THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Folding PDZ2 Domain Using the Molecular Transfer Model

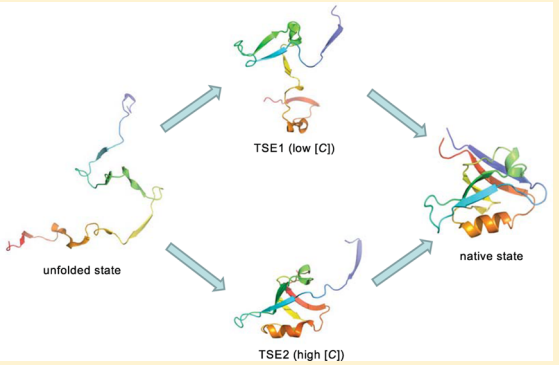
Zhenxing Liu,*[,†] Govardhan Reddy,[‡] and D. Thirumalai[§]

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

[‡]Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore, Karnataka 560012, India

[§]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 major challenge in molecular simulations is to describe denaturant-dependent folding of proteins in order to make direct comparisons with *in vitro* experiments. We use the molecular transfer model (MTM), currently the only method that accomplishes this goal, albeit phenomenologically, to quantitatively describe urea-dependent folding of the PDZ2 domain, a protein that is important in molecular recognition and signaling. Experiments show that urea-dependent unfolding rates of the PDZ2 domain exhibit a downward curvature at high urea concentrations ([*C*]'s), which has been interpreted as indicating the presence of a sparsely populated high energy intermediate. Simulations using the MTM and a coarse-grained self-organized polymer (SOP) representation of PDZ2 are used to show that the intermediate (I_{EQ}), which has some native-like character, is present in equilibrium both in the presence and absence of urea. The free energy profiles as a function of the structural overlap order parameter show that there are two barriers separating the folded and unfolded states. Structures of the transition state ensembles (TSE1 separating the unfolded and I_{EQ} and TSE2 separating I_{EQ} and the native state), determined using the P_{fold} method, show that TSE1 is greatly expanded while TSE2 is compact and native-like. Folding trajectories reveal that PDZ2 folds by parallel routes. In one pathway, folding occurs exclusively through I_1, which resembles I_{EQ}. In a fraction of trajectories, constituting the second pathway, folding occurs through a combination of I_1 and a kinetic intermediate. We establish that the radius of gyration (R_g^U) of the unfolded state is more compact (by ∼9%) under native conditions. Theory and simulations show that the decrease in R_g^U occurs on the time scale on the order of at most ∼20 μs. The modest decrease in R_g^U and the rapid collapse suggest that high spatial and temporal resolution, currently beyond the scope of most small-angle X-ray scattering experiments, are needed to detect compaction in finite-sized proteins. The present work further establishes that MTM is efficacious in producing nearly quantitative predictions for folding of proteins under conditions used in experiments.



INTRODUCTION

The molecular transfer model (MTM),[1] based on the statistical mechanical theory of liquid mixtures,[2] is currently the only available computational method that accurately predicts the outcomes of experiments and provides the structural basis of folding as a function of denaturants[3,4] and pH.[5] Using the MTM in conjunction with coarse-grained (CG) representation of polypeptide chains, we have quantitatively predicted the thermodynamics and kinetics of folding as a function of denaturants for a number of small proteins (protein L, cold shock protein, srcSH3 domain, and ubiquitin),[1,2,4,6] and GFP, a large single domain protein.[3] Because the effects of denaturants are taken into account naturally within the MTM framework,[2] it has been possible to calculate chevron plots for src SH3 domain producing quantitative agreement with experiments for the slopes of folding and unfolding arms.[7] Although MTM can be implemented in conjunction with atomically detailed simulations, we have so far used CG models for proteins. The virtue of CG models[8−11] is that exhaustive sampling over a wide range of external conditions allows us to calculate the thermodynamic and kinetic properties accurately. This is needed for detailed comparison with experiments.[12] These studies illustrate that simulations based on the MTM provide concrete predictions for *in vitro* experiments, thus enabling us to go beyond general concepts (framework model, nucleation-collapse mechanism, diffusion-collision mechanism, and funneled landscape) used to understand generic aspects of protein folding.[13−23] Here, we investigate the folding mechanism of PDZ2 domain using CG simulations within the theoretical framework of the MTM.

PDZ domains are a large family of globular proteins that mediate protein−protein interactions and play an important role in molecular recognition.[24−26] These proteins generally consist of 80−100 amino acids. The folding mechanism of the 94 residue PDZ2 domain, with six β strands and two α helices (Figure 1a), has been studied experimentally[27] using classical

Special Issue: J. Andrew McCammon Festschrift

Received: January 11, 2016
Revised: February 26, 2016
Published: March 1, 2016

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

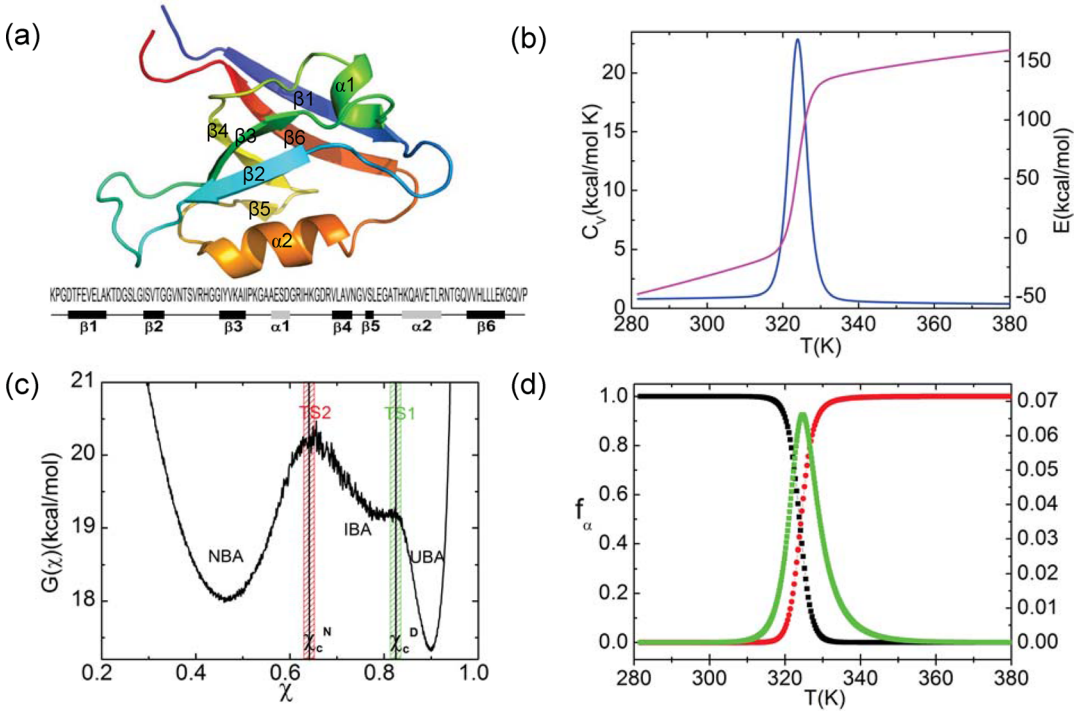


Figure 1. Thermodynamics of folding. (a) Ribbon diagram representation of PDZ2 (PDB code: 1GM1). (b) Temperature dependence of specific heat (blue) and total energy (magenta). (c) Free energy profile at T_m as a function of χ. The values χ_c^N and χ_c^D are used to classify the major equilibrium states. The shaded areas give putative regions for the two transition state ensembles. (d) Fraction of molecules in NBA (black), UBA (red), and IBA (green) as functions of temperature.

chemical kinetics methods. The key findings in these experiments are the following: (i) In urea-induced equilibrium denaturation experiments, the observed transition is cooperative, which is well-described by an apparent two-state model.[27,28] (ii) Typically, if proteins fold thermodynamically in a two-state manner, a similar behavior is observed kinetically in ensemble experiments. However, urea-dependent unfolding rates exhibit a downward curvature at high urea concentrations at pH > 5.5. On the basis of the observation that the folding kinetics is monophasic, with no detectable burst phase in the initial fluorescence of the initial unfolding time course, it was surmised that there is no low energy intermediate in the unfolding of PDZ2. Rather, the data were used to suggest the presence of a high energy on-pathway intermediate, which does not accumulate significantly in equilibrium.[27] (iii) The high energy intermediate is not detected under stabilizing conditions, achievable in the PDZ2 domain by addition of a modest amount of sodium sulfate. Under these conditions PDZ2 folds thermodynamically and kinetically in a two-state manner. (iv) The structures of the two transition state ensembles were also inferred using measured Φ values as constraints in all atom molecular dynamics simulations.[29] Unlike the results summarized by points i−iii, the predicted structures of the transition state ensembles are not as reliable for reasons explained later in this work.

In order to provide a comprehensive picture of folding of PDZ2, with potential implications for other single domain proteins, we performed molecular simulations of a coarse-grained off-lattice model with side chains[9,30] and used MTM[1,3,7] to account for denaturant effects. The calculated free energy profiles as a function of the structural overlap order parameter (χ) at different urea concentration, [C], and temperature T show that the folding mechanism of PDZ2 can be altered by changing the stability of the folded state. In accord with experiments, we demonstrate directly the existence of the fleeting obligatory intermediate both in equilibrium (I_{EQ}) and kinetics (I_1).[31] The structures of I_{EQ} and I_1 are similar. However, the fraction of molecules in intermediate basin of attraction (IBA), f_{IBA}, is small, thus explaining the difficulty in detecting it in standard denaturation experiments. In addition to I_1, a kinetic intermediate, I_2, is consistently populated in ∼53% of the folding trajectories. Guided by the free energy profiles, we identified two transition state ensembles, TSE1 and TSE2. The computed values of the Tanford-like β parameters, using the solvent accessible surface area as a surrogate, for the two transition state ensembles (one connecting the NBA and the IBA and the other involving transition between the IBA and the UBA) are in qualitative agreement with those obtained from experiments.[27,31] The current work further establishes that simulations based on the MTM are efficacious in providing a nearly quantitative picture of folding of single domain proteins.

