Effects of Crowding and Confinement on the Structures of the Transition State Ensemble in Proteins

Margaret S. Cheung[†] and D. Thirumalai*[,‡,§]

Department of Physics, University of Houston, Houston, Texas 77204, and Biophysics Program, Institute for Physical Science and Technology, and Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

Received: November 29, 2006; In Final Form: February 28, 2007

Characterization of the structures of the transition state ensemble is a key step in describing the folding reaction. Using two variants of a coarse-grained model of the three-stranded β-sheet WW domain and a fully automated progress variable clustering (PVC) algorithm, we have dissected the effect of macromolecular crowding and confinement on the changes in the transition state structures in comparison to bulk. Each amino acid is represented using a C_α atom and a side chain. The distance between the C_α atom and center of mass of the side chain is taken to be its effective van der Waals radius. For the bulk case, we predict using the PVC algorithm, which does not assume knowledge of the underlying folding reaction coordinate, that there are two classes of structures in the transition state ensemble (TSE). The structures in both of the classes are compact. The dominant cluster is more structured than the structures in the less populated class. In accord with bulk experiments, the residues in strands β_2 and β_3 and the interactions at the $\beta_2-\beta_3$ interface are structured. When only excluded volume interactions between the crowding particles and the WW domain are taken into account or when the protein is confined to an inert spherical pore, the overall structure of the TSE does not change dramatically. However, in this entropy dominated regime, the width of the TSE decreases and the structures become more oblate and less spherical as the volume fraction of crowding particle increases or when the pore radius decreases. It suggests that the shape changes, which are computed using the moment of inertia tensor, may represent the slow degrees of freedom during the folding process. When non-native interactions between side chains and interactions with the cavity of the pores are taken into account, the TSE becomes considerably broader. Although the topology in the transition has a fold similar to the native state, the structures are far more plastic than in the bulk. The TSE is sensitive to the size of the pore as well as interactions between the pore and the protein. The differences between the two cases (confinement in an inert pore and when pore−protein interactions are considered) arise due to the increased importance of enthalpic interactions in the transition state as the strength of the protein−pore interaction increases.

Introduction

Cellular medium is crowded with lipids, nucleic acids, cytoskeletons, and other macromolecules. Estimates show that the volume fraction (Φ_c) of these macromolecules, collectively referred to as crowding agents, can exceed $\approx$0.2.[1,2] Spontaneous folding of nascent proteins in the crowded environment can be different from in vitro experiments that are typically conducted under infinite dilution conditions. Protein folding under crowded conditions is complicated because the nature of the water-mediated interactions between the crowding agents and the polypeptide chains is not fully understood. However, to a first approximation, the dominant effect of crowding agents is to exclude the volume occupied by them to the proteins. If excluded volume interactions dominate, then the stability of the folded state of the protein, compared to the $\Phi_c = 0$ case, is enhanced. Because of the entropically derived depletion interaction,[3] due to the crowding particles, the folded structure is stabilized with respect to the $\Phi_c = 0$ case. The entropic stabilization of the folded state, in the presence of crowding agents, has been firmly established in pioneering studies by Minton.[4,5] Although the extent to which crowding agents stabilize proteins is sometimes debated,[6] the theoretical predictions regarding crowding-induced stability has been confirmed in a number of in vitro experiments.[6−9] More recently, we[10] showed that the folding rates of proteins increase non-monotonically as Φ_c is increased, reaching a maximum at Φ_c^* that approximately corresponds to the overlap volume fraction of the protein. When excluded volume interactions dominate, then the observed rate enhancement can be explained in terms of the entropic destabilization of the denatured state ensemble (DSE).[10] The entropically destabilized DSE idea was first used to rationalize rate enhancement of proteins in a related problem of folding confined inert pores.[11] We showed previously[10] that folding in the presence of crowding agents is approximately equivalent to folding in a Φ_c-dependent spherical pore.

Although many aspects of folding in restricted spaces (or in the presence of crowding agents) are well understood, the complete characterization of crowding (or confinement)-induced folding also requires knowledge of the transition state ensemble (TSE). Here, we determine the extent to which the TSE changes compared to the bulk case due to confinement or when Φ_c is nonzero. In order to rationalize the crowding-induced rate enhancement, we previously assumed[10] that the location of the

[†] University of Houston.
[‡] Institute for Physical Science and Technology, University of Maryland.
[§] Department of Chemistry and Biochemistry, University of Maryland.

