JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

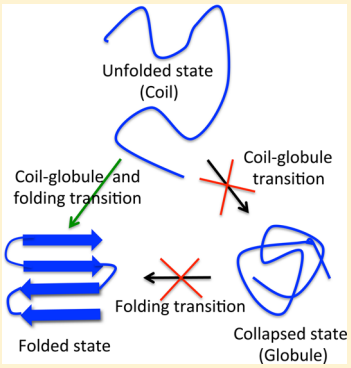
Folding of Protein L with Implications for Collapse in the Denatured State Ensemble

Hiranmay Maity and Govardhan Reddy*

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

Supporting Information

ABSTRACT: A fundamental question in protein folding is whether the coil to globule collapse transition occurs during the initial stages of folding (burst phase) or simultaneously with the protein folding transition. Single molecule fluorescence resonance energy transfer (FRET) and small-angle X-ray scattering (SAXS) experiments disagree on whether Protein L collapse transition occurs during the burst phase of folding. We study Protein L folding using a coarse-grained model and molecular dynamics simulations. The collapse transition in Protein L is found to be concomitant with the folding transition. In the burst phase of folding, we find that FRET experiments overestimate radius of gyration, R_g, of the protein due to the application of Gaussian polymer chain end-to-end distribution to extract R_g from the FRET efficiency. FRET experiments estimate ≈6 Å decrease in R_g when the actual decrease is ≈3 Å on guanidinium chloride denaturant dilution from 7.5 to 1 M, thereby suggesting pronounced compaction in the protein dimensions in the burst phase. The ≈3 Å decrease is close to the statistical uncertainties of the R_g data measured from SAXS experiments, which suggest no compaction, leading to a disagreement with the FRET experiments. The transition-state ensemble (TSE) structures in Protein L folding are globular and extensive in agreement with the Ψ-analysis experiments. The results support the hypothesis that the TSE of single domain proteins depends on protein topology and is not stabilized by local interactions alone.



INTRODUCTION

There is an ongoing debate[1−5] on whether the denatured ensemble of single domain proteins undergoes a coil to globule transition during the burst phase of folding as the denaturant concentration is diluted to lower values. Proteins are heteropolymers and behave like random coils at high temperatures or denaturant concentrations.[6−8] An interesting question is whether proteins akin to polymers undergo a collapse transition in the burst phase of folding as the conditions are made conducive for folding.[9−11] Single domain proteins unlike polymers are finite sized and are composed of a specific sequence of amino acids which are hydrophobic and hydrophilic in character. The finite size effects and heteropolymer character are the reasons attributed to the marginal stability of proteins and the near overlap of the collapse and folding transition temperatures, which makes them fold efficiently.[12,13]

The collapse transition in proteins is generally studied using single molecule fluorescence resonance energy transfer (FRET) and small-angle X-ray scattering (SAXS) experiments. Although FRET and SAXS experiments agree that proteins like Cytochrome c[14−17] and Monellin[18−20] collapse during the burst phase of folding, their results disagree for Protein L. FRET experiments for Protein L[21−23] infer collapse, whereas SAXS experiments[24,25] conclude no collapse in the burst phase of folding on dilution of guanidine hydrochloride [GuHCl]. Both FRET and SAXS estimate the radius of gyration, R_g, of the protein to infer the size of the protein in the unfolded ensemble. The difference in the R_g predictions of FRET[21,22] and SAXS[24,25] experiments for Protein L during the burst phase of folding is statistically significant. The reasons for the disagreement between these experiments for Protein L are not completely clear. Understanding the impact of various approximations used in these methods to estimate the size of the protein can not only aid in resolving the disagreement between FRET and SAXS but also to understand the problem of protein collapse better.

Single domain proteins close to the melting temperature or the midpoint denaturant concentration generally fold in a two-state manner through an ensemble of transition-state structures (TSE). ϕ-Analysis experiments for Protein L[26−29] predict that the TSE is polarized with only the N-terminal β-hairpin present. Whereas the ψ-analysis experiments[30] predict that the TSE is globular and extensive with both the N- and C-termini β-hairpins present along with some non-native interactions in the C-terminal β-hairpin. The ψ-analysis experiments support that the TSE and the folding pathways for the single domain proteins depend on the protein topology,[31] whereas the ϕ-analysis experiments conclude that the TSE is mostly stabilized by local interactions.[32]

Various aspects of Protein L folding such as folding pathways, transition-state structures, and properties of the unfolded ensemble are studied[33−40] using both coarse-grained

Received: October 28, 2015
Published: February 2, 2016

DOI: 10.1021/jacs.5b11300
J. Am. Chem. Soc. 2016, 138, 2609−2616

and atomistic simulations. In this manuscript, we study the burst-phase folding of Protein L to understand the origin of discrepancy between the FRET[21,22] and SAXS[24,25] experiments using the native-centric self-organized polymer model with side chains (SOP-SC)[41,42] and molecular dynamics simulations. The effect of [GuHCl] on Protein L conformations is taken into account using the molecular transfer model (MTM).[11,43]

The computed FRET efficiency, $\langle E \rangle$, for Protein L in the burst phase of folding is in quantitative agreement with the FRET experiments of Eaton et al.[22] and only in partial agreement with the experiments of Haran et al.[21] The FRET experiments are found to overestimate R_g compared to the actual values computed directly from the simulations owing to the use of the Gaussian polymer chain end-to-end distribution function to extract R_g from $\langle E \rangle$. The deviation between FRET-extracted and actual R_g increased with [GuHCl]. As a result, FRET experiments[22] estimate ≈6 Å decrease in R_g and infer protein collapse in the burst phase, when the actual decrease is ≈3 Å as [GuHCl] is diluted from 7.5 M to 1 M.

The equilibrium R_g computed as a function of [GuHCl] is in near quantitative agreement with the SAXS experiments.[25] The SAXS experiments[25] infer no protein collapse as the burst-phase R_g at [GuHCl] = 4.0 and 0.67 M is not statistically different. In the simulations, the burst-phase R_g at [GuHCl] = 4 and 1 M is 25.1 ± 4.0 and 23.8 ± 3.9 Å, respectively, a difference of ≈1.3 Å, which is well within the standard deviation $\sigma_{R_g} \approx 4$ Å. From this analysis, which is similar to the SAXS analysis, we can infer no collapse as the change in protein dimensions are not statistically significant, leading to a disagreement with the FRET experiments.