METHODS

SOP-Side Chain Model. We performed simulations using the self-organized polymer (SOP) model[32] in which each amino acid is represented using both C_α atoms as well as side chains (SCs).[2,7] In the SOP-SC model, the stability of the folded state is determined by backbone−backbone (bb), side chain−side chain (ss), and backbone−side chain (bs) interactions. Neglect of non-native interactions, which do not significantly alter the folding mechanism beyond the global collapse of the protein,[33−36] is nevertheless a limitation of the model.

The energy of a conformation, describing the intrapeptide interactions, is

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

$$E_P(\{r_i\}) = V_{FENE} + V_{LJ}^{NAT} + V^{NEI} + V_{LJ}^{NN} \quad (1)$$

The nonlinear elastic potential (FENE), V_{FENE}, accounting for the chain connectivity between backbones and side chains, is given by

$$V_{FENE} = V_{FENE}^{bb} + V_{FENE}^{bs} = -\sum_{i=1}^{N-1} \frac{k}{2} R_o^2 \log\left(1 - \frac{(r_{i,i+1\_bb} - r^o_{i,i+1\_bb})^2}{R_o^2}\right) - \sum_{i=1}^{N} \frac{k}{2} R_o^2 \log\left(1 - \frac{(r_{i,i\_bs} - r^o_{i,i\_bs})^2}{R_o^2}\right) \quad (2)$$

The nonbonded native interaction, V_{LJ}^{NAT} in eq 1 is taken to be

$$V_{LJ}^{NAT} = V_{LJ\_NAT}^{bb} + V_{LJ\_NAT}^{ss} + V_{LJ\_NAT}^{bs} = \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} e_{bb}\left[\left(\frac{r^o_{i,j\_bb}}{r_{i,j\_bb}}\right)^{12} - 2\left(\frac{r^o_{i,j\_bb}}{r_{i,j\_bb}}\right)^{6}\right]\Delta_{ij}^{bb} + \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} e_{ss}|\epsilon_{ij} - 0.7|\left[\left(\frac{r^o_{i,j\_ss}}{r_{i,j\_ss}}\right)^{12} - 2\left(\frac{r^o_{i,j\_ss}}{r_{i,j\_ss}}\right)^{6}\right]\Delta_{ij}^{ss} + \sum_{i=1,j=1/|i-j|\geq 3}^{N} e_{bs}\left[\left(\frac{r^o_{i,j\_bs}}{r_{i,j\_bs}}\right)^{12} - 2\left(\frac{r^o_{i,j\_bs}}{r_{i,j\_bs}}\right)^{6}\right]\Delta_{ij}^{bs} \quad (3)$$

In eqs 2 and 3 the superscript ^{o} refers to distances in the native state. A native contact exists if the distance between two noncovalently linked beads, r_{ij} ($|i - j| \geq 3$), in the PDB structure is less than a cutoff distance R_c (=8 Å). In this case $\Delta_{ij} = 1$. If r_{ij} exceeds R_c, then $\Delta_{ij} = 0$. The strengths of the nonbonded interactions e_{bb}, e_{ss}, e_{bs} are constants. Sequence effects are taken into account using the SC-dependent interactions, which are parametrized using the Betancourt–Thirumalai (BT)[37] statistical potential matrix with elements ϵ_{ij}.

Excluded volume interactions between neighboring beads are modeled using a repulsive potential. We computed the repulsive potential using

$$V^{NEI} = V_{NEI}^{bb} + V_{NEI}^{ss} + V_{NEI}^{bs} = \sum_{i=1}^{N-2} e_l\left(\frac{\sigma_{bb}}{r_{i,i+2\_bb}}\right)^6 + \sum_{i=1}^{N-1} e_l\left(\frac{\sigma_{i,i+1\_ss}}{r_{i,i+1\_ss}}\right)^6 + \sum_{i=1}^{N-2} e_l\left(\frac{\sigma_{i,i+2\_ss}}{r_{i,i+2\_ss}}\right)^6 + \sum_{i=1,j=1/0<|i-j|<3}^{N} e_l\left(\frac{\sigma_{j\_bs}}{r_{i,j\_bs}}\right)^6 \quad (4)$$

where e_l is the strength of the repulsion, and σ_{bb}, σ_{i,j_ss}, σ_{j_bs} are the ranges for bb, ss, and bs interactions, respectively. The choice of $\sigma_{bb} = a = 3.8$ Å (a is average distance between neighboring C_α atoms), $\sigma_{i,j\_ss} = f(\sigma_i + \sigma_j)$ (σ_i, σ_j are the van der Waals radii of the side chains and $f = 0.5$), $\sigma_{j\_bs} = f(a + \sigma_j)$.

The nonbonded non-native interactions are given by

$$V_{LJ}^{NN} = V_{LJ\_NN}^{bb} + V_{LJ\_NN}^{ss} + V_{LJ\_NN}^{bs} = \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} e_l\left(\frac{\sigma_{bb}}{r_{i,j\_bb}}\right)^6 (1 - \Delta_{ij}^{bb}) + \sum_{i=1}^{N-3}\sum_{j=i+3}^{N} e_l\left(\frac{\sigma_{i,j\_ss}}{r_{i,j\_ss}}\right)^6 (1 - \Delta_{ij}^{ss}) + \sum_{i=1,j=1/|i-j|\geq 3}^{N} e_l\left(\frac{\sigma_{j\_bs}}{r_{i,j\_bs}}\right)^6 (1 - \Delta_{ij}^{bs}) \quad (5)$$

In eqs 4 and 5, we have used a longer range repulsion for two reasons: (a) The longer range accounts for hydration of side chains.[38] (b) The softer potential allows us to take larger time steps in integrating the equations of motion, which is necessary because simulations are being carried out for a range of urea concentrations. We note that as the interaction is short ranged, which is the case even with the softer potential, none of the results would change qualitatively.

The SOP-SC energy function $E_P(\{r_i\})$ has seven parameters: $R_o = 2$ Å, $k = 20$ kcal/mol/Å^2, $R_c = 8$ Å, $e_{bb} = 0.73$ kcal/mol, $e_{bs} = 0.17$ kcal/mol, $e_{ss} = 0.3$ kcal/mol, $e_l = 1$ kcal/mol. The results would not be significantly affected by the precise choice of parameters enforcing the integrity of the polypeptide chain, determined by R_o and k. Thus, in effect there are five parameters in the SOP-SC model. The first and third terms in eq 3 account for bb and bs interactions, which are found in the folded states of proteins.[39] Inclusion of these terms is needed to accurately describe packing effects. The experimental melting temperature of the protein is used to determine the strengths of the native contacts, specified by e_{bb}, e_{bs}, e_{ss}.

Molecular Transfer Model. According to the MTM theory, the free energy of transferring a protein from water ($[C] = 0$) to a denaturant solution with ($[C] \neq 0$) is

$$\Delta G(\{r_i\}, [C]) = \sum_i \delta g(i, [C])\alpha_i/\alpha_{Gly-i-Gly} \quad (6)$$

where the sum is over bb and sc, $\delta g(i, [C])$ is the measured transfer free energy for group i, α_i is the solvent accessible surface area (SASA), and $\alpha_{Gly-i-Gly}$ is the SASA of the ith group in the tripeptide $Gly-i-Gly$. At finite $[C]$ the total free energy for a protein is

$$H_P(\{r_i\}, [C]) = E_P(\{r_i\}) + \Delta G(\{r_i\}, [C]) \quad (7)$$

For computational expediency we combined converged simulations at finite temperature using $E_P(\{r_i\})$ and computed the partition function at $[C] \neq 0$ to obtain thermodynamic properties using a weighted histogram analysis method,[40] which takes into account the effects of $\Delta G(\{r_i\}, [C])$. Such an approximation, whose validity for obtaining thermodynamic properties has been previously established,[7] is used to obtain thermodynamic properties.

