

Coarse-Grained Strategy for Modeling Protein Stability in Concentrated Solutions

Jason K. Cheung* and Thomas M. Truskett*[†]

*Department of Chemical Engineering, and [†]Institute of Theoretical Chemistry, The University of Texas at Austin, Austin, Texas

ABSTRACT We present a coarse-grained approach for modeling the thermodynamic stability of single-domain globular proteins in concentrated aqueous solutions. Our treatment derives effective protein-protein interactions from basic structural and energetic characteristics of the native and denatured states. These characteristics, along with the intrinsic (i.e., infinite dilution) thermodynamics of folding, are calculated from elementary sequence information using a heteropolymer collapse theory. We integrate this information into Reactive Canonical Monte Carlo simulations to investigate the connections between protein sequence hydrophobicity, protein-protein interactions, protein concentration, and the thermodynamic stability of the native state. The model predicts that sequence hydrophobicity can affect how protein concentration impacts native-state stability in solution. In particular, low hydrophobicity proteins are primarily stabilized by increases in protein concentration, whereas high hydrophobicity proteins exhibit richer nonmonotonic behavior. These trends appear qualitatively consistent with the available experimental data. Although factors such as pH, salt concentration, and protein charge are also important for protein stability, our analysis suggests that some of the nontrivial experimental trends may be driven by a competition between destabilizing hydrophobic protein-protein attractions and entropic crowding effects.

INTRODUCTION

Proteins in their native states play an important role in many biological processes and pharmaceutical applications. They participate in almost every aspect of the biochemical transport and regulation required for living organisms. They also serve as therapeutic drugs for targeting infectious diseases and cancer. However, most proteins, under most conditions, exhibit only marginal thermodynamic stability. As a result, minor sequence mutations or even small perturbations to solution parameters (e.g., pH, temperature, concentration, etc.) can result in protein denaturation (1).

This has important practical consequences. Unfolded or misfolded proteins look and behave differently than they do in their native states (2). They lack the highly specific molecular structure necessary for biological activity. Moreover, since unfolded proteins generally expose a significant number of hydrophobic core residues to the aqueous solvent, they have a tendency to associate and form non-native aggregates in solution. Unwanted protein aggregation and subsequent precipitation pose enormous problems in biological and pharmaceutical contexts (3,4). These processes are connected, although in a manner still imperfectly understood, to a number of debilitating diseases such as Parkinson's, Alzheimer's, Huntington's, and Down's syndrome (5–9). They are also known to cause rapid degradation of pharmaceutical formulations, reducing the shelf life of promising new drugs and restricting the strategies available for purification, handling, and delivery of therapeutics (10–15).

Given the technological and practical importance of protein stability, there is an urgent need to develop a generic understanding of the thermodynamic driving forces for protein unfolding, aggregation, crystallization, and phase separation. One way to facilitate this understanding is to build models that can account for, at various levels of sophistication, three vital aspects of stability for the native and denatured states: the relationship between protein sequence and structure, protein-protein interactions, and the global phase behavior of protein solutions. This is a formidable challenge because proteins are inherently large and complex molecules. Moreover, proteins encounter a wide variety of solution environments in biological and pharmaceutical processes, exposing them to thermal, mechanical, osmotic, and chemical stresses. To describe protein behavior under these conditions, one must also have a reliable method for accounting for hydrophobic interactions (16–23), which are a dominant force (24,25) in biomolecular folding and assembly events. These interactions have been particularly challenging to model because they exhibit subtle dependencies on both the state of the solution and the size and shape of the participating solutes.

Although each of the aforementioned aspects of protein stability have been long appreciated, theoretical investigations have focused more on their independent study than on devising strategies for integrating them into a single model. For example, models developed to investigate the single-molecule protein folding problem (26–30) are almost exclusively too complicated to be extended, either theoretically or via computer simulation, to investigate the collective behavior of thousands of proteins and millions of water molecules in solution. On the other hand, many recent theoretical models introduced to study the thermodynamics of

Submitted February 28, 2005, and accepted for publication July 15, 2005.

Address reprint requests to Thomas M. Truskett, Dept. of Chemical Engineering, The University of Texas at Austin C0800, Austin, TX 78712. Tel.: 512-471-6308; Fax: 512-471-7060; E-mail: truskett@che.utexas.edu.

© 2005 by the Biophysical Society

0006-3495/05/10/2372/13 \$2.00

doi: 10.1529/biophysj.105.062067

crystallization or liquid-liquid phase separation in protein solutions (31–40), although insightful in many respects, do not consider sequence information, the polymeric character of the individual proteins, or the possibility of protein unfolding. Despite the development of some powerful coarse-grained models that address some of these issues (41–60), theoretical methods that can provide a comprehensive understanding of the protein stability problem are still lacking.

One goal of this article is to introduce a new theoretical strategy for treating, in a simple integrated way, basic aspects of the single-molecule protein folding problem, protein-protein interactions, and the global thermodynamics of protein solutions. A second goal is to use this approach to study the thermal stability of protein molecules in concentrated aqueous environments. In particular, we would like to gain some physical insights into why the native-state stabilities of three commonly studied single-domain globular proteins—ribonuclease A, lysozyme, and metmyoglobin—display different experimental dependencies on protein concentration. Calorimetric studies indicate that, for the conditions investigated, ribonuclease A shows slightly increasing native-state stability with increasing protein concentration (61), whereas lysozyme exhibits decreasing stability (62). Moreover, metmyoglobin shows nonmonotonic behavior: increasing protein concentration decreases native-state stability in dilute solutions but increases stability at high concentrations (63). The nonmonotonic trend for metmyoglobin persists over a wide range of thermodynamic conditions, with the concentration of minimum stability falling in the range 10–50 mg/ml depending on the pH and salt concentration of the solution.

To our knowledge, there is not yet a conceptual framework for understanding these experimental findings or relating them to the biophysical characteristics of the individual protein molecules. However, since hydrophobic interactions are a central component for both protein stability and protein-protein interactions, it is natural to ask whether one should also expect sequence hydrophobicity to be an important factor here. If one designates amino acids Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, and Val as hydrophobic (see, e.g., Shen et al. (64)), then protein sequences for ribonuclease A, lysozyme, and metmyoglobin are 33%, 44%, and 51% hydrophobic, respectively. We will use this hydrophobicity scale throughout the present article, but other commonly used hydrophobicity scales (see, e.g., Tanford (65) and Dill et al. (66)) show the same trend. Given these differences and the experimental observations discussed above, it is reasonable to hypothesize that the hydrophobic content of a protein's sequence may help to determine the qualitative concentration dependence of its native-state stability in solution. In this article, we use our coarse-grained strategy to explore this idea, and whether predicted trends are consistent with the available experimental data for single-domain globular proteins.

In short, we find that calculations based on our model suggest a simple physical picture for understanding the con-

centration dependencies of native-state stability that is consistent with the experiments. Specifically, they predict that globular proteins with lower sequence hydrophobicity, such as ribonuclease A, will tend to be stabilized by increases in protein concentration due to entropic crowding effects. They also predict that globular proteins with higher sequence hydrophobicity, such as metmyoglobin, will display concentration destabilization at lower protein concentrations due to protein-protein interactions and crowding-induced restabilization at higher protein concentrations. Of course, there are still important open questions about the generality of these conclusions that need to be systematically explored, including quantifying the precise role of solution pH, salt concentration, and protein charge. A more comprehensive set of experimental data for the concentration-dependencies of globular protein stabilities in solution also seems necessary to fully decouple and analyze the roles of these various factors. Nonetheless, we believe that this study provides a promising starting point for predicting, and ultimately understanding, the thermodynamic consequences of protein concentration on native-state stability in solution.

GENERAL COARSE-GRAINED MODELING STRATEGY

Our general strategy for modeling the various aspects of native-state protein stability in solution is displayed in bold-face type below, followed by a brief description of how we implement this approach to develop a specific model.

1. Employ a heteropolymer collapse (HPC) model to determine thermodynamic and structural characteristics of native and denatured proteins. Here, we utilize the HPC theory introduced by Dill and co-workers (66,67) to predict the temperature-dependent free energy, radius of gyration, and number of hydrophobic surface residues of both the native and denatured states of a single protein molecule in aqueous solution.
2. Use the molecular characteristics determined from the HPC model to derive approximate center-to-center interprotein potentials. In this step, we determine the effective diameters and energies associated with native-native (*NN*), native-denatured (*ND*), and denatured-denatured (*DD*) contacts. We then integrate this information into a state-dependent protein-protein interaction potential.
3. Predict how environmental conditions and sequence information affect the native-state stability of proteins in solution. Specifically, we utilize the information from steps 1 and 2 to perform Reactive Canonical Monte Carlo (RCMC) simulations, which sample both protein molecule translations and folding/unfolding events. These simulations permit us to explore how sequence hydrophobicity, temperature, and protein concentration impact native-state stability in solution.

Single protein properties from the HPC model

We derive the single-protein properties required for our approach from the HPC theory developed by Dill and co-workers (66,67). Since the details of the HPC theory are explained in the original publications, we primarily discuss the inputs, outputs, and the essential physics of the approach here. Basically, proteins are modeled as heteropolymers of hydrophobic and polar monomers in an infinitely dilute aqueous solution. Depending on the size of the protein, the number of hydrophobic residues, and the temperature, either a collapsed (native-like) or expanded (denatured-like) state of the protein is predicted to be thermodynamically stable. The corresponding polymer collapse transition (i.e., protein folding) is thus an equilibrium process driven by the formation of favorable hydrophobic contacts and opposed by the loss of chain conformational entropy. Fig. 1 illustrates the collapse process for the heteropolymer chain. Note that the native state exposes fewer hydrophobic residues to the solvent than the denatured state, in accord with Kauzmann's original hydrophobic-core picture of protein folding (24).

