatured-denatured ε_{DD} , and native-native ε_{NN} contact attractions are given by²⁴

$$\varepsilon_{ND} = \frac{N_s \chi \Phi \Theta k_B T}{12} \left(\frac{f_e(\rho_s^*)}{[1 + \rho_s^{*-1/3}]^2} + \frac{f_e(1)}{[1 + \rho_s^{*1/3}]^2} \right), \quad (3)$$

$$\varepsilon_{DD} = \frac{N_s \chi f_e(\rho_s^*) \Phi^2 k_B T}{24}, \quad (4)$$

$$\varepsilon_{NN} = \frac{N_s \chi f_e(1) \Theta^2 k_B T}{24}, \quad (5)$$

where $f_i(\rho_s^*) = [1 - (4\pi\rho_s^*/\{3N_s\})^{1/3}]^3$ is the fraction of residues in the interior of a denatured protein, $f_e(\rho_s^*) = 1 - f_i(\rho_s^*)$ is the fraction of residues on the exterior of a denatured protein, $f_i(1)$ is the fraction of residues in the native protein core (where $\rho_s^* = 1$), and $f_e(1)$ is the fraction of residues on the native protein surface.

Because $\Phi > \Theta$ (i.e., the solvent-exposed surface of the denatured state is, on average, more hydrophobic than that of the native state), the contact attractions involving denatured proteins will generally be stronger than those involving the native state (e.g., $\varepsilon_{DD} > \varepsilon_{NN}$). These more favorable interactions, which can serve as a driving force for protein denaturation in solution, generally increase in magnitude with increasing sequence hydrophobicity Φ . We will return to this point in the next section when discussing the predictions of our model.

Having established this, recall that one of the main goals of this section is to use information derived from HPC theory to estimate *square-well* protein-protein interaction parameters σ_{ij}^{eff} , u_{ij} , and λ . The advantage of adopting the square-well description of interactions over that of Eq. (2) is that there is an exact one-dimensional theory²⁶ for square-well mixtures that we can readily generalize to make qualitative predictions about the effects of protein concentration on native-state stability in solution. Nonetheless, Eq. (2) and its previously derived connections to HPC theory^{24,25} are still very helpful here because Noro and Frenkel⁴⁰ have introduced a simple method that allows one to map a continuous potential of this kind onto a square-well potential such that the two systems satisfy an extended corresponding states principle.

The first step in the Noro-Frenkel mapping is to calculate the effective protein diameter σ_{ij}^{eff} . This is accomplished using an expression suggested by Barker and Henderson,⁴¹ $\sigma_{ij}^{\text{eff}} = \int_0^\infty dr \{1 - \exp[-V_{ij,\text{rep}}^c(r)/k_B T]\}$, where $V_{ij,\text{rep}}^c(r)$ is the repulsive part of the Weeks-Chandler-Andersen decomposition⁴² of Eq. (2). The second step is to set the strength of attraction of the square-well fluid u_{ij} equal to ε_{ij} , the minimum of the pair potential in Eq. (2). The third and final step is to determine λ by requiring equality of the second virial coefficients of the effective interprotein potentials given by Eqs. (1) and (2). Figure 2 provides sample comparisons of the native-native and denatured-denatured interprotein potentials of Eq. (2) and their respective square-well counterparts obtained via the Noro-Frenkel mapping procedure.

Having determined square-well protein-protein interaction parameters from HPC theory, we now require a molecular thermodynamic model to estimate the effects of protein concentration on native-state stability. Specifically, a model is needed that can calculate the equilibrium equation of state of a binary mixture of native (N) and denatured (D) square-well particles along with the chemical potentials of the two species, μ_N and μ_D , respectively. The folding reaction can then be readily accounted for by solving the equation of state of the mixture simultaneously with the constraint $\mu_N + \Delta G_f^0 = \mu_D$, where ΔG_f^0 is determined from our generalized HPC theory.²¹

As mentioned earlier, Monson has derived the exact statistical mechanical theory for equilibrium one-dimensional (1D) square-well fluid mixtures, and that model is able to describe many of the qualitative trends of three-dimensional (3D) binary square-well fluids. One exception, of course, is that it cannot reproduce the first-order phase transitions of the 3D system because phase transitions do not occur in 1D fluids with finite range interparticle attractions.⁴³ This does not pose a particular problem in the present work because our focus is on the conformational stability of proteins in homogeneous, single-phase solutions.