Langevin and Brownian Dynamics Simulations. We assume that the Langevin equation, with a friction coefficient ζ and a Gaussian random force Γ, describes the dynamics of the protein. The equation of motion is $m\ddot{r}_i = -\zeta\dot{r}_i + F_c + \Gamma$ where m is the mass of a bead, $F_c = -\partial E_P(\{r_i\})/\partial r_i$. In the discrete form the statistics for $\Gamma(t)$ is $\left\langle \Gamma(t)\Gamma(t + nh)\right\rangle = \frac{2\zeta k_B T}{h}\delta_{0,n}$[41] where

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

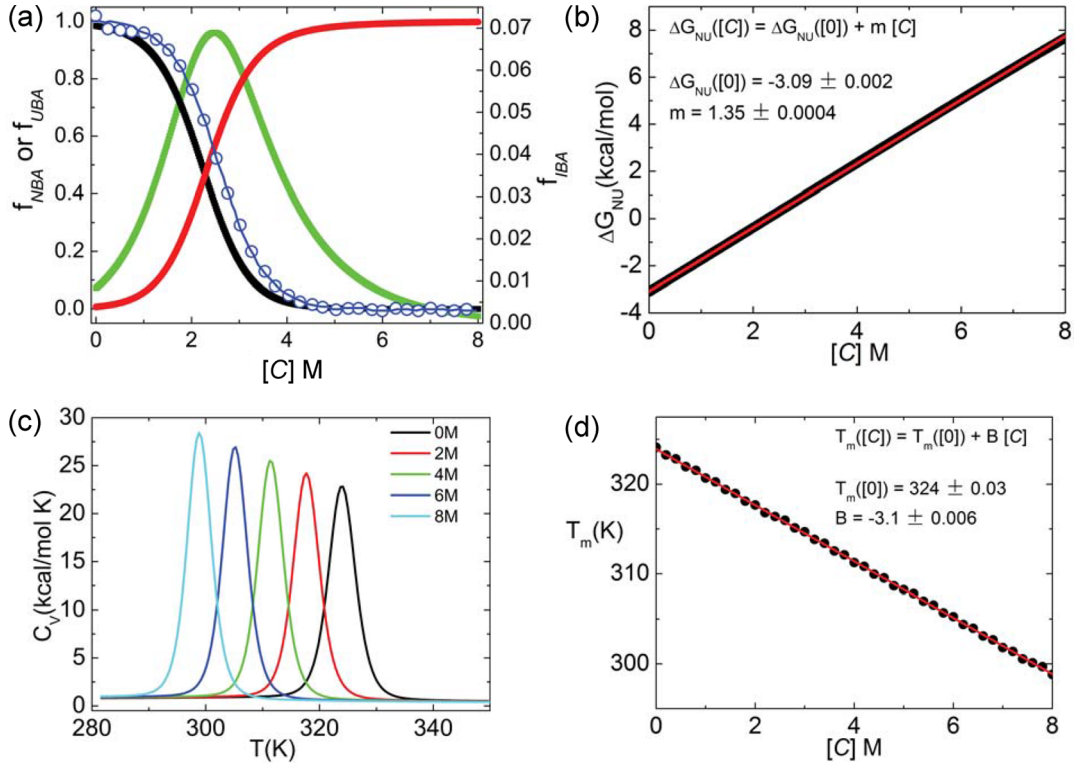


Figure 2. Denaturation effects. (a) Fraction of molecules in the NBA (black), UBA (red), and IBA (green) as a function of urea concentration $[C]$. For comparison, the experimental curve for $f_{NBA}[C]$ (blue) is shown. (b) $[C]$ dependence of free energy of stability of the native state with respect to the unfolded state. Fit to a linear function yields $\Delta G_{NU} = \Delta G_{NU}(0) + m[C]$ where $\Delta G_{NU}(0) = -3.09$ kcal/mol and $m = 1.35$ kcal/mol/M. (c) Heat capacity versus temperature for different values of $[C]$. (d) The $[C]$ dependence of the melting temperature. The line is a fit to $T_m[C] = T_m(0) - B[C]$ where $T_m(0) = 324$ K and $B = -3.1$ KM^{-1}.

$\delta_{0,n}$ is the Kronecker delta function and $n = 0, 1, 2, \ldots$, and h is the time step.

To obtain enhanced sampling, we used the replica exchange molecular dynamics (REMD)[42−45] to perform thermodynamics sampling using a low value of $\zeta = 0.05\ m/\tau_L$.[38] The Verlet leapfrog algorithm

$$v_i(t + h/2) = \frac{2m - h\zeta}{2m + h\zeta}\cdot v_i(t - h/2) + \frac{2h}{2m + h\zeta}[F_c(t) + \Gamma(t)] \quad (8)$$

$$r_i(t + h) = r_i(t) + h\cdot v_i(t + h/2) \quad (9)$$

is used to integrate the equations of motion.

In order to simulate the kinetics of folding, we set $\zeta = 50\ m/\tau_L$ corresponding to the value in water.[41] In this limit, we integrated the equations of motion using[46]

$$r_i(t + h) = r_i(t) + \frac{h}{\zeta}(F_c(t) + \Gamma(t)) \quad (10)$$

Time Scales. In the high ζ case the unit of time is $\tau_H \approx \frac{\zeta_H a^2}{k_B T_s} = \frac{(\zeta_H \tau_L / m)e_l}{k_B T_s}\tau_L$. In order to convert the simulation time to real time, we chose $e_l = 1$ kcal/mol, average mass $m = 1.8 \times 10^{-22}$ g,[41] $a = 4$ Å, which makes $\tau_L = 2$ ps. For $\zeta_H = 50\ m/\tau_L$, we obtain $\tau_H = 159$ ps. For thermal folding simulations, the integration time step, h, is 0.005 τ_L. In the kinetic folding simulations, h, in eq 10, is 0.02 τ_H. Because of coarse-graining there is effectively a decrease in the number of degrees of freedom leading to underestimate the actual time scales. The difference in the simulation time scale and the real times has no consequence on the thermodynamic properties.

Data Analyses. The peak in the heat capacity is associated with the $[C]$-dependent melting temperature of the protein. The structural overlap function $\chi = 1 - \frac{N_k}{N_T}$[47] is used to monitor the folding reaction, where

$$N_k = \sum_{i=1}^{N-3}\sum_{j=i+3}^{N}\Theta(\delta - |r_{i,j\_bb} - r^o_{i,j\_bb}|) + \sum_{i=1}^{N-3}\sum_{j=i+3}^{N}\Theta(\delta - |r_{i,j\_ss} - r^o_{i,j\_ss}|) + \sum_{i=1,j=1/|i-j|\geq 3}^{N}\Theta(\delta - |r_{i,j\_bs} - r^o_{i,j\_bs}|) \quad (11)$$

In eq 11, $\Theta(x)$ is the Heavyside function. If $|r_{i,j} - r^o_{i,j}| \leq \delta$ (= 2 Å), there is a contact. The number of contacts in the kth conformation is N_k, and N_T is the total number in the folded state. The microscopic order parameter, χ, distinguishes between the folded and unfolded state unambiguously.[47] In addition, intermediate states would have χ values that are different from those in the unfolded and folded states. As we show here, the free energy profiles as functions of χ allow us to identify high energy states as well.

RESULTS

Thermal Denaturation. The temperature dependence of the heat capacity, $C_v\left(= \frac{\langle E^2\rangle - \langle E\rangle^2}{k_B T^2}\right)$, where $\langle E\rangle$ and $\langle E^2\rangle$ are the

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

mean and mean square averages of the energy, respectively, demonstrates that PDZ2 folds cooperatively in a two-state manner (Figure 1b). The melting temperature, identified with the peak in $C_v(T)$, is T_m = 324 K. The value of T_m obtained in our simulations agrees with the experimentally measured T_m = 321 K.[48]

In order to identify the NBA, UBA, and the intermediate basin of attraction (IBA) representing the I_{EQ} state, we plot in Figure 1c the free energy ($G(\chi)$) profile as a function of χ at T_m = 324 K. All the conformations can be classified into three states specified by the black vertical lines based on the χ values. If $\chi \leq \chi_c^N = 0.64$, the conformations are in the NBA, conformations with $\chi \geq \chi_c^D = 0.825$ belong to the UBA, and the rest of the conformations are in the IBA. The fractions of molecules in the NBA, $f_{NBA}([0], T)$ (the first argument shows the value of the denaturant concentration), in the UBA, $f_{UBA}([0], T)$, and in the IBA, $f_{IBA}([0], T)$ as a function of temperature are plotted in Figure 1d. Both $f_{NBA}([0], T)$ and $f_{UBA}([0], T)$ show that the folding or unfolding is cooperative. The value of $f_{IBA}([0], T)$ is negligible compared to $f_{NBA}([0], T)$ and $f_{UBA}([0], T)$, suggesting that, a two-state description is adequate, reflecting the cooperative transition in the heat capacity curve (Figure 1b). The sparse population of the IBA explains why the I_{EQ} state is hard to detect in experiments although its presence appears as a shoulder in the free energy profile (Figure 1c). The value of T_m computed using $f_{NBA}(T_m)$ = 0.5 yields T_m = 324 K, which coincides with the peak in the heat capacity (Figure 1b).

Chemical Denaturation. Following our previous studies,[1,3,7] we choose a value for T_s, at which the calculated free energy difference between the native state (N) and the unfolded state (U), $\Delta G_{NU}(T_s)$ ($G_N(T_s) - G_U(T_s)$), and the measured free energy $\Delta G_{NU}(T_E)$ at T_E (=298 K) coincide. For PDZ2, $\Delta G_{NU}(T_E = 298\ K) = -3.1$ kcal/mol at $[C]$ = 0 M,[28] which results in T_s = 317 K. Besides T_s, no other parameter is parametrized to obtain agreement with experiments for any property.

With T_s = 317 K fixed, we calculated the dependence of $f_{NBA}([C], T_s)$, $f_{UBA}([C], T_s)$, and $f_{IBA}([C], T_s)$ on $[C]$ (Figure 2a). The agreement between the measured and simulated results for $f_{NBA}([C], T_s)$ as a function of $[C]$ is excellent (Figure 2a). We find, just as in thermal denaturation, that $f_{IBA}([C], T_s)$ is small.[27,28] The midpoint concentration, C_m, obtained using $f_{NBA}([C_m], T_s) = 0.5$ is $[C]$ = 2.3 M, agrees well with the experimentally measured value of 2.6 M (see Figure 6D in ref 28).