To make predictions about a particular protein, HPC theory requires the following information: the total number of amino acid residues in the protein sequence N_r , the fraction of those residues considered to be hydrophobic Φ (e.g., based on thermodynamic data from oil-water partitioning experiments (65)), and the temperature T . The interactions between hydrophobic residues and solvent, or between hydrophobic and polar residues, are less favorable than those between polar residues and solvent, or between two hydrophobic or two polar residues. The effective energetic difference is assumed to scale with $\chi(T)k_B T$, a term parameterized by Dill et al. (66) to approximate the experimental free energy associated with the transfer of a typical hydrophobic amino acid from its own pure phase into water. The conformational entropic contributions to the free energy of the protein are treated using standard statistical mechanical arguments from polymer physics (67).

HPC theory predicts the following coarse structural information about the native (N) and denatured (D) states of the protein, respectively: the radii of gyration R_N and R_D , the related fractions of residues in contact with the solvent $f_c(R_N)$ and $f_c(R_D)$, and the fractions of these partially solvated

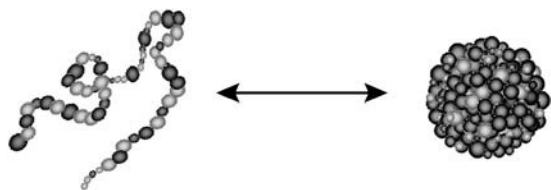


FIGURE 1 Schematic of the protein folding process predicted by HPC theory. Hydrophobic residues are light, and polar residues are dark. The denatured state is more expanded than the native state and exposes more hydrophobic amino acid residues to the solvent. Adapted from Dill et al. (66).

residues that are considered hydrophobic Θ and Φ . Based on these parameters, the theory predicts the intrinsic, i.e., infinite dilution, free energy of folding ΔG_f^0 .

Note that HPC theory captures the fact that the structure and thermodynamics of proteins can be affected both by protein sequence (e.g., sequence hydrophobicity Φ) and by solution conditions (e.g., temperature). Since these properties will impact protein-protein interactions, they will prove vital for understanding how protein concentration modifies native-state stability. Much of the relevant physics can be understood in terms of the quality or compatibility of the aqueous solvent for the protein. For instance, all other factors being equal, an increase in the number of hydrophobic residues in a protein sequence can lead to a more compact denatured state (68). This behavior is in accord with basic notions of polymer physics. Since water is a poor solvent for apolar molecules, proteins with higher hydrophobic sequence content tend to adopt more compact and structured conformations in their denatured states, increasing the number of hydrophobic-hydrophobic residue contacts. This trend has been studied indirectly in protein systems by examining how mutations affect the solvent exposure and the free energy of denatured proteins using both experiments (69) and statistical mechanical models (70,71). On the other hand, increases in temperature or denaturant concentration tend to produce more expanded denatured states since they tip the thermodynamic balance in favor of chain conformational entropy over weakened hydrophobic contacts (66).

Though the amino acid interactions and the protein sequence information are treated at a rudimentary level in HPC theory, it still gives a reasonable overall accounting of the protein folding process. For example, Fig. 2 shows that if one chooses sequence characteristics ($N_r = 154$, $\Phi = 0.5$) typical of single-domain globular proteins (see, e.g., Shen et al. (64)), HPC theory predicts free energies of folding in good qualitative agreement with the experiments. HPC theory also qualitatively reproduces many other experimental aspects of the thermodynamics of protein folding, which are discussed in detail elsewhere (66,72). However, since HPC theory assumes an infinitely dilute protein solution, one must keep in mind that the predictions only pertain to protein stability behavior in the absence of any protein-protein interactions.

Protein-protein interactions

The single-protein information provided by the HPC theory allows us to derive approximate protein-protein interaction potentials. We assume that the attractive part of the protein-protein interaction is primarily due to the driving force of proteins to desolvate their hydrophobic surface residues by burying them into a hydrophobic patch on a neighboring protein (see Fig. 3). In contrast, the repulsive part of the interaction accounts for the volume that each individual protein excludes to the centers of mass of other protein molecules in the solution.

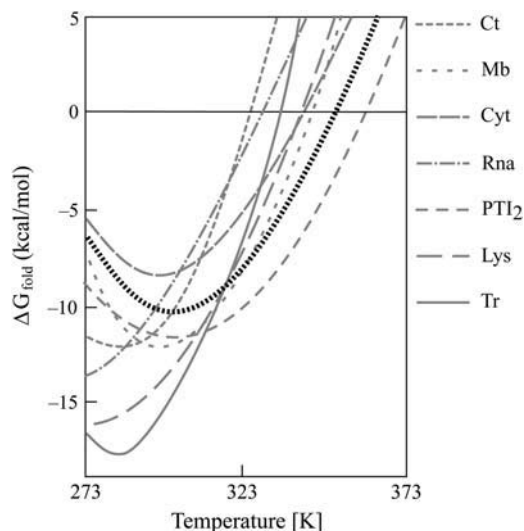


FIGURE 2 Free energy of folding for a model protein ($N_r = 154$, $\Phi = 0.5$ dotted line) from HPC theory compared to that of real proteins (*Mb*, metmyoglobin; *PTI*₂, dimer of pancreatic trypsin inhibitor; *Cyt*, cytochrome c; *Ct*, α -chymotrypsin; *Lys*, lysozyme; *Tr*, β -trypsin; and *RNA*, ribonuclease A). Reproduced with permission from Dill et al. (66). Copyright 1989 American Chemical Society.

HPC theory correctly predicts that denatured protein molecules exclude more volume ($R_D \geq R_N$) to other proteins and display higher surface hydrophobicity ($\Phi \geq \Theta$) than their native-state counterparts (66). As a result, the effective exclusion diameters and the strength of attractions between proteins should be quantitatively different for each of the various types of protein-protein interactions (i.e., *NN*, *ND*, *DD*). In fact, a simple mean-field argument suggests that the magnitudes of those various contact attractions should scale in the following way:

$$\varepsilon_{NN}(T) \propto \chi(T)\Theta^2 k_B T, \quad (1)$$

$$\varepsilon_{DD}(T) \propto \chi(T)\Phi^2 k_B T, \quad (2)$$

$$\varepsilon_{ND}(T) \propto \chi(T)\Phi\Theta k_B T. \quad (3)$$

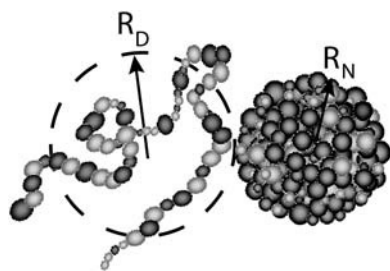


FIGURE 3 Schematic of a native-denatured protein interaction. Hydrophobic residues are light, and polar residues are dark. The denatured and native states have radii of gyration R_D and R_N , respectively. The fraction of surface residues that are hydrophobic on the denatured protein is assumed to be equal to the hydrophobicity of the protein sequence Φ . The analogous quantity for the native state, Θ (determined from HPC theory) is generally smaller (i.e., $\Theta < \Phi$). Derivation of the various types of protein-protein contact attractions are presented in Appendix A.

The proportionality constants for Eqs. 1–3 are derived in Appendix A.

We incorporate the state-dependent interaction strengths presented above and the protein sizes derived from the HPC theory into a protein-protein potential V_{ij} (34) that has been used to qualitatively capture many aspects of protein solution thermodynamics (see, e.g., Petsev et al. (73)),

$$V_{ij}(r) = \infty \quad r < \sigma_{ij}$$

$$V_{ij}(r) = \frac{4\varepsilon_{ij}}{\alpha^2} \left\{ \frac{1}{\left[\left(\frac{r}{\sigma_{ij}} \right)^2 - 1 \right]^6} - \frac{\alpha}{\left[\left(\frac{r}{\sigma_{ij}} \right)^2 - 1 \right]^3} \right\} \quad r \geq \sigma_{ij}. \quad (4)$$

Here, α is a potential range parameter; ε_{ij} is the potential well-depth; σ_{ij} is the protein-protein diameter; and $ij \in (NN, ND, DD)$. If one chooses $\alpha = 50$, the potential mimics, in an average sense, the type of relatively short-range attractions exhibited by globular proteins (34). We calculate the temperature-dependent parameters ε_{ij} from Eqs. 7–9 found in Appendix A, whereas the ratios $\sigma_{DD}/\sigma_{NN} = R_D/R_N$ and $\sigma_{ND}/\sigma_{NN} = (R_N + R_D)/2R_N$ are direct outputs from HPC theory (66).

By assuming that the above potential can describe *NN*, *ND*, and *DD* interactions, we are adopting the simplistic view that denatured proteins, like their native counterparts, are on average spherical in shape with effective sizes and attractive interactions that can be derived from the structural and energetic predictions of HPC theory. Clearly, a refined picture of the denatured state will be needed to develop a more comprehensive description of these non-native interactions. However, we note that the general strategy presented at the beginning of this section is flexible enough to accommodate the more accurate interprotein potential forms that are likely to emerge from continuing experimental studies (74,75). Thus, although our preliminary model potential is necessarily a simplistic one, future implementations will be able to readily incorporate state-of-the-art ideas about the effective interactions between protein molecules in their various states.