10.1021/jp068201y CCC: $37.00 © 2007 American Chemical Society
Published on Web 06/22/2007

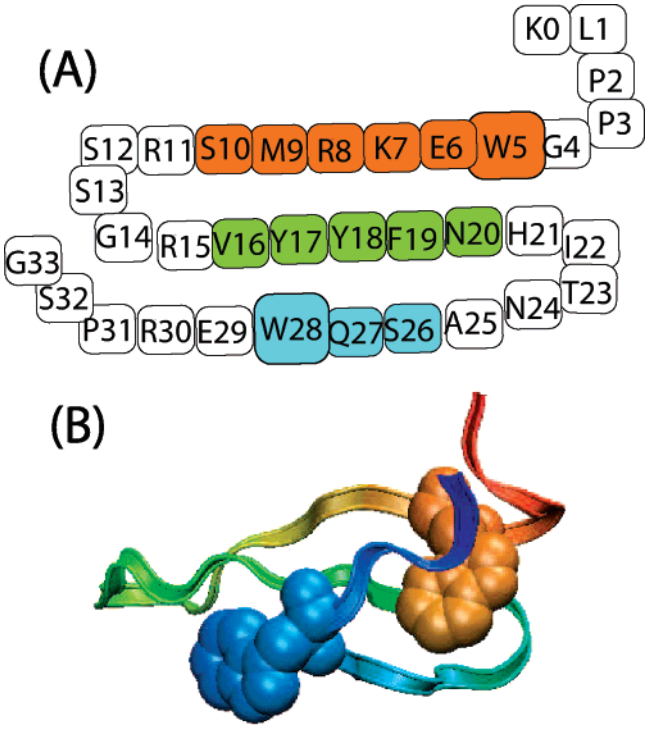

Figure 1. (A) The amino acid sequence of the 1PIN WW domain labeled with one-letter code and numbering starting from 0. (B) The native structure of the 1PIN domain. Orange color represents β1, β2 is colored in green, and blue corresponds to β3 in part A. In part B, the ribbon is colored from red (N-terminus) to blue (C-terminus). The two tryptophans (W5 and W28) are represented using spheres.

TSE does not change significantly as Φ_c is increased from zero. In other words, at low values of Φ_c (or modest confinement), the decrease in the free energy barrier is largely due to the entropic destabilization of the DSE. It is also likely that there are changes in the characteristics of the TSE such as its width or the heterogeneity of structures. In the presence of crowding agents, the polypeptide chain would prefer to be localized in a region that is devoid of crowding agents. The localization not only decreases the conformational entropy of the DSE but also should lessen the structural fluctuations of the TSE. We infer that the TSE width should decrease as Φ_c increases. Similarly, the width of the TSE should be smaller in confined spaces compared to the bulk. In order to expand on these arguments, we address the following specific questions:

(1) What are the changes in the TSE of a protein in the presence of crowding agents? How does the TSE change in the related problem of folding in confined space? These questions are relevant when considering the excluded volume effects on the TSE changes.

(2) Does the inclusion of non-native interactions between the side chains and interactions with the cavity walls alter the nature of the TSE? The latter issue is of particular interest when considering protein folding in the GroEL cavity[12] or in the ribosome tunnel.[13] In these situations, the polypeptide chain interacts with the cavity walls whose strength can be modulated by protein sequence.

The questions posed above are addressed using a coarse-grained minimal model of proteins in which each amino acid is represented using its C_α atom and its side chain. The resulting C_α−SCM (SCM stands for side-chain model) representation of the WW domain (Figure 1) is used to probe the TSE movements in the presence of crowding agents. Determination of the structures of the TS has been a topic of much discussion that has centered largely on whether a set of low dimensional reaction coordinates suffices to obtain rates for a complex reaction such as biomolecular folding.[14−24] The limitations and the advantages of the methods that do not assume underlying reaction coordinates have been described elsewhere.[24,25]