The TSE of Protein L at the melting temperature is inferred using P_{fold} calculations.[44] The TSE is found to be globular and extensive with both the N- and C-termini β-hairpins present resembling a topology similar to that of the folded structure. The results are in agreement with the Ψ-analysis experiments[30] and only in partial agreement with the ϕ-analysis experiments.[26−29] The inferred TSE supports the hypothesis that the transition-state structures of single domain homologous proteins are extensive and depend on the protein topology.[45]

METHODS

Self-Organized Polymer Side Chain (SOP-SC) Model. We used the SOP-SC model[41,42] in which each amino acid residue is represented by two beads. One bead is at the C_α position representing the backbone atoms, and the other bead is at the center of mass of the side chain representing the side chain atoms. The effective energy of a protein conformation in the SOP-SC model is a sum of bonded and nonbonded interactions. The bonded interactions (E_B) are present between a pair of connected beads. The nonbonded interactions are a sum of native (E_{NB}^N) and non-native (E_{NB}^{NN}) interactions (see Supporting Information (SI) for more details). The native interactions for protein L are identified using the crystal structure[46] (Protein Data Bank ID: 1HZ6) (Figure 1A). The number of residues in the crystal structure, $N_{res} = 64$. The native interactions between the beads representing the amino acid side chains interact via a residue dependent Betancourt−Thirumalai statistical potential.[47]

The coarse-grained force-field in the SOP-SC model for a protein conformation given by the coordinates $\{\mathbf{r}\}$ in the absence of denaturants, $[C] = 0$, is

$$E_{CG}(\{\mathbf{r}\}, 0) = E_B + E_{NB}^N + E_{NB}^{NN} \quad (1)$$

Description of the various energy terms in eq 1 and the parameters used in the energy function are given in the SI. These parameters are identical to the values previously used to successfully study the folding

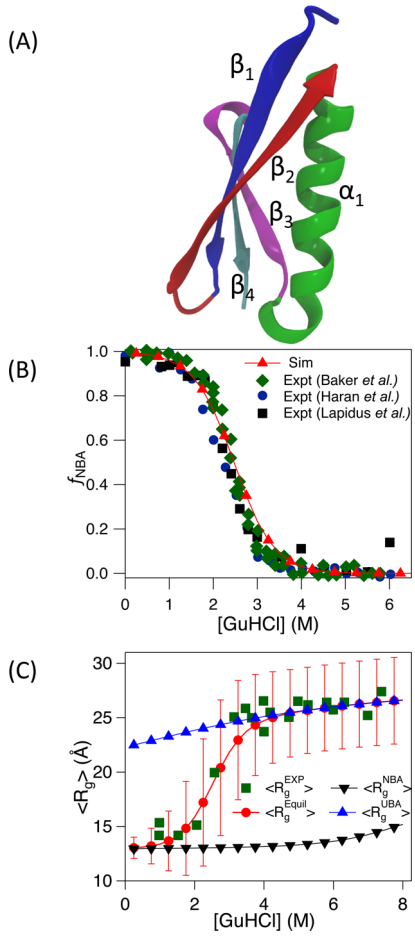


Figure 1. (A) Crystal structure of Protein L (PDB ID: 1HZ6). The α-helix is in green (α_1), and the four β-strands are in blue (β_1), red (β_2), magenta (β_3), and cyan (β_4). (B) The fraction of the protein in the native basin of attraction, f_{NBA}, as a function of [GuHCl]. Data in red triangles are from simulations. Data in blue circles, black squares, and green diamonds are from the experiments of Haran et al.,[59] Lapidus et al.,[23] and Baker et al.,[29] respectively. (C) The radius of gyration, R_g as a function of [GuHCl]. Data in red circles and green squares are from simulations and experiments,[25] respectively. $\langle R_g \rangle$ of UBA and NBA basins computed from simulations is shown in blue triangles and black inverted triangles, respectively.

properties of the proteins ubiquitin[48] and GFP.[49] We used the same force field to study the properties of different proteins, and as a result this force field satisfies the criterion of a transferable force field.

Molecular Transfer Model (MTM). To simulate Protein L folding thermodynamics and kinetics in the presence of [GuHCl], we used the MTM.[11,43] In the presence of a denaturant of concentration $[C]$, the effective coarse-grained force field for the protein using MTM is given by

$$E_{CG}(\{\mathbf{r}\}, [C]) = E_{CG}(\{\mathbf{r}\}, 0) + \Delta G_{tr}(\{\mathbf{r}\}, [C]) \quad (2)$$

where $E_{CG}(\{\mathbf{r}\}, 0)$ is given by eq 1, $\Delta G_{tr}(\{\mathbf{r}\},[C])$ is the protein−denaturant interaction energy in a solution with denaturant concentration $[C]$ and is given by

$$\Delta G_{tr}(\{\mathbf{r}\}, [C]) = \sum_{k=1}^{N} \delta g_{tr,k}([C]) \alpha_k(\{\mathbf{r}\}) / \alpha_{Gly\text{-}k\text{-}Gly} \quad (3)$$

where $N(= N_{res} \times 2 = 128)$ is the number of beads in coarse-grained Protein L, $\delta g_{tr,k}([C])$ is the transfer free energy of bead k, $\alpha_k(\{\mathbf{r}\})$ is the solvent accessible surface area (SASA) of the bead k in a protein conformation described by positions $\{\mathbf{r}\}$, $\alpha_{Gly\text{-}k\text{-}Gly}$ is the SASA of the bead k in the tripeptide Gly-k-Gly. The radii for amino acid side chains

DOI: 10.1021/jacs.5b11300
J. Am. Chem. Soc. 2016, 138, 2609−2616

to compute $\alpha_k(\{\mathbf{r}\})$ are given in Table S2 in ref 48. The experimental[11,50,51] transfer free energies $\delta g_{tr,i}([C])$, which depend on the chemical nature of the denaturant, for backbone and side chains are listed in Table S3 in ref 42. The values for $\alpha_{Gly\text{-}k\text{-}Gly}$ are listed in Table S4 in ref 42.

Simulations and Data Analysis. Low-friction Langevin dynamics simulations[52] are used to generate protein conformations as a function of T in $[C]$ = 0 M conditions. To compute thermodynamic properties of the protein in a denaturant solution of concentration $[C]$, $\Delta G_{tr}(\{\mathbf{r}\},[C])$ is treated as perturbation to $E_{CG}(\{\mathbf{r}\},0)$ in eq 2, and weighted histogram method[11,43,53] is used to compute average value of various physical quantities at any [C]. Brownian dynamics simulations[54] are used with the full Hamiltonian (eq 2) to simulate the burst-phase folding kinetics of the protein in a denaturant solution of concentration $[C]$ (see SI for details).

We computed structural overlap function,[55] χ, and radius of gyration, R_g, to monitor Protein L folding kinetics. The structural overlap function is defined as $\chi = 1 - \frac{1}{N_{tot}}\sum_{i=1}^{N_{tot}}\Theta(\delta - |r_i - r_i^0|)$. Here, N_{tot} (= 777) is the number of pairs of beads in the SOP-SC model of Protein L assuming that the bead centers are separated by at least two bonds, r_i is the distance between the i^{th} pair of beads, r_i^0 is the corresponding distance in the folded state, Θ is the heaviside step function, and δ = 2 Å. Using χ as an order parameter, we calculated the fraction of molecules in the native basin of attraction (NBA), f_{NBA} as a function of [GuHCl] (see SI for details and Figure S2). R_g, is calculated using $R_g = (1/2N^2)(\sum_{i,j}\vec{r}_{ij}^2)^{1/2}$, where $\vec{r}_{ij}$ is the vector connecting the beads i and j. The extent of long-range contacts in the TSE structures compared to the coarse-grained PDB structure is analyzed using the relative contact order,[56,57] RCO, which is defined as $RCO = \frac{1}{N_{res}N_{nat}}\sum_{i=1}^{N_{nat}} L_i$, where N_{nat} is the number of pairs of beads with native interactions in the protein conformation (see SI), and L_i is the number of residues separating the contact pair i.

