THE JOURNAL OF PHYSICAL CHEMISTRY B
Article
pubs.acs.org/JPCB

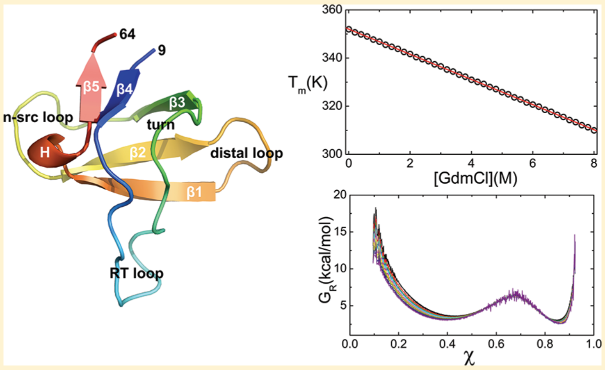
Theory of the Molecular Transfer Model for Proteins with Applications to the Folding of the src-SH3 Domain

Zhenxing Liu,[†] Govardhan Reddy,*[,‡] and D. Thirumalai[‡,§]

[†]Department of Physics, Beijing Normal University, Beijing 100875, China

[‡]Biophysics Program, Institute for Physical Science and Technology and [§]Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, United States

ABSTRACT: A theoretical basis for the molecular transfer model (MTM), which takes into account the effects of denaturants by combining experimental data and molecular models for proteins, is provided. We show that the MTM is a mean field-like model that implicitly takes into account denaturant-induced many body interactions. The MTM in conjunction with the coarse-grained self organized polymer model with side chains (SOP-SC) for polypeptide chains is used to simulate the folding of the src-SH3 domain as a function of temperature (T) and guanidine hydrochloride (GdmCl) concentration $[C]$. Besides reproducing the thermodynamic aspects of SH3 folding, the SOP-SC also captures the cooperativity of the folding transitions. A number of experimentally testable predictions are also made. First, we predict that the melting temperature $T_m([C])$ decreases linearly as $[C]$ increases. Second, we show that the midpoints $C_{m,i}$ and melting temperatures $T_{m,i}$ at which individual residues acquire 50% of their native contacts differ from the global midpoint ($C_m \approx 2.5$ M) and melting temperature ($T_m = 355$ K) at which the folded and unfolded states coexist. Dispersion in $C_{m,i}$ is greater than that found for $T_{m,i}$. Third, folding kinetics at $[C] = 0$ M shows that the acquisition of contacts between all the secondary structural elements and global folding occur nearly simultaneously. Finally, from the free energy profiles as a function of the structural overlap function and the radius of gyration of the protein, we find that at a fixed T the transition state moves toward the folded state as $[C]$ increases in accord with the Hammond postulate. In contrast, we predict that along the locus of points $T_m([C])$ the location of the transition state does not change. The theory and the models used here are sufficiently general for studying the folding of other single domain proteins.



INTRODUCTION

In the transition from an ensemble of unfolded conformations to the folded native state, proteins reach their native basins of attraction (NBAs) by multiple pathways. These predictions, which were made on the basis of theory and simulations of precisely soluble models,[1−9] have been validated by experiments, especially those done at the single molecule level.[10−16] Thus, instead of the restricted view that folding occurs in discrete steps, the intrinsic heterogeneity of the folding pathways demands that it is best to describe the unfolded state, the transition state, and even the folded states as ensembles. Each state should be described in terms of distributions of features that best capture the structural characteristics of the protein. The simplest two-state folding reaction is often interpreted as conversion of unfolded states to a low free energy native state without populating any discernible intermediate states as the denaturant concentration ($[C]$) is decreased. From a statistical mechanics perspective, such a folding process should be viewed as changes in the distribution of quantities that describe the protein of interest. Distribution (preferably joint distribution functions obtained using multiple probes) functions of various quantities (for example, radius of gyration (R_g), secondary structure content, and extent of tertiary contact formation) are broad in the unfolded basin of attraction (UBA) and narrow in the folded native basins of attraction (NBA). Thus, a more general description of the two-state folding reaction is

$$\{\mathrm{UBA}\} \leftrightarrows \{\mathrm{TSE}\} \leftrightarrows \{\mathrm{NBA}\} \qquad (1)$$

where the curly brackets indicate the ensemble of conformations, and TSE is the transition state ensemble with a higher free energy than either UBA or NBA. The more general description based on the statistical mechanical viewpoint and polymer concepts have led to predictions for the dependence of folding rates and stability on the size of proteins,[17−19] links between co-operativity, stability gap, and folding kinetics,[20−22] and collapse and folding.[23] In addition, scenarios such as downhill folding, which can also be explained as a special case of the nucleation collapse mechanism,[24] were anticipated on

Special Issue: B: Harold A. Scheraga Festschrift

Received: December 12, 2011
Revised: April 11, 2012
Published: April 12, 2012

the basis of the energy landscape perspective.[2] Many of these predictions have found broad experimental support.[10,25]

In order to quantitatively describe the folding reaction (eq 1), one has to characterize in detail the statistical properties of the UBA, NBA, and TSE. In principle, conformations that are sampled during the folding reaction as a function of [C] can be obtained using all-atom molecular dynamics simulations (MD). Although MD simulations have provided useful insights into the folding mechanisms[26−30] of proteins, difficulties in effectively sampling the conformational space of proteins and uncertainties in the force fields for all atom MD simulations have made it difficult to obtain thermodynamic properties that compare favorably with experiments. The notable exception is the study by Garcia and co-workers[31] on the small protein Trp-cage. In contrast, CG models have provided considerable insight into the mechanisms of protein folding.[9,32−34] A clear drawback is that many of the CG models are constructed on the basis of the folded structure with emphasis only on the stabilizing interactions between contacts present in the native state. In addition, simulations using the popular Go-like[34,35] and self-organized polymer (SOP)[36,37] models are carried out using temperature to initiate folding or unfolding. In contrast, in a large number of experiments, changes in [C] are used to trigger folding and unfolding.

In order to solve some of the problems alluded to above, we developed the molecular transfer model (MTM)[38,39] for which simulations in the presence of osmolytes and denaturants can be carried out so that direct comparisons with ensemble and single molecule experiments can be made. The MTM combines simulations using a description of polypeptide chains at any level (all-atom detail or CG) and the effect of denaturants using experimentally measured transfer free energies for protein backbone and side chains to describe the folding reaction. Here, we describe the theory underlying the MTM, which not only reveals the approximations but also exposes its limitations. As an application, we use the MTM to fully characterize the folding of the src-SH3 domain, which has been extensively characterized using experiments as well as computations.[40−45] The present work has led to a number of experimentally testable predictions.

■ THEORY, MODELS, AND SIMULATIONS

MTM Theory. Consider a ternary system consisting of monomer protein (p), solvent (s), and denaturant (d). The energy for the system can be written as

$$\begin{aligned} E = {} & E_{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\}) + E_{\mathrm{s}}(\{\mathbf{r}_j^{\mathrm{s}}\}) + E_{\mathrm{d}}(\{\mathbf{r}_k^{\mathrm{d}}\}) + E_{\mathrm{ps}}(\{\mathbf{r}_i^{\mathrm{p}}\}, \{\mathbf{r}_j^{\mathrm{s}}\}) \\ & + E_{\mathrm{pd}}(\{\mathbf{r}_i^{\mathrm{p}}\}, \{\mathbf{r}_k^{\mathrm{d}}\}) + E_{\mathrm{ds}}(\{\mathbf{r}_k^{\mathrm{d}}\}, \{\mathbf{r}_j^{\mathrm{s}}\}) \end{aligned} \tag{2}$$

where $\{\mathbf{r}_i^{\mathrm{p}}\}$, $\{\mathbf{r}_j^{\mathrm{s}}\}$, and $\{\mathbf{r}_k^{\mathrm{d}}\}$ are the coordinates of the protein atoms, solvent, and denaturant, respectively, $E_{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})$ corresponds to interactions between the protein atoms, $E_{\mathrm{s}}(\{\mathbf{r}_j^{\mathrm{s}}\})$ is the energy of the solvent (water), and $E_{\mathrm{d}}(\{\mathbf{r}_k^{\mathrm{d}}\})$ represents the interactions involving the denaturant molecules. Interactions between protein and solvent are represented by $E_{\mathrm{ps}}(\{\mathbf{r}_i^{p}\},\{\mathbf{r}_j^{\mathrm{s}}\})$, $E_{\mathrm{pd}}(\{\mathbf{r}_i^{\mathrm{p}}\}, \{\mathbf{r}_k^{\mathrm{d}}\})$ denotes the interactions between protein and denaturant, and $E_{\mathrm{ds}}(\{\mathbf{r}_k^{\mathrm{d}}\}, \{\mathbf{r}_j^{\mathrm{s}}\})$ represents the interactions between the denaturant and solvent. The partition function for the system is

$$Z = \iiint \mathrm{d}\{\mathbf{r}_i^{\mathrm{p}}\}\, \mathrm{d}\{\mathbf{r}_j^{\mathrm{s}}\}\, \mathrm{d}\{\mathbf{r}_k^{\mathrm{d}}\}\, \mathrm{e}^{-\beta E(\{\mathbf{r}_i^{\mathrm{p}}\},\{\mathbf{r}_j^{\mathrm{s}}\},\{\mathbf{r}_k^{\mathrm{d}}\})} \tag{3}$$