There are several ways to define the transition state (TS). The TSE can be defined as the locus of structures along the stochastic separatrix that have equal probability (P_{fold}) of leading to the folded or unfolded states.[16] Such a definition is best suited when folding can be described using two-state kinetics but may become problematic if there are intermediates along the folding pathway.[24] The computationally demanding P_{fold} method has been applied to a number of systems.[26] Prior to the introduction of the P_{fold} method in the context of protein folding, Guo and Thirumalai[27] used a pattern recognition type clustering algorithm and ideas familiar in the context of crystallization of simple liquids to investigate the nucleation mechanism in protein folding. Here, we use the basic algorithm that has been automated and modified in subsequent studies[25,28,29] to answer the questions raised above. The automated progress variable clustering (PVC) algorithm also does not assume any underlying reaction coordinate, and is computationally far less demanding than the P_{fold} method.[25,28] In the PVC method, that has been numerically shown to be equivalent to the P_{fold} algorithm,[28,29] the TSE is computed from kinetic folding trajectories. Using PVC, we show that the bulk and TSEs in the presence of crowding agents are similar at $\Phi_c = 0.1$ and 0.15. As Φ_c increases, the width (measured using the root mean square deviation (rmsd) of the TS structures) of the TSE decreases. The TSE changes substantially when non-native interactions between the side chains are taken into account and when the polypeptide chain is in a hydrophobic cavity. The width of the TSE is broader. More interestingly, as the interaction strength (λ) between the protein and the cavity increases, certain long-range non-native contacts become enthalpically stabilized. Our simulations show that by modulating the value of λ one can change the nature of the barrier from being entropic (λ small) to being enthalpic ($\lambda > 1$).

Methods

Coarse-Grained Protein Models. Following our earlier studies,[10,30] we use the coarse-grained C_α−SCM representation of the three-stranded β-sheet WW domain with PDB ID 1PIN whose sequence and structure are given in Figure 1. In order to compare the effects of crowding and confinement on the nature of the TSE, we used model A which is a modified Go representation of the PIN WW domain.[10] The TSE movements in a nanopore were probed using model B[30] in which interactions between the proteins and the cavity interior are taken into account. Each amino acid (excluding Gly) is represented using two interaction sites, one corresponding to the C_α atom and the other the center of mass of the side chain. The sizes of the side chains correspond to the van der Waals radii of the side chains. The C_α-SCM has $(2N - N_G)$ beads where N_G is the number of glycines. Because the details of models A[10] and B[30] are given elsewhere, here we give only a brief description.

Model A. The polypeptide chain is represented using a modified Go potential in which nonbonded interactions between the side chains are modeled using statistical potentials. The structural Hamiltonian that includes bond-length potential, side chain−backbone connectivity potential, bond-angle potential, and dihedral potential follows a Go-like potential in which their equilibrium values are determined from the native structure.[30,31]

The interaction potential of a native side-chain contact pair is

$$E_{ij}^{NB} = \epsilon_{ij}\left\{\left(\frac{\sigma_{ij}}{r}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r}\right)^{6}\right\} \quad (1)$$

where a native contact between side chain i and side chain j, $|i - j| > 1$, is determined from the native structure using the CSU program.[32] $\sigma_{ij} = f(\sigma_i + \sigma_j)$, and σ_i and σ_j are the effective van der Waals radii of side chains in units of σ where σ is the distance between two adjacent C_α atoms. We take σ_i to be the distance from the center of mass of a given side chain to the C_α atom. The center of mass of the side chain is calculated using the PDB coordinates of the heavy atoms. With our definition, the effective van der Waals radii depend on the side chain as well its position in the native state. The values of σ_i are given in Table 1. To avoid clashes between bulky side chains, we use $f = 0.9$. For the native contact energies (ϵ_{ij}), we use the Betancourt−Thirumalai statistical potential.[33] For hydrogen bond interactions, we use the following anisotropic potential to account for the directional property of hydrogen bonds:

$$E_{ij}^{HB} = A(\rho)E_{ij}^{NB} \quad (2)$$

where $A(\rho)$ is

$$A(\rho) = \frac{1}{1 + B\left[(1 + \cos\rho)(1 - \cos\rho)\left(1 - \frac{\cos\rho}{\cos\rho_\alpha}\right)\right]^2} \quad (3)$$