RESULTS AND DISCUSSION

Thermodynamics of Protein L Folding. The protein in the folded state has one α-helix (α_1) and four β-strands (β_1–β_4) (Figures 1A and S1A). Low-friction Langevin dynamics simulations performed at different temperatures, T, ranging from 300 to 430 K show that folding occurs in a two-state manner (see Figure S1). The melting temperature, T_M, of Protein L obtained from the heat capacity, C_v, plot is 374.5 K (Figure S1), and the value observed in experiments[58] is 348.5 K. The difference in T_M between experiments and simulations can be attributed to the simplified coarse-grained SOP-SC model. At T_M, the protein transitions between the NBA and the unfolded basin of attraction (UBA) (Figure S2). Protein L folding thermodynamics and kinetics in the presence of the denaturant GuHCl are studied using the MTM.[11,43] In order to compare the denaturant-dependent folding properties of the protein computed from simulations with the experiments, a simulation temperature T_S (= 357.7 K) at which theoretically obtained free energy difference between the NBA and UBA, ΔG_{NU}^{Sim} (= $G_N(T_S) - G_U(T_S)$) matches with the experimentally[29] measured value, ΔG_{NU}^{Exp} (= −4.6 kcal/mol), at [GuHCl] = 0 M is used. This is the only adjustable parameter in the model, which is equivalent to matching the energy scales between the simulations and experiments.

The structural overlap parameter, χ (see Methods), is used to distinguish between the NBA and UBA protein conformations (Figure S2). The protein conformations with $\chi \leq 0.47$ belong to the NBA, and conformations with $\chi > 0.47$ belong to the UBA (Figure S2B). The fraction of molecules in NBA, f_{NBA}, as a function of [GuHCl] computed from simulations is in quantitative agreement with the experiments[23,29,59] (Figure 1B). The midpoint [GuHCl] at which the protein unfolds is ≈2.5 M. The average radius of gyration, $\langle R_g\rangle$, of Protein L as a function of [GuHCl] is in quantitative agreement with the SAXS experiments[25] (Figure 1C). The standard deviation of R_g in the protein unfolded state, $\sigma_{R_g} \approx 4$ Å, indicates that R_g fluctuates between $22 \lesssim R_g \lesssim 30$ Å. As [GuHCl] is diluted from 8 to 4M, the $\langle R_g\rangle$ decreases from ≈26.5 to ≈24.7 Å almost linearly with a slope of 0.42 Å M^{-1}. The experimental data[25] fit equally well with a horizontal line or a line with slope 0.33 ± 0.35 Å M^{-1}. The average R_g of the protein conformations in the UBA basin, $\langle R_g^{UBA}\rangle$, as a function of [GuHCl] shows that the size of the protein decreases from ≈26.5 to ≈22.5 Å as [GuHCl] is diluted from 7.5 to 0.25 M.

Denaturant-Dependent FRET Efficiency. Average FRET efficiency, $\langle E\rangle$, as a function of [GuHCl] is computed from the Langevin dynamics simulations (Figure 2A). In FRET experiments,[21,22] the donor (AlexaFluor 488) and acceptor (AlexaFluor 594) dyes are attached near the N- and C-termini of Protein L. All-atom simulations[60] have shown that the dyes have negligible effect on the size of disordered protein structures. The $\langle E\rangle$ is calculated using

$$\langle E\rangle = \int_0^L \frac{P(R_{ee})}{1 + (\frac{R_{ee}}{R_0})^6} dR_{ee} \quad (4)$$

where $P(R_{ee})$ is the end-to-end distance, R_{ee}, probability distribution function of the protein, L(= 248 Å) is the contour length of the protein, and R_0(= 54 Å) is the Forster radius for the donor–acceptor dyes used in the experiments.[21,22]

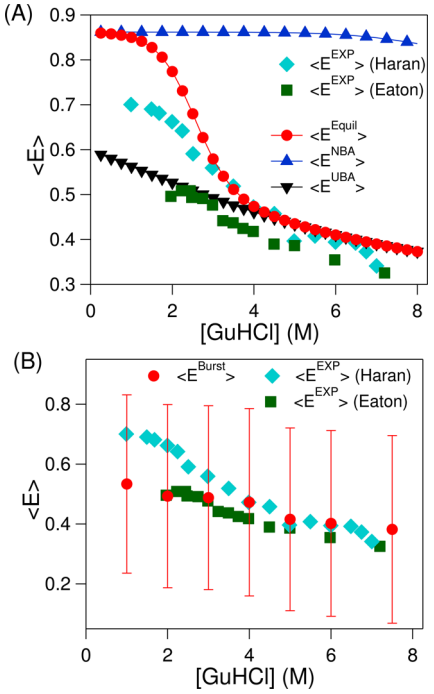


Figure 2. (A) Equilibrium FRET efficiency, $\langle E^{Equil}\rangle$, as a function of [GuHCl] is in red circles. Experimental data are shown in green squares[22] and cyan diamonds.[21] $\langle E\rangle$ for the protein conformations in the NBA and UBA basins is shown in blue triangles and black inverted triangles, respectively. (B) $\langle E^{Burst}\rangle$ shown in red circles is computed from the initial 0.25 ms of Protein L Brownian dynamics folding trajectories at T = 357.7 K in various [GuHCl]. Data in green squares and cyan diamonds are the same as in (A).

DOI: 10.1021/jacs.5b11300
J. Am. Chem. Soc. 2016, 138, 2609–2616

The equilibrium $\langle E \rangle$ transitions from lower values (<0.5) to higher values ($\approx$ 0.85) as the protein folds from an unfolded state upon [GuHCl] dilution (Figure 2A). The large standard deviation, $\sigma_E \approx 0.3$, for [GuHCl] > 2.5 M (Figure S3) indicates that the protein in the UBA basin samples conformations with large size fluctuations is in agreement with the R_g data (Figure 1C). The average FRET effeciency computed for the UBA ensemble, $\langle E^{UBA} \rangle$, to study whether the protein collapses in the early stages of folding is in quantitative agreement with the experiments of Eaton et al.,[22] where as they are in disagreement with the experiments of Haran et al.[21] for the denaturant concentrations $1 \lesssim$ [GuHCl] $\lesssim 3$ M (Figure 2A). $\langle E^{UBA} \rangle$ gradually increases from 0.37 to 0.6 as [GuHCl] is diluted from 8 to 0.25 M, pointing to an average decrease in the size of the protein (Figure 2A).