Here, $\beta = 1/k_{\mathrm{B}}T$, where k_{B} is the Boltzmann constant and T is the temperature. If we formally integrate over the coordinates of the denaturant molecules, we obtain

$$Z \equiv \iint \mathrm{d}\{\mathbf{r}_i^{\mathrm{p}}\}\, \mathrm{d}\{\mathbf{r}_j^{\mathrm{s}}\}\, \mathrm{e}^{-\beta[E_{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})+E_{\mathrm{s}}(\{\mathbf{r}_j^{\mathrm{s}}\})+E_{\mathrm{ps}}(\{\mathbf{r}_i^{\mathrm{p}}\},\{\mathbf{r}_j^{\mathrm{s}}\})+\Delta G_{\mathrm{d}}(\{\mathbf{r}_i^{\mathrm{p}}\},\{\mathbf{r}_j^{\mathrm{s}}\})]} \tag{4}$$

where

$$\begin{aligned} & \mathrm{e}^{-\beta[\Delta G_{\mathrm{d}}(\{\mathbf{r}_i^{\mathrm{p}}\}),\{\mathbf{r}_j^{\mathrm{s}}\})]} \\ & \quad = \int \mathrm{d}\{\mathbf{r}_k^{\mathrm{d}}\}\, \mathrm{e}^{-\beta[E_{\mathrm{d}}(\{\mathbf{r}_k^{\mathrm{d}}\})+E_{\mathrm{pd}}(\{\mathbf{r}_i^{\mathrm{p}}\},\{\mathbf{r}_k^{\mathrm{d}}\})+ E_{\mathrm{ds}}(\{\mathbf{r}_k^{\mathrm{d}}\},\{\mathbf{r}_j^{\mathrm{s}}\})]} \end{aligned}$$

By writing $\Delta G_{\mathrm{d}}(\{\mathbf{r}_i^{\mathrm{p}}\}, \{\mathbf{r}_j^{\mathrm{s}}\}) = \Delta G_{\mathrm{d}}^{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\}) + \Delta G_{\mathrm{d}}^{\mathrm{s}}(\{\mathbf{r}_j^{\mathrm{s}}\}) + \Delta G_{\mathrm{d}}^{\mathrm{ps}}(\{\mathbf{r}_i^{\mathrm{p}}\}, \{\mathbf{r}_j^{\mathrm{s}}\})$, eq 4 becomes

$$\begin{aligned} Z = {} & \int \mathrm{d}\{\mathbf{r}_i^{\mathrm{p}}\}\, \mathrm{e}^{-\beta[E_{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})+\Delta G_{\mathrm{d}}^{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})]} \\ & \times \int \mathrm{d}\{\mathbf{r}_j^{\mathrm{s}}\}\, \mathrm{e}^{-\beta[E_{\mathrm{s}}(\{\mathbf{r}_j^{\mathrm{s}}\})+E_{\mathrm{ps}}(\{\mathbf{r}_i^{\mathrm{p}}\},\{\mathbf{r}_j^{\mathrm{s}}\})+\Delta G_{\mathrm{d}}^{\mathrm{s}}(\{\mathbf{r}_j^{\mathrm{s}}\})+\Delta G_{\mathrm{d}}^{\mathrm{ps}}(\{\mathbf{r}_i^{\mathrm{p}}\},\{\mathbf{r}_j^{\mathrm{s}}\})]} \\ \equiv {} & \int \mathrm{d}\{\mathbf{r}_i^{\mathrm{p}}\}\, \mathrm{e}^{-\beta[E_{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})+\Delta G_{\mathrm{d}}^{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})+\Delta G_{\mathrm{s}}(\{\mathbf{r}_i^{\mathrm{p}}\})]} \end{aligned} \tag{5}$$

where

$$\begin{aligned} & \mathrm{e}^{-\beta[\Delta G_{\mathrm{s}}(\{\mathbf{r}_i^{\mathrm{p}}\})]} \\ & \quad = \int \mathrm{d}\{\mathbf{r}_j^{\mathrm{s}}\}\, \mathrm{e}^{-\beta[E_{\mathrm{s}}(\{\mathbf{r}_j^{\mathrm{s}}\})+E_{\mathrm{ps}}(\{\mathbf{r}_i^{\mathrm{p}}\},\{\mathbf{r}_j^{\mathrm{s}}\})+\Delta G_{\mathrm{d}}^{\mathrm{s}}(\{\mathbf{r}_j^{\mathrm{s}}\})+\Delta G_{\mathrm{d}}^{\mathrm{ps}}(\{\mathbf{r}_i^{\mathrm{p}}\},\{\mathbf{r}_j^{\mathrm{s}}\})]} \end{aligned}$$

The manipulations leading to eq 5 are formally exact. To make progress, one has to construct the energy functions such as $E_{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})$ and other terms in eq 5. In order to devise practical strategies, we use the MTM, which can be formally obtained from eq 5 using approximations, to solve the problem of including osmolytes (denaturants or stabilizing agents) in a natural manner. In order to render the formal expressions in eq 5 practical, we make two crucial approximations: (1) We assume that $E_{\mathrm{CG}}(\{\mathbf{r}_i^{\mathrm{p}}\}) \equiv E_{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\}) + \Delta G_{\mathrm{s}}(\{\mathbf{r}_i^{\mathrm{p}}\})$, the solvent averaged interaction between the various protein interaction centers, can be represented using a suitable CG model. (2) The effect of denaturants, $\Delta G_{\mathrm{d}}^{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})$, is included using the molecular transfer model (MTM). According to MTM, the free energy cost of transferring a polypeptide chain with conformation $\{\mathbf{r}_i^{\mathrm{p}}\}$ (includes all interactions of the polypeptide chain) to an aqueous denaturant solution at concentration [C] is approximated as

$$\begin{aligned} \Delta G_{\mathrm{d}}^{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\}) & = \sum_l \delta g_{\mathrm{tr},l}(\{\mathbf{r}_i^{\mathrm{p}}\}, [C]) \\ & = \sum_l \delta g_{\mathrm{tr},l}^{\mathrm{exp}}([C]) \frac{\alpha_l(\{\mathbf{r}_i^{\mathrm{p}}\})}{\alpha_{l,\mathrm{G}\text{-}l\text{-}\mathrm{G}}} \end{aligned} \tag{6}$$

where the summation is over all the amino acid side chains and the backbone peptide groups, $\delta g_{\mathrm{tr},l}^{\mathrm{exp}}([C])$ is the experimentally measured transfer free energy of side chain or backbone l, α_l is the solvent exposed area of l, and $\alpha_{l,\mathrm{G}\text{-}l\text{-}\mathrm{G}}$ is the solvent exposed area of l in the tripeptide Gly-l-Gly. Although the MTM expression for $\Delta G_{\mathrm{d}}^{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})$ appears to be a sum of the single particle terms, the dependence of $\delta g_{\mathrm{tr},l}^{\mathrm{exp}}([C])$ and $\alpha_l(\{\mathbf{r}_i^{\mathrm{p}}\})$ on all the coordinates of the polypeptide chains shows that many body interactions are implicitly taken into account. In addition, the contributions to $\Delta G_{\mathrm{d}}^{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})$ even for two identical residues (Ala, for example) in a polypeptide chain will depend on the extent of solvent exposure of the two residues, which in turn would be determined by the residues that are in the neighborhood of the two residues. By way of contrast, it is worth noting that, in the standard transfer model used to calculate m-values[46] for proteins, $\delta g_{\mathrm{tr},l}(\{\mathbf{r}_i^{\mathrm{p}}\}, [C])$ is taken to be independent of $\{\mathbf{r}_i^{\mathrm{p}}\}$. The values of the various parameters such as van der Waal's radii of the protein bead required to estimate $\alpha_l(\{\mathbf{r}_i^{\mathrm{p}}\})$, $\alpha_{l,\mathrm{G}\text{-}l\text{-}\mathrm{G}}$ values for all the amino acid residues, and m_k and b_k values to calculate $\delta g_{\mathrm{tr},l}^{\mathrm{exp}}([C])$ $(=m_k[C] + b_k)$ are given in the Supporting Information of ref 47.

Coarse-Grained Model for $E_{\mathrm{CG}}(\{\mathbf{r}_i^{\mathrm{p}}\})$. In order to sample the conformational space thoroughly and to ensure that the protein can make multiple transitions between folded and unfolded states, we use the SOP-SC (self-organized polymer-side chain) model. Each amino acid residue is represented using two interaction centers, one at the C_α atom and the other centered at the center of mass of the side chain.[47] The stability of the protein in the SOP-SC model is guaranteed by taking attractive interactions between side chains (SCs) and backbone atoms that are in contact in the native structure. Although neglect of non-native interactions can have important consequences (see below), the folding mechanism and potential thermodynamics are not greatly compromised, as shown in recent applications of MTM.