The native state stability, $\Delta G_{NU}([C])(=G_N([C]) - G_U([C]))$, is computed using $\Delta G_{NU}([C]) = -k_B T_s \ln\left(\frac{f_{NBA}}{f_{UBA}}\right)$. The linear fit, $\Delta G_{NU}([C]) = \Delta G_{NU}(0)+m[C]$, yields $\Delta G_{NU}([0]) = -3.09$ kcal/mol and m = 1.35 kcal/mol/M (Figure 2b). The experimentally inferred m = 1.20 kcal/mol/M compares well with the simulations, which establishes again that simulations based on MTM predict the thermodynamic properties of proteins accurately. The $[C]$-dependent heat capacity curves (Figure 2c) show that the peaks corresponding to $T_m([C])$ decrease as $[C]$ increases (Figure 2c,d). We conclude that the simulations show that the equilibrium folding induced by temperature or denaturants is cooperative.

Urea-Dependent Changes in the Shape of PDZ2. The dependence of the radius of gyration, $\langle R_g([C])\rangle$, on urea concentration (black line in Figure 3a) shows a transition from an expanded to a collapsed state as $[C]$ decreases. The radii of

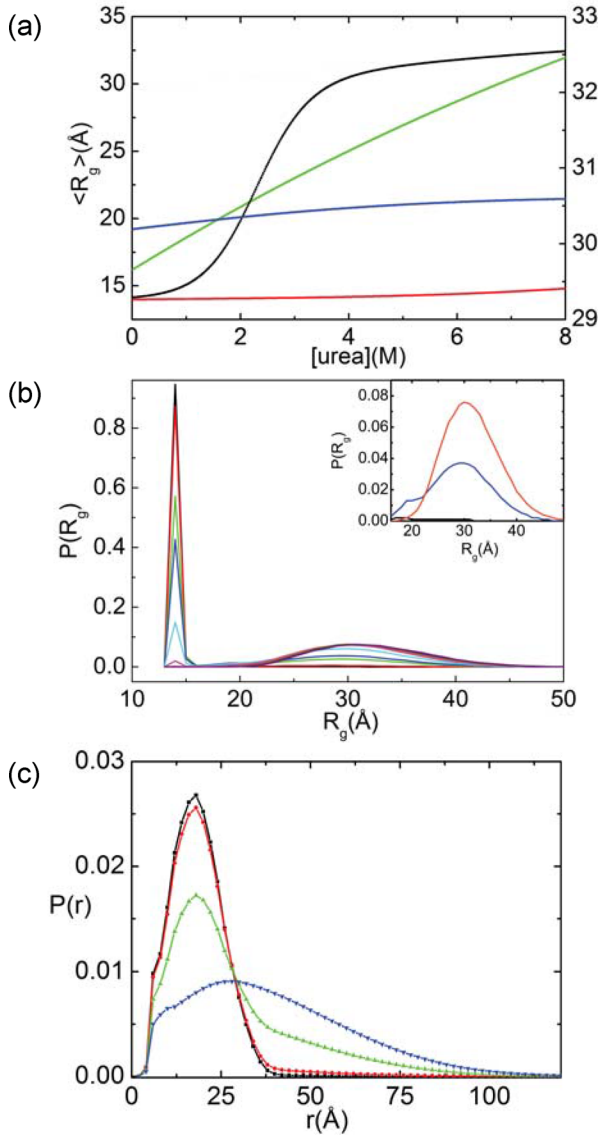


Figure 3. Equilibrium collapse. (a) Average $\langle R_g\rangle$ (black) as a function of $[C]$. Red, green, and blue curves correspond to $\langle R_g\rangle$ of the folded, unfolded, and I_{EQ} states, respectively. The scale for the unfolded state is on the right. (b) Distribution $P(R_g)$ of R_g for various concentrations of urea. The inset shows $P(R_g)$ for $[C]$ = 0 M (black), $[C]$ = 2.3 M (blue), and $[C]$ = 5.0 M (orange) corresponding to the extended conformations (R_g > 16 Å). (c) Distance distribution function $P(r)$, the inverse Fourier transform of the scattering intensity, for 0 M (black), 1.0 M (red), 2.3 M (green), and 5.0 M (blue) urea. Here, r is the distance between all noncovalently linked beads.

gyration of the I_{EQ} and the native state are virtually constant at all urea concentrations. However, the radius of gyration of the UBA decreases as $[C]$ decreases implying that the ensemble of structures of the unfolded state is more compact under native conditions than at 8 M urea. The decrease in the R_g of the UBA structures in going from 8 M urea to 1 M is about 9%, which should be measurable in a high precision small angle X-ray scattering (SAXS) experiment (see below for additional discussion). The $P(R_g)$ distributions at various urea concentrations show the expected behavior (Figure 3b). The protein is largely in the NBA at $[C]$ less than 2.3 M (mid point of the folding transition) and is expanded at higher values of $[C]$. The distance distributions, which can be measured as the inverse

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

Fourier transform of the wave-vector-dependent scattering intensity, are plotted in Figure 3c. These results constitute one of the predictions of this work.

Free Energy Profiles, $G(\chi)$, Reveal Lowly Populated Intermediate. To illustrate how urea changes the folding landscape, we plotted $G(\chi)$ versus χ at different $[C]$ at $T = T_m$ = 324 K in Figure 4a and at $T = T_s$ = 317 K in Figure 4b. Figure

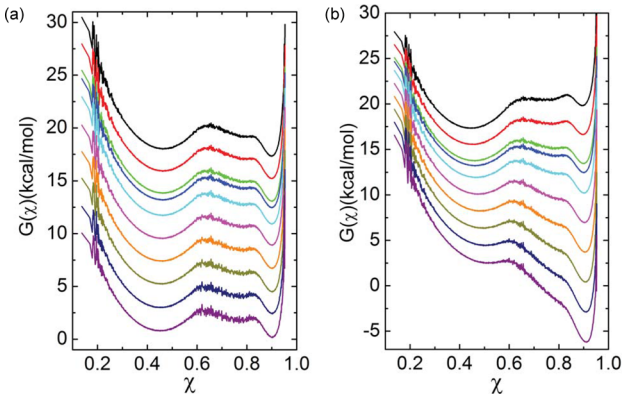


Figure 4. Free energy profiles versus χ at different $[C]$. (a) $T = T_m$. (b) $T = T_s$. The values of $[C]$ measured in M from top to bottom are 0, 1, 2, 2.3, 3, 4, 5, 6, 7, and 8.

4a shows that, at all urea concentrations, the conformations could be partitioned into three states, which is consistent with the results in Figure 1c displaying $G(\chi)$ at T_m with $[C]$ = 0 M. The basin of attraction corresponding to I_{EQ} becomes deeper as $[C]$ increases but is shallow compared to the NBA and UBA. The barrier for TSE1 is lower than for TSE2, implying that the transition from the intermediate states to the UBA is more facile than transition to the NBA. The number of distinct minima is unchanged at the lower temperature, T_s = 317 K (Figure 4b). However, the basin for the intermediate states becomes shallower as $[C]$ increases, and the barrier for TSE1 almost disappears, especially when $[C] > C_m$ = 2.3 M, indicating that the intermediate, if it can be detected at all, is unstable. The absence of I_{EQ} at the lower temperature suggests that stabilization of the native fold of PDZ2 leads to a two-state thermodynamic transition. Our finding supports the same observation in experiments showing that native state stabilization upon addition of sodium sulfate results in the high energy I_{EQ} being undetectable.

Structures of the Transition State Ensembles (TSE1 and TSE2). We identified the transition state structures using the folding trajectories generated at T_m. We picked the putative transition state from the barrier regions of the free energy profile as a function of χ, using the conditions $0.815 < \chi < 0.835$ for TSE1 and $0.63 < \chi < 0.65$ for TSE2, which are represented as shaded areas in Figure 1c. Starting from these structures, we calculated the commitment probability, P_{fold}, of reaching the NBA.[49] The sets of structures with $0.4 \leqslant P_{fold} \leqslant 0.6$ are identified as the transition state ensemble. The characteristics of the two transition state ensembles are displayed in Figure 5a, and the distribution of the structural order parameter $P(\chi)$ of the TSE1 along with the contact map is shown in Figure 5b,c, respectively. Distribution of $P(\chi)$ and the structural details for TSE2 are given in Figure 5d,e, respectively.

The characteristics of the TSEs are experimentally described using the Tanford β parameter, which is related to the buried SASA in the transition state. By fitting the measured Chevron plots as linear function of $[C]$, it has been shown that $\beta_E^1 = 0.53$ for TSE1, and $\beta_E^2 = 0.89$ for TSE2.[27] In a subsequent paper[31] it was reported that $\beta_E^1 = 0.35$ for TSE1 and $\beta_E^2 = 0.85$ for TSE2.[31] For the simulated TSEs, we calculated the distribution $P(\Delta_R)$ (see Figure 5a), where $\Delta_R = (\Delta_U - \Delta_{TSE})/(\Delta_U - \Delta_N)$ with Δ_U, Δ_{TSE}, Δ_N being the SASAs in the DSE ($[C]$ = 8.0 M), TSE, and the NBA ($[C]$ = 0.0 M), respectively. We found that the average $\langle\Delta_R\rangle = 0.23 = \beta_s^1$ for TSE1 and $\langle\Delta_R\rangle = 0.68 = \beta_s^2$ for TSE2, which qualitatively agree with the values inferred from experiments. The small value of β_1 suggests that the TSE1 structures are more UBA-like whereas the TSE2 structures resemble the native state.

The structural details are revealed in Figure 5c, which shows the contact map for TSE1. The extent of structure formation in TSE1 (Figure 5c) is modest (green) to low (blue). The contacts between β strands 1−6−4 have low formation probabilities as indicated by the two black circled regions. The secondary structures, β_{23}, and the two helices, are ordered to a greater extent. A representative structure for TSE1 displayed on the left of Figure 5a shows that the structure is expanded. Only β_{23} and the two helices are formed with high probabilities.