The average FRET efficiency $\langle E^{Burst} \rangle$ is also computed from the initial 0.25 ms of the Brownian dynamics simulations performed to study the folding kinetics of Protein L. The initial 0.25 ms of the folding trajectories are used to check whether the protein decreases in size in the burst phase of folding when [GuHCl] is diluted from 7.5 M to lower concentrations. The initial unfolded protein conformation to initiate the folding simulations in various [GuHCl] is obtained from simulations performed at [GuHCl] = 7.5 M. Twenty independent simulations starting from different initial protein conformations are performed for each [GuHCl]. $\langle E^{Burst} \rangle$ computed for the early stages of folding is in quantitative agreement with the experiments of Eaton et al.[22] for all [GuHCl] and deviates from the values obtained from the experiments of Haran et al.[21] for the [GuHCl] range $1 \lesssim$ [GuHCl] $\lesssim 3$ M (Figure 2B). $\langle E^{Burst} \rangle$ increases from 0.38 to 0.53 as [GuHCl] is diluted from 7.5 to 1.0 M, signifying that the protein on an average decreases in size. The standard deviation of burst-phase FRET efficiency, $\sigma_E \approx 0.3$, shows that $\langle E^{Burst} \rangle$ varies between 0.2 and 0.8, indicating that the protein in the initial stages of folding samples conformations with a significant variation in size (Figure 2B). The R_g plot as a function of time shows that it varies in the range $15 \lesssim R_g \lesssim 30$ Å during the initial hundreds of microseconds after folding is initiated for [GuHCl] = 1 and 2 M conditions (Figure S4).

FRET Overestimates Radius of Gyration in High [GuHCl]. We mimicked the FRET experiments[21,22] to estimate $\langle R_g^{FRET} \rangle$ from $\langle E^{Burst} \rangle$. The Gaussian polymer chain end-to-end probability distribution is given by

$$P(R_{ee}) = 4\pi R_{ee}^2 \left(\frac{3}{2\pi \langle R_{ee}^2 \rangle} \right)^{3/2} \exp\left(-\frac{3R_{ee}^2}{2\langle R_{ee}^2 \rangle} \right) \quad (5)$$

The $P(R_{ee})$ given by eq 5 is used in eq 4 to estimate the average end-to-end distance square, $\langle R_{ee}^2 \rangle$, from $\langle E \rangle$. $\langle R_g \rangle$ is calculated using the relation,[61] $\langle R_g \rangle = \sqrt{\langle R_{ee}^2 \rangle / 6}$. $\langle R_g^{FRET} \rangle$ values estimated from $\langle E^{Burst} \rangle$ (Figure 2B) using eqs 4 and 5 at different [GuHCl] are in near quantitative agreement with the experimentally[22] estimated values (Figure 3A). On diluting [GuHCl] from 7.5 to 1 M, $\langle R_g^{FRET} \rangle$ decreases from $\approx$30 to $\approx$24 Å, nearly a 6 Å change in the size of the protein, which is in agreement with the experiments[22] of Eaton et al. However, $\langle R_g^{FRET} \rangle$ deviates from the $\langle R_g^{Burst} \rangle$ values computed directly from the protein conformations obtained from the simulation trajectories (Figure 3A). $\langle R_g^{Burst} \rangle$ decreases from $\approx$26.7 to $\approx$23.8 Å, a decrease of only $\approx$3 Å upon [GuHCl] dilution from 7.5 to 1 M (Figure 3A). This shows that FRET overestimates the size

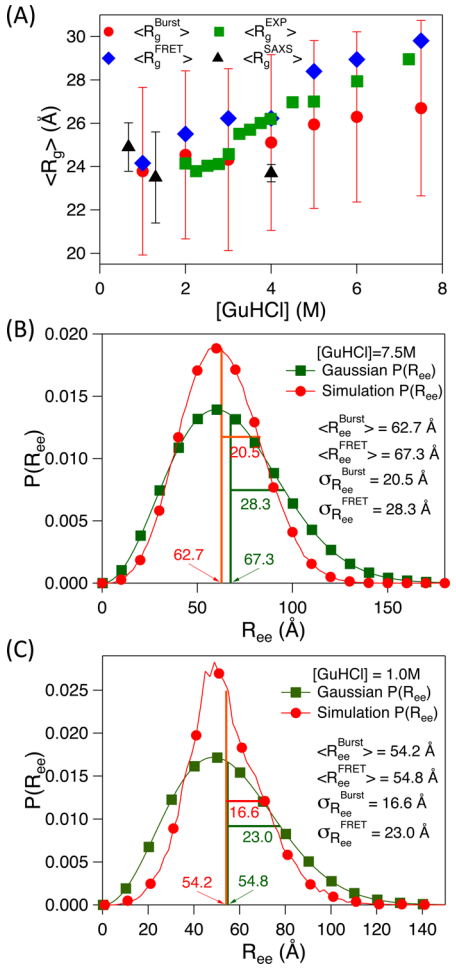


Figure 3. (A) $\langle R_g^{FRET} \rangle$ estimated from $\langle E^{Burst} \rangle$ is in blue diamonds. $\langle R_g^{Burst} \rangle$ computed from the initial 0.25 ms of the Protein L folding trajectories is in red circles. Data in green squares and black triangles are from FRET[22] and SAXS[25] experiments, respectively. (B) The end-to-end distance, R_{ee}, probability distribution function $P(R_{ee})$ during the burst phase (initial 0.25 ms) of protein L folding at T = 357.7 K and [GuHCl] = 7.5 M is in red circles. $P(R_{ee})$ estimated from $\langle E^{Burst} \rangle$ and Guassian polymer chain statistics in green squares. (C) Same as in (B) except that [GuHCl] = 1.0 M.

of the protein especially in higher [GuHCl], and this gives rise to the appearance of pronounced compaction in the dimensions of the protein in the burst phase as [GuHCl] is diluted. The standard deviation, $\sigma_{R_g} \approx 4$ Å, of $\langle R_g^{Burst} \rangle$ shows that at all [GuHCl], the protein samples conformations with R_g varying from $\approx$22 to $\approx$28 Å (Figure 3A), and the 3 Å decrease in $\langle R_g^{Burst} \rangle$ upon [GuHCl] dilution are within the σ_{R_g}. The deviation between $\langle R_g^{FRET} \rangle$ and $\langle R_g^{Burst} \rangle$ at high [GuHCl] was also emphasized in the work of O'Brien et al.,[62] and the reasons for the deviation are attributed to the use of Gaussian chain $P(R_{ee})$ to extract $\langle R_g^{Burst} \rangle$. The problems associated with the use of Gaussian chain $P(R_{ee})$ to extract information about protein dimensions are also highlighted in previous studies.[63−66] The deviation between $\langle R_g^{FRET} \rangle$ and $\langle R_g^{Burst} \rangle$ increases with [GuHCl], and the reasons for the discrepancy can be understood using the relation $\langle R_{ee}^2 \rangle = \langle R_{ee} \rangle^2 + \sigma_{R_{ee}}^2$, where $\sigma_{R_{ee}}$ is the standard deviation in R_{ee}.