Previous studies showed that m-values (defined using $\Delta G_{\mathrm{NU}}[C] = \Delta G_{\mathrm{NU}}[0] + m[C]$, with $\Delta G_{\mathrm{NU}}[C]$ $(=G_{\mathrm{N}}[C] - G_{\mathrm{U}}[C])$ being the free energy of stability of the native (N) state with respect to the unfolded state (U) are accurately reproduced, using the strict assumption of complete lack of correlation between interaction centers.[46] Because of the connectivity of the backbone C_α atoms and the covalent linkage between the SC and the C_α atoms, the α_l (eq 6) for the lth group depends on $(\{\mathbf{r}_i^{\mathrm{p}}\})$. As a result, $\Delta G_{\mathrm{d}}^{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})$ captures approximately correlations between various SC and the backbone C_α atoms even though eq 6 contains only single particle terms.

With these two approximations, the partition function of MTM (eq 5) is written as

$$Z_{\mathrm{MTM}} = \int \mathrm{d}\{\mathbf{r}_i^{\mathrm{p}}\}\; \mathrm{e}^{-\beta[E_{\mathrm{SOP-SC}}(\{\mathbf{r}_i^{\mathrm{p}}\}) + \sum_l \delta g_{\mathrm{tr},l}(\{\mathbf{r}_i^{\mathrm{p}}\},[C])]} \tag{7}$$

In the first applications[38,48] of MTM used to predict equilibrium properties of protein L and cold shock proteins, we further rewrote eq 7 as

$$Z_{\mathrm{MTM}} = Z_{\mathrm{SOP\text{-}SC}} \langle \mathrm{e}^{-\beta \Delta G_{\mathrm{d}}^{\mathrm{p}}(\{\mathbf{r}_i^{\mathrm{p}}\})} \rangle_{\mathrm{SOP\text{-}SC}} \tag{8}$$

For computational expediency, we used low friction Langevin simulations (see below) to get thermodynamic properties of the protein in various denaturant concentrations, $[C]$, using the conformations of the protein obtained at $[C] = 0$. We further used the weighted histogram technique (WHAM)[49,50] to combine the simulation data obtained at $[C] = 0$ to calculate protein thermodynamic properties at $[C] \neq 0$ condition. The WHAM[38] equation used to calculate the thermodynamic property $A([C], T)$ is

$$\langle A([C], T) \rangle = \frac{1}{Z([C], T)} \sum_{k=1}^{R} \sum_{t=1}^{n_k} \frac{A_{k,t}([0]) \mathrm{e}^{-\beta(E_{\mathrm{SOP\text{-}SC}}^{k,t}([0],\{\mathbf{r}_i^{\mathrm{p}}\}) + \Delta G_{\mathrm{d}}^{k,t}([C],\{\mathbf{r}_i^{\mathrm{p}}\}))}}{\sum_{m=1}^{R} n_m \mathrm{e}^{f_m - \beta_m E_{\mathrm{SOP\text{-}SC}}^{k,t}([0],\{r_i^{\mathrm{p}}\})}} \tag{9}$$

where $Z([C], T)$ is the partition function, R is the number of independent simulations, n_k is the number of conformations from the kth simulation, and n_m and f_m correspond to the number of conformations and free energy in the mth simulation, respectively. $A_{k,t}([0])$ is the value of property A, $E_{\mathrm{SOP\text{-}SC}}^{k,t}([0], \{\mathbf{r}_i^{\mathrm{p}}\})$ is the potential energy of the SOP-SC model (eq 10) of the protein, and $G_{\mathrm{d}}^{k,t}([C], \{\mathbf{r}_i^{\mathrm{p}}\})$ is the MTM free energy of transferring (eq 6) protein conformation from $[C] = 0$ solution to a solution with denaturant concentration $[C]$. The subscripts or superscripts k and t refer to the tth protein conformation of kth simulation. The use of conformations generated at $[C] = 0$ to estimate eq 8, although not necessary, was made solely for computational expediency. We show explicitly here that eq 8 provides a good estimate of Z_{MTM}. Indeed, eq 8 is exact, provided the conformational space can be extensively sampled at finite T, and merely states that if the entire space of conformations of a system is known then the partition at any arbitrary external condition can be computed.

SOP Side Chain Model for Polypeptide Chains. The SOP-SC model of the 56 residue src-SH3 protein studied here is constructed using the crystal structure (Protein Data Bank ID: 1SRL). The effective energy $E_{\mathrm{SOP\text{-}SC}}(\{\mathbf{r}_i^{\mathrm{p}}\})$ $(i = 1, 2, ..., 112)$ of the protein is a sum of bonded (B) and nonbonded (NB) terms, which are a sum of native (N) and non-native (NN) interactions. Interaction between two sites separated by at least two other sites is native, if the distance between them is less than a cutoff distance, R_c, in the SOP-SC contact map of the crystal structure. The functional form of $E_{\mathrm{SOP\text{-}SC}}$ is

$$\begin{aligned} E_{\mathrm{SOP\text{-}SC}}(\{r_i^{\mathrm{p}}\}) &= E_{\mathrm{B}} + E_{\mathrm{NB}}^{\mathrm{N}} + E_{\mathrm{NB}}^{\mathrm{NN}} \\ &= -\sum_{i=1}^{N_{\mathrm{B}}} \frac{k}{2} R_{\mathrm{o}}^{\;2} \log\left(1 - \frac{(r_i - r_{\mathrm{cry},i})^2}{R_{\mathrm{o}}^{\;2}}\right) \\ &+ \sum_{i=1}^{N_{\mathrm{N}}^{\mathrm{bb}}} \varepsilon_{\mathrm{h}}^{\mathrm{bb}} \left[\left(\frac{r_{\mathrm{cry},i}}{r_i}\right)^{12} - 2\left(\frac{r_{\mathrm{cry},i}}{r_i}\right)^{6}\right] \\ &+ \sum_{i=1}^{N_{\mathrm{N}}^{\mathrm{bs}}} \varepsilon_{\mathrm{h}}^{\mathrm{bs}} \left[\left(\frac{r_{\mathrm{cry},i}}{r_i}\right)^{12} - 2\left(\frac{r_{\mathrm{cry},i}}{r_i}\right)^{6}\right] \\ &+ \sum_{i=1}^{N_{\mathrm{N}}^{\mathrm{ss}}} \lambda|\varepsilon_i^{\mathrm{ss}} - 0.7| \left[\left(\frac{r_{\mathrm{cry},i}}{r_i}\right)^{12} - 2\left(\frac{r_{\mathrm{cry},i}}{r_i}\right)^{6}\right] \\ &+ \sum_{i=1}^{N_{\mathrm{NN}}} \varepsilon_l \left(\frac{\sigma_i}{r_i}\right)^{6} + \sum_{i=1}^{3N-4} \varepsilon_l \left(\frac{\sigma_{i,i+2}}{r_{i,i+2}}\right)^{6} \end{aligned} \tag{10}$$

In eq 10, N_{B} $(=2N - 1)$ is the number of bonds in the coarse grained model of protein and N_{NN} is the number of non-native interactions; $N_{\mathrm{N}}^{\mathrm{bb}}$, $N_{\mathrm{N}}^{\mathrm{bs}}$, and $N_{\mathrm{N}}^{\mathrm{ss}}$ are the number of backbone–backbone, backbone–side chain, and side chain–side chain native interactions, respectively. The distance between the ith pair of residues that are either bonded or interact natively or non-natively is r_i, and $r_{\mathrm{cry},i}$ is the corresponding distance in the crystal structure. The sum of the radii of the ith pair of residues is σ_i, $\sigma_{i,i+2}$ is the sum of the radii of the interaction sites i and $i + 2$, and $r_{i,i+2}$ is the distance between the sites i and $i + 2$. The radii of the side chain, $\sigma^{\mathrm{ss}}/2$, is defined as the distance between the position of the backbone bead and the side chain bead in the PDB structure. The strength of interaction between a pair of SC beads i, $\varepsilon_i^{\mathrm{ss}}$, is taken from the Betancourt–Thirumalai statistical interaction potential,[51] and scaled by $\lambda = 0.3$. The other interaction parameters in the energy function are given in Table 1. The effective energy function for the protein at $[C] \neq 0$ is the sum of $E_{\mathrm{SOP\text{-}SC}}$, where $E_{\mathrm{SOP\text{-}SC}}$ is given by eq 10, and the

Table 1. Parameters for the SOP Side Chain Model Described in eq 10

parameter	protein
R_o	2.0 Å
k	20 kcal/(mol Å^2)
R_c	8 Å
ε_h^{bb}	0.55 kcal/mol
ε_h^{bs}	0.40 kcal/mol
ε_l	1.0 kcal/mol
σ^{bb}	3.8 Å

$\{\mathbf{r}_i^p\}$-dependent transfer free energy $\Delta G_d^p(\{\mathbf{r}_i^p\})$ is calculated using eq 6.

$$E_{\mathrm{MTM}} = E_{\mathrm{SOP\text{-}SC}} + \Delta G_d^p(\{\mathbf{r}_i^p\}) \tag{11}$$

Simulations. Extensive thermodynamic sampling of the protein is achieved using low friction Langevin dynamics simulations and the SOP-SC model of the src-SH3. The equations of motion for the position of a protein bead, r_i, are integrated using the equation

$$m\ddot{r}_i = -\zeta\dot{r}_i + F_c + \Gamma \tag{12}$$

Here, m is the mass of the protein bead, $F_c = -\partial E_{\mathrm{MTM}}/\partial r_i$, and $\Gamma(t)$ is random force with a white noise spectrum satisfying the autocorrelation function in the discretized form $\langle\Gamma(t)\Gamma(t + nh)\rangle = (2\zeta k_B T/h)\delta_{0,n}$,[52] where $\delta_{0,n}$ is the Kronecker delta function and $n = 0, 1, 2, ...$

We used the Verlet leapfrog algorithm to integrate the equation of motion. The velocity, v_i, at time $t + h/2$ and the position, r_i, at time $t + h$ of a bead are given by

$$v_i(t + h/2) = \frac{2m - h\zeta}{2m + h\zeta}\cdot v_i(t - h/2) + \frac{2h}{2m + h\zeta}[F_c(t) + \Gamma(t)] \tag{13}$$

$$r_i(t + h) = r_i(t) + h\cdot v_i(t + h/2) \tag{14}$$

The value of the time step used to integrate eqs 13 and 14 is $h = 0.005\tau_L$, and a low friction, $\zeta = 0.05m/\tau_L$ (τ_L is the unit of time) is used to obtain enhanced sampling and converged equilibrium thermodynamic data.