Here, we introduce two *ad hoc* modifications to Monson’s 1D theory so that it serves as a better approximate model for the 3D system that we are studying. The first modification that we employ accounts for the fact that the 1D mixture packing fraction η_{1D} , given by

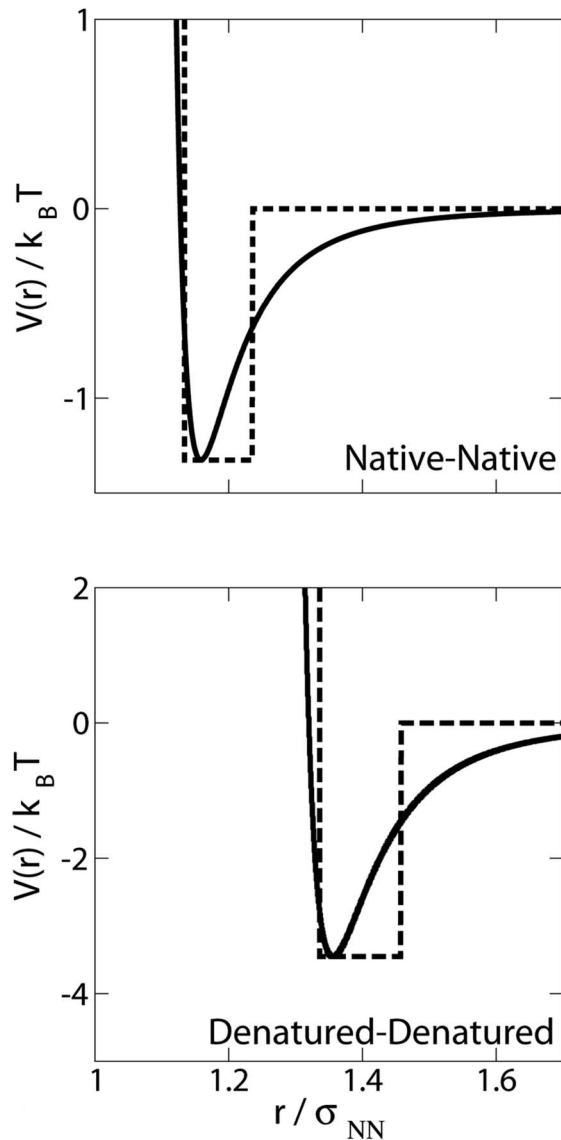


FIG. 2. Effective native-native (top panel) and denatured-denatured (bottom panel) interprotein potentials defined by Eq. (2) (solid) compared to the square well potential (dashed) determined using the Noro-Frenkel mapping (Ref. 40) for a model protein with $N_r=154$ residues and a fractional sequence hydrophobicity of $\Phi=0.5$ at $T=300$ K.

$$\eta_{1D} = \rho_{1D} \sigma_{NN}^{\text{eff}} \left[f_N + \frac{\sigma_{DD}^{\text{eff}}(1-f_N)}{\sigma_{NN}^{\text{eff}}} \right], \quad (6)$$

varies between 0 and 1, while the actual 3D mixture packing fraction η , given by

$$\eta = \frac{\pi \rho (\sigma_{NN}^{\text{eff}})^3}{6} \left[f_N + \left(\frac{\sigma_{DD}^{\text{eff}}}{\sigma_{NN}^{\text{eff}}} \right)^3 (1-f_N) \right], \quad (7)$$

should vary between 0 and η_{MRJ} . In the above, ρ_{1D} is the 1D particle number density, ρ is the 3D particle number density, f_N is the fraction of particles in the native (N) state, and η_{MRJ} is the highest attainable packing fraction of the 3D binary square-well fluid (i.e., its Maximally Random Jammed state⁴⁴). In particular, here we assume that the packing fraction of our 3D protein solution η can be calculated directly from that of the 1D mixture η_{1D} obtained via Monson's theory via the ansatz $\eta/\eta_{\text{MRJ}} = \eta_{1D}$. We also assume that the

result $\eta_{\text{MRJ}}=0.64$, valid for monodisperse hard spheres,⁴⁴ also approximately applies for the binary mixture, independent of its composition. This type of simple mapping between 1D and 3D packing fractions is similar in spirit to the one used to convert the exact 1D equation of state for hard spheres⁴⁵ to the 3D “hard-sphere” contribution to the equation of state of a van der Waals fluid.⁴⁶