The contact map for TSE2 (Figure 5e) shows that the formation probabilities of contacts even between residues that are distant in sequence are high, resulting in the ensemble of TSE2 structures being compact. The major blue region in the contact map indicates that β_{16} (a β sheet formed between β_1 and β_6) is still largely unstructured. Four superimposed representative structures from TSE2 are shown on the right of Figure 5a. The structures are native-like except that the strand β_1 is not as well-packed as in the native state. The lack of stabilizing interactions in β_{16} found in our simulations disagrees with the inferences from the all atom molecular dynamics simulations using measured Φ values as constraints[29] (see below for additional discussion).

Folding and Collapse Kinetics. We calculated the collapse and folding rates at zero urea concentration from the folding trajectories, generated using Brownian dynamics simulations.[41] From 93 folding trajectories, the fraction of unfolded molecules at time t is computed using $P_u(t) = 1 - \int_0^t P_{fp}(s)\,ds$, where $P_{fp}(s)$ is the distribution of first passage times. We fit $P_u(t) = e^{-tk_f}$ (Figure 6a) with k_f = 172 s^{-1}, a value that is about 60 times larger than found in the experiments (2.8 s^{-1}).[27] The kinetics of collapse of the PDZ2 domain, shows that $\langle R_g(t)\rangle$ decays with a single rate constant, $k_c([C])$, the rate of collapse. The extracted values of $k_c([C])$ = 244 s^{-1} from the data in Figure 6b are greater than $k_f([C])$, which shows compaction occurs ahead of folding.

Thermodynamic and Kinetic Intermediates. Analyses of the folding trajectories show that the folded state is reached through a kinetic intermediate state, I_1 (Figure 7a−d). This intermediate state, while not directly detected in experiment, is likely responsible for the downward curvature observed in the unfolding arm of the chevon plots.[27] We also find the presence of other kinetic intermediate states (I_2) in some of the trajectories. A quantitative analysis allows us to classify the occurrences of I_1 and I_2 in the 93 folding trajectories. In 49 trajectories both I_1 and I_2 are found. Two trajectories illustrating this finding are shown in Figure 7a,b. In Figure 7a, PDZ2 samples I_1 and I_2 only once before reaching the native state, whereas in Figure 7b, it samples I_1 and I_2 more than once. In the second class of trajectories (44 out of 93),

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

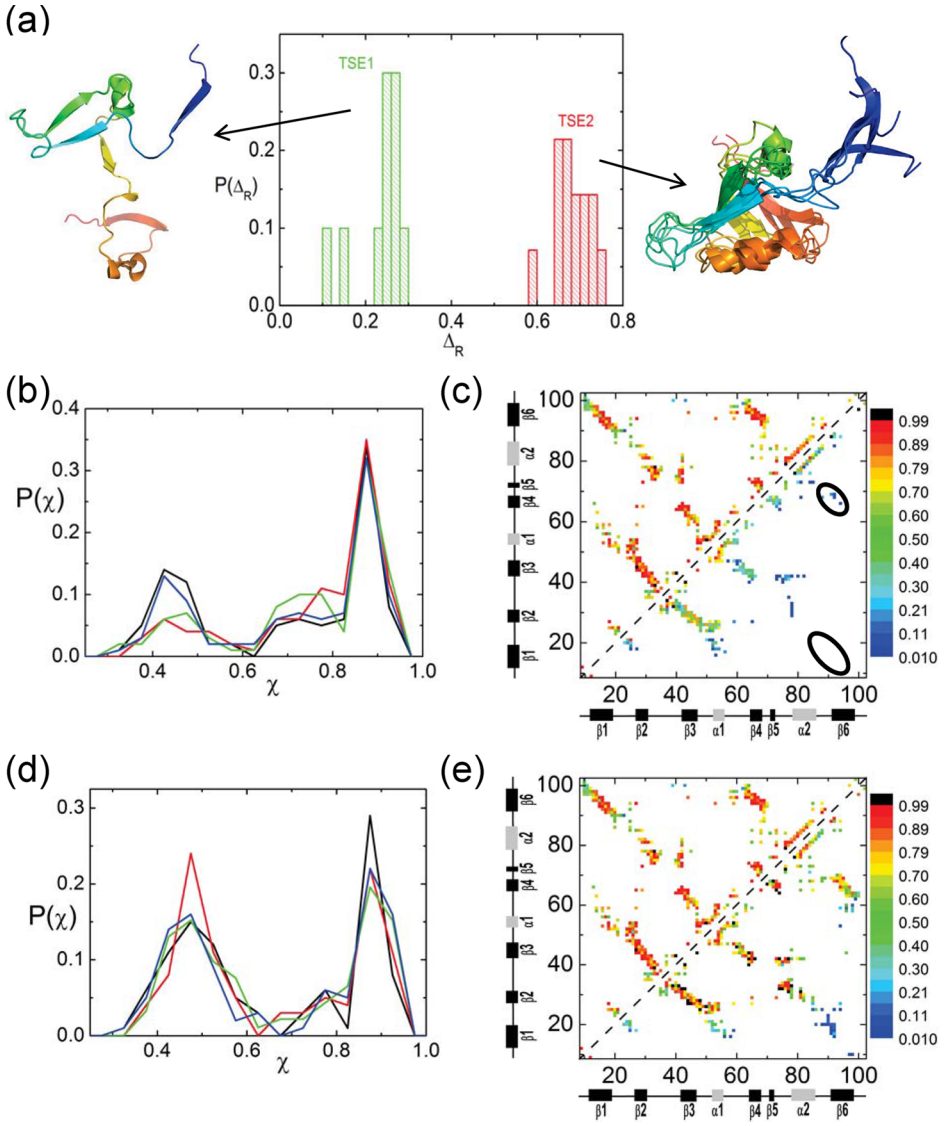


Figure 5. Quantifying the transition state ensembles. (a) Distribution $P(\Delta_R)$ of the $\Delta_R = (\Delta_U - \Delta_{TSE})/(\Delta_U - \Delta_N)$, which is the fraction of buried solvent accessible surface area relative to the unfolded structures. The average $\langle\Delta_R\rangle = 0.23$ for TSE1, and $\langle\Delta_R\rangle = 0.68$ for TSE2. These values coincide qualitatively with Tanford β parameters extracted from the observed Chevron plot. A few of the TSE1 (TSE2) structures are displayed on the left (right), respectively. (b) The distributions of χ computed from 400 simulation trajectories spawned from the transition state structures in TSE1. Data are shown for the four different structures. The distribution shows that roughly half of these trajectories go to the folded basin of attraction ($P_{fold} \approx 0.5$). (c) Contact map of the native state ensemble (upper left) and the one for the TSE1 (lower right). The scale on the right gives the probability of contact formation. (d and e) Same as parts b and c, respectively, except the results are for TSE2.

PDZ2 samples only the I_1 (see Figure 7c,d for examples). Structures of I_1 and I_2 are displayed in Figure 7e.

I_{EQ} and I_1 Are Structurally Similar. To illustrate the structural similarity between I_{EQ}, identified in the free energy profiles (Figures 1 and 4), I_1 in 100% of the kinetic folding trajectories, and I_2 sampled in some of the folding trajectories, we computed the average fraction of native contacts formed by every residue, f_Q, for the three states (Figure 8a,b). The correlation between I_{EQ} and I_1, shown in Figure 8c, is very high ($R = 0.995$). Therefore, we surmise that I_1 and I_{EQ} are structurally identical. Quantitative results in Figure 8b,d and sample structures (Figure 8e) show that the structure of I_2 differs from I_{EQ}. These results show that both thermodynamic and kinetic intermediates can be sampled during the folding process although the former is far more prevalent.

DISCUSSION

Minimum Energy Compact Structures (MECS). The folding trajectories reveal that two major intermediates are sampled as PDZ2 folds. The equilibrium intermediate is found in all the trajectories whereas the kinetic intermediate is found in only a fraction of the trajectories. Two aspects of these findings, which are of general validity for folding, are the following: (i) Both I_1 and I_2 form on the time scale of collapse. In those trajectories in which I_2 forms there frequent transitions between I_2 and I_1 (see Figure 7). In the process of making such transitions, PDZ2 undergoes considerable expansion. (ii) The intermediates, I_{EQ} and I_1, are compact and contain native-like features. The major difference between I_1 and the folded state is in the extent of structure involving the strands β_1, β_4, and β_6 as well as the helix α_2. The structure of I_1 is similar to that found in TSE1, which follows from the Hammond postulate. These

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

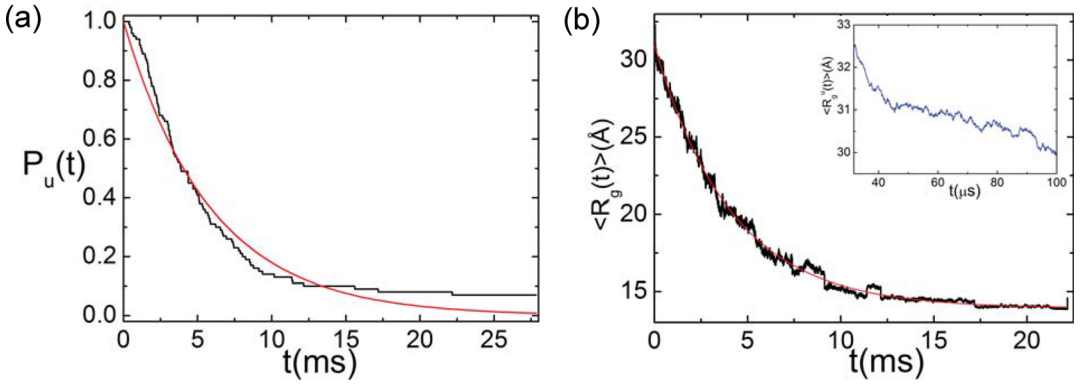


Figure 6. Folding and collapse kinetics. (a) Fraction of molecules that have not folded ($[C]$ = 0 M) as a function of time. The line is an exponential fit ($P_u(t) = \exp(-(t/\tau_F))$ with τ_F = 5.8 ms) to the data. (b) Collapse kinetics monitored by the time-dependent average $\langle R_g(t)\rangle$ as a function of t with the line giving an exponential fit to the data. The collapse time of PDZ2 is τ_c = 4.1 ms.