Brownian dynamics simulations are used to simulate the folding kinetics of the protein. The simulation algorithm is a straightforward implementation of the Ermak and McCammon algorithm[53] without hydrodynamic interactions. The equation of motion involving the positions of the interaction centers in the protein, r_i, are integrated using

$$r_i(t + h) = r_i(t) + \frac{h}{\zeta_H}(F_c(t) + \Gamma(t)) \tag{15}$$

We simulated the kinetics of folding by generating a large number of trajectories with $\zeta_H = 50m/\tau_L$, which represents the overdamped limit and corresponds to the friction in water.[52] The time step used to integrate eq 15 is $0.01\tau_H$. The natural unit of time for overdamped condition at the simulation temperature T_s is $\tau_H \approx \zeta_H a^2/k_B T_s = (((\zeta_H\tau_L/m)e_l)/k_B T_s)\tau_L$. To convert simulation time to real time, we chose e_l = 1 kcal/mol, average mass $m = 1.8 \times 10^{-22}$ g,[52] and a = 4 Å, which makes τ_L = 2 ps. For $\zeta_H = 50m/\tau_L$, we obtain τ_H = 148 ps.

Data Analysis. We defined a structural overlap function[54]

$$\chi = 1 - \frac{N_k}{N_T} \tag{16}$$

as an order parameter to monitor the folding reaction; N_k is defined as

$$N_k = \sum_{i=1}^{N-3}\sum_{j=i+3}^{N}\Theta(\delta - |r_{i,j\mathrm{bb}} - r_{i,j\mathrm{bb}}^{o}|) + \sum_{i=1}^{N-3}\sum_{j=i+3}^{N}\Theta(\delta - |r_{i,j\mathrm{ss}} - r_{i,j\mathrm{ss}}^{o}|) + \sum_{\substack{i=1,j=1\\|i-j|\geq 3}}^{N}\Theta(\delta - |r_{i,j\mathrm{bs}} - r_{i,j\mathrm{bs}}^{o}|) \tag{17}$$

In eq 17, $r_{i,j}$ is the distance between interaction centers i and j, $r_{i,j}^o$ is the corresponding value in the native conformation, $\Theta(x)$ is the Heavyside function, and δ = 2 Å. N_k is the number of interacting pairs that are within δ = 2 Å in the kth conformation and N_T (=5724) for src-SH3 in the folded state, and it is obtained by setting all the $\Theta(x)$ in eq 17 to unity.

To determine whether a protein conformation belongs to NBA or UBA, we defined χ_c such that conformations with $\chi \leq \chi_c$ belong to the NBA. In order to determine χ_c, we calculated the distribution $P(\chi)$ at the melting temperature $T_m([C])$. At $T_m([C] = 0)$, $P(\chi)$ as a function of χ shows a bimodal distribution due to the frequent transitions of the protein from low χ (NBA) to high χ (UBA). From the observed bimodal distribution, we surmised that $\chi_c = 0.65$ (see the Supporting Information in ref 47 for additional details).

RESULTS

Melting Temperature Decreases Linearly as [*C*] Increases. In order to test the applicability of using eq 8 in estimating the exact partition function, we calculated the specific heat as a function of temperature, T, at different denaturant concentrations, $[C]$ (Figure 1). The results obtained from the exact partition function (eq 7) and the approximate partition function (eq 8) are displayed in Figure 1. At the two

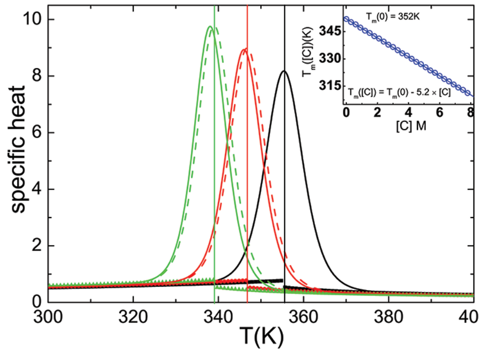


Figure 1. Specific heat as a function of temperature for various GdmCl concentrations, $[C]$. The data in black, red, and green are for $[C]$ = 0, 1.0, and 2.5 M, respectively. The data in solid and dashed lines are obtained from the exact partition function (eq 7) and the approximate partition function (eq 8), respectively. The good agreement between the solid and dashed lines shows that thermodynamic properties at nonzero $[C]$ can be obtained by extensive conformational sampling at $[C]$ = 0. The inset shows the linear decrease in the melting temperature of the protein as $[C]$ increases. The equation of the line is also displayed.

values of $[C]$, the entire shapes of C_v are in quantitative agreement, which implies that as long as the conformations are extensively sampled at a given temperature with $[C] = 0$ then the partition function at nonzero $[C]$ can be accurately estimated. The excellent comparison in Figure 1 justifies using Z_{MTM} to evaluate thermodynamic properties at finite $[C]$. The melting temperature of the SH3 domain, T_m, which is identified with the peak in C_v, decreases in a linear fashion (inset of Figure 1) as $[C]$ increases. The predicted linearity has previously been observed in the melting of S6.[55]

Since the energy fluctuations in the unfolded states are greater than the folded state, we expect that the heat capacity at high T corresponding to the unfolded state must exceed the low temperature value.[56−58] However, the opposite trend is seen in our model and other coarse-grained models.[59] The most likely reason is that water-mediated interactions, which are known to contribute to this difference in the specific heat between folded and unfolded states,[56] are not captured in CG models. Despite this limitation, the extent of cooperativity is not severely compromised (see below).

To assess if thermal melting is a two-state process, we calculated the ratio $\lambda = \Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}}$, where ΔH_{vH} is the van't Hoff enthalpy and ΔH_{cal} is the calorimetric enthalpy. If $\lambda \sim 1$, then folding is co-operative, although in a strict sense it merely shows that folding behaves thermodynamically as a two-state system. We calculated van't Hoff enthalpy ΔH_{vH} using $\Delta H_{\mathrm{vH}} = 4k_BT_m^{\,2}|\mathrm{d}f_{\mathrm{NBA}}/\mathrm{d}T|_{T_m}$, where f_{NBA} is the fraction of molecules in the NBA (Figure 2a). It is well-known that, in order to calculate the calorimetric enthalpy, ΔH_{cal}, accurate care must be taken to ensure that possible drifts in the baselines of the C_v are accounted for.[59,58] Since the Jackson−Brandts baseline is theoretically exact for a two-state process, we calculated the calorimetric enthalpy using

$$\Delta H_{\mathrm{cal}} = \int_{T_{\mathrm{I}}}^{T_{\mathrm{II}}} C_v \,\mathrm{d}T - \int_{T_{\mathrm{I}}}^{T_m} C_v^{\mathrm{NBA}} \,\mathrm{d}T - \int_{T_m}^{T_{\mathrm{II}}} C_v^{\mathrm{DSE}} \,\mathrm{d}T \tag{18}$$

where T_{I} is the lowest temperature in the heat capacity, at which the system is assumed to be entirely in NBA. Similarly, T_{II} is the highest temperature, at which the system is assumed to belong entirely to the UBA. The melting temperature, T_m, is given by the peak in (solid lines in Figure 1)

$$C_v = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_BT^2} \tag{19}$$

C_v^{NBA} is given by the specific heat at low temperatures. C_v^{NBA} is taken to be

$$C_v^{\mathrm{NBA}} = \frac{\langle E_{\mathrm{NBA}}^{\;2} \rangle - \langle E_{\mathrm{NBA}} \rangle^2}{k_BT^2} \tag{20}$$

and C_v^{DSE} is

$$C_v^{\mathrm{DSE}} = \frac{\langle E_{\mathrm{DSE}}^{\;2} \rangle - \langle E_{\mathrm{DSE}} \rangle^2}{k_BT^2} \tag{21}$$

The ratio λ for $[C] = 0.0$, 1.0, and 2.5 M is 1.00, 0.95, and 0.91, respectively, which shows that according to the traditionally accepted definition the SH3 domain folds in a two-state manner. These results show that SOP-SC can accurately predict λ values.