The second *ad hoc* modification that we employ is that we multiply the square-well interaction potential of Eq. (1) in the 1D theory by a factor of 3. This approximately accounts for the difference that spatial dimension has on the number of nearest-neighbor interactions—the average coordination number of dense 1D fluids is ≈ 2 , while that of 3D fluids is ≈ 6 .⁴⁷ Since our overall approach focuses on the balance of intrinsic protein stability and protein-protein interactions, this modification is essential in that it allows the model to provide a more reasonable accounting for the latter in a 3D solution environment.

Clearly, the “three-dimensionalized” version of Monson's exact 1D theory produced by our modifications will not be able to provide a quantitative thermodynamic description of 3D square-well mixtures. However, that is not the goal here. Rather, our aim is to develop a simple analytical approach that can capture some of the main qualitative consequences of packing and particle interactions in our model system, and thus ultimately, when integrated with the information from HPC theory described above, aid in understanding protein stability in concentrated solutions.

IV. RESULTS

In this section, we use our analytical modeling strategy to explore a few of the combined effects of temperature, pressure, and protein concentration on the native-state stability behaviors of four model proteins with $N_r=154$ residues and sequence hydrophobicities $\Phi=0.4, 0.445, 0.473,$ and 0.5 , parameters typical of medium-sized globular proteins.³⁷ The property that we focus on specifically is f_N , the average fraction of proteins in solution that are in their native (i.e., folded) state. We also examine the loci of coordinates in the temperature-pressure plane associated with the midpoint folding transitions for these model proteins (i.e., conditions for which $f_N=0.5$).

We begin by plotting in Fig. 3 how protein concentration affects f_N for our four model proteins at $T=325$ K, contrasting the behaviors at ambient ($P=1$ bar) and at an elevated ($P=2$ kbar) pressure. Focusing on either pressure, at infinite dilution, we see that f_N simply reflects sequence hydrophobicity Φ . As discussed earlier, all other factors being equal, proteins with lower Φ typically have less intrinsic thermodynamic stability due to the fact that the hydrophobic cores of their native states contain more destabilizing polar residues. We also see that increasing the pressure from 1 bar to 2 kbar significantly reduces the intrinsic stability of the proteins, as is reflected by the decrease in f_N at infinite dilution. This qualitative trend is expected based on the inset of Fig. 1, which shows that applied pressure reduces $\chi(T,P)$, the strength of the hydrophobic interactions that stabilize the native state. One interesting point is that even very small