$\langle R_{ee}^{FRET} \rangle$ and $\sigma_{R_{ee}}^{FRET}$ estimated from the Gaussian polymer chain $P(R_{ee})$ to compute $\langle R_g^{FRET} \rangle$ deviate from the values

DOI: 10.1021/jacs.5b11300
J. Am. Chem. Soc. 2016, 138, 2609−2616

$\langle R_g^{Burst}\rangle$ computed directly from the initial 0.25 ms of the Protein L folding trajectories (Figure 3B,C). At high [GuHCl]-(= 7.5 M), $\langle R_{ee}^{FRET}\rangle$ and $\sigma_{R_{ee}}^{FRET}$ estimated from the Gaussian chain $P(R_{ee})$ are 67.3 and 28.3 Å, respectively, which deviate from the $\langle R_{ee}^{Burst}\rangle$ and $\sigma_{R_{ee}}^{Burst}$ values 62.7 and 20.5 Å, respectively (Figure 3B), computed directly from the simulations. As a result FRET overestimates $\langle R_g^{FRET}\rangle (= \sqrt{\langle (R_{ee}^{FRET})^2\rangle/6} = \sqrt{[\langle R_{ee}^{FRET}\rangle^2 + (\sigma_{R_{ee}}^{FRET})^2]/6})$ compared to $\langle R_g^{Burst}\rangle$ in high [GuHCl] (Figure 3A). As [GuHCl] decreases, the deviation between $\langle R_g^{FRET}\rangle$ and $\langle R_g^{Burst}\rangle$ decreases (Figure S5). In low [GuHCl](= 1.0 M), the $\langle R_{ee}^{FRET}\rangle$ values computed from the Gaussian chain $P(R_{ee})$ and $\langle R_{ee}^{Burst}\rangle$ computed from simulations are in good agreement, where as the $\sigma_{R_{ee}}^{FRET}$ values deviate from $\sigma_{R_{ee}}^{Burst}$ (Figure 3C). Due to this, the deviation between $\langle R_g^{Burst}\rangle$ and $\langle R_g^{FRET}\rangle$ is small in low [GuHCl], and increases with [GuHCl] (Figure 3 and S5).

To conclude, during the burst phase of folding, $\langle R_g^{Burst}\rangle$ for Protien L decreases from ≈26.7(±4) to ≈23.8(±4)Å, a decrease of ≈3 Å upon [GuHCl] dilution from 7.5 to 1 M (Figure 3A). However, FRET overestimates the size of the protein in high [GuHCl] (≈7.5 M) due to the application of the Gaussian polymer chain $P(R_{ee})$ to estimate $\langle R_g^{FRET}\rangle$. During the burst phase $\langle R_g^{FRET}\rangle$ estimated from FRET decreases from ≈30 to ≈24 Å upon [GuHCl] dilution from 7.5 to 1 M (Figure 3A). Due to the ≈6 Å decrease in $\langle R_g^{FRET}\rangle$, FRET experiments[21,22] suggest pronounced compaction in the protein size during the burst phase of folding.

Disagreement between FRET and SAXS Experiments on Protein L Compaction in Burst-Phase Folding. In simulations, the average radius of gyration in the burst-phase folding, $\langle R_g^{Burst}\rangle$, decreased by ≈3 Å when [GuHCl] is diluted from 7.5 to 1 M. The ≈3 Å decrease in $\langle R_g^{Burst}\rangle$ is close to statistical uncertainties of the R_g data obtained from SAXS experiments[25] for Protein L. The R_g data from SAXS experiments for $3 \lesssim$ [GuHCl] $\lesssim 7$ M can fit a horizontal line or a line with slope 0.33 ± 0.35 Å M^{-1} equally well.[25] Using this slope to compute the R_g change upon [GuHCl] dilution from ≈7.5 to 1 M gives a R_g decrease of 2.1 ± 2.3 Å. The SAXS experiments[25] report that R_g of Protein L in the burst phase upon [GuHCl] dilution to 1.3 and 0.67 M is ≈23.5 ± 2.1 and 24.9 ± 1.12 Å, respectively, which is statistically not different from the value 23.7 ± 0.4 at [GuHCl] = 4.0 M. In the simulations, $\langle R_g^{Burst}\rangle$ at [GuHCl] = 4 and 1 M is 25.1 ± 4.0 and 23.8 ± 3.9 Å, respectively, a difference of ≈1.3 Å, which is well within $\sigma_{R_g} \approx 4$ Å. This analysis, similar to the SAXS analysis, leads to the conclusion that on [GuHCl] dilution, Protein L compaction is minimal and within statistical uncertainties. This conclusion is in agreement with the SAXS experiments[25] and in disagreement with the FRET experiments.[21,22]

Recent FRET experiments[67] on polyethylene glycol (PEG) showed that FRET efficiency decreased as [GuHCl] is increased when hydrophilic PEG is unlikely to expand on increasing [GuHCl]. This led to questions about the interpretation of the FRET data to study protein collapse in low [GuHCl]. We find that the computed variation in $\langle E\rangle$ and $\langle R_g\rangle$ as a function of [GuHCl] for Protein L is in quantitative agreement with at least one of the FRET experiments[22] (Figure 2) and also in agreement with SAXS experiments[25] within the statistical uncertainties (Figures 1C and 3A). The results point to the use of Gaussian polymer chain statistics to extract R_g from FRET efficiency data to be the cause for the discrepancy between the SAXS and FRET experiments in estimating R_g.

The Coil−Globule Transition in Protein L is Concomitant with the Folding Transition. In polymers the ratio of the radius of gyration to the hydrodynamic radius, R_g/R_h, can point to the coil−globule collapse transition. The R_g/R_h ratio for a polymer in a good solvent[68] is ≈1.56, where as the ratio in a poor solvent[61] is ≈0.77. We used the Kirkwood−Riseman approximation[69] to compute the hydrodynamic radius of the protein, which is given by $R_h = (1/2N^2)\sum_{i\neq j} 1/|\vec{r_i} - \vec{r_j}|$, where N is the number of beads in the coarse-grained protein, $\vec{r_i}$ and $\vec{r_j}$ are the position vectors of beads i and j. The R_g/R_h ratio for the burst-phase folding decreases from ≈1.31 to ≈1.28 as [GuHCl] is diluted from 7.5 to 1 M, indicating that this is not a coil−globule transition observed in polymers (Figure 4). The single domain proteins which are finite in size compared to polymers are predicted to have a near overlap of the collapse and folding transition temperatures.[12,13] In agreement, the equilibrium ratio of R_g/R_h decreases from ≈1.3 to ≈0.98 as the folding transition occurs (Figure 4). The R_g/R_h ratio does not approach 0.77 as the protein folds because the Kirkwood−Riseman approximation[69] used to compute R_h does not hold for the protein in the folded state, as it assumes all the beads are equally bathed by the solvent. The absence of coil−globule transition in the burst phase of protein L folding does not imply that the collapse transition or significant protein compaction is universally absent in the burst-phase folding of all single domain proteins. Both FRET and SAXS experiments agree that the protein Monellin[18−20] shows compaction during the burst phase of folding. Although both the experimental techniques observe compaction in the case of Cytochrome c,[14−17] the FRET experiments show that this compaction, a sub-100 μs event, is barrier limited, and it is due to the formation of marginally stable partially folded structures.[14] Experiments[8] show that for the protein Cyclophilin A, the R_g/R_h ratio decreases from a value between 1.1 and 1.2 to a value between 0.9 and 1.0 as [GuHCl] is diluted from 8 to 0 M, indicating a coil−globule transition in the burst phase of folding.