Fig. 4 shows the three interaction potentials for a model protein ($N_r = 154$, $\Phi = 0.5$) at 362 K, and the inset displays the temperature dependence of the corresponding potential well depths. Note that, as expected, *DD* protein contacts are more favorable than *ND* and *NN* contacts, since denatured proteins have a greater number of solvent-exposed hydrophobic residues.

Using the above potential, one can also calculate a dimensionless second virial coefficient $B_{22,ij}^*$ (76):

$$B_{22,ij}^* = \frac{12}{\sigma_{ij}^3} \int \left[1 - \exp\left(-\frac{V_{ij}(r)}{k_B T}\right) \right] r^2 dr. \quad (5)$$

The second virial coefficient is an experimentally measurable quantity (see, e.g., Tessier et al. (77)) that has been linked to protein phase behavior (76,78). Vliegthart and

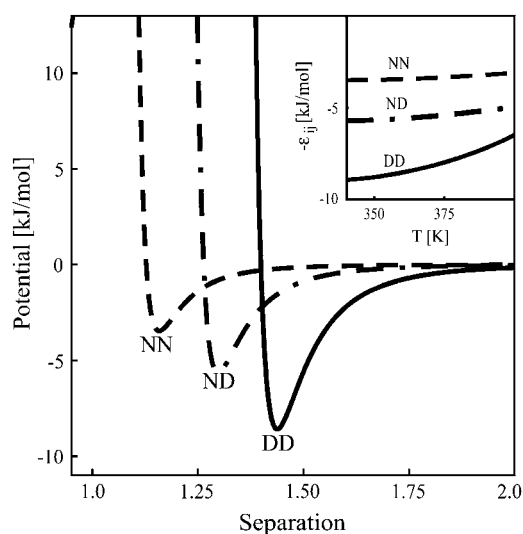


FIGURE 4 Effective protein-protein potentials, calculated using Eq. 4 for a model protein ($N_r = 154$, $\Phi = 0.5$) at $T = 362$ K. There are three different types of potentials (NN, ND, and DD) because proteins can exist in either native (N) or denatured (D) states. Note that the DD interactions are stronger and have larger exclusion volumes than the NN or ND interactions, since denatured proteins are more expanded and tend to bury more hydrophobic amino acid surface residues upon contact. (Inset plot) Potential well depths for ND, DD, and NN interactions as a function of temperature, calculated using Eqs. 7–9.

Lekkerkerker (76) noticed that liquid-liquid phase separation generally occurs in protein solutions when $B_{22,ij}^* < -6$. Fig. 5 shows $B_{22,ij}^*$ as a function of temperature for a model protein ($N_r = 154$, $\Phi = 0.5$), obtained by applying HPC theory, Eq. 4, and Eq. 5. As expected, the $B_{22,ij}^*$ values are consistently more negative for interactions involving denatured proteins since they tend to bury more hydrophobic

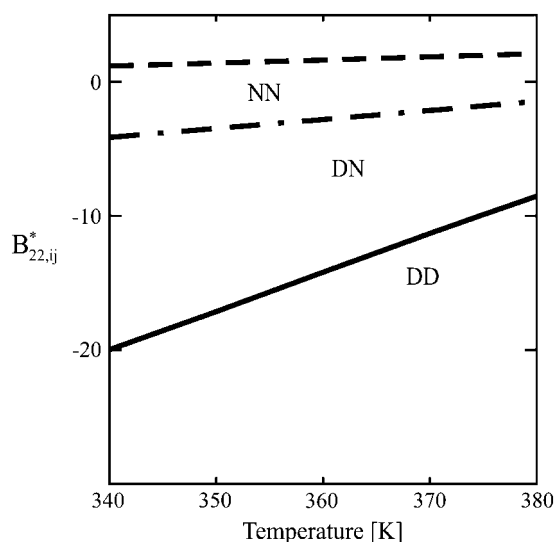


FIGURE 5 B_{22}^* calculated using Eq. 5 for the model protein ($N_r = 154$, $\Phi = 0.5$). Since DD attractions are stronger, the corresponding B_{22}^* is more negative, which generally indicates different solubility behaviors for the denatured and native states.

residues when making protein-protein contacts (see Appendix A). Moreover, since $B_{22,DD}^* < -6$ for this model protein and $B_{22,NN}^* > -6$, one might expect significant differences in the solubilities (79,80), and perhaps the aggregation behavior, of the native and denatured states. In the Results and Discussion sections of this article, we explore how these pronounced differences in protein interactions can affect native-state stability in concentrated solutions. We are also currently investigating how these differences can impact the global phase diagram of protein solutions (see Conclusions and Future Work).

Simulation details

Our ultimate aim is to understand how protein sequence information, protein-protein interactions, and protein solution variables (e.g., temperature and protein concentration) combine to determine native-state stability. To accomplish this within the context of our model, we need to integrate the single-protein thermodynamics from HPC theory and our derived protein-protein interactions into a simulation technique that can generate representative equilibrium states of solutions at finite protein concentration. However, this task is perfectly suited for the Reactive Canonical Monte Carlo (RCMC) method developed by Johnson, Panagiotopoulos, and Gubbins (81,82). This algorithm allows one to readily compute the equilibrium properties of chemically reacting systems by taking advantage of both molecular translational moves and forward/reverse reaction steps. In our case, the reaction of interest is simply the unimolecular folding/unfolding of the coarse-grained protein molecules in their finite-concentration solution environment.

The inputs required for the canonical-ensemble implementation of the RCMC simulations are the total number of protein molecules N , the dimensionless protein concentration $\rho\sigma_{NN}^3 = N\sigma_{NN}^3/V$ (where V is the volume), the temperature T , the effective protein-protein potentials from Eq. 4, and the intrinsic free energy of folding $\Delta G_f^0(T)$ from HPC theory. Given this information, the algorithm samples the phase space relevant to the equilibrium reacting system, converging relatively quickly to the equilibrium fraction f_N of native-state proteins. Since earlier articles (81–83) provide the detailed expressions for the reaction move probabilities along with strategies for efficient implementation of the RCMC algorithm, we do not elaborate on those issues here.

For each state point studied in this work, $N = 256$ proteins were simulated in a cubic box with periodic boundary conditions. To ensure that significant system-size effects were not present, we reproduced our simulations results for a number of state points using both smaller ($N = 128$) and larger ($N = 500$) systems. We have also reproduced the results for a number of state points with a grand canonical ensemble version of the algorithm. Fifteen percent of attempted moves in our simulations were denatured to native reactions (i.e., protein folding), and 15% were attempted native to denatured

reactions (i.e., protein unfolding). The remaining 70% of attempted moves were protein displacements. Initially, the proteins were placed in an open FCC lattice configuration with either all native ($f_N = 1$) or denatured ($f_N = 0$) states. They were then equilibrated for the number of MC cycles required to displace each protein at least a distance of $5\sigma_{NN}$ from its original position. The potential energy, the fraction of native-state proteins f_N , and the osmotic pressure Π were tracked closely to make sure that equilibration was completed before subsequent production runs were initiated.

To verify that stable, or at least metastable, homogeneous fluid phases were being sampled by our simulations, standard translational and bond-orientational order parameters for detecting crystallization (84,85) were monitored during the equilibration and production stages. We found that in all but a few higher-concentration state points (see Appendix B), convergence to the same fluid state could be reproducibly achieved by initiating from either all-denatured ($f_N = 0$) or all-native ($f_N = 1$) protein lattices. Of course, without knowledge of the complete phase diagram of a system, one cannot unambiguously determine, either from simulations or experiments, whether a given state is thermodynamically stable or metastable with respect to a phase change (e.g., crystallization). In this initial study, we have focused on understanding native-state stability of homogeneous fluid states of model protein solutions, and we have not attempted the larger problem of calculating rigorous thermodynamic phase boundaries for the solution. However, we are currently studying the global phase diagram of our model protein systems, and we will report those results in a future publication (see Conclusions and Future Work).

To explore how sequence hydrophobicity might qualitatively impact the concentration dependence of native-state stability using our approach, we have studied two different model proteins with the same number of amino acid residues ($N_r = 154$) but different hydrophobic contents ($\Phi = 0.4$ and $\Phi = 0.5$). As was illustrated recently (64), these hydrophobicities span the typical range of those seen in single-domain globular proteins with sizes between 125 and 175 residues, and so they reasonably bracket the types of behaviors that one might expect to see in experiments. As will be shown in Results, the difference in midpoint folding temperatures at infinite dilution for these two model proteins is ~ 26 K, with the higher hydrophobicity protein showing more thermal stability. As a rough comparison, it can be seen in Fig. 2 that the folding temperatures of ribonuclease A and metmyoglobin also differ by ~ 20 K (66). Although the model proteins ($\Phi = 0.4$ and $\Phi = 0.5$) exhibit both cold and warm denaturation, we focused exclusively on destabilization by the latter mechanism in the present work.

RESULTS

In this section, we briefly present the main results of our study. We divide our data into two main parts that pertain to the two

model proteins that were explored. In the Discussion, we explore the physical insights that can be gained from our results.