ρ_α is the pseudo-dihedral angle of a canonical helical turn, 0.465 999 (rad) and $B = 1$. A native pair of hydrogen bonding between C_α^i and C_α^j is determined using DSSP.[34] For backbone hydrogen bonding between C_α^i and C_α^i, the value of ϵ_{ij} is 0.6 kcal/mol. The variable ρ, that is defined in ref 31, is the angular alignment between two interacting strands of backbones, and σ_{ij} (=1.22σ) is the hydrogen bond length. Following our earlier work,[10] we model the crowding agents as soft spheres. The size of the crowding particles is identical to the radius of gyration of the WW domain in the native state. All other interactions such as those between backbones and side chains, crowding agents and sites on the protein, or the nanopore−protein are repulsive (see ref 10 for additional details).

Model B. In this model, interactions between nonbonded side chains are explicitly taken into account. In other words, interactions between i and j, regardless of whether they are in contact in the native conformation, are given by eqs 1 and 2 depending on whether they are side-chain interactions or hydrogen bonding interactions. Features of the potential functions, such as the dihedral angle, hydrogen bond potentials are the same as in model A. Interaction between the side chains and the backbone C_α is repulsive. In order to include the influence of the pore, we also include interaction of the polypeptide chain with the walls of the spherical cavity that can be can be attractive depending on the chain conformation. The energy of interaction between the hydrophobic residues (F, L, W, V, I, M, Y, and A) and the pore is taken to be

$$E_{wp}^{s} = \epsilon\left[\left(\frac{\sigma}{|R_s - r_i|}\right)^{12} - \lambda\left(\frac{\sigma}{|R_s - r_i|}\right)^{6}\right] \quad (4)$$

where $\sigma \approx 3.8$ Å is the distance between two neighboring α-carbon atoms, r_i is the position of the ith hydrophobic chain, R_s is the radius of the nanopore, and the dimensionless parameter λ specifies the strength of attraction between the polypeptide chain and the cavity interior.

TABLE 1: Effective van der Waals Radii of Side Chains (eq 1) in Units of σ, where σ is the Distance between Two Adjacent C_α Atoms

residue index	residue type	σ	residue index	residue type	σ
0	LYS	0.734166	17	TYR	1.03927
1	LEU	0.685561	18	TYR	1.00308
2	PRO	0.494669	19	PHE	0.903388
3	PRO	0.488243	20	ASN	0.658735
4	GLY	0	21	HIS	0.836876
5	TRP	1.00771	22	ILE	0.62932
6	GLU	0.923236	23	THR	0.514048
7	LYS	0.995708	24	ASN	0.676364
8	ARG	1.02806	25	ALA	0.402245
9	MET	0.891618	26	SER	0.518079
10	SER	0.51835	27	GLN	0.919156
11	ARG	0.980229	28	TRP	0.975422
12	SER	0.516018	29	GLU	0.76824
13	SER	0.527387	30	ARG	1.18646
14	GLY	0	31	PRO	0.491916
15	ARG	1.02886	32	SER	0.524727
16	VAL	0.511977	33	GLY	0

Simulation Methods. We assume that the dynamics of the chain can be described by the Langevin equation. To enhance the conformational sampling of the WW domain, we used the replica exchange method (REM).[35,36] The conformations, in each replica, were generated by integrating the equations of motion in the low friction limit. For models A and B, we chose a broad range of temperatures (245−425 K) using 12−16 replicas. The rate of acceptance (based on the Metropolis criterion) of exchange between each pair of replicas varies from 30 to 50%. The Metropolis criterion was used in the exchange of replicas. The use of the REM and Langevin equation in the low friction limit[37,38] ensures that the thermodynamic properties, which are calculated using the weighted histogram analysis method (WHAM),[39] are converged.

For kinetic folding trajectories in model A, the initial structures are equilibrated at $T = 1.2T_f^b$ and 200 folding trajectories are generated for each condition at $0.83T_f^b$ (T_f^b, the folding temperature in bulk for model A is ≈342 K). For kinetic folding trajectories in model B, the initial structures are equilibrated with interactions with the pore at $T = 1.2T_f^b$ and 400−500 folding trajectories are simulated for each condition at $0.85T_f^b$ (T_f^b for model B in the bulk is ≈335 K). Brownian dynamics with a friction constant corresponding to water viscosity is used for the equations of motion. The natural time unit is 4.2 ns.