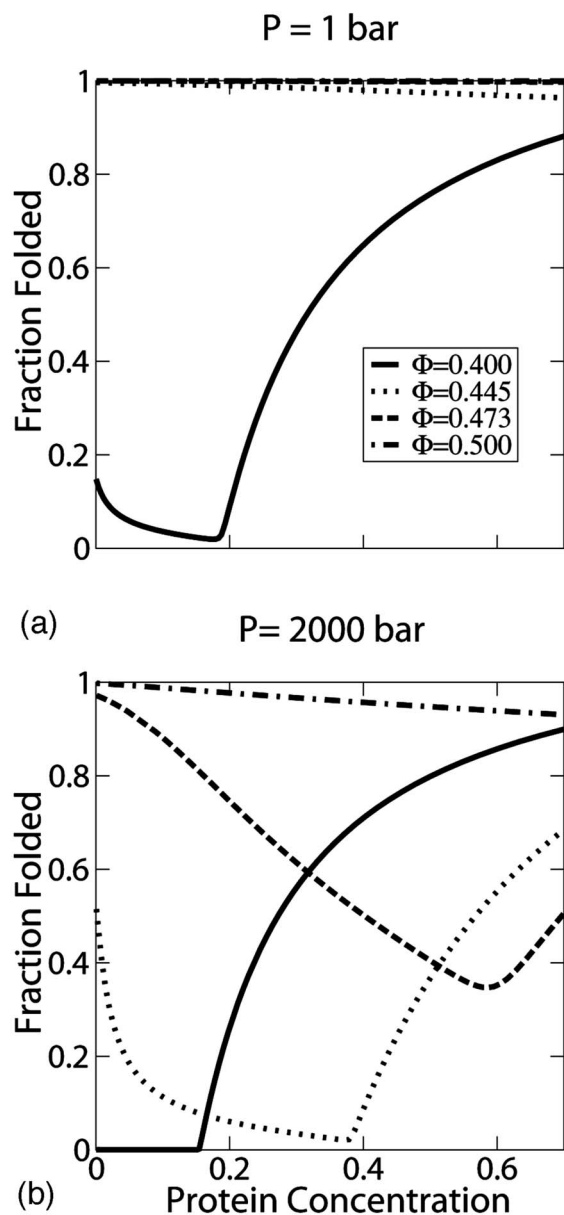


FIG. 3. Fraction of folded (native) proteins in solution as a function of protein concentration $\rho\sigma_{NN}^3$ for proteins of chain length $N_c=154$ and hydrophobicities $\Phi=0.400$ (solid), 0.445 (dotted), 0.473 (dashed), and 0.500 (dot-dashed) at $T=325$ K for (a) $P=1$ bar and (b) $P=2000$ bar.

changes in $\chi(T, P)$ (e.g., a fraction of a $k_B T$) can significantly destabilize the folded state because of the sheer number of interactions required to hold the hydrophobic core of a protein together.

Sequence hydrophobicity is not only reflected in the behaviors of proteins at infinite dilution. It also plays a role in determining how the stability of the native state in solution responds to increases in protein concentration. Notice in Fig. 3 that many of the protein solutions show the following trend: the fraction of native proteins decreases with increasing protein concentration at low protein concentrations but increases with increasing protein concentrations at high concentrations. The predicted concentration-induced destabilization of the native state at low bar concentrations is due to the fact that denatured proteins can form more favorable protein-

protein interactions than their native counterparts. This trend is in agreement with the experimentally observed stability behavior of the globular proteins myoglobin³⁰ and lysozyme,³¹ and also with computer simulations of solutions of coarse-grained model proteins.^{24,25} The subsequent restabilization of the native state at higher concentrations occurs when protein self-crowding effects (which favor the more compact native fold) outweigh the destabilizing protein-protein attractions.^{24,25} Our simple modeling approach predicts that protein sequence hydrophobicity Φ , plays an important role in the determining balance of these two opposing factors for protein stability.

To understand why, consider that a marginally stable native protein is more likely to unfold in solution if two local “destabilizing criteria” are met: (i) it has enough local free volume to accommodate the unfolding transition from a compact native state to the more expanded denatured configuration and (ii) it can simultaneously form enough favorable interprotein hydrophobic contacts upon denaturing to overcome the intrinsic free energy penalty for unfolding. The latter is certainly aided by the fact, discussed earlier, that the effective interprotein attractions involving the denatured state are much stronger than those involving only native molecules. Clearly, increasing protein concentration decreases the probability of (i) but increases the likelihood of (ii). Thus, a competition arises, and the specific properties of the native and denatured states determine the “winner” under a given set of thermodynamic conditions.

If a protein has a sufficiently high sequence hydrophobicity Φ , then it will also generally have very high intrinsic stability (i.e., it will not be “marginally stable”), and thus increases in protein concentration will have a negligible effect on f_N . This can explain the behaviors of the higher hydrophobicity ($\Phi=0.445, 0.473, 0.5$) proteins at $P=1$ bar shown in Fig. 3. In contrast, the lowest hydrophobicity protein ($\Phi=0.4$) is marginally unstable under these conditions (making it vulnerable to the effects of protein-protein interactions), and thus it shows the nonmonotonic concentration dependency mentioned above.