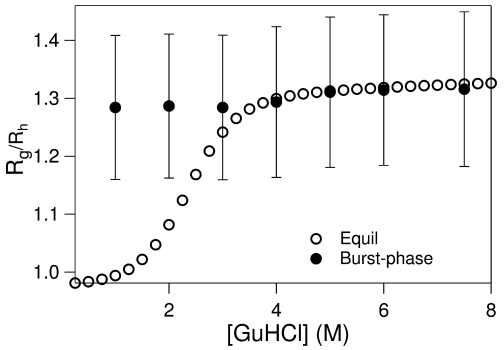


Figure 4. Ratio of the radius of gyration to the hydrodynamic radius, R_g/R_h, for the burst phase (solid circles) and equilibrium (empty circles) conditions.

Transition-State Ensemble (TSE). The TSE of Protein L at the melting temperature, T_M, is identified using the P_{fold} analysis[44] (see SI for details). Twelve out of 108 putative transition-state structures which satisfy the condition, $0.4 < P_{fold} < 0.6$, are labeled as TSE (Figure S6). The structures in the TSE are globular, extensive, and homogeneous, with most of the secondary and tertiary contacts formed (Figure 5). The Ψ-analysis experiments[30] predict that TSE contains all the four β-

DOI: 10.1021/jacs.5b11300
J. Am. Chem. Soc. 2016, 138, 2609−2616

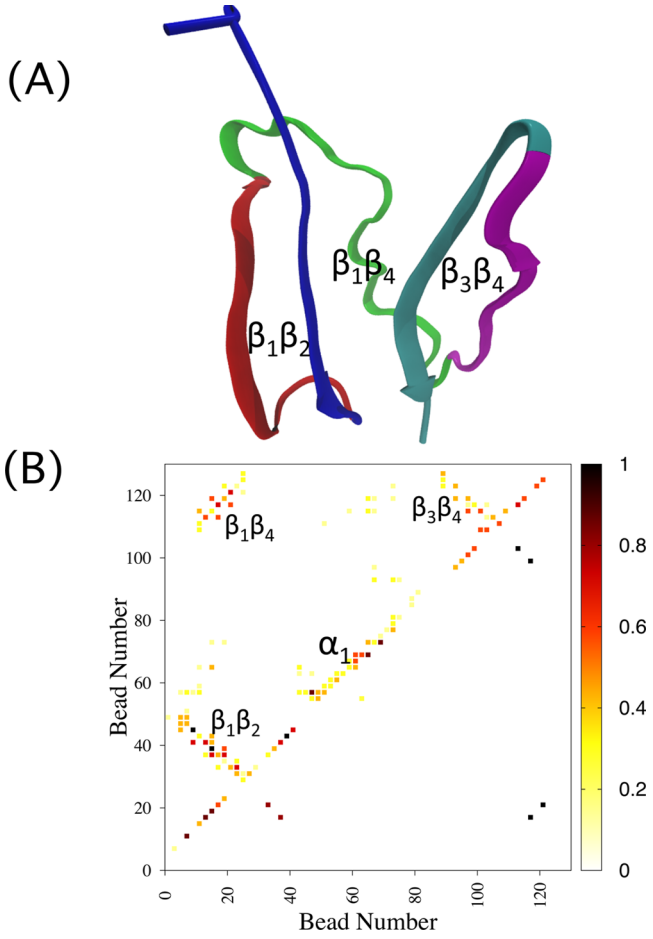


Figure 5. (A) Representative transition-state structure from the TSE obtained using the P_{fold} analysis. (B) The contact map of the TSE is shown in the upper half of the diagonal. The experimental[30] Ψ-values for the transition-state structure are shown in the lower half of the diagonal. The Ψ-values for the α-helix residue pairs K28-E32 and A35-T39 from experiments[30] are 0.26 and $\ll$0, respectively. The Ψ-value used for the residue pair A35-T39 is 0 in the plot. The small Ψ-values for α-helix are not visible in the plot.

sheet strands (β_1–β_4). The TSE from simulations show that both the N- and C-termini hairpins $\beta_1\beta_2$ and $\beta_3\beta_4$ and the contacts between the strands $\beta_1\beta_4$ are present in the structures in agreement with the Ψ-analysis experiments[30] (Figure 5B).

The Ψ-analysis experiments on two residue pairs, K28-E32 and A35-T39, present in the helical region of the protein gave Ψ-values 0.26 and $\ll$0, respectively, indicating that contacts between these pairs of residues is largely absent and concluded that helix α_1 is mostly not present in the TSE.[30] The contact map of the TSE obtained from the simulations shows that the side chains of the residue pairs K28-E32 and A35-T39 form contacts with a probability of 0.41 and 0.08, respectively. The simulations further indicate that a cluster of residues between S31 and A37 present approximately at the center of the helix containing 3 Ala residues (A33, A35 and A37) can form stable contacts in the TSE (Figure 5B).

The Ψ-analysis experiments[31,70] predict a relationship between the relative contact order, RCO (see Methods), of the native protein topology and TSE, $RCO^{TSE} \approx 0.7\ RCO^{Native}$, which shows the extent of the long-range contacts present in the TSE compared to the native state. The TSE structures extracted from the simulations show $RCO^{TSE}/RCO^{Native} = 0.77$, which is in reasonable agreement with the value of 0.75 estimated from Ψ-analysis experiments.[30] The simulations using the coarse-grained protein model support the basic topology of the TSE structures predicted by the Ψ-analysis experiments.

The folding simulations of only the C-terminal hairpin ($\beta_3\beta_4$) using atomistic models predicted the presence of non-native contacts, a two amino acid register shift, in the TSE.[30] We do not observe this two amino acid register shift in the C-terminal hairpin because the SOP-SC model includes only native interactions. The predicted TSE is only in partial agreement with the Φ-analysis experiments[27−29,71] which predicted a polarized structure with only $\beta_1\beta_2$ hairpin. The results support the hypothesis that folding pathways and TSE of single domain proteins are influenced by the topology of the folded structure in agreement with the experiments.[45]

CONCLUDING REMARKS

In summary, we have studied Protein L folding in the presence of the denaturant guanidine hydrochloride using the SOP-SC coarse-grained model and molecular dynamics simulations. The effect of [GuHCl] on the protein is taken into account using the MTM.[11,43] The study mainly focused on whether there is a coil−globule collapse transition in the burst phase of folding after the denaturant concentration is diluted to lower values. The main finding of this study is that the coil−globule transition in Protein L is concomitant with the folding transition. It is not observed during the burst phase. The FRET experiments overestimate the R_g of the protein at high [GuHCl] concentrations, owing to the use of the Gaussian polymer chain end-to-end distribution function to extract R_g from FRET efficiency. As a result, in the burst phase of folding, FRET observes pronounced compaction in the size of the protein as [GuHCl] is diluted. The actual decrease in the size of the protein (≈3 Å) observed during the burst phase is close to statistical uncertainties of the R_g data measured from SAXS experiments,[25] and these experiments conclude that there is no collapse leading to a discrepancy with the FRET experiments.