High hydrophobicity protein ($N_r = 154$, $\Phi = 0.5$)

In Fig. 6, we plot the equilibrium fraction of folded (i.e., native-state) proteins f_N as a function of temperature T for a series of protein concentrations. As is expected, for low concentrations, our simulations are extremely close to the predictions from HPC theory (66). The midpoint folding temperature T_m , i.e., the temperature that yields $f_N = 0.5$, is ~ 364 K for this protein at infinite dilution. By fitting the simulation points obtained at each protein concentration to sigmoidal curves, we can deduce the concentration dependence of T_m for our model protein. The main trends are as follows. If one starts with a protein solution at low concentrations, increasing protein concentration results in protein destabilization (i.e., T_m decreases as concentration increases). However, as one continues to increase protein concentration, the protein is ultimately restabilized (i.e., T_m rebounds to higher values). At this point, we simply assert that this non-monotonic stability behavior is due to a competition between destabilizing attractive protein-protein interactions that dominate at low protein concentration and stabilizing entropic crowding effects (see, e.g., Minton (86)) that prevail at high protein concentration. We will provide quantitative justification for this statement in the Discussion.

In Fig. 7, T_m is plotted as a function of protein concentration, yielding a protein stability “phase diagram” for the high hydrophobicity protein. The white (nonshaded) area comprises state points for which the native state is thermodynamically favored ($f_N > 0.5$), whereas the shaded area

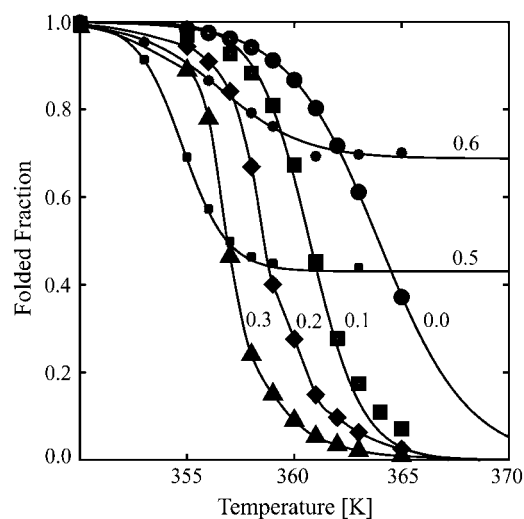


FIGURE 6 RCMC simulation results showing the fraction of folded proteins f_N as a function of temperature T for the high hydrophobicity model protein ($N_r = 154$, $\Phi = 0.5$). At low concentrations, increasing concentration destabilizes native-state proteins, whereas the opposite trend occurs at high protein concentration. The protein concentrations $\rho\sigma_{NN}^3$ are shown on the graph nearest to their curve. The lines are present as a guide to the eye.

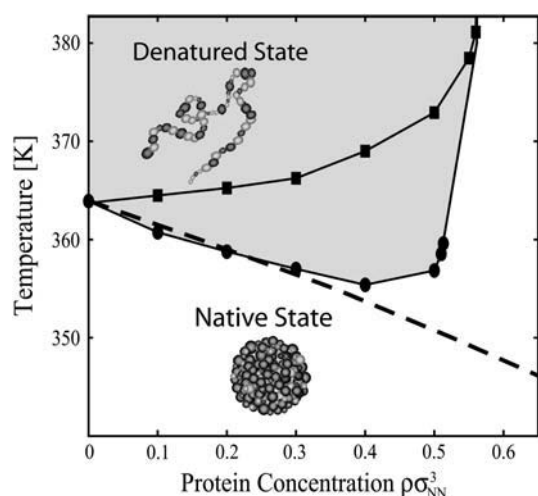


FIGURE 7 Protein stability phase diagram for the high hydrophobicity model protein ($N_r = 154$, $\Phi = 0.5$) in the temperature T -protein concentration $\rho\sigma_{NN}^3$ plane. The shaded region indicates the temperatures and concentrations that favor the denatured state ($f_N < 0.5$). Circles represent T_m from the RCMC protein simulations, the squares show T_m from the nonattracting high hydrophobicity protein simulations, and the dashed line indicates the estimated T_m from Eq. 6.

above the curve indicates the temperature-concentration coordinates where the denatured state prevails ($f_N < 0.5$). As we demonstrate later, very high protein concentrations cannot be attained in this model if a majority of proteins are to remain in the denatured state, simply due to the fact that denatured proteins exclude more volume than proteins in the more compact native state. However, attractive protein-protein interactions also play an important role at lower protein concentrations, where they induce significant concentration destabilization. For example, T_m at $\rho\sigma_{NN}^3 = 0.4$ is nearly 10 K below the infinite dilution value.

Fig. 8 provides a clearer picture for how increasing protein concentration destabilizes the native state at low concentrations and restabilizes it at high concentrations. To quantify this effect, one can define a midpoint concentration for unfolding at low concentrations, and a midpoint concentration for refolding at high concentrations. Since the intrinsic stability of a protein molecule decreases upon heating (i.e., $\Delta G_f^0(T)$ becomes less negative), proteins become more vulnerable to destabilizing effects like attractive protein-protein interactions, and thus the midpoint concentration for unfolding decreases with increasing temperature. Similarly, the midpoint concentration for refolding increases with increasing temperature, indicating that more crowding is required to refold proteins with lower intrinsic thermal stability. A figure containing all of the simulated state points for this protein can be found in Appendix B.

Low hydrophobicity protein ($N_r = 154$, $\Phi = 0.4$)

Here we explore protein concentration effects on the thermodynamic stability of a low hydrophobicity protein. In Fig.

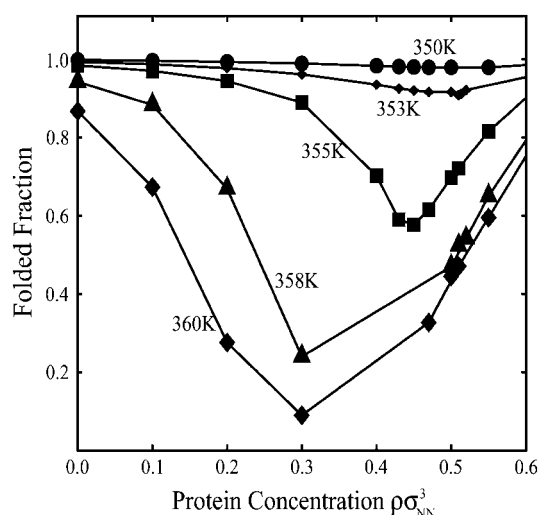


FIGURE 8 Fraction of folded proteins f_N as a function of protein concentration $\rho\sigma_{NN}^3$ for several isotherms of the high hydrophobicity model protein ($N_r = 154$, $\Phi = 0.5$). At temperatures far below the intrinsic folding temperature (~ 364 K), concentration effects are almost negligible. As we approach the intrinsic folding temperature at low concentration, increasing concentration shows a pronounced destabilizing effect. At higher concentrations, the native state is restabilized due to crowding effects. The solid lines are present as a guide to the eye.

9, we plot the fraction of native-state proteins as a function of temperature for a series of different protein concentrations. Again, as expected, the low concentration results closely agree with HPC theory. Moreover, T_m is ~ 338 K at infinite dilution, which is logically lower in value than that of the

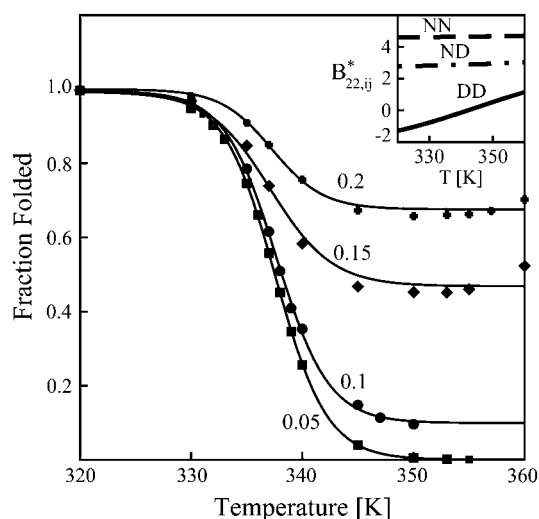


FIGURE 9 RCMC simulation results showing the fraction of folded proteins f_N as a function of temperature T for the low hydrophobicity model protein ($N_r = 154$, $\Phi = 0.4$). The general trend is increasing native-state stability with increasing protein concentration. The protein concentrations $\rho\sigma_{NN}^3$ are shown on the graph nearest to their curve. The lines are present as a guide to the eye. (Inset) The inset plot is the reduced second virial coefficient, calculated using Eq. 5, as a function of temperature. The lower hydrophobicity of this model protein leads to weaker protein-protein interactions.

high hydrophobicity protein (~ 364 K). The most striking feature of this data set is that the concentration destabilization effects seen for the high hydrophobicity protein are essentially negligible here. In short, for this model protein we find that increasing protein concentration shifts the folding equilibrium to the native state for essentially all finite protein concentrations. Fig. 10 shows the corresponding protein stability phase diagram, which highlights the stabilizing role that protein concentration plays for the low hydrophobicity protein. A figure displaying all of the simulated state points for this protein can also be found in Appendix B.

In the next section, we discuss the physical origins of the relationship between sequence hydrophobicity, protein interactions, and the different stability behaviors for the high- and low-hydrophobicity proteins.