A more robust way to assess how co-operative a folding transition is to calculate the dimensionless measure,[60] with a much larger dynamic range,

$$\Omega_c = \frac{T_m^{\;2}}{\Delta T}\left(\frac{\mathrm{d}f_{\mathrm{NBA}}}{\mathrm{d}T}\right) \tag{22}$$

where ΔT is the full width at half-maximum of the derivative of the probability of being in the f_{NBA} with respect to T, and $\mathrm{d}f_{\mathrm{NBA}}/\mathrm{d}T$ is to be evaluated at T_m. We calculated f_{NBA} using

$$f_{\mathrm{NBA}} = \frac{\sum_i \delta(\chi_c - \chi_i)\mathrm{e}^{-\beta E_{\mathrm{MTM}}}}{Z_{\mathrm{MTM}}} \tag{23}$$

In the above equation, χ_i is the overlap function for the ith conformation and χ_c (=0.65) is the boundary between the NBA and the unfolded state (see the Supporting Information in ref 47 for details). The extent of cooperativity increases as Ω_c increases. We find that Ω_c values are 1084, 1158, and 1203, respectively, at $[C] = 0$, 1.0, and 2.5 M, respectively. These results are not inconsistent with the van't Hoff criterion for src-SH3.

GdmCl-Dependent Thermodynamics. The changes in stability of the folded state with respect to the unfolded state are calculated using a two-state description, which is valid for src-SH3 folding. We calculated $\Delta G_{\mathrm{NU}}[C]$ using

$$\Delta G_{\mathrm{NU}} = -RT \ln\left(\frac{f_{\mathrm{NBA}}}{1 - f_{\mathrm{NBA}}}\right) \tag{24}$$

The dependence of f_{NBA} as a function of T at $[C] = 0$, 1.0, and 2.5 M shows (Figure 2a) that the probability of being in the NBA shifts to lower values as $[C]$ increases at a fixed T. The decrease in the melting temperatures, T_m, obtained from

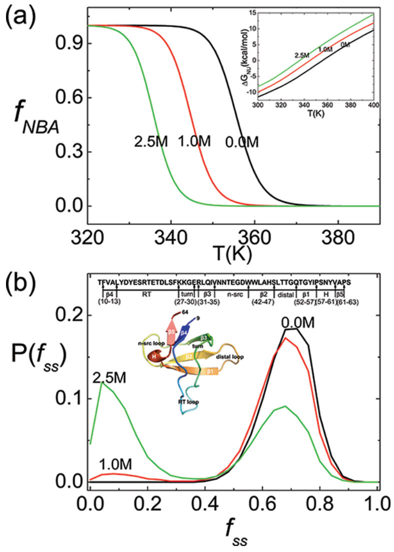


Figure 2. (a) Fraction of molecules in the native state as a function of temperature, T, for GdmCl concentrations $[C] = 0$, 1.0, and 2.5 M calculated using the exact partition function (eq 7). The inset shows the free energy of the native state with respect to the unfolded state as a function of T at three GdmCl concentrations. (b) Distribution $P(f_{ss})$ of secondary structure content, f_{ss} (see text for definition of f_{ss}), for $[C] = 0$, 1.0, and 2.5 M at $T = 339$ K. At $[C] = 2.5$ M, the areas under the two peaks coincide. The ribbon diagram of src-SH3, the secondary structural elements (SSEs), along with the sequence and location of SSE is displayed.

$f_{\mathrm{NBA}}([C], T_{\mathrm{m}}) = 0.5$ shows, in accord with experiments, that GdmCl destabilizes src-SH3. In the inset, we show the T dependence of ΔG_{NU} at the three concentrations of GdmCl. The predictions for the combined T and $[C]$ dependence of ΔG_{NU} can be validated using standard ensemble experiments.

Dependence of Secondary Structure Formation on Denaturant Concentration. The distributions $P(f_{\mathrm{ss}})$ of the fraction of native contacts formed by the secondary structural elements (displayed in the inset in Figure 2b) for $[C] = 0$, 1, and 2.5 M at $T = 339$ K are shown in Figure 2b. The fraction of native contacts, f_{ss}, in the secondary structural elements (helix, β_{12}, β_{23}, β_{34}, β_{45}) is defined as $f_{\mathrm{ss}} = N_{\mathrm{ss}}/N_{\mathrm{ss}}^{\mathrm{o}}$, where $N_{\mathrm{ss}}^{\mathrm{o}}$ ($=N_{\mathrm{helix}}^{\mathrm{o}} + N_{\beta_{12}}^{\mathrm{o}} + N_{\beta_{23}}^{\mathrm{o}} + N_{\beta_{34}}^{\mathrm{o}} + N_{\beta_{45}}^{\mathrm{o}}$) is the total number of contacts in the secondary structural elements in the native state. For src-SH3, $N_{\mathrm{helix}}^{\mathrm{o}} = 12$, $N_{\beta_{12}}^{\mathrm{o}} = 60$,$N_{\beta_{34}}^{\mathrm{o}} = 49$, $N_{\beta_{34}}^{\mathrm{o}} = 19$, and $N_{\beta_{45}}^{\mathrm{o}} = 37$ are the numbers of contacts in the native state in the helix, β_{12}, β_{23}, β_{34}, and β_{45}, respectively. As shown in Figure 2b, the probability of formation of the fraction of native contacts decreases as $[C]$ increases. At $[C] = 2.5$ M, which is the midpoint GdmCl concentration ($f_{\mathrm{NBA}}([C_{\mathrm{m}}], T = 339\ \mathrm{K}) = 0.5$), $P(f_{\mathrm{ss}})$ exhibits a bimodal behavior. At C_{m}, the probabilities of formation and disruption of contacts involving the secondary structural elements are equal.

Variations in the Residue Dependent Unfolding Transition Midpoints. The dependence in the fraction of native contacts that each residue i forms, $\langle Q_i \rangle$, as a function of $[C]$ at $T = 339$ K for all 56 residues in the protein is shown in Figure 3a. The midpoint concentration, $C_{\mathrm{m},i}$, of the melting transition of residue i is defined as the peak position of $\mathrm{d}\langle Q_i \rangle/\mathrm{d}[C]$. The range of $C_{\mathrm{m},i}$ is between 2.2 and 2.7 M (Figure 4a). It follows from Figures 3a and 4a that the transition midpoints depend on i, which means that various residues acquire their native contacts at $[C]$ values that are different from C_{m}, the midpoint at which $f_{\mathrm{NBA}}([C_{\mathrm{m}}]) = 0.5$. The dispersion in $C_{\mathrm{m},i}$ is a consequence of finite size effects and is determined by the specific context-dependent interactions associated with various residues.[61] A similar result is found for i-dependent melting temperatures (Figures 3b and 4b). However, the dispersion in $T_{\mathrm{m},i}$ for the temperature induced transition in $\langle Q_i \rangle$ at $[C] = 0$ (Figure 4b) is smaller (Figure 4a), indicating that the thermal melting of the protein is more cooperative than the denaturant-induced unfolding. In a previous paper, we showed that for a number of proteins Ω_{c} values are larger upon thermal unfolding compared to unfolding at high $[C]$. The molecular reason for this observation is unclear. Thermal perturbation acts uniformly on all parts of the protein, whereas a more specific set of interactions between denaturants and proteins results in global unfolding.[62] This could be the reason for reduced cooperativity in denaturants.

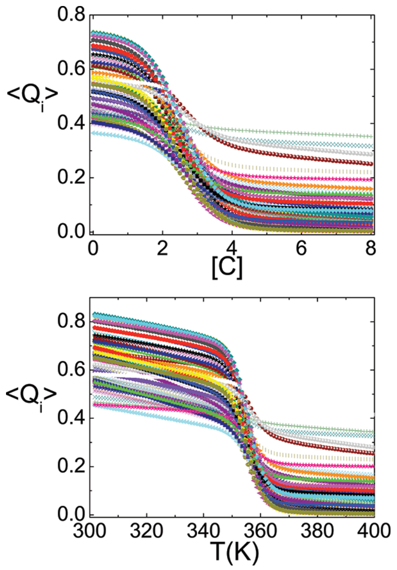


Figure 3. (a) Average fraction of native contacts of residue i in src-SH$_3$, $\langle Q_i \rangle$, as a function of $[C]$ at $T = 339$ K. Each color represents a different residue i, and data are shown for all 56 residues. (b) Same as part a except here $\langle Q_i \rangle$ as a function of T at $[C] = 0$ M for the 56 residues are shown.

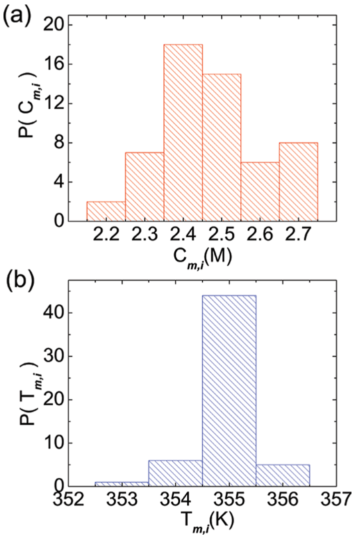


Figure 4. (a) Distribution of the residue dependent midpoints $C_{\mathrm{m},i}$ at $T = 339$ K. (b) Same as part a except these correspond to melting temperatures $T_{\mathrm{m},i}$ for the 56 residues at $[C] = 0$ M.