It is highly desirable to formulate a method to accurately extract the distances between the donor and acceptor dyes used in the FRET experiments. However, it is a nontrivial inverse problem as we seek to accurately extract a probability distribution of the distances between the dyes from the average FRET efficiency measured in experiments, especially in cases like Protein L where the compaction in protein dimensions is small on denaturant dilution. Previous studies[62] have shown that even other polymer models such as the self-avoiding chain or the worm-like chain model are also not very accurate quantitatively to predict the small subtle changes in the protein dimensions.

The results presented in this manuscript clearly point out the aspects of the Gaussian chain model, which leads to over estimating the size of the protein when used to analyze the FRET data. The results show that the Gaussian chain model fails in accurately capturing the width of the protein end-to-end probability distribution, which is essential to compute the radius of gyration. For any method to be quantitatively accurate, it should capture the peak position as well as the width of the probability distribution accurately, and this is a challenging task because we need to estimate probability distribution from an average value, and also the method should be reliable enough to work on proteins with different amino acid composition and native folds.

DOI: 10.1021/jacs.5b11300
J. Am. Chem. Soc. 2016, 138, 2609−2616

To check the accuracy of the distance between the dyes extracted from the FRET efficiency data using the Gaussian polymer model assumption, a self-consistency check can be performed to see if the assumption is valid or not for the protein under study.[62] If the dyes are attached at locations i and j in the protein, and $\langle R_{ij}^2 \rangle$ is the average distance square extracted from FRET efficiency, and similarly if $\langle R_{kl}^2 \rangle$ is the average distance square extracted from FRET efficiency with dyes at positions k and l, then the relation $\langle R_{ij}^2 \rangle / \langle R_{kl}^2 \rangle = |j - i|/|l - k|$ should hold if the protein behaves as a Gaussian chain. If the relation is not satisfied, then one should be cautious in quantitatively inferring results about the protein dimensions assuming that the protein in the unfolded state behaves as a Gaussian chain.

The magnitude of protein compaction in the burst-phase folding of single domain proteins upon denaturant dilution is not uniform, and it should depend on protein length, sequence, and composition of amino acids. For example, SAXS experiments on Protein L[25] and ubiquitin[72] infer no compaction in the protein dimensions in the burst phase, while experiments on Cytochrome c[14] and Monellin[20] observe compaction. The key features in the single domain proteins responsible for compaction in protein dimensions on denaturant dilution need to be identified. In addition to temperature and denaturants, force can also be used to unfold proteins and study protein folding. Experiments[73] and simulations[74,75] show that a protein unfolded by force when allowed to refold in the presence of lower quenching forces undergoes a rapid compaction in the initial stages of folding. This compaction of the protein is driven by entropy because the protein in the stretched state is in a low entropic state and upon force quench undergoes rapid compaction in the first stage of folding until entropy is maximized.[75] The extent of protein compaction in the initial stages of folding also depends on the experimental probes used to study protein folding.

The transition-state structures inferred from the P_{fold} analysis are globular and extensive with both the C and N-termini hairpins $\beta_1\beta_2$ and $\beta_3\beta_4$ and interactions between the strands $\beta_1\beta_4$. These results are in agreement with the Ψ-analysis experiments[30] and support the hypothesis that for single domain globular proteins, the transition-state structures depend on the protein native-state topology and are not stabilized by local interactions alone.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b11300.

Description of the simulation methods; Table S1; Figures S1–S6 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*greddy@sscu.iisc.ernet.in

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank D. Thirumalai, Gilad Haran, Shachi Gosavi and D. D. Sarma for valuable comments and discussion. G.R. acknowledges startup grant from Indian Institute of Science-Bangalore and funding from Nano mission, Department of Science and Technology, India. H.M. acknowledges a research fellowship from Indian Institute of Science-Bangalore.

■ REFERENCES

(1) Ziv, G.; Thirumalai, D.; Haran, G. *Phys. Chem. Chem. Phys.* **2009**, *11*, 83–93.
(2) Sosnick, T. R.; Barrick, D. *Curr. Opin. Struct. Biol.* **2011**, *21*, 12–24.
(3) Haran, G. *Curr. Opin. Struct. Biol.* **2012**, *22*, 14–20.
(4) Thirumalai, D.; Liu, Z.; O'Brien, E. P.; Reddy, G. *Curr. Opin. Struct. Biol.* **2013**, *23*, 22–29.
(5) Udgaonkar, J. B. *Arch. Biochem. Biophys.* **2013**, *531*, 24–33.
(6) Tanford, C.; Kawahara, K.; Lapanje, S. *J. Biol. Chem.* **1966**, *241*, 1921.
(7) Kohn, J.; Millett, I.; Jacob, J.; Zagrovic, B.; Dillon, T.; Cingel, N.; Dothager, R.; Seifert, S.; Thiyagarajan, P.; Sosnick, T.; Hasan, M.; Pande, V.; Ruczinski, I.; Doniach, S.; Plaxco, K. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12491–12496.
(8) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16155–16160.
(9) Bryngelson, J. D.; Wolynes, P. G. *Biopolymers* **1990**, *30*, 177–188.
(10) Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 447–490.
(11) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13403–13408.
(12) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6369–6372.
(13) Li, M. S.; Klimov, D. K.; Thirumalai, D. *Phys. Rev. Lett.* **2004**, *93*, 268107.
(14) Kathuria, S. V.; Kayatekin, C.; Barrea, R.; Kondrashkina, E.; Graceffa, R.; Guo, L.; Nobrega, R. P.; Chakravarthy, S.; Matthews, C. R.; Irving, T. C.; Bilsel, O. *J. Mol. Biol.* **2014**, *426*, 1980–1994.
(15) Lapidus, L. J.; Yao, S.; McGarrity, K. S.; Hertzog, D. E.; Tubman, E.; Bakajin, O. *Biophys. J.* **2007**, *93*, 218–224.
(16) Hagen, S. J.; Eaton, W. A. *J. Mol. Biol.* **2000**, *301*, 1019–1027.
(17) Shastry, M. C. R.; Sauder, J. M.; Roder, H. *Acc. Chem. Res.* **1998**, *31*, 717–725.
(18) Goluguri, R. R.; Udgaonkar, J. B. *Biochemistry* **2015**, *54*, 5356–5365.
(19) Jha, S. K.; Udgaonkar, J. B. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12289–12294.
(20) Kimura, T.; Uzawa, T.; Ishimori, K.; Morishima, I.; Takahashi, S.; Konno, T.; Akiyama, S.; Fujisawa, T. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2748–2753.
(21) Sherman, E.; Haran, G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11539–11543.
(22) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1528–1533.
(23) Waldauer, S. A.; Bakajin, O.; Ball, T.; Chen, Y.; DeCamp, S. J.; Kopka, M.; Jaeger, M.; Singh, V. R.; Wedemeyer, W. J.; Weiss, S.; Yao, S.; Lapidus, L. J. *HFSP J.* **2008**, *2*, 388–395.
(24) Plaxco, K. W.; Millett, I. S.; Segel, D. J.; Doniach, S.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 554–556.
(25) Yoo, T. Y.; Meisburger, S. P.; Hinshaw, J.; Pollack, L.; Haran, G.; Sosnick, T. R.; Plaxco, K. *J. Mol. Biol.* **2012**, *418*, 226–236.
(26) Scalley, M. L.; Yi, Q.; Gu, H. D.; McCormack, A.; Yates, J. R.; Baker, D. *Biochemistry* **1997**, *36*, 3373–3382.
(27) Gu, H. D.; Kim, D.; Baker, D. *J. Mol. Biol.* **1997**, *274*, 588–596.
(28) Kim, D. E.; Yi, Q.; Gladwin, S. T.; Goldberg, J. M.; Baker, D. *J. Mol. Biol.* **1998**, *284*, 807–815.
(29) Kim, D. E.; Fisher, C.; Baker, D. *J. Mol. Biol.* **2000**, *298*, 971–984.
(30) Yoo, T. Y.; Adhikari, A.; Xia, Z.; Huynh, T.; Freed, K. F.; Zhou, R.; Sosnick, T. R. *J. Mol. Biol.* **2012**, *420*, 220–234.
(31) Baxa, M. C.; Freed, K. F.; Sosnick, T. R. *J. Mol. Biol.* **2008**, *381*, 1362–1381.
(32) Naganathan, A. N.; Munoz, V. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8611–8616.
(33) Koga, N.; Takada, S. *J. Mol. Biol.* **2001**, *313*, 171–180.