TSE Determination Using the Progress Variable Clustering (PVC) Algorithm. We determined the structures of the TSE using the clustering algorithm based on a pattern recognition search from a large number of folding trajectories. Since the details of the PVC algorithm are described in a number of previous studies,[25,28,29] we only provide a brief description here. The clustering stage is preceded by a search for the approximate location of the transition duration which is expressed in terms of a progress variable, δ. For the ith folding trajectory with the first passage time (τ_{1i}), the variable is $\delta = t/\tau_{1i}$. Let t_{TS} be the time after which the polypeptide chain reaches the native basin of attraction (NBA) with overwhelming probability. Because it satisfies $0 \le t_{TS} \le \tau_{1i}$, the range of δ is between 0 and 1. In order to locate the putative value of δ at the TS, namely, δ_{TS}, we computed the probability, $P(\delta)$, of forming native contacts averaged over all folding trajectories, as a function of δ. The putative TS region is associated with the value of δ_{TS} when $dP(\delta)/d\delta$ starts to rise rapidly. In practice, we find that the results

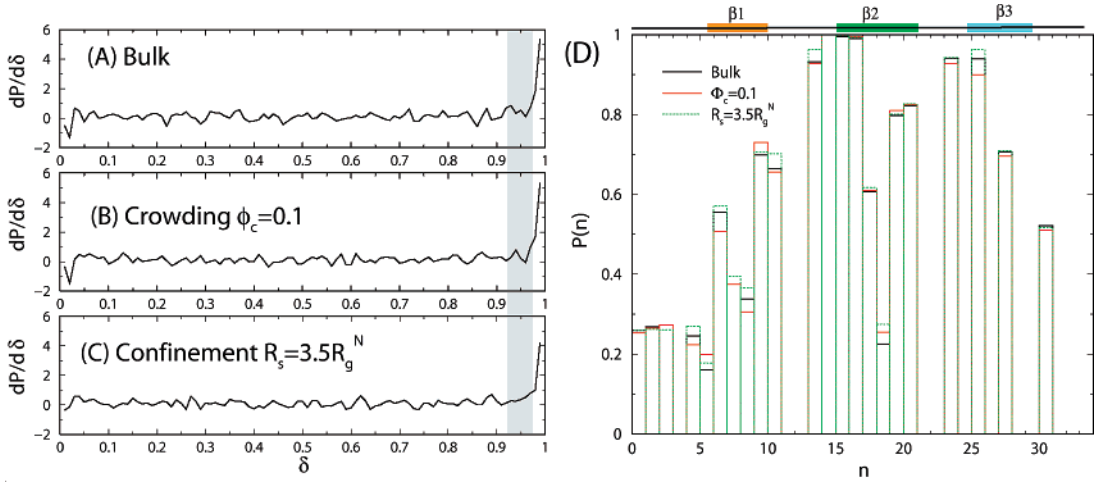

Figure 2. P is the probability of forming native contacts averaged over hundreds of folding trajectories along a dimensionless temporal axis δ that monitors the progress of the folding reaction. The data are for model A. Here, we show dP/dδ that gives a putative value of δ_{TS}, the region where a sharp increase in P takes place. Folding simulations were performed (A) in the bulk, (B) in the presence of crowding agents at a volume fraction of crowders, $\Phi_c = 0.1$, and (C) in a pure, spherical confinement with $R_s = 3.5R_g^N$. The TSE structures are collected at the onset of a steep raise in dP/dδ, which is boxed in a gray area ($0.93 < \delta < 0.97$). (D) Comparison of the probabilities of side-chain contact formation, $P(n)$, in the transition state region for each residue, n, under different conditions that are explicitly labeled. Data for $P(n)$ are computed from the gray area in parts A−C. The three colors on the top identify the three strands in the 1PIN WW domain (see Figure 1). Most of the residues in β_2 and β_3 and the interfaces between them are highly structured in the TSE. Crowding and confinements do not significantly alter the global nature of the TSE for model A.

are insensitive to the precise value of δ_{TS} as long as it is chosen in the neighborhood of a sharp increase in d$P(\delta)$/dδ.