Deviations from the global melting temperature or C_{m} have been previously reported in experiments beginning with the report by Holtzer and co-workers[63] who showed there is a dispersion in the melting temperature in leucine zipper as assessed by shifts in one-dimensional NMR chemical shifts. Various residues order at temperatures that are different from the global T_{m}. In a much more exhaustive study, it has recently been shown that for BBL[64] there is a great dispersion in the melting temperature around T_{m}.

Distance Distribution Functions as a Function of [*C*]. In order to further characterize the structural changes as $[C]$ is altered, we calculated the distance distribution function, the inverse Fourier transform of the scattering intensity, which is being now routinely measured using small-angle X-ray scattering (SAXS) experiments. The normalized distance distribution function, $P(r)$, as a function of r for various $[C]$ at $T = 339$ K is plotted in Figure 5a. The function $P(r)$, which is the distribution of distances between all noncovalently linked atoms, broadens as $[C]$ increases. As the protein unfolds with increasing $[C]$, the distance between different interaction centers increases, leading to the broadening seen in $P(r)$. The R_{g} of SH3 calculated using the relation $R_{\mathrm{g}} = [\int r^2 P(r)\,\mathrm{d}r/2]^{1/2}$ for $[C] = 0$, 1.0, 2.5, and 5.0 M is 11.03, 11.75, 17.3, and 23.19 Å, respectively. These are in good agreement with the values 11.57, 11.994, 16.56, and 22.94 Å calculated using $R_{\mathrm{g}} = ((1/2N^2)\sum r_{ij}^{\,2})^{1/2}$. The predictions for $P(r)$ can in principle be tested using SAXS experiments.

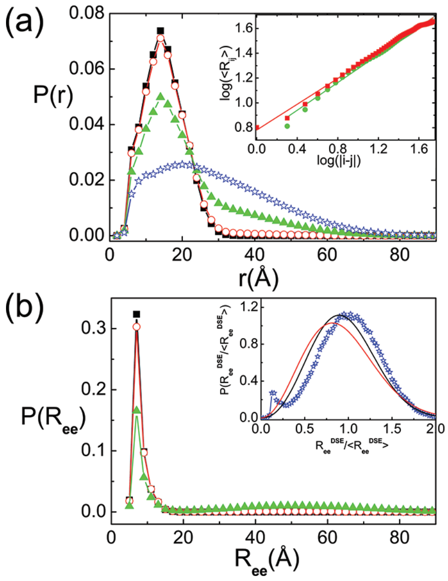


Figure 5. (a) Distance distribution function $P(r)$, the inverse Fourier transform of the scattering intensity, for 0 M (black), 1.0 M (red), 2.5 M (green), and 5.0 M (blue) GdmCl. Here, r is the distance between all noncovalently linked atoms. The inset shows a plot of $\log\langle R_{ij}\rangle$ as a function of the separation, $\log|i - j|$, between the interaction centers i and j. The symbols are obtained from simulations at 5.0 M. Green (red) corresponds to backbone atoms (backbone atoms and side chains). Solid lines are fits to $\langle R_{ij}\rangle \sim |i - j|^{\nu_{\text{eff}}}$, where ν_{eff} is 0.56 and 0.52 for green and red, respectively. Both of these values are less than the Flory exponent $\nu \sim 0.6$, which implies that 5.0 M is not a good solvent for the src-SH3 domain. (b) End-to-end distribution $P(R_{\text{ee}})$ at three GdmCl concentrations. The color code is the same as in part a. The inset shows the distributions of $y = R_{\text{ee}}^{\text{DSE}}/\langle R_{\text{ee}}^{\text{DSE}}\rangle$ at $[C] = 5.0$ M. The black line is the expected distribution for a self-avoiding walk (see text for the analytic expression), and the red line is the result for a Gaussian chain.

In the inset to Figure 5a, we show plots of $\log\langle R_{ij}\rangle$ as a function of $\log|i - j|$ at $[C] = 5.0$ M, where $\langle R_{ij}\rangle$ is the distance between interaction centers i and j. If high GdmCl is a good solvent for src-SH3 protein, then we expect that $\langle R_{ij}\rangle \sim |i - j|^{\nu}$ with $\nu \approx 0.6$. The inset shows that $\langle R_{ij}\rangle$ does increase as a power law for both backbone atoms (green) and for backbone and side chain (red). The effective fractal dimension exponent is 0.56 for the green line and 0.52 for the red line. Both of these values are less than what is expected for a random coil, which suggests that even 5.0 M GdmCl is not a good solvent for src-SH3 (see below).

In order to assess whether SH3 behaves as a random coil at high $[C]$, we also calculated the distribution $P(R_{\text{ee}})$ of the end-to-end distance R_{ee}. At concentrations below $C_{\text{m}} = 2.5$ M, $P(R_{\text{ee}})$ is sharply peaked at $R_{\text{ee}} \sim 10$ Å (Figure 5b) because SH3 is predominantly in the native state. The distribution at $[C] = 5.0$ M is broad, reflecting the heterogeneous nature of the unfolded states. If $[C] = 5.0$ M ($\approx 2C_{\text{m}}$) is a good solvent for src-SH3, then we expect that the distribution function should take the form

$$P(y) = c_1 y^{2+\theta} \exp(-c_2 y^{1/(1-\nu)}) \qquad (25)$$

where $y = R_{\text{ee}}^{\text{DSE}}/\langle R_{\text{ee}}^{\text{DSE}}\rangle$ ($\langle R_{\text{ee}}^{\text{DSE}}\rangle = 55.0$ Å), $\theta = (\gamma - 1)/\nu \approx 0.28$, $\nu \approx 0.6$ is the Flory exponent, $c_1 = 3.7$, and $c_2 = 1.2$. Comparison of the plot using eq 25 and the simulated result for $P(y)$ (blue symbols in the inset in Figure 5b) shows that the simulated $P(y)$ deviates from the expected universal behavior for a random coil even at 5 M GdmCl. This finding that even at $[C]$ values that greatly exceed C_{m} proteins are not true random coils has also been reported for other proteins.[38] Thus, it is not surprising that even at high $[C]$ there is residual structure in the DSE.

Kinetic Cooperativity in Folding. In our previous study,[47] we showed at nonzero $[C]$ SH3 folds in a single step characterized by near simultaneous acquisition of all the contacts involving the various secondary structural elements and the global contacts characterizing the tertiary structure. In addition, the formation of secondary structural elements was coincident with the global collapse of the protein. We repeated these calculations for $[C] = 0$ M by generating folding trajectories starting from initial conformations generated at a high temperature and subsequently quenching T to 339 K. From these folding trajectories, we probed the collapse kinetics using $\langle R_{\text{g}}(t)\rangle$ as a function of time, t. The time evolution of $\langle R_{\text{g}}(t)\rangle$ can be fit using a single exponential function with the rate of collapse, $k_{\text{c}}([C] = 0.0) = 815\ \text{s}^{-1}$ (red line in Figure 6a). The collapse time, $\tau_{\text{c}} = k_{\text{c}}([C])^{-1} \sim 1.2$ ms, compares favorably with theoretical prediction ($\tau_{\text{c}} \approx \tau_0 N^{2.2}$, with $\tau_0 \approx 1\ \mu$s) for heteropolymer collapse.[17]

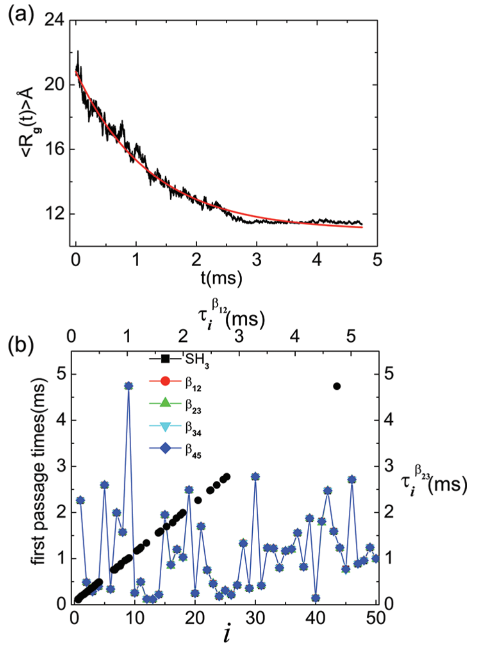


Figure 6. (a) Collapse kinetics monitored by the time-dependent average $\langle R_{\text{g}}(t)\rangle$ as a function of t. The decrease in $\langle R_{\text{g}}(t)\rangle$ can be fit using a single exponential function shown in red. (b) First passage time for the formation of contacts between various secondary structural elements from 50 kinetic folding trajectories at $[C] = 0$ M and $T = 339$ K. The solid black symbols show the correlation between $\tau_i^{\beta_{12}}$ and $\tau_i^{\beta_{23}}$, which are first passage times for formation of interactions β_1–β_2 and β_2–β_3, respectively.

Interestingly, during the process of folding, interactions between various secondary structural elements are consolidated nearly simultaneously. The first passage times for the formation of various secondary structural elements for 50 folding trajectories at $[C] = 0.0$ M and $T = 339$ K are shown in Figure 6b. The overlap of the first passage times for contacts between the secondary structural elements for 50 different trajectories demonstrates that the protein folds into the native state with all the secondary structures forming and coming together at the same time. This is further demonstrated by the perfect correlation (black dots in Figure 6b) between the first passage times for the formation of contacts between $\beta1-\beta2$ and $\beta2-\beta3$ (see inset in Figure 2 for secondary structure labels β-sheets) for the 50 trajectories.