DISCUSSION

The simulation results suggest that the thermodynamic stability of a high hydrophobicity protein can exhibit a non-monotonic dependence on concentration. At low total volume fractions, increasing protein concentration has a destabilizing effect for the native state. This finding is consistent with the low-concentration destabilization that is observed experimentally for the high hydrophobicity protein metmyoglobin ($\Phi = 0.51$) (63), and to a lesser extent with the medium hydrophobicity protein lysozyme ($\Phi = 0.44$) (62). Here, as discussed in the Introduction, we have adopted Φ values

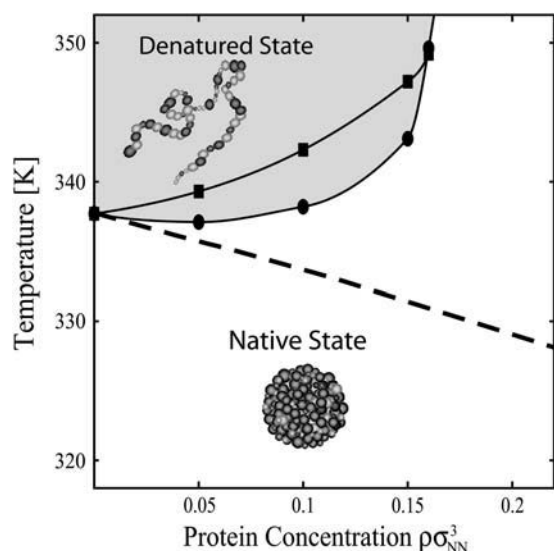


FIGURE 10 Protein stability phase diagram for the low hydrophobicity model protein ($N_r = 154$, $\Phi = 0.4$) in the temperature T -protein concentration $\rho\sigma_{NN}^3$ plane. The shaded region indicates the temperatures and concentrations that favor the denatured state ($f_N < 0.5$). Circles represent T_m from the RCMC protein simulations, the squares show T_m from the nonattracting low hydrophobicity protein simulations, and the dashed line indicates the estimated T_m from Eq. 6. Because of weaker intraprotein attractions, the denatured state is more expanded than the high hydrophobicity protein (see Fig. 11). Thus, concentration effects are mostly controlled by crowding-induced stabilization of the native-state.

calculated from a hydrophobicity scale where Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp, and Val are considered hydrophobic (64). Protein concentration destabilization of the native state is also consistent with the increase in non-native aggregation rate that has been experimentally observed with increasing protein concentration for both reversible (87,88) and irreversible protein aggregation (see, e.g., Cueto et al. (62) and Krishnamurthy and Manning (89)), although for irreversible processes it is generally difficult to determine whether a thermodynamic driving force actually underlies the kinetically controlled laboratory phenomenon.

At higher volume fractions, however, our model predicts that increasing protein concentration shifts the folding equilibrium back toward the native state. This second trend is also in agreement with the so-called entropic crowding of proteins in solution that has been explained theoretically and observed experimentally (51,52,54,86). To understand which of these two concentration effects will dominate for different proteins under various solution conditions, we require some additional insight into the molecular mechanisms.

One physical picture that is consistent with the above observations is the following. Native proteins may unfold to the denatured state if by doing so 1), they can form enough favorable attractions with neighboring proteins to outweigh their intrinsic thermal stability (i.e., their negative ΔG_f^0); and 2), they do not simultaneously create new highly unfavorable repulsive interactions with neighboring proteins due to their expanded denatured configuration. This type of picture is consistent with experimental results that indicate that reversible formation of non-native oligomers can induce unfolding (90). On the other hand, even intrinsically unstable proteins (i.e., exhibiting positive ΔG_f^0) may refold due to crowding if the protein concentration is so high that there is not enough free volume to accommodate them in their denatured states. This is quite similar in nature to the experimental protein stabilization induced by confinement (91). Below, we use our data to subject this physical picture to some basic quantitative tests.

Our strategy in this regard is to devise simple and transparent means for separately calculating the contributions of destabilizing protein-protein attractions and stabilizing crowding effects to native-state stability. Armed with the information that we obtain from these calculations, we will be in a better position to interpret the RCMC results presented in the previous section for our two model proteins. We begin by considering the destabilizing protein-protein interactions for both model proteins, and then we return to the balancing effects of crowding.

As noted in Results, the T_m for the high hydrophobicity protein ($N_r = 154$, $\Phi = 0.5$) decreases from ~ 364 K at infinite dilution to ~ 354 K at $\rho\sigma_{NN}^3 = 0.4$. To check whether destabilizing protein-protein attractions can explain this trend, we simply balance ΔG^0 (calculated from HPC theory) with the attractive protein-protein interactions that we expect the protein to gain by unfolding in solution:

$$\Delta G^0 \approx -4\pi\sigma_{\text{ND}}^2\Delta_{\text{ND}}\rho_{\text{N}}\frac{\varepsilon_{\text{ND}}}{2} - 4\pi\sigma_{\text{DD}}^2\Delta_{\text{DD}}\rho_{\text{D}}\frac{\varepsilon_{\text{DD}}}{2}. \quad (6)$$

Here, Δ_{ND} and Δ_{DD} characterize the range of the ND and DD attractions (calculated from the effective pair potentials), whereas ρ_{N} and ρ_{D} are the number densities of native and denatured proteins, respectively. In particular, the quantity Δ_{ij} represents the width of the attractive potential well, and it is defined to be the distance between the protein-protein hard-core separation σ_{ij} and the larger separation $r_{ij,u}$ where the potential energy (Eq. 4) has risen to $V(r_{ij,u}) = -0.05\varepsilon_{ij}$. Taking a look at the first term on the right-hand side of Eq. 6, the quantity $4\pi\sigma_{\text{ND}}^2\Delta_{\text{ND}}\rho_{\text{N}}$ represents the number of new contacts a protein makes with native-state proteins upon unfolding, and $-\varepsilon_{\text{ND}}/2$ is approximately the average energy of that contact interaction. Similarly, the second term quantifies the total energy of the new contacts that a protein makes with neighboring denatured proteins upon unfolding. For simplicity, we have assumed here that the local concentrations of protein states can be reasonably approximated by the corresponding bulk concentrations. To test whether the balancing idea presented above has merit, Eq. 6 is plotted along with the simulation data in Fig. 7. Note that it correctly, and even semiquantitatively, predicts the decrease in T_{m} exhibited by the RCMC simulation data for the high hydrophobicity protein.

We have also used Eq. 6 to analyze the attractive destabilizing forces for the low hydrophobicity protein solution. As can be seen by the dashed curve in Fig. 10, Eq. 6 predicts that, in the absence of entropic crowding, attraction destabilization should also be a noticeable effect for this protein. However, this prediction is clearly not born out by the full RCMC results of the low hydrophobicity protein solution. To understand why attractions give rise to a net destabilizing effect for the high hydrophobicity protein at low concentrations and not for the low hydrophobicity protein, we need to examine the concentration dependencies of the stabilizing crowding effects for the two protein solutions.

To isolate crowding effects from destabilizing attractions, we have performed additional RCMC simulations of our two model proteins, identical in all respects to the ones discussed previously except that we have now “turned off” all interprotein attractions. For the nonattracting high hydrophobicity protein, we find that the T_{m} is almost unchanged for concentrations below $\rho\sigma_{\text{NN}}^3 = 0.3$ (see the *squares* in Fig. 7). This absence of a strong crowding effect at low concentrations basically explains why the protein-protein attraction destabilization dominates under these conditions. Conversely, the native state of the nonattracting high hydrophobicity protein shows a very pronounced increase in stability for concentrations above $\rho\sigma_{\text{NN}}^3 = 0.3$. Since crowding stabilization (*squares* in Fig. 7) outweighs attraction destabilization (*dashed line* in Fig. 7) for high protein concentrations but not for low concentrations, we get a nonmonotonic concentration dependence for stability of the high hydrophobicity protein.

We also plot the T_{m} of the nonattracting low hydrophobicity protein solution (*squares*) in Fig. 10, along with Eq. 6 (*dashed curve*) and the T_{m} of the regular (i.e., attracting) low hydrophobicity protein solution (*circles*). Notice that the attraction destabilization contribution essentially exactly balances the crowding effect for low concentrations, whereas the crowding effect dominates for higher concentrations. The net result is that the native state of the low hydrophobicity protein shows nearly monotonic stabilization with increasing protein concentration.

Why does the crowding stabilization effect play a more important role, even at low concentrations, for the low hydrophobicity protein? As was hinted at above, it is not because the interprotein attractions have a weaker destabilizing effect for this protein. In fact, because lower hydrophobicity means both weaker protein-protein interactions and lower intrinsic thermal stability (i.e., less negative ΔG_{f}^0), Eq. 6 predicts that the low hydrophobicity protein would still be significantly destabilized in the absence of the entropic crowding effect. However, the weaker intraprotein attractions give rise to a more expanded denatured state for the low hydrophobicity protein (see Fig. 11). As a result, there is a larger entropic penalty for unfolding of low hydrophobicity proteins in concentrated solutions, which manifests itself as larger relative thermodynamic stability of the native state. Because of these physical factors, one might generally expect low hydrophobicity proteins to be stabilized (or, at worst, unaffected) by increases in protein concentration despite their comparatively smaller intrinsic stability. As was discussed in the Introduction, this analysis is consistent with the slightly increasing experimental stability of the low hydrophobicity protein ribonuclease A ($\Phi = 0.33$) with increasing protein concentration (61).

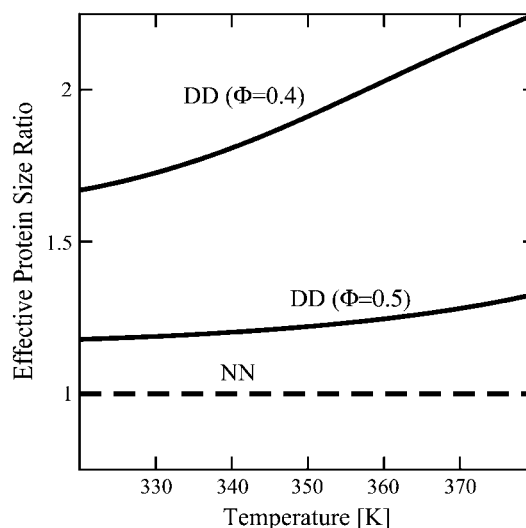


FIGURE 11 Ratio of the effective hard-core diameter ($\sigma_{\text{DD}}/\sigma_{\text{NN}}$) as a function of temperature for both model proteins: $N_{\text{r}} = 154$, $\Phi = 0.4$ and $N_{\text{r}} = 154$, $\Phi = 0.5$. Note that the lower hydrophobicity protein has a much larger effective size in the denatured state, implying an increased excluded volume.