Having determined the range for δ_{TS}, the conformations of the polypeptide chain in this region are clustered using a self-organized neural net. A fully automated progress variable clustering (PVC) algorithm,[25] that does not depend on the choice of reaction coordinates, is used to determine the structures of the TSE. Each conformation, obtained from 200 to 500 folding trajectories at the chosen value of δ_{TS}, is characterized by a vector, T, with elements $x_j = [x_{1j}, x_{2j}, \ldots, x_{Tj}]$, where j labels the conformation. The elements x_{kj}, $k = 1, 2, \ldots, T$, are the Euclidean distances between side chains in the conformation j whose knowledge uniquely determines the conformation of the polypeptide chain. The conformations for the ith trajectory are partitioned into clusters. A given conformation is assigned to cluster l if the Euclidean distance between the vector x_j and the center of mass of the lth cluster

$$C_l = \frac{1}{T_l}\sum_{j=1}^{T_l} x_j \qquad (5)$$

is less than a cutoff value, R_c. The assignment of the putative TS structures are clustered iteratively until there is no additional partitioning of the conformations among the various clusters which is indicated by the convergence of the C_k values. As shown elsewhere,[40] the stability of the clustering procedure is determined by two parameters, namely, R_c and the tolerance (D_c) used in defining contacts. Although one can choose R_c and D_c that depend on a particular folding trajectory, it is sufficient to use identical values for all of the trajectories. We explored a wide range of R_c and D_c to determine the robustness of the PVC procedure. For model A, the PVC method is stable as long as $1.1 \le D_c \le 1.6$ and $12\sigma \le R_c \le 16\sigma$; we choose $R_c = 15\sigma$ and $D_c = 1.3$. The stability ranges for model B are $1.1 \le D_c \le 1.6$ and $15\sigma \le R_c \le 18\sigma$; we choose $R_c = 16\sigma$ and $D_c = 1.3$. Additional details of the PVC algorithm are given elsewhere.[25,27,28,40] In order to obtain the TSE structures, we used the folding trajectories that were generated to obtain the results reported in our previous studies.[10,30] Because a large number of trajectories was generated, the statistical errors in the predicted transition state structures are small.

Results and Discussion

Residues at the Interface of β_2 and β_3 Are Structured in the TSE in the Bulk. We first characterized the TSE structures in the bulk ($\Phi_c = 0$) which not only serves as a reference but also can be used to validate the model by comparison to experiments.[41] By averaging all of the folding trajectories, we find (Figure 2A) that a value of $\delta_{TS} \approx 0.93$. Indeed, the value does not change in the presence of crowding agents ($\Phi_c = 0.1$) or under pure confinement in the absence of attractive interaction between the wall and the polypeptide chain (Figure 2B and C). Structures at the onset of rising P against δ ($0.93 < \delta < 0.97$) are collected for the TSE analysis. The TSE structures are obtained using the PVC algorithm (Methods) using the conformations in the δ_{TS} region. The extent to which the WW domain is structured is shown in Figure 2D which shows the probability that the nth side chain forms native contacts in the transition state. Somewhat surprisingly, the $P(n)$ distributions in the bulk, in the presence of crowding agents, and under confinement are nearly identical. This observation shows that when excluded volume interactions dominate, which can be mimicked by a Φ_c-dependent confinement in a spherical pore,[10] the global structural features of the TSE are not significantly altered at low Φ_c values (however, see below).

The TSE structures in the bulk compare favorably with experiments[42] that have used temperature-dependent Φ_T values to infer the overall transition state movement in a related WW domain of the Yes-associated protein (hYAP). On the basis of the Φ_T values for the wild type protein and the mutant W39F of the hYAP domain, it was shown[42] that the residues at the interface of the β_2 and β_3 strands (see Figure 1) are structured in the TSE at room temperature. The residues in this region form the hydrophobic core. For the 1PIN domain considered here, we find that the residues in the β_2 and β_3 strands form with high probability in the TSE (Figure 2D). Residues in other regions of the three-stranded β-sheet proteins are much less structured in the transition state. The predicted similarity