Free Energy Profiles and the Movement of the Transition State. In order to assess the response of the transition state location SH3 to changes in temperature and denaturant, we calculated the free energy profiles $G(\chi)$ and $G(R_g)$. Although the use of restricted order parameters may not accurately represent the multidimensional nature of the underlying folding landscape, the kinetic cooperativity in the folding of SH3 suggests that both χ (eq 16) and R_g could capture the salient features of the energy landscape. The calculated profiles $G(\chi)$ and $G(R_g)$ at a fixed $T = 339$ K are shown in Figure 7. At $[C] = 0$ M, the native state is stable compared to the unfolded state (black curves in Figure 7a and b). The free energy difference between the folded and unfolded states obtained from $G(\chi)$ at $[C] = 0$ is $\Delta G_{NU} = -3.8$ kcal/mol. This value, which is in good agreement with experiment, also coincides with our previous estimate.[47] As $[C]$ increases, the free energy difference between the native and unfolded states increases. At $[C] > 2.5$ M, src-SH3 unfolds (Figure 7a), and the unfolded state is more stable compared to the native state.

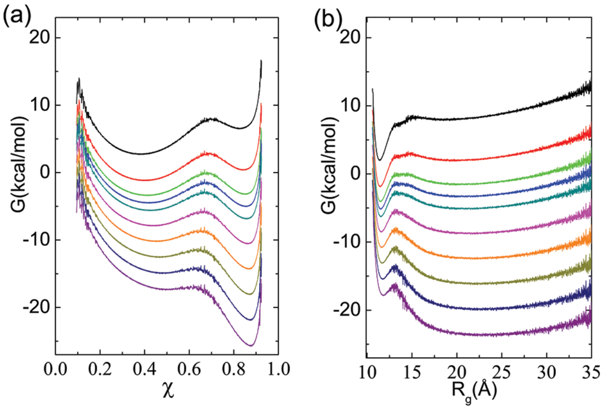


Figure 7. (a) Free energy profiles $G(\chi)$ as a function of the structural overlap function χ at different GdmCl concentrations. (b) Same as part a except the free energy profiles are calculated as a function of the radius of gyration, R_g. In both parts a and b, $T = 339$ K. The values of $[C]$ measured in M from top to bottom are 0, 1, 2, 2.5, 3, 4, 5, 6, 7, and 8.

Interestingly, the width of the NBA is broad as measured by χ, whereas, as expected, it is narrow as assessed by R_g (Figure 7). The free energy profile $G(\chi)$ is a projection of a multidimensional landscape onto a collective dimensional coordinate. Hence, this intrinsic averaging masks the ruggedness of the NBA, which is reflected in a broad native basin in $G(\chi)$. The ruggedness is most likely due to disorder in the side chains in the NBA at finite temperatures. The profiles in Figure 7 can also be used to estimate the Tanford β_F parameter, which can be estimated using $(\chi_{TS} - \chi_{NBA})/(\chi_{UBA} - \chi_{NBA})$, where χ_{TS} is the location of the transition state with respect to the minimum, χ_{NBA}, in the NBA and χ_{UBA} is the corresponding minimum in the UBA. At $[C] = 0$, $\chi_{TS} = 0.69$, $\chi_{UBA} = 0.82$, and $\chi_{NBA} = 0.38$. Thus, the calculated value of β_F is 0.7, which agrees well with the experimental estimate based on Chevron plots[40,41] and also coincides with our previous value obtained by an entirely different method.[47] As $[C]$ increases, there is a discernible increase in χ_{NBA}, which leads to a decrease in β_F. Thus, upon increasing $[C]$, the global location of the transition state ensemble moves toward the more stable state (NBA), which is in accord with the Hammond postulate.

A different picture emerges in analyzing the free energy profiles (Figure 8) calculated along the phase boundary in the $T-[C]$ plane. As $[C]$ is altered, the melting temperature T_m changes linearly (Figure 1). The calculated $G(\chi)$ and $G(R_g)$ at different $T_m[C]$ are shown in Figure 8. Because all the profiles are computed at $T_m[C]$, they display the expected bistable behavior. The differences in these profiles are in the values of the stabilities and the barrier heights. We constructed rescaled free energies $G_R(\chi)$ and $G_R(R_g)$ such that the values at the transition state for $[C] \neq 0$ coincide. This is achieved using $G_R(\chi) = G(\chi) - [G(\chi^{TS}) - G_0(\chi^{TS})]$, where χ^{TS} is the transition state location and $G_0(\chi^{TS})$ is the value of the free energy profile at χ^{TS} at 0 M. A similar rescaling was done for $G(R_g)$. We find that $[G(S^{TS}) - G_0(S^{TS})] = -(1.02 + 2.61[C])$ kcal/mol, where S is either χ or R_g. All the rescaled profiles fall on a single curve (see the bottom panels in Figure 8). The rescaled free energies show that β_F does not change. From this observation, we conclude that the transition state location does not change by moving across the phase boundary in the $T-[C]$ plane. Thus, the response of the transition state expressed in terms of the stiffness parameter β_F depends on the conditions at which the experiments are carried out.

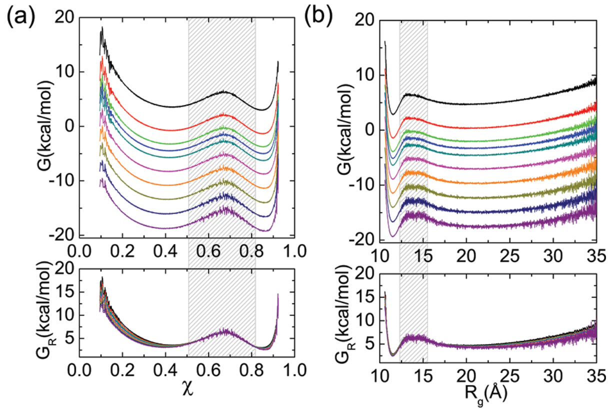


Figure 8. (a) Dependence of $G(\chi)$ at various GdmCl concentrations. These profiles are calculated along the boundary, $T_m([C])$, at which the folded and unfolded states are of equal stability. In the bottom panel, we show $G_R(\chi)$ obtained by scaling the profiles above in such a way that the values at the transition state coincide. (b) Same as part a except these profiles correspond to $G(R_g)$ and $G_R(R_g)$. The GdmCl concentrations are the same as in Figure 7.

■ CONCLUSIONS

The theoretical basis for the MTM, which is a physically motivated way of including the effects of denaturants and osmolytes, shows that it is a mean-field-like model. Because we have combined the experimental measured transfer free energies with the underlying polymeric nature of the polypeptide chain, the theory implicitly accounts for multiparticle interactions. In addition, the amino acid context dependence of the response of denaturants is also implicitly taken into account. Perhaps, it is for these reasons that the MTM is successful in predicting a number of aspects of denaturant-dependent folding in applications to a variety of single domain proteins. Here, we have shown that the MTM accurately reproduces the cooperativity of the folding transition when used in conjunction with the SOP-SC representation of src-SH3. The near coincidence of the van't Hoff and calorimetric enthalpies attests to the two-state nature of SH3 folding. Moreover, the calculated probability of being in the NBA as a function of GdmCl concentration at a fixed T agrees with experimental measurements.

In addition to merely reproducing known experimental results, the current application of MTM to folding of the src-SH3 domain has produced a number of predictions. (1) We have shown that the melting temperature decreases linearly as the GdmCl concentration is increased. (2) Computing the temperature-dependent changes in the fraction of molecules in the NBA at different GdmCl concentrations (Figure 2) allows us to obtain a phase diagram for the SH3 domain as a function of T and $[C]$. (3) We have predicted that both the midpoints and the melting temperatures should be residue dependent. Interestingly, thermal melting is more cooperative than unfolding by GdmCl (smaller dispersion in the melting temperatures compared to dispersion in the denaturant midpoints). (4) We have predicted changes in the distance distribution functions as a function of $[C]$. The characterization of the UBA shows that SH3 is not a Flory random coil even at 5 M GdmCl. (5) The transition state ensemble moves in accord with the Hammond postulate at a fixed T when $[C]$ is increased. However, the transition state location is invariant along the locus of points $T_m([C])$. In other words, at the $[C]$-dependent melting temperatures, the Tanford β_F does not change. All of these novel predictions are amenable to experimental test using standard experimental techniques (SAXS, measuring chemical shifts using one-dimensional NMR, stopped flow experiments).

We note in closing that the theory outlined here is not without limitations. First, the MTM itself does not explicitly take into account atomic details of the denaturants. It is worth emphasizing that not including water explicitly is solely for computational reasons, and is not an inherent limitation of MTM. In the MTM, what is needed is extensive sampling of protein conformations at a fixed temperature. Such simulations can be performed using all atom molecular dynamics simulations in explicit water. The resulting ensemble of conformations can be used to obtain thermodynamic properties as a function of denaturants using the method proposed here and elsewhere.[38] Thus, there is freedom in choosing the level of description for both water and the proteins. The use of transfer free energy implies that in the MTM theory denaturants are treated implicitly.