Figure 3 (bottom panel) demonstrates that the elevated pressure of 2 kbar has sufficiently weakened the hydrophobic interactions so that the $\Phi=0.445$ and 0.473 proteins are now only marginally stable at infinite dilution. The consequence of this is that both proteins show nonmonotonic concentration dependencies of f_N , reflecting the competition between destabilizing protein-protein attractions (which dominate at low concentrations) and crowding effects (which dominate at high concentrations). However, the $\Phi=0.4$ protein has clearly become unstable ($f_N \approx 0$) at infinite dilution and $P=2$ kbar. Since it has fewer hydrophobic residues, and since $\chi(T, P)$ is weakened under these conditions, protein-protein attractions play a negligible role in determining its stability behavior. Instead, the $\Phi=0.4$ protein is monotonically stabilized by protein concentration, reflecting the fact that the more expanded denatured proteins become increasingly crowded out of solution as protein concentration is raised. This latter self-crowding trend is closely related to the observed increase in stability of native proteins due to the

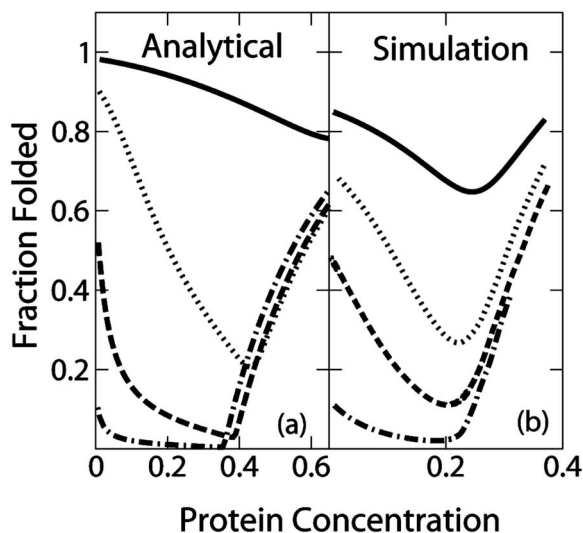


FIG. 4. Fraction of folded (native) proteins in solution as a function of protein concentration $\rho\sigma_{\text{NN}}^3$ for proteins of chain length $N_r=154$ and hydrophobicity $\Phi=0.445$ at $P=1$ bar. (a) Predictions of the present analytical model for $T=330, 335, 340,$ and 345 K. (b) Results of molecular simulations of a solution of coarse-grained model proteins for $T=348, 350, 352,$ and 356 K. Simulation data is from Shen *et al.* (Ref. 25).

lack of available free volume in confined environments^{48–50} and in solutions with a concentration of inert macromolecules.^{27–29}

It is also apparent from Fig. 3 that there is a systematic trend involving the concentration where the minimum in f_N for a given protein solution is located; i.e., where destabilizing interprotein attractions give way to the self-crowding effect. The trend is that this “turnaround” concentration increases with increasing sequence hydrophobicity Φ . This behavior results from HPC theory’s prediction that, all other factors being equal, decreasing sequence hydrophobicity results in a more expanded unfolded state. Since this type of more expanded denatured conformation takes up more space, it is crowded out of solution at a lower total protein concentration as compared to the denatured state of a higher hydrophobicity protein. This is in accord with basic principles of polymer physics: water is a poor solvent for the apolar residues, so the denatured state of high Φ proteins corresponds to a more compact ensemble of conformations than that of low Φ proteins due to the increased number of hydrophobic residue contacts that the higher Φ denatured protein forms in solution.

Behavior similar to that shown in Fig. 3 is observed if one studies the effect of temperature on the concentration dependence of f_N [see Fig. 4(a)]. Specifically, for conditions near the warm denaturation transition, increasing temperature has a destabilizing effect analogous to that of reducing the sequence hydrophobicity of the protein. The reason is that the former weakens the hydrophobic interactions that stabilize the native state, while the latter reduces the number of stabilizing hydrophobic interactions. As should be expected, the predictions of the theory shown in Fig. 4(a) are in good qualitative, but not quantitative, agreement with the results of recent simulations²⁵ of coarse-grained proteins in solution presented in Fig. 4(b).

Finally, we return to Fig. 1 to investigate to what extent the global stability of the $\Phi=0.4$ native protein in the P - T plane is impacted by protein concentration (similar trends, not shown explicitly here, are found for solutions of the other three model proteins). The main point is that this modeling strategy predicts that protein concentration can, in principle, have a significant effect on the size of the region in the P - T plane where the native state is predicted to be more stable than the unfolded state. We are not aware of any experimental studies that probe this type of global effect of protein concentration on the midpoint folding transition of a protein. This is no doubt due, in part, to the fact that concentration-induced changes in native-state protein stability can lead to irreversible aggregation, which complicates (and in many cases prevents) measurements of equilibrium properties. Nonetheless, effects as large as those shown in Fig. 1 might also have important implications for, and thus may be reflected in, protein aggregation rates. However, as the focus here is the introduction of a modeling strategy for probing the thermodynamic stability of proteins in concentrated solutions, we reserve further exploration of the protein aggregation issue for future studies.