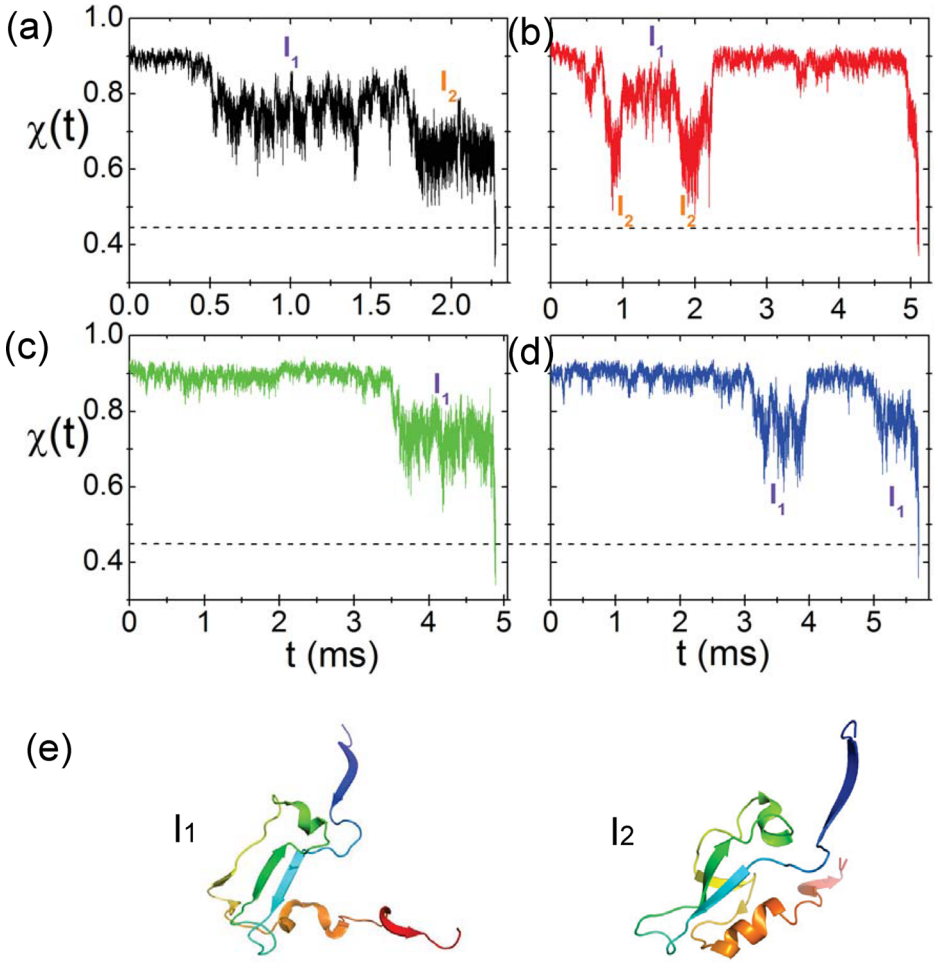


Figure 7. Two major folding pathways monitored by χ as a function of t. (a and b) Two representative trajectories showing that the native state is reached by sampling both I_1 and I_2. (c and d) Two folding trajectories in which the polypeptide chain samples only I_1, often multiple times, before reaching the folded state. (e) Structures for I1 and I2.

intermediates, which are like MECS (minimum energy compact structures),[50] facilitate rapid folding. Although they are difficult to detect in ensemble experiments, single molecule pulling experiments using cycles of force increase, and force quench can be used to detect MECS, as has been done for ubiquitin.[51] It would be interesting to do similar experiments on PDZ2 to directly observe I_1 and I_2 using force as a perturbation.

Collapse and Folding Transition. Is the size of the polypeptide chain of the unfolded state under folding conditions ($[C] < C_m$) less than that at $[C] > C_m$? Theoretical arguments,[47,52,53] simulations,[1] and a number of single molecule FRET and SAXS experiments[19,54] have answered the question in the affirmative whereas some SAXS experiments on protein L suggest that there is no evidence for polypeptide chain collapse, which is manifestly unphysical. The arguments

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

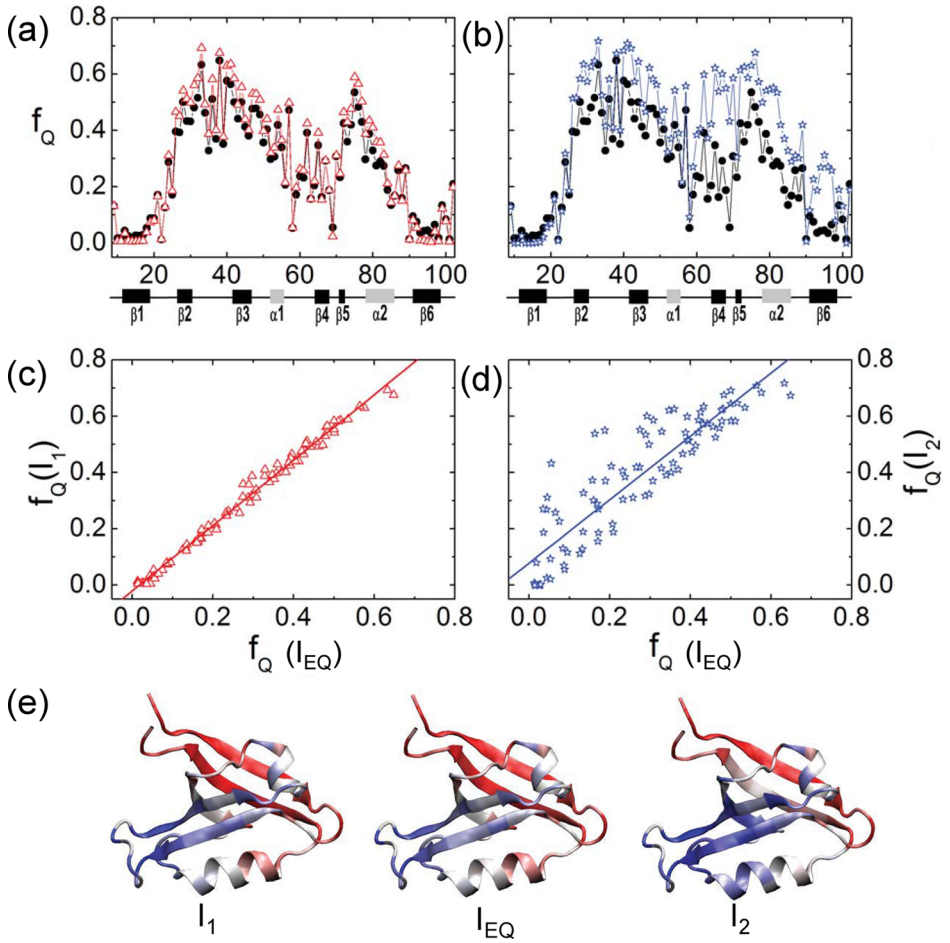


Figure 8. Comparison of I_{EQ}, I_1, and I_2. (a) Average fraction of native contacts formed for residues, f_Q, for I_{EQ} (black) and I_1 (red). (b) f_Q for I_{EQ} (black) and I_2 (blue). (c) Correlation between f_Q's for I_{EQ} and I_1. The correlation coefficient is near unity. (d) Relation between f_Q's for I_{EQ} and I_2 with correlation coefficient ≈0.9. (e) Structures of the three intermediates.

in favor of collapse preceding folding are based on the following observations. The random coil state of a polypeptide chain with N residues is expected to be $R_g^U \sim a_R N^{0.6}$ with $a_R \approx 2.0$ Å. This estimate is likely to be an upper bound because 8 M urea is not a good solvent because even at these elevated concentrations the unfolded state has residual structure. As the denaturant concentration decreases, the maximum decrease in R_g is likely to have a lower bound $a_D N^{0.5}$. It cannot be maximally compact because if it were so then enthalpic interactions would drive these structures to the folded state for which $R_g \approx a_N N^{1/3}$ with $a_N = 3.0$ Å.[39] Thus, we surmise that $a_D N^{0.5} < R_g^U < a_R N^{0.6}$. The reduction in R_g of the unfolded state for $N = 64$ (protein L) is predicted to be between 5% and 10% depending on the values of a_R and a_D. In PDZ2 ($N = 94$) we find that the unfolded state R_g decreases by only 9% as $[C]$ decreases (Figure 3a). Thus, for small N only a modest decrease in the unfolded state R_g is expected. Detection of small changes in R_g^U requires high precision SAXS measurements as $[C]$ decreases below C_m. The errors in SAXS for protein L are far too large to accurately estimate R_g^U, especially under native conditions.