DOI: 10.1021/jacs.5b11300
J. Am. Chem. Soc. 2016, 138, 2609–2616

(34) Karanicolas, J.; Brooks, C. L., III *Protein Sci.* **2002**, *11*, 2351−2361.
(35) Clementi, C.; Garcia, A. E.; Onuchic, J. N. *J. Mol. Biol.* **2003**, *326*, 933−954.
(36) Ejtehadi, M. R.; Avall, S. P.; Plotkin, S. S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 15088−15093.
(37) Brown, S.; Head-Gordon, T. *Protein Sci.* **2004**, *13*, 958−970.
(38) Yang, Q.; Sze, S.-H. *BMC Bioinf.* **2008**, *9*, 320.
(39) Voelz, V. A.; Singh, V. R.; Wedemeyer, W. J.; Lapidus, L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 4702−4709.
(40) Chen, T.; Chan, H. S. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6460−6479.
(41) Hyeon, C. B.; Dima, R. I.; Thirumalai, D. *Structure* **2006**, *14*, 1633−1645.
(42) Liu, Z.; Reddy, G.; O'Brien, E. P.; Thirumalai, D. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 7787−7792.
(43) Liu, Z.; Reddy, G.; Thirumalai, D. *J. Phys. Chem. B* **2012**, *116*, 6707−6716.
(44) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. *J. Chem. Phys.* **1998**, *108*, 334−350.
(45) Baxa, M. C.; Yu, W.; Adhikari, A. N.; Ge, L.; Xia, Z.; Zhou, R.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 8302−8307.
(46) O'Neill, J. W.; Kim, D. E.; Baker, D.; Zhang, K. Y. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 480−487.
(47) Betancourt, M. R.; Thirumalai, D. *Protein Sci.* **1999**, *8*, 361−369.
(48) Reddy, G.; Thirumalai, D. *J. Phys. Chem. B* **2015**, *119*, 11358−11370.
(49) Reddy, G.; Liu, Z.; Thirumalai, D. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17832−17838.
(50) Auton, M.; Bolen, D. W. *Biochemistry* **2004**, *43*, 1329−1342.
(51) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. *Biochemistry* **2009**, *48*, 3743−3754.
(52) Veitshans, T.; Klimov, D.; Thirumalai, D. *Folding Des.* **1997**, *2*, 1−22.
(53) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(54) Ermak, D. L.; Mccammon, J. A. *J. Chem. Phys.* **1978**, *69*, 1352−1360.
(55) Guo, Z.; Thirumalai, D. *J. Mol. Biol.* **1996**, *263*, 323−343.
(56) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985−994.
(57) Klimov, D. K.; Thirumalai, D. *J. Mol. Biol.* **1998**, *282*, 471−492.
(58) Gillespie, B.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 12014−12019.
(59) Sherman, E.; Itkin, A.; Kuttner, Y. Y.; Rhoades, E.; Amir, D.; Haas, E.; Haran, G. *Biophys. J.* **2008**, *94*, 4819−4827.
(60) Zerze, G. H.; Best, R. B.; Mittal, J. *Biophys. J.* **2014**, *107*, 1654−1660.
(61) Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press: Oxford, 2003.
(62) O'Brien, E. P.; Morrison, G.; Brooks, B. R.; Thirumalai, D. *J. Chem. Phys.* **2009**, *130*, 124903.
(63) Sinha, K. K.; Udgaonkar, J. B. *J. Mol. Biol.* **2007**, *370*, 385−405.
(64) Laurence, T. A.; Kong, X. X.; Jager, M.; Weiss, S. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 17348−17353.
(65) Goldenberg, D. P. *J. Mol. Biol.* **2003**, *326*, 1615−1633.
(66) Zhou, H.-X. *J. Phys. Chem. B* **2002**, *106*, 5769−5775.
(67) Watkins, H. M.; Simon, A. J.; Sosnick, T. R.; Lipman, E. A.; Hjelm, R. P.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 6631−6636.
(68) Oono, Y.; Kohmoto, M. *J. Chem. Phys.* **1983**, *78*, 520−528.
(69) Kirkwood, J. G.; Riseman, J. *J. Chem. Phys.* **1948**, *16*, 565−573.
(70) Pandit, A. D.; Jha, A.; Freed, K. F.; Sosnick, T. R. *J. Mol. Biol.* **2006**, *361*, 755−770.
(71) Scalley, M. L.; Yi, Q.; Gu, H. D.; McCormack, A.; Yates, J. R.; Baker, D. *Biochemistry* **1997**, *36*, 3373−3382.
(72) Jacob, J.; Krantz, B.; Dothager, R. S.; Thiyagarajan, P.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *338*, 369−382.
(73) Fernandez, J.; Li, H. *Science* **2004**, *303*, 1674−1678.
(74) Best, R.; Hummer, G. *Science* **2005**, *308*, 498b.
(75) Hyeon, C.; Morrison, G.; Pincus, D. L.; Thirumalai, D. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 20288−20293.

DOI: 10.1021/jacs.5b11300
J. Am. Chem. Soc. 2016, 138, 2609−2616