V. CONCLUSIONS

We have introduced an elementary approach for studying the conformational stability of globular proteins in aqueous solution as a function of temperature, pressure, and protein concentration that is based on analytical models of various physical phenomena. The intrinsic (i.e., infinite dilution) thermodynamic stability and the average structural properties of the proteins, as well as the effective protein-protein interactions, are estimated from a mean-field heteropolymer collapse theory that accounts for temperature- and pressure-dependent hydrophobic interactions. The effects of protein concentration are calculated using this information and a molecular thermodynamic model, which is a *ad hoc* generalization of the exact equilibrium theory of a one-dimensional binary mixture of square-well particles that interconvert through an isomerization (i.e., folding) reaction. This modeling approach predicts that native proteins exhibit a closed-loop region of stability in the pressure-temperature plane and that protein concentration has a nonmonotonic effect on protein stability. Both results are qualitatively consistent with trends observed in experiments of protein solutions and simulations of coarse-grained protein models.

We understand, though, that our approach treats proteins in a highly simplified manner and lacks specific structural details due to the coarse resolution of the adopted HPC model. Certainly a more detailed model may help bridge the gap between experimental observations and simulation/theoretical predictions at the expense of computational tractability. However, our simple approach appears to capture some of the relevant physical driving forces associated with pressure-driven unfolding, and therefore we believe it offers a first step toward understanding some of the possible denaturation mechanisms in concentrated protein solutions.

ACKNOWLEDGMENTS

P. S. R. would like to thank the Texas Research Experience (TRES) program of The University of Texas at Austin for financial support. T. M. T. and J. K. C. gratefully acknowledge the financial support of the Merck Co. Foundation. T. M. T. also acknowledges financial support of the David and Lucile Packard Foundation, the Alfred P. Sloan Foundation, and the National Science Foundation Grant No. CTS-0448721.