CONCLUSIONS AND FUTURE WORK

In this article, we have introduced a new coarse-grained approach for modeling protein stability in concentrated solution environments. Our treatment calculates state-dependent center-to-center interactions between proteins from basic structural and energetic characteristics of their native and denatured states. These characteristics are each derived from simple protein sequence information using an insightful heteropolymer collapse theory. We have integrated the thermodynamic and structural information obtained from the heteropolymer collapse theory and the resulting inter-protein potentials into RCMC simulations to investigate the connections between protein sequence hydrophobicity, protein-protein interactions, protein concentration, and the thermodynamic stability of the native state.

The calculations based on this approach outline a simple physical picture for understanding the concentration dependencies of native-state stability. In particular, they predict that globular proteins with lower sequence hydrophobicity, such as ribonuclease A, should be stabilized by increases in protein concentration due to entropic crowding effects. They also predict that globular proteins with higher sequence hydrophobicity, such as metmyoglobin, should display concentration destabilization at lower protein concentrations due to protein-protein interactions and crowding-induced destabilization at higher protein concentrations. These findings are consistent with the available experimental data for these proteins. However, as is discussed below, we also recognize that there are still important open questions about the generality of our predicted trends that need to be explored through new calculations, including understanding the role of solution pH, salt concentration, and protein charge. A more comprehensive set of experimental data for the concentration dependencies of globular protein stabilities in solution also seems necessary for this effort.

Finally, we think that there are three promising directions for future research based on this work that will help to broaden our understanding of protein stability. The first is the calculation of the thermodynamic phase boundaries for our model protein solutions. In particular, we are interested in understanding whether molecular unfolding can induce liquid-liquid immiscibility between protein-rich and protein-poor solution phases. These types of thermodynamic transitions could give insights into the driving forces for aggregation or precipitation in protein systems. Moreover, we are interested in investigating how sequence hydrophobicity and protein size can impact the relative location of the liquid-liquid and solid-liquid (protein crystallization) transitions on the phase diagram of protein solutions. A second future area of inquiry is to use our approach to explore the precise molecular mechanisms for concentration-induced protein destabilization. For instance, how does the mechanism for protein destabilization (e.g., the reversible formation of non-native dimers and trimers) depend on the protein

characteristics and solution conditions? The third direction is to account for protein charge, solution pH, and salt concentration in our model calculations. Alonso, Dill, and Stigter (92,93) have outlined a systematic methodology for incorporating these effects into the heteropolymer collapse theory used in this work, and we are currently investigating how this type of treatment could be extended into our general coarse-grained simulation strategy.

APPENDIX A: ATTRACTIVE WELL-DEPTH DERIVATION

The effective protein-protein contact attractions in our model depend on the molecular characteristics of the proteins involved. Two of those characteristics come directly from the protein sequences, e.g., the total number of hydrophobic $N_f\Phi$ and polar $N_r(1 - \Phi)$ amino acid residues (66). Other properties are structural and can be derived from HPC theory. These structural quantities, which generally take on different values for native (N) and denatured (D) proteins, include the radius of gyration (R_N or $R_D(T)$), the fraction of residues on the protein surface ($f_e(R_N)$ or $f_e(R_D)$), where $f_e(r) = 1 - (1 - r^{-1})^3$, and the fraction of surface residues that are hydrophobic ($\Theta(T)$ or Φ for N or D states, respectively). The protein-protein interactions also depend on the free energy associated with the hydration of a hydrophobic amino acid residue $\chi(T)k_B T$, and this term represents the only place where solvent effects enter the theory. The derivation and physical interpretation of each of the quantities mentioned above are described in detail in the original HPC articles (66,67). Here, we use the quantities to derive approximate expressions for the attractive well depths (see Fig. 4, *inset*) of our coarse-grained protein-protein potentials using geometric approximations and mean-field energetic arguments. These expressions are not intended to quantitatively reproduce the interactions for specific proteins, but rather to provide simple and transparent relations between the strength of globular protein-protein attractions and the molecular characteristics of the proteins themselves.

Here, we explicitly derive the general relationship for the ND contact attraction, and the analytical forms for the NN and DD attractions can be easily deduced from this expression. As a first step, we approximate the fraction of the denatured protein's surface desolvated by contact with a native protein to be $\pi R_N^2 / 4\pi (R_N + R_D)^2 = (4[1 + \{R_D/R_N\}]^2)^{-1}$, which is simply the projected surface area of the native protein divided by the total ND interaction area. If one analyzes this expression in the limit of $R_N = R_D$, then one finds that it makes the reasonable, but not quantitatively exact, prediction that one protein would be completely desolvated if it were closely surrounded by eight neighboring proteins. Of course there should also be an $\mathcal{O}(1)$ prefactor included to correct this formula to account for the specific packing properties of the proteins under study. However, to keep the ideas simple and general, we do not attempt to describe that level of detail here. To determine the surface fraction of a native protein that is desolvated by contact with a denatured protein, we use a similar argument: $\pi R_D^2 / 4\pi (R_N + R_D)^2 = (4[1 + \{R_N/R_D\}]^2)^{-1}$.

The magnitude of the effective energy change upon making the aforementioned ND contact is then given by

$$\begin{aligned} \epsilon_{ND} = & \frac{N_s}{4[1 + \frac{R_D}{R_N}]^2} \chi(T) f_e(R_D) (1 - \sigma) \Phi \Theta k_B T \\ & + \frac{N_s}{4[1 + \frac{R_N}{R_D}]^2} \chi(T) f_e(R_N) (1 - \sigma) \Phi \Theta k_B T. \end{aligned} \quad (7)$$

In this expression, $N_s = N_r/1.4$ is the total number of segments per protein (1.4 amino acid residues per segment), and $\sigma = 2/3$ is the fractional area of each surface residue that is not in direct contact with solvent (67). The product $\Phi\Theta$ that appears in both terms accounts for the probability of

forming a contact between two hydrophobic residues in the mutually desolvated area between proteins. This approximation implicitly assumes that hydrophobic residues are uniformly distributed on the surface of each protein molecule.

Using analogous arguments, the magnitudes of the effective NN and DD contact energies directly follow:

$$\epsilon_{DD} = \frac{N_s}{8} \chi(T) f_c(R_D) (1 - \sigma) \Phi^2 k_B T, \quad (8)$$

$$\epsilon_{NN} = \frac{N_s}{8} \chi(T) f_c(R_N) (1 - \sigma) \Theta^2 k_B T. \quad (9)$$

The contact energies derived above are used in our model for the well depths of the effective protein-protein potentials given by Eq. 4.

The derivations provided in this Appendix yield, at best, average contact interactions, and for simplicity they have not addressed some potentially important structural and energetic features of real proteins. Clearly, the development of more accurate and comprehensive theoretical expressions for protein-protein interactions and their experimental verification are important areas for future scientific inquiry.

APPENDIX B: OCCURRENCE OF ORDERED NON-NATIVE AGGREGATES

We noted in Simulation Details that a small number of higher-concentration state points did not converge to an equilibrium fluid solution with the same fraction of native proteins f_N if they were initialized from all native ($f_N = 1$) versus all denatured ($f_N = 0$) FCC configurations. These state points corresponded to concentrations for which an FCC crystalline lattice of denatured proteins was very favorable energetically, with each particle taking full advantage of its neighbors' hydrophobic contacts. In the context of our simple model, one might view such a structure as ordered non-native aggregates (perhaps very roughly analogous to fibrillar structures). Although we were primarily interested in, and only report properties of, equilibrium

fluid protein solutions in Results, we show below that it is also relatively straightforward to predict the general region of the phase diagram where an ordered FCC arrangement of non-native proteins would be favorable in this model.

From Eq. 4, one can determine the ratio of the minimum-energy separation $\sigma_{\min, DD}(T)$ to the hard-core diameter $\sigma_{DD}(T)$ for denatured proteins $\sigma_{\min, DD}(T)/\sigma_{DD}(T) \approx 1.158$. However, the nearest-neighbor separation for an FCC crystal of denatured proteins is also given by

$$\rho \sigma_{\min, DD}^3(T) = 2^{\frac{1}{2}}, \quad (10)$$

where ρ is the number density. By combining these two ideas, we can determine the ideal energetic conditions for forming an ordered aggregate of denatured proteins from the relation

$$\rho \sigma_{NN}^3 = \frac{2^{\frac{1}{2}}}{1.158^3} \left[\frac{\sigma_{NN}}{\sigma_{DD}(T)} \right]^3. \quad (11)$$

Figs. 12 and 13 show that this relationship very closely approximates the conditions where ordered arrangements of denatured proteins actually occurred in our simulations.

We thank Prof. W. R. Smith and Prof. B. Triska for their help in the development of an efficient RCMC code. We are grateful to Prof. Venkat Ganesan for his useful comments on the model and Prof. Christopher Roberts for providing insights into various aspects of protein stability. We thank Sabrina Bromberg for providing Fig. 2. We also thank Dr. Allen Minton, Dr. Vincent Shen, and Prof. Jeffrey Errington for their comments on an early version of this manuscript, and their ideas about future studies using this model. Finally, we thank Prof. Ken Dill for fruitful discussions regarding the HPC theory and for several useful references.