Second, we have used a coarse-grained SOP-SC native-centric model for the SH3 domain. As such, it is open to criticism that non-native interactions could affect the folding of SH3 domains. Several previous studies have taken partial account of non-native interactions in simulations.[65−68] In all these insightful studies, only a subset of non-native interactions among a subset of residues are taken into account often with reduced strength and/or altered interaction range. Thus, the extent to which non-native interactions affect the folding thermodynamics of proteins that ostensibly fold in a two-state manner remains an open question. It is extremely unlikely that the presence of persistent non-native interactions contributes significantly to the thermodynamics of folding, at least for the proteins that we have studied. Had this been the case, the good agreement between simulations and experiments across the board would not be possible. The effect of non-native interactions on the kinetics is harder to assess. Undoubtedly, they are relevant in at least the early stages (time scales on the order of collapse times) of folding, as has been known for some time through detailed studies of simple models[69] and more recently in the all atom simulations of cytochrome c by Elber.[70] It remains unclear whether the effect of non-native interactions can be adequately accounted for using a roughness-dependent diffusion coefficient at least in rationalizing the rates of folding. Although studies based on lattice models show that to a large extent the folding mechanisms are unaltered by non-native interactions,[71,72] additional work is required to fully quantify their effect on the folding reaction.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: greddy1@umd.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Edward P. O'Brien, Mikael Oliveberg and George Makhatadze for useful discussions. This work was supported in part by a grant from the National Science Foundation through CHE 14-09433. Z.L. acknowledges partial financial support from the National Natural Science Foundation of China NNSFC under Grant No. 11104015.

■ REFERENCES

(1) Dill, K. A.; Bromberg, S.; Yue, K. Z.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. *Protein Sci.* **1995**, *4*, 561−602.
(2) Bryngelson, J.; Onuchic, J.; Socci, N.; Wolynes, P. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167−195.
(3) Thirumalai, D.; Woodson, S. A. *Acc. Chem. Res.* **1996**, *29*, 433−439.
(4) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 10−19.
(5) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70−75.
(6) Shakhnovich, E. *Chem. Rev.* **2006**, *106*, 1559−1588.
(7) Dill, K. A.; Ozkan, B. S.; Shell, M.; Weikl, T. R. *Ann. Rev. Biophys.* **2008**, *37*, 289−316.
(8) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. *Annu. Rev. Biophys.* **2010**, *39*, 159−183.
(9) Liwo, A.; He, Y.; Scheraga, H. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16890−16901.
(10) Schuler, B.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16−26.
(11) Gebhardt, J. C. M.; Bornschloegla, T.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2013−2018.
(12) Stigler, J.; Ziegler, F.; Gieseke, A.; Gebhardt, J. C. M.; Rief, M. *Science* **2011**, *334*, 512−516.
(13) Garcia-Manyesa, S.; Dougana, L.; Badillaa, C. L.; Brujic, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10534Ð10539.
(14) Fernandez, J. M.; Li, H. *Science* **2004**, *303*, 1674−1678.
(15) Peng, Q.; Fang, J.; Wang, M.; Li, H. *J. Mol. Biol.* **2011**, *412*, 698−709.
(16) Mickler, M.; Dima, R. I.; Dietz, H.; Hyeon, C.; Thirumalai, D.; Rief, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20268−20273.
(17) Thirumalai, D. *J. Phys. I* **1995**, *5*, 1457−1467.
(18) Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14249−14255.
(19) Thirumalai, D.; Hyeon, C. *Biochemistry* **2005**, *44*, 4957−4970.
(20) Guo, Z. Y.; Thirumalai, D.; Honeycutt, J. D. *J. Chem. Phys.* **1992**, *97*, 525−535.
(21) Goldstein, R. A.; Luthey-Schulten, Z. A.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4918−4922.

(22) Klimov, D. K.; Thirumalai, D. *J. Chem. Phys.* **1998**, *109*, 4119−4125.
(23) Camacho, C. J.; Thirumalai, D. *Phys. Rev. Lett.* **1993**, *71*, 2505−2508.
(24) Chen, J.; Bryngelson, J.; Thirumalai, D. *J. Phys. Chem. B* **2008**, *112*, 16115−16120.
(25) Liu, F.; Maynard, C.; Scott, G.; Melnykov, A.; Hall, K. B.; Gruebele, M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3542−3549.
(26) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341−346.
(27) Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. *Nat. Phys.* **2010**, *6*, 751−758.
(28) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. *Biophys. J.* **2008**, *94*, L75−L77.
(29) Bowman, G. R.; Voelz, V. A.; Pande, V. *J. Am. Chem. Soc.* **2011**, *133*, 664−667.
(30) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.
(31) Canchi, D. R.; Pascheck, D.; Garcia, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 2338−2344.
(32) Hyeon, C.; Thirumalai, D. *Nat. Commun.* **2011**, *2*, 487.
(33) Maisuradze, G. G.; Liwo, A.; Oldziej, S.; Scheraga, H. A. *J. Am. Chem. Soc.* **2010**, *132*, 9444−9452.
(34) Clementi, C.; Nymeyer, H.; Onuchic, J. *J. Mol. Biol.* **2000**, *298*, 937−953.
(35) Klimov, D. K.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7254−7259.
(36) Hyeon, C.; Dima, R. I.; Thirumalai, D. *Structure* **2006**, *14*, 1633−1645.
(37) Hyeon, C.; Lorimer, G. H.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18939−18944.
(38) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13403−13408.
(39) O'Brien, E. P.; Morrison, G.; Brooks, B. R.; Thirumalai, D. *J. Chem. Phys.* **2009**, *130*, 124903.
(40) Grantcharova, V. P.; Riddle, D. S.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7084−7089.
(41) Riddle, D. S.; Grantcharova, V. P.; Santiago, J. V.; Alm, E.; Ruczinski, I.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 1016−1024.
(42) Martinez, J.; Serrano, L. *Nat. Struct. Biol.* **1999**, *6*, 1010−1016.
(43) Ding, F.; Guo, W. H.; Dokholyan, N. V.; Shakhnovich, E. I.; Shea, J. E. *J. Mol. Biol.* **2005**, *350*, 1035−1050.
(44) Klimov, D.; Thirumalai, D. *J. Mol. Biol.* **2002**, *317*, 721−737.
(45) Fernandez-Escamilla, A.; Cheung, M.; Vega, M.; Wilmanns, M.; Onuchic, J.; Serrano, L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2834−2839.
(46) Auton, M.; Holthauzen, L. M. F.; Bolen, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15317−15322.
(47) Liu, Z.; Reddy, G.; O'Brien, E. P.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 7787−7792.
(48) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. *Biochemistry* **2009**, *48*, 3743−3754.
(49) Kumar, S.; J.M., R.; Bouzida, D.; Swendsen, R.; Kollman, P. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(50) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(51) Betancourt, M.; Thirumalai, D. *Protein Sci.* **1999**, *8*, 361−369.
(52) Veitshans, T.; Klimov, D.; Thirumalai, D. *Folding Des.* **1997**, *2*, 1−22.
(53) Ermak, D. L.; Mccammon, J. A. *J. Chem. Phys.* **1978**, *69*, 1352−1360.
(54) Guo, Z.; Thirumalai, D. *J. Mol. Biol.* **1996**, *263*, 323−343.
(55) Otzen, D.; Oliveberg, M. *Protein Sci.* **2004**, *13*, 3253−3263.
(56) Makhatadze, G. I. *Biophys. Chem.* **1998**, *71*, 133−156.
(57) Naganathan, A.; Perez-Jimenez, R.; Sanchez-Ruiz, J.; Munoz, V. *Biochemistry* **2005**, *44*, 7435−7449.
(58) Kaya, H.; Chan, H. *Proteins* **2000**, *40*, 637−661.
(59) Zhou, Y.; Hall, C.; Karplus, M. *Protein Sci.* **1999**, *8*, 1064−1074.
(60) Klimov, D.; Thirumalai, D. *Fold. Des.* **1998**, *3*, 127−139.
(61) Klimov, D.; Thirumalai, D. *J. Comput. Chem.* **2002**, *23*, 161−165.
(62) O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. *J. Am. Chem. Soc.* **2007**, *129*, 7346−7353.
(63) Holtzer, M.; Lovett, E.; dAvignon, D.; Holtzer, A. *Biophys. J.* **1997**, *73*, 1031−1041.
(64) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2006**, *442*, 317−321.
(65) Wallin, S.; Zeldovich, K. B.; Shakhnovich, E. I. *J. Mol. Biol.* **2007**, *368*, 884−893.
(66) Zarrine-Afsar, A.; Wallin, S.; Neculai, A. M.; Neudecker, P.; Howell, P. L.; Davidson, A. R.; Chan, H. S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9999−10004.
(67) Chan, H. S.; Zhang, Z.; Wallin, S.; Liu, Z. *Annu. Rev. Phys. Chem.* **2011**, *62*, 301−326.
(68) Azia, A.; Levy, Y. *J. Mol. Biol.* **2009**, *393*, 527−542.
(69) Camacho, C.; Thirumalai, D. *Proteins: Struct., Funct., Genet.* **1995**, *22*, 27−40.
(70) Cardenas, A.; Elber, R. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 245−257.
(71) Klimov, D. K.; Thirumalai, D. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 465−475.
(72) Gin, B. C.; Garrahan, J. P.; Geissler, P. L. *J. Mol. Biol.* **2009**, *392*, 1303−1314.