The absence of detectable contracted form of the polypeptide chain in time-resolved SAXS experiments on protein L was used as further evidence that compact states are not formed in the folding of any single domain protein. Using theory[55−57] it can be shown that collapse of the unfolded state occurs on time scales on the order of $\tau_c \approx \tau_0 N^{\beta}$ where β is between 1.5 and 2.2, and $\tau_0 \approx O(10^{-9})$ s. For PDZ2 we estimate that $\tau_c \sim 20$ μs. Our simulations show that the contraction of the unfolded state occurs on a time scale that does not exceed a maximum of ≈50 μs (see the inset in Figure 6b). Theoretical estimate based on the scaling law given above for collapse of the unfolded protein L ($N = 64$) yields a maximum value of $\tau_c \sim 4$ μs. Reconfiguration time extracted using fluorescence correlation spectroscopy[58,59] indicates that τ_c for small proteins could be less by an order of magnitude compared to the theoretical estimate. A larger value of 20 μs, still much shorter than the folding time, for reconfiguration time has also been reported for protein L.[60] For the larger DHFR ($N = 154$) there is nearly a 23% reduction in R_g^U in ∼300 μs,[61] which is also considerably shorter than the folding time. All these studies show that the time scale for collapse of the unfolded state, which clearly increases with N, is on the order of at most tens of microseconds. Thus, we conclude that the time resolution in the most recent SAXS experiments on protein L (4 ms),[62] which is comparable to the folding time, is too long to shed light on compaction of the polypeptide chain. The presence of 15 His tags in the first study[63] makes it difficult to ascertain if the results are relevant for protein L. Indeed, the inability to accurately determine the characteristics of the unfolded state also leads to erroneous conclusions about equilibrium collapse and kinetic foldability,[64] which has been corrected recently using smFRET experiments.[65]

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

Fast mixing experiments that simultaneously detect compaction and acquisition of structure on a number of proteins[66−68] have produced ample evidence that collapse is an integral process of the folding process, as predicted by theory. Although it is likely that the simplified analysis of smFRET measurements overestimates the extent of collapse,[69] fast mixing SAXS experiments on a variety of proteins leave no doubt that the unfolded state is indeed more compact under native conditions despite persistent claims to the contrary based on SAXS data based largely on one protein (protein L) with large errors. It is a pity that the erroneous conclusion has resulted in unnecessary obfuscation. What is needed are high precision data for single domain proteins spanning a range of N (say between 50 and 250), which cannot be easily obtained by SAXS alone[70] but is more readily available in smFRET experiments.

Detection of I_{EQ} in PDZ2 in Single Molecule Pulling Experiments. The downward curvature in the unfolding rate as a function $[C]$ in PDZ2 implies that an intermediate is populated.[27] Single molecule pulling experiments are well-suited to explore this finding more readily. On the basis of the free energy profile computed here and postulated elsewhere,[31] we suggest that unfolding by mechanical force (f) would give rise to downward curvature in a plot of log $k_u(f)$ versus f. At low forces the inner barrier separating the NBA and I_{EQ} would dominate whereas at high forces the second outer barrier is relevant for mechanical unfolding. The two sequential barrier picture implies that there would be two transition state distances with a switch between the two occurring as f is increased. A plausible support for this argument comes from the observation that the Tanford β_T as a function of $[C]$ exhibits a sigmoidal behavior (see Figure 4 in ref 31) with β_T changing from 0.35 at low $[C]$ to 0.85 at high $[C]$. The scenario postulated here is distinct from that in the SH3 domain in which the $[\log k_u(f), f]$ plot exhibits an upward curvature. This finding is a signature of parallel unfolding pathways with a switch between the transition state ensembles[71] as opposed to the predicted sequential barrier model for PDZ2. Laser optical tweezer experiments are ideally suited to test our prediction.

Structures of the TSEs. It is interesting to compare the structures of TSE1 and TSE2 obtained in this work with those reported earlier.[29] The structure of TSE1 (Figure 5) shows interactions involving the β_2, β_3 strands, and formation of the two helices. This is in sharp contrast to the conclusion in ref 29 suggesting that in TSE1 β strands 1−6−4 are structured whereas the other secondary structural elements are essentially disordered. Although the agreement between the predicted structures in our work and those reported in ref 29 for TSE2 is better in that we find that it is native-like, there are crucial differences as well. In particular, the lack of stabilizing interactions involving the strands β_1 and β_6 found in our simulations disagrees with the inferences drawn from the all atom molecular dynamics simulations using measured Φ values as constraints.[29] The differences are likely to be related to the completely different approaches used in the two studies. In ref 29, the measured Φ values were used as constraints in standard all atom molecular dynamics simulations, which use inaccurate force fields. The procedure, while interesting, is not systematic in the sense the accumulation of errors both from the Φ values as well as the MD simulations is nearly impossible to quantify. More importantly, the putative TSE structures were not used to obtain the P_{fold} values in the earlier study,[29] which casts doubt on whether the identified TSEs accurately represent the actual TSEs. We believe additional experiments, including perhaps double mutant cycles, would be needed to ascertain the nature of the TSEs in PDZ2.

It has been shown[72] that if the connectivity of secondary structures is altered to produce a symmetric version of PDZ2, then the TSE2 ensemble is remains roughly unchanged. However, TSE1 is altered suggesting that the early folding pathways are malleable. Although these findings are consistent with general theoretical arguments,[73] it would be most interesting to use our model to produce a quantitative picture of the symmetric variant of PDZ2.

CONCLUDING REMARKS

We have used a phenomenological theory based on the MTM to simulate the folding of PDZ2 domain as a function of temperature and urea. In addition to providing support to the folding mechanism discovered in experiments,[27] we have made a few testable predictions. (1) We have obtained a precise dependence of the melting point as a function of urea concentration, which can be tested using standard calorimetry experiments. (2) The presence of a high energy intermediate in the absence of added salt can be characterized using single molecule pulling experiments as shown using simulations for the srcSH3 domain[74] demonstrating that the excited state is sparsely (≈2−5)% populated, which coincided with the findings based on NMR experiments.[75]

AUTHOR INFORMATION

Corresponding Author

*E-mail: zxliu@bnu.edu.cn.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to Ben Schuler and Gilad Haran for their interest and useful comments. This work was supported in part by a grant from the National Science Foundation to D.T. through CHE 13-61946. Z.L. acknowledges partial financial support from the National Natural Science Foundation of China (11104015) and the Fundamental Research Funds for the Central Universities (2012LYB08). G.R. acknowledges startup grant from Indian Institute of Science-Bangalore, and funding from Nano mission, Department of Science and Technology, India.

REFERENCES

(1) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13403−13408.

(2) Liu, Z. X.; Reddy, G.; Thirumalai, D. Theory of the Molecular Transfer Model for Proteins with Applications to the Folding of the src-SH3 Domain. *J. Phys. Chem. B* **2012**, *116*, 6707−6716.

(3) Reddy, G.; Liu, Z. X.; Thirumalai, D. Denaturant-dependent folding of GFP. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17832−17838.

(4) Reddy, G.; Thirumalai, D. Dissecting Ubiquitin Folding Using the Self-Organized Polymer Model. *J. Phys. Chem. B* **2015**, *119*, 11358−11370.

(5) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. Effects of pH on Proteins: Predictions for Ensemble and Single-Molecule Pulling Experiments. *J. Am. Chem. Soc.* **2012**, *134*, 979−987.

(6) Chen, T.; Chan, H. S. Effects of desolvation barriers and sidechains on local-nonlocal coupling and chevron behaviors in coarse-grained models of protein folding. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6460−6479.