- ¹P. L. Privalov, *Adv. Protein Chem.* **33**, 167 (1979).
- ²A. L. Fink, *Folding Des.* **3**, R9 (1998).
- ³R. Wetzel, *Trends Biotechnol.* **12**, 193 (1994).
- ⁴S. J. Shire, Z. Shahrokh, and J. Liu, *J. Pharm. Sci.* **93**, 1390 (2004).
- ⁵J. L. Cleland, M. F. Powell, and S. J. Shire, *Crit. Rev. Ther. Drug Carrier Syst.* **10**, 307 (1993).
- ⁶B. S. Kendrick, T. Li, and B. S. Chang, *Rational Design of Stable Protein Formulations* (Kluwer Academic/Plenum, Dordrecht, 2002), Chap. 3, pp. 61–83.
- ⁷J. D. Harper and P. T. Lansbury, *Annu. Rev. Biochem.* **66**, 385 (1997).
- ⁸C. M. Dobson, *Philos. Trans. R. Soc. London, Ser. B* **356**, 133 (2001).
- ⁹K. A. Dill, *Biochemistry* **29**, 7133 (1990).
- ¹⁰P. L. Privalov and S. J. Gill, *Adv. Protein Chem.* **39**, 191 (1988).
- ¹¹W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959).
- ¹²P. L. Privalov, *J. Chem. Thermodyn.* **29**, 10163 (1997).
- ¹³S. Kunugi and N. Tanaka, *Biochim. Biophys. Acta* **1595**, 329 (2002).
- ¹⁴S. T. Whitten, A. J. Kurtz, M. S. Pometun, A. J. Wand, and V. J. Hilsner, *Biochemistry* **45**, 10163 (2006).
- ¹⁵G. Hummer, S. Garde, A. E. Garcia, M. E. Paulaitis, and L. R. Pratt, *J. Phys. Chem. B* **102**, 10469 (1998).
- ¹⁶T. Ghosh, A. E. Garcia, and S. Garde, *J. Am. Chem. Soc.* **123**, 10997 (2001).
- ¹⁷G. Hummer, S. Garde, A. E. Garcia, A. Pohorille, and L. R. Pratt, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8951 (1996).
- ¹⁸K. A. Dill and D. Shortle, *Annu. Rev. Biochem.* **60**, 795 (1991).
- ¹⁹M. E. Goldberg, R. Rudolph, and R. Jaenicke, *Biochemistry* **30**, 2790 (1991).
- ²⁰C. J. Tsai, S. L. Lin, H. J. Wolfson, and R. Nussinov, *Protein Sci.* **6**, 53 (1997).
- ²¹J. K. Cheung, P. Shah, and T. M. Truskett, *Biophys. J.* **91**, 2427 (2006).
- ²²K. A. Dill, *Biochemistry* **24**, 1501 (1985).
- ²³K. A. Dill, D. O. V. Alonso, and K. Hutchinson, *Biochemistry* **28**, 5439 (1989).
- ²⁴J. K. Cheung and T. M. Truskett, *Biophys. J.* **89**, 2372 (2005).
- ²⁵V. K. Shen, J. K. Cheung, J. R. Errington, and T. M. Truskett, *Biophys. J.* **90**, 1949 (2005).
- ²⁶P. A. Monson, *Mol. Phys.* **70**, 401 (1990).
- ²⁷S. B. Zimmerman and A. P. Minton, *Annu. Rev. Biophys. Biomol. Struct.* **22**, 27 (1993).
- ²⁸D. Hall and A. P. Minton, *Biochim. Biophys. Acta* **1649**, 127 (2003).
- ²⁹A. P. Minton, *Biophys. J.* **88**, 971 (2004).
- ³⁰P. Tomicki, R. L. Jackman, and D. W. Stanley, *Lebensm.-Wiss. Technol.* **29**, 547 (1996).
- ³¹M. Cueto, M. J. Dorta, O. Munguia, and M. Llabres, *Int. J. Pharm.* **252**, 159 (2003).
- ³²Y. Nozaki and C. Tanford, *J. Biol. Chem.* **1971**, 2211 (1971).
- ³³T. M. Truskett, P. G. Debenedetti, and S. Torquato, *J. Chem. Phys.* **114**, 2401 (2001).
- ³⁴G. Hummer, S. Garde, A. E. Garcia, M. E. Paulaitis, and L. R. Pratt, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1552 (1998).
- ³⁵S. Rajamani, T. M. Truskett, and S. Garde, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9475 (2005).
- ³⁶L. Smeller, *Biochim. Biophys. Acta* **1595**, 11 (2002).
- ³⁷M. Shen, F. Davis, and A. Sali, *Chem. Phys. Lett.* **405**, 224 (2005).
- ³⁸P. R. ten Wolde and D. Frenkel, *Science* **277**, 1975 (1997).
- ³⁹D. N. Petsev, X. Wu, O. Galkin, and P. G. Vekilov, *J. Phys. Chem. B* **107**, 3921 (2003).
- ⁴⁰M. G. Noro and D. Frenkel, *J. Chem. Phys.* **113**, 2941 (2000).
- ⁴¹J. A. Barker and D. Henederson, *Rev. Mod. Phys.* **48**, 587 (1976).
- ⁴²H. C. Andersen, J. D. Weeks, and D. Chandler, *Phys. Rev. A* **4**, 1597 (1971).
- ⁴³L. Van Hove, *Physica (Amsterdam)* **16**, 137 (1950).
- ⁴⁴S. Torquato, T. M. Truskett, and P. G. Debenedetti, *Phys. Rev. Lett.* **84**, 2064 (2000).
- ⁴⁵L. Tonks, *Phys. Rev.* **50**, 955 (1936).
- ⁴⁶J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic, London, 1986).
- ⁴⁷M. D. Rintoul and S. Torquato, *Phys. Rev. E* **58**, 532 (1998).
- ⁴⁸D. K. Eggers and J. S. Valentine, *Protein Sci.* **10**, 250 (2001).
- ⁴⁹R. Ravindra, S. Zhao, H. Gies, and R. Winter, *J. Am. Chem. Soc.* **126**, 12224 (2004).
- ⁵⁰H. X. Zhou and K. A. Dill, *Biochemistry* **40**, 11289 (2001).