This work was generously supported by The Merck Company Foundation and the David and Lucile Packard Foundation.

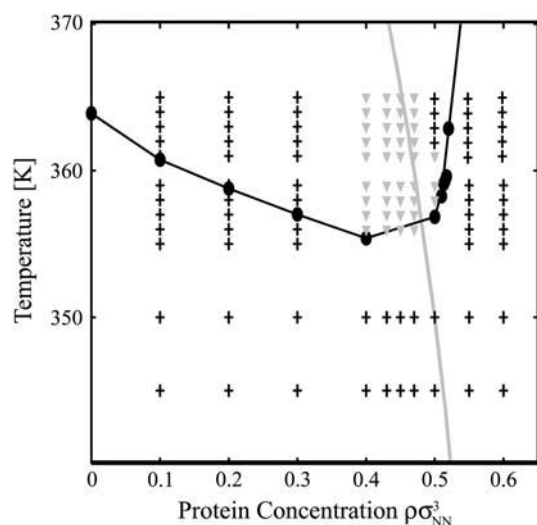


FIGURE 12 Complete set of RCMC simulation data in the temperature T -protein concentration $\rho\sigma_{NN}^3$ plane for the higher hydrophobicity protein ($N_r = 154$, $\Phi = 0.5$). Solid circles represent T_m , plus signs indicate equilibrium fluid protein solutions, shaded triangles indicate the occurrence of ordered non-native aggregates, and the shaded line represents the predicted ordered non-native aggregate curve (see Eq. 11). The solid curve is a guide to the eye.

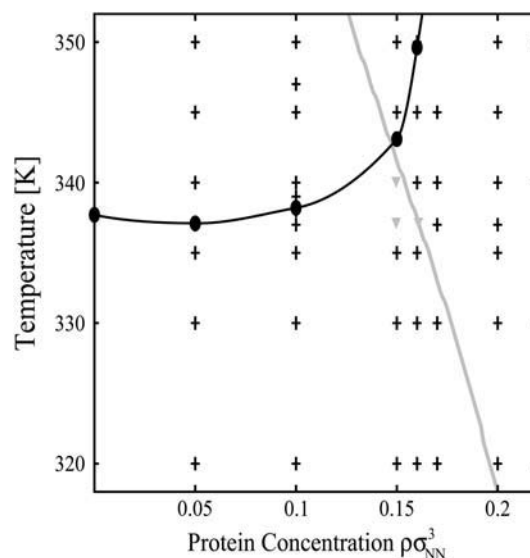


FIGURE 13 Complete set of simulation data and results in the temperature T -protein concentration $\rho\sigma_{NN}^3$ plane for the lower hydrophobicity protein ($N_r = 154$, $\Phi = 0.4$). Solid circles represent T_m , plus signs indicate equilibrium fluid protein solutions, shaded triangles indicate the occurrence of ordered non-native aggregates, and the shaded line represents the predicted ordered non-native aggregate curve (see Eq. 11). The solid curve is a guide to the eye.

REFERENCES

- Privalov, P. L. 1979. Stability of proteins: small globular proteins. *Adv. Protein Chem.* 33:167–241.
- Shortle, D. 1999. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* 10:27–34.
- Wetzel, R. 1994. Mutations and off-pathway aggregation of proteins. *Trends Biotechnol.* 12:193–198.
- Fink, A. L. 1998. Protein aggregation: folding aggregates, inclusion bodies, and amyloids. *Fold. Des.* 3:R9–R23.
- Aguzzi, A., and C. Haass. 2003. Games played by rogue proteins in prion disorders and Alzheimer's disease. *Science.* 302:814–818.
- Hsu, A., C. T. Murphy, and C. Kenyon. 2003. Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science.* 300:1142–1145.
- Kelly, J. W. 2002. Towards an understanding of amyloidogenesis. *Nat. Struct. Biol.* 9:323–325.
- Dobson, C. M. 1999. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* 24:329–332.
- Dobson, C. M. 2001. The structural basis of protein folding and its links with human disease. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356:133–145.
- Cleland, J. L., M. F. Powell, and S. J. Shire. 1993. The development of stable protein formulations: a close look at protein aggregation, deamidation, and oxidation. *Crit. Rev. Thera. Drug Carr. Sys.* 10:307–377.
- Kendrick, B. S., J. F. Carpenter, J. L. Cleland, and T. W. Randolph. 1998. A transient expansion of the native state precedes aggregation of recombinant human interferon- γ . *Proc. Natl. Acad. Sci. USA.* 95:14142–14146.
- Rey, L., and J. C. May, editors. 1999. Mechanisms of protein stabilization during freeze-drying and storage: the relative importance of thermodynamic stabilization and glassy state relaxation dynamics. In *Freeze-Drying/Lyophilization of Pharmaceutical and Biological Products*. Marcel Dekker, NY. 161–198.
- Bummer, P. M., and S. Koppenol. 2000. Chemical and physical considerations in protein and peptide stability. *Drugs Pharm. Sci.* 99:5–69.
- Chi, E. Y., S. Krishnan, T. Randolph, and J. Carpenter. 2003. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm. Res.* 20:1325–1336.
- Roberts, C. J. 2003. Kinetics of irreversible protein aggregation: analysis of extended Lumry-Eyring models and implications for predicting protein shelf life. *J. Phys. Chem. B.* 107:1194–1207.
- Stillinger, F. H. 1973. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J. Sol. Chem.* 2:141–158.
- Pratt, L. R., and D. Chandler. 1977. Theory of the hydrophobic effect. *J. Chem. Phys.* 67:3683–3704.
- Hummer, G., S. Garde, A. E. Garcia, A. Pohorille, and L. R. Pratt. 1996. An information theory model of hydrophobic interactions. *Proc. Natl. Acad. Sci. USA.* 93:8951–8955.
- Lum, K., D. Chandler, and J. D. Weeks. 1999. Hydrophobicity at small and large length scales. *J. Phys. Chem. B.* 103:4570–4577.
- Ashbaugh, H. S., T. M. Truskett, and P. G. Debenedetti. 2001. A simple molecular thermodynamic theory of hydrophobic hydration. *J. Chem. Phys.* 116:2907–2921.
- Pratt, L. R. 2002. Molecular theories of hydrophobic effects: “She is too mean to have her name repeated.” *Annu. Rev. Phys. Chem.* 53:409–436.
- Southall, N. T., K. A. Dill, and A. D. J. Haymet. 2002. A view of the hydrophobic effect. *J. Phys. Chem. B.* 106:521–533.
- Rajamani, S., T. M. Truskett, and S. Garde. 2005. Hydrophobic hydration from small to large lengthscales: understanding and manipulating the crossover. *Proc. Natl. Acad. Sci. USA.* 102:9475–9480.
- Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.
- Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry.* 29:7133–7155.
- Dill, K. A., S. Bromberg, K. Z. Yue, K. M. Fiebig, D. Yee, P. D. Thomas, and H. Chan. 1995. Principles of protein folding—a perspective from simple exact models. *Protein Sci.* 4:561–602.
- Duan, Y., and P. A. Kollman. 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science.* 282:740–744.
- Snow, C. D., H. Nguyen, V. S. Pande, and M. Gruebele. 2002. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature.* 420:102–106.
- Garcia, A. E., and J. N. Onuchic. 2003. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proc. Natl. Acad. Sci. USA.* 100:13898–13903.
- Herges, T., and W. Wenzel. 2004. An all-atom force field for tertiary structure prediction of helical proteins. *Biophys. J.* 87:3100–3109.
- Rosenbaum, D. F., and C. F. Zukoski. 1996. Protein interactions and crystallization. *J. Cryst. Growth.* 169:752–758.
- Neal, B. L., D. Asthagiri, and A. M. Lenhoff. 1998. Molecular origins of osmotic second virial coefficients of proteins. *Biophys. J.* 75:2469–2477.
- Leckband, D., and S. Sivasanker. 1999. Forces controlling protein interactions: theory and experiment. *Colloids Surf. B.* 14:83–97.
- ten Wolde, P. R., and D. Frenkel. 1997. Enhancement of protein crystal nucleation by critical density fluctuations. *Science.* 277:1975–1978.
- Foffi, G., G. D. McCullagh, A. Lawlor, E. Zaccarelli, K. A. Dawson, F. Sciortino, P. Tartaglia, D. Pini, and G. Stell. 2002. Phase equilibria and glass transition in colloidal systems with short-ranged attractive interactions: application to protein crystallization. *Phys. Rev. E.* 65:031407-1–031407-17.
- Dixit, N. M., and C. F. Zukoski. 2003. Competition between crystallization and gelation: a local description. *Phys. Rev. E.* 67:061501-1–061501-13.
- Hloucha, M., J. F. M. Lodge, A. M. Lenhoff, and S. I. Sandler. 2001. A patch-antipatch representation of specific protein interactions. *J. Cryst. Growth.* 232:195–203.
- Lomakin, A., N. Asherie, and G. B. M. Benedek. 1999. Aeolotopic interactions of globular proteins. *Proc. Natl. Acad. Sci. USA.* 96:9465–9468.
- Wu, J. Z., D. Bratko, H. W. Blanch, and J. M. Prausnitz. 2000. Interaction between oppositely charged micelles or globular proteins. *Phys. Rev. E.* 62:5273–5280.
- Curtis, R. A., C. Steinbrecher, A. Heinemann, H. W. Blanch, and J. M. Prausnitz. 2002. Hydrophobic forces between protein molecules in aqueous solutions of concentrated electrolyte. *Biophys. Chem.* 98:249–265.
- Minton, A. P. 1981. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers.* 20:2093–2120.
- Fields, G. B., D. Alonso, D. Stigter, and K. A. Dill. 1992. Theory for the aggregation of proteins and copolymers. *J. Phys. Chem.* 96:3974–3981.
- Zimmermann, S. B., and A. P. Minton. 1993. Macromolecular crowding: biochemical, biophysical and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* 22:27–65.
- Minton, A. P. 1998. Molecular crowding: analysis of effects of high concentrations of inert co-solutes on biochemical equilibria and rates in terms of volume exclusion. *Meth. Enzymol.* 295:127–149.
- Smith, A. V., and C. K. Hall. 2001. α -helix formation: discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins Struct. Funct. Genet.* 44:344–360.
- Smith, A. V., and C. K. Hall. 2001. Assembly of a tetrameric α -helical bundle: computer simulations on an intermediate-resolution protein model. *Proteins Struct. Funct. Genet.* 44:376–391.