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

(7) Liu, Z. X.; Reddy, G.; O'Brien, E. P.; Thirumalai, D. Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 7787−7792.
(8) Tozzini, V. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144−150.
(9) Hyeon, C.; Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2011**, *2*, 487.
(10) Elber, R.; Cardenas, A. E. In *Comprehensive Biophysics, Vol 9: Simulation and Modeling*; Egelman, E. H., Ed.; Academic Press, 2011; pp 2−26.
(11) Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. Biomolecular dynamics: order-disorder transitions and energy landscapes. *Rep. Prog. Phys.* **2012**, *75*, 076601.
(12) Thirumalai, D.; Liu, Z.; O'Brien, E. P.; Reddy, G. Protein folding: from theory to practice. *Curr. Opin. Struct. Biol.* **2013**, *23*, 22−29.
(13) Kim, P. S.; Baldwin, R. L. Intermediates in the Folding Reactions of Small Proteins. *Annu. Rev. Biochem.* **1990**, *59*, 631−660.
(14) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the Folding Routes. *Science* **1995**, *267*, 1619−1620.
(15) Dill, K. A.; Chan, H. S. From Levinthal to Pathways to Funnels. *Nat. Struct. Biol.* **1997**, *4*, 10−19.
(16) Daggett, V.; Fersht, A. R. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **2003**, *28*, 18−25.
(17) Onuchic, J. N.; Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70−75.
(18) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys.* **2008**, *37*, 289−316.
(19) Schuler, B.; Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16−26.
(20) Shakhnovich, E. Protein Folding Thermodynamics and Dynamics: Where Physics, Chemistry, and Biology Meet. *Chem. Rev.* **2006**, *106*, 1559−1588.
(21) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. Theoretical Perspectives on Protein Folding. *Annu. Rev. Biophys.* **2010**, *39*, 159−183.
(22) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042−1046.
(23) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98−105.
(24) Harris, B. Z.; Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* **2001**, *114*, 3219−3231.
(25) Kim, E. J.; Sheng, M. PDZ domain proteins of synapses. *Nat. Rev. Neurosci.* **2004**, *5*, 771−781.
(26) Jemth, P.; Gianni, S. PDZ domains: Folding and binding. *Biochemistry* **2007**, *46*, 8701−8708.
(27) Gianni, S.; Calosci, N.; Aelen, J. M. A.; Vuister, G. W.; Brunori, M.; Travaglini-Allocatelli, C. Kinetic folding mechanism of PDZ2 from PTP-BL. *Protein Eng., Des. Sel.* **2005**, *18*, 389−395.
(28) Sicorello, A.; Torrassa, S.; Soldi, G.; Gianni, S.; Travaglini-Allocatelli, C.; Taddei, N.; Relini, A.; Chiti, F. Agitation and High Ionic Strength Induce Amyloidogenesis of a Folded PDZ Domain in Native Conditions. *Biophys. J.* **2009**, *96*, 2289−2298.
(29) Gianni, S.; Geierhaas, C. D.; Calosci, N.; Jemth, P.; Vuister, G. W.; Travaglini-Allocatelli, C.; Vendruscolo, M.; Brunori, M. A PDZ domain recapitulates a unifying mechanism for protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 128−133.
(30) Pincus, D. L.; Cho, S. S.; Hyeon, C. B.; Thirumalai, D. Minimal Models for Proteins and RNA: From Folding to Function. *Prog. Mol. Biol. Transl. Sci.* **2008**, *84*, 203−250.
(31) Ivarsson, Y.; Travaglini-Allocatelli, C.; Jemth, P.; Malatesta, F.; Brunori, M.; Gianni, S. An on-pathway intermediate in the folding of a PDZ domain. *J. Biol. Chem.* **2007**, *282*, 8568−8572.
(32) Hyeon, C.; Dima, R. I.; Thirumalai, D. Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure* **2006**, *14*, 1633−1645.
(33) Camacho, C. J.; Thirumalai, D. Modelling Disulfide Bonds in Globular Proteins: Entropic Barriers and Pathways. *Proteins: Struct., Funct., Genet.* **1995**, *22*, 27−40.
(34) Klimov, D. K.; Thirumalai, D. Multiple protein folding nuclei and the transition state ensemble in two-state proteins. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 465−475.
(35) Gin, B. C.; Garrahan, J. P.; Geissler, P. L. The Limited Role of Nonnative Contacts in the Folding Pathways of a Lattice Protein. *J. Mol. Biol.* **2009**, *392*, 1303−1314.
(36) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874−17879.
(37) Betancourt, M.; Thirumalai, D. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* **1999**, *8*, 361−369.
(38) Honeycutt, J. D.; Thirumalai, D. The Nature of Folded States of Globular-Proteins. *Biopolymers* **1992**, *32*, 695−709.
(39) Dima, R. I.; Thirumalai, D. Asymmetry in the shapes of folded and denatured states of proteins. *J. Phys. Chem. B* **2004**, *108*, 6564−6570.
(40) Ferrenberg, A. M.; Swendsen, R. H. New Monte Carlo Technique for Studying Phase Transitions. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.
(41) Veitshans, T.; Klimov, D.; Thirumalai, D. Protein folding kinetics: Timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding Des.* **1997**, *2*, 1−22.
(42) Hansmann, U.; Okamoto, Y.; Eisenmenger, F. Molecular Dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem. Phys. Lett.* **1996**, *259*, 321−330.
(43) Okamoto, Y.; Hansmann, U. Thermodynamics of helix-coil transitions studied by multicanonical algorithms. *J. Phys. Chem.* **1995**, *99*, 11276−11287.
(44) Zhou, R. H.; Berne, B. J.; Germain, R. The free energy landscape for beta hairpin folding in explicit water. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 14931−14936.
(45) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(46) Ermak, D. L.; McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **1978**, *69*, 1352−1369.
(47) Camacho, C. J.; Thirumalai, D. Kinetics and Thermodynamics of Folding in Model Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6369−6372.
(48) Ivarsson, Y.; Travaglini-Allocatelli, C.; Morea, V.; Brunori, M.; Gianni, S. The folding pathway of an engineered circularly permuted PDZ domain. *Protein Eng., Des. Sel.* **2008**, *21*, 155−160.
(49) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. I. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998**, *108*, 334.
(50) Camacho, C. J.; Thirumalai, D. Minimum Energy Compact Structures of Random Sequences of Heteropolymers. *Phys. Rev. Lett.* **1993**, *71*, 2505−2508.
(51) Garcia-Manyes, S.; Dougan, L.; Badilla, C.; Brujic, J.; Fernandez, J. M. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 10534−10539.
(52) Ziv, G.; Haran, G. Protein Folding, Protein Collapse, and Tanford's Transfer Model: Lessons from Single-Molecule FRET. *J. Am. Chem. Soc.* **2009**, *131*, 2942−2947.
(53) Ziv, G.; Thirumalai, D.; Haran, G. Collapse transition in proteins. *Phys. Chem. Chem. Phys.* **2009**, *11*, 83−93.
(54) Haran, G. How, when and why proteins collapse: the relation to folding. *Curr. Opin. Struct. Biol.* **2012**, *22*, 14−20.
(55) de Gennes, P. G. Kinetics of collapse of a flexible coil. *J. Phys., Lett.* **1985**, *46*, L639−L642.

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101

(56) Pitard, E.; Orland, H. Dynamics of the swelling or collapse of a homopolymer. *Europhys. Lett.* **1998**, *41*, 467−472.
(57) Thirumalai, D. From Minimal Models to Real Proteins: Time Scales for Protein Folding Kinetics. *J. Phys. I* **1995**, *5*, 1457−1467.
(58) Nettels, D.; Gopich, I. V.; Hoffmann, A.; Schuler, B. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2655−2660.
(59) Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R.; Mueller-Spaeth, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17800−17806.
(60) Waldauer, S. A.; Bakajin, O.; Lapidus, L. J. Extremely slow intramolecular diffusion in unfolded protein L. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 13713−13717.
(61) Arai, M.; Kondrashkina, E.; Kayatekin, C.; Matthews, C. R.; Iwakura, M.; Bilsel, O. Microsecond hydrophobic collapse in the folding of Escherichia coli dihydrofolate reductase, an alpha/beta-type protein. *J. Mol. Biol.* **2007**, *368*, 219−229.
(62) Yoo, T. Y.; Meisburger, S. P.; Hinshaw, J.; Pollack, L.; Haran, G.; Sosnick, T. R.; Plaxco, K. Small-Angle X-ray Scattering and Single-Molecule FRET Spectroscopy Produce Highly Divergent Views of the Low-Denaturant Unfolded State. *J. Mol. Biol.* **2012**, *418*, 226−236.
(63) Plaxco, K.; Millett, I.; Segel, D.; Doniach, S.; Baker, D. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat. Struct. Biol.* **1999**, *6*, 554−556.
(64) Millet, I.; Townsley, L.; Chiti, F.; Doniach, S.; Plaxco, K. Equilibrium collapse and the kinetic 'foldability' of proteins. *Biochemistry* **2002**, *41*, 321−325.
(65) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16155−16160.
(66) Chan, C.; Hu, Y.; Takahashi, S.; Rousseau, D.; Eaton, W.; Hofrichter, J. Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 1779−1784.
(67) Akiyama, S.; Takahashi, S.; Kimura, T.; Ishimori, K.; Morishima, I.; Nishikawa, Y.; Fujisawa, T. Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1329−1334.
(68) Kimura, T.; Uzawa, T.; Ishimori, K.; Morishima, I.; Takahashi, S.; Konno, T.; Akiyama, S.; Fujisawa, T. Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2748−2753.
(69) O'Brien, E. P.; Morrison, G.; Brooks, B. R.; Thirumalai, D. How accurate are polymer models in the analysis of Forster resonance energy transfer experiments on proteins? *J. Chem. Phys.* **2009**, *130*, 124903.
(70) Trewhella, J.; Hendrickson, W. A.; Kleywegt, G. J.; Sali, A.; Sato, M.; Schwede, T.; Svergun, D. I.; Tainer, J. A.; Westbrook, J.; Berman, H. M. Report of the wwPDB Small-Angle Scattering Task Force: Data Requirements for Biomolecular Modeling and the PDB. *Structure* **2013**, *21*, 875−881.
(71) Guinn, E. J.; Jagannathan, B.; Marqusee, S. Single-molecule chemo- mechanical unfolding reveals multiple transition state barriers in a small single- domain protein. *Nat. Commun.* **2015**, *6*, 6861.
(72) Ivarsson, Y.; Travaglini-Allocatelli, C.; Brunori, M.; Gianni, S. *J. Am. Chem. Soc.* **2009**, *131*, 11727−11733.
(73) Klimov, D. K.; Thirumalai, D. Symmetric Connectivity of Secondary Structure Elements Enhances the Diversity of Folding Pathways. *J. Mol. Biol.* **2005**, *353*, 1171−1186.
(74) Zhuravlev, P. I.; Reddy, G.; Straub, J. E.; Thirumalai, D. Propensity to Form Amyloid Fibrils Is Encoded as Excitations in the Free Energy Landscape of Monomeric Proteins. *J. Mol. Biol.* **2014**, *426*, 2653−2666.
(75) Neudecker, P.; Robustelli, P.; Cavalli, A.; Walsh, P.; Lundstrom, P.; Zarrine-Afsar, A.; Sharpe, S.; Vendruscolo, M.; Kay, L. E. Structure of an Intermediate State in Protein Folding and Aggregation. *Science* **2012**, *336*, 362−366.

DOI: 10.1021/acs.jpcb.6b00327
J. Phys. Chem. B 2016, 120, 8090−8101