47. Smith, A. V., and C. K. Hall. 2001. Protein refolding versus aggregation: computer simulations on an intermediate-resolution protein model. *J. Mol. Biol.* 312:187–202.
48. Dima, R. I., and D. Thirumalai. 2002. Exploring protein aggregation and self-propagation using lattice models: phase diagrams and kinetics. *Protein Sci.* 11:1036–1049.
49. Kinjo, A. R., and S. Takada. 2002. Effects of macromolecular crowding on protein folding and aggregation studied by density functional theory: dynamics. *Phys. Rev. E.* 66:051902–1–051902–10.
50. Kinjo, A. R., and S. Takada. 2002. Effects of macromolecular crowding on protein folding and aggregation studied by density functional theory: statics. *Phys. Rev. E.* 66:031911–1–031911–9.
51. Hall, D., and A. P. Minton. 2002. Effects of inert volume-excluding macromolecules on protein fiber formation. I. Equilibrium models. *Biophys. Chem.* 98:93–104.
52. Hall, D., and A. P. Minton. 2003. Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges. *Biochim. Biophys. Acta.* 1649:127–139.
53. Kinjo, A. R., and S. Takada. 2003. Competition between protein folding and aggregation with molecular chaperones in crowded solutions: insight from mesoscopic simulations. *Biophys. J.* 85:3521–3531.
54. Hall, D., and A. P. Minton. 2004. Effects of inert volume-excluding macromolecules on protein fiber formation. II. Kinetic models for nucleated fiber growth. *Biophys. Chem.* 107:299–316.
55. Jang, H., C. K. Hall, and Y. Zhou. 2004. Assembly and kinetic folding pathways of a tetrameric β -sheet complex: molecular dynamics simulations on simplified off-lattice protein models. *Biophys. J.* 86:31–49.
56. Jang, H., C. K. Hall, and Y. Zhou. 2004. Thermodynamics and stability of a β -sheet complex: molecular dynamics simulations on simplified off-lattice protein models. *Protein Sci.* 13:40–53.
57. Nguyen, H. D., and C. K. Hall. 2004. Phase diagrams describing fibrillization by polyalanine peptides. *Biophys. J.* 87:4122–4134.
58. Nguyen, H. D., and C. K. Hall. 2004. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. USA.* 101:16180–16185.
59. Braun, F. N. 2002. Adhesion and liquid-liquid phase separation in globular protein solutions. *J. Chem. Phys.* 116:6826–6830.
60. Sear, R. P. 2004. Solution stability and variability in a simple model of globular proteins. *J. Chem. Phys.* 120:998–1005.
61. Sochava, I. V., T. V. Belopolskaya, and O. I. Smirnova. 1985. DSC study of reversible and irreversible thermal denaturation of concentrated globular protein solutions. *Biophys. Chem.* 22:323–336.
62. Cueto, M., M. J. Dorta, O. Munguia, and M. Llabres. 2003. New approach to stability assessment of protein solution formulations by differential scanning calorimetry. *Int. J. Pharm.* 252:159–166.
63. Tomicki, P., R. L. Jackman, and D. W. Stanley. 1996. Thermal stability of metmyoglobin in a model system. *Lebensm. Wiss. U. Technol.* 29:547–551.
64. Shen, M., F. Davis, and A. Sali. 2005. The optimal size of a globular protein domain: a simple sphere-packing model. *Chem. Phys. Lett.* 405: 224–228.
65. Tanford, C. 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* 84:4240–4247.
66. Dill, K. A., D. O. V. Alonso, and K. Hutchinson. 1989. Thermal stability of globular proteins. *Biochemistry.* 28:5439–5449.
67. Dill, K. A. 1985. Theory for the folding and stability of globular proteins. *Biochemistry.* 24:1501–1509.
68. Dill, K. A., and D. Shortle. 1991. Denatured states of proteins. *Annu. Rev. Biochem.* 60:795–825.
69. Shortle, D., W. Stites, and A. Meeker. 1990. Contributions of large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry.* 29:8033–8041.
70. Shortle, D., H. S. Chan, and K. A. Dill. 1992. Modeling the effects of mutations on the denatured states of proteins. *Protein Sci.* 1:201–215.
71. Alonso, D. O. V., and K. A. Dill. 1991. Solvent denaturation and stabilization of globular proteins. *Biochemistry.* 20:5974–5985.
72. Dill, K. A., and D. Stigter. 1995. Modeling protein stability as heteropolymer collapse. *Adv. Protein Chem.* 46:59–104.
73. Petsev, D. N., X. Wu, O. Galkin, and P. G. Vekilov. 2003. Thermodynamic functions of concentrated protein solutions from phase equilibria. *J. Phys. Chem. B.* 107:3921–3926.
74. Shortle, D. 2002. The expanded denatured state: an ensemble of conformations trapped in a locally encoded topological space. *Adv. Protein Chem.* 62:1–23.
75. Millet, I. S., S. Doniach, and K. W. Plaxco. 2002. Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv. Protein Chem.* 62:241–262.
76. Vliegthart, G. A., and H. N. W. Lekkerkerker. 2000. Predicting the gas-liquid critical point from the second virial coefficient. *J. Chem. Phys.* 112:5364–5369.
77. Tessier, P., A. M. Lenhoff, and S. I. Sandler. 2002. Rapid measurement of protein osmotic second virial coefficients by self-interaction chromatography. *Biophys. J.* 82:1620–1631.
78. George, A., and W. W. Wilson. 1994. Predicting protein crystallization from dilute solution property. *Acta Crystallogr.* 50:361–365.
79. Ruppert, S., S. I. Sandler, and A. M. Lenhoff. 2001. Correlation between the osmotic second virial coefficient and the solubility of proteins. *Biotechnol. Prog.* 17:182–187.
80. Sear, R. P. 2002. Distribution of the second virial coefficients of globular proteins. *Europhys. Lett.* 60:938–944.
81. Johnson, J. K., A. Z. Panagiotopoulos, and K. E. Gubbins. 1994. Reactive canonical Monte Carlo. *Mol. Phys.* 81:717–733.
82. Johnson, J. K. 1999. Reactive canonical Monte Carlo. *Adv. Chem. Phys.* 105:461–481.
83. Smith, W. R., and B. Triska. 1993. The reaction ensemble method for the computer simulation of chemical and phase equilibria. I. Theory and basic examples. *J. Chem. Phys.* 4:3019–3027.
84. Truskett, T. M., S. Torquato, and P. G. Debenedetti. 2000. Towards a quantification of disorder in materials: distinguishing equilibrium and glassy sphere packings. *Phys. Rev. E.* 62:993–1001.
85. Torquato, S., T. M. Truskett, and P. G. Debenedetti. 2000. Is random close packing of sphere well defined? *Phys. Rev. Lett.* 84:2064–2067.
86. Minton, A. P. 2005. Models for excluded volume interaction between an unfolded protein and rigid macromolecular co-solutes: macromolecular crowding and protein stability revisited. *Biophys. J.* 88:971–985.
87. Young, L. R. D., K. A. Dill, and A. L. Fink. 1993. Aggregation and denaturation of apomyoglobin in aqueous urea solutions. *Biochemistry.* 32:3877–3886.
88. Young, L. R. D., A. L. Fink, and K. A. Dill. 1993. Aggregation of globular proteins. *Acc. Chem. Res.* 26:614–620.
89. Krishnamurthy, R., and M. C. Manning. 2002. The stability factor: importance in formulation development. *Curr. Pharm. Biotechnol.* 3:361–371.
90. Casares, S., M. Sadqi, O. Lopez-Mayorga, F. Conejero-Lara, and N. A. J. van Nuland. 2004. Detection and characterization of partially unfolded oligomers of the SH3 domain of α -spectrin. *Biophys. J.* 86: 2403–2413.
91. Eggers, D. K., and J. S. Valentine. 2001. Molecular confinement influences protein structure and enhances thermal protein stability. *Protein Sci.* 10:250–261.
92. Stigter, D., D. O. V. Alonso, and K. A. Dill. 1991. Protein stability: electrostatic and compact denatured states. *Proc. Natl. Acad. Sci. USA.* 88:4176–4180.
93. Alonso, D. O. V., K. A. Dill, and D. Stigter. 1991. The three states of globular proteins: acid denaturation. *Biopolymers.* 31:1631–1649.