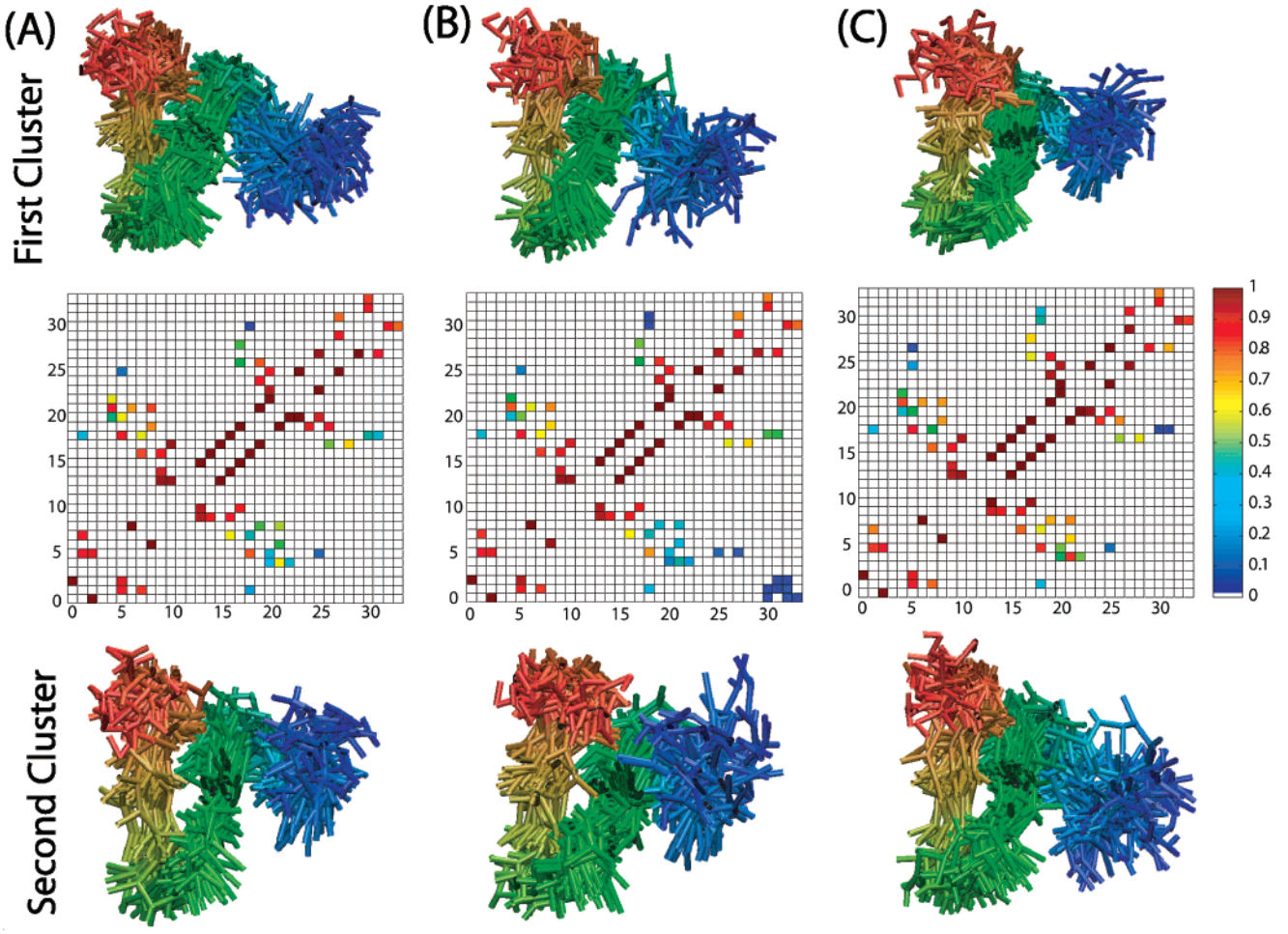

Figure 3. The top panel gives a superposition of the TSE structures from the first dominant cluster for model A of the 1PIN WW domain. TSE structures are for (A) the bulk case, (B) in the presence of crowding particles with $\Phi_c = 0.1$, and (C) the purely spherical confinement case with $R_s = 3.5R_g^N$. The middle panel gives maps of contact formation for TSE structures for each system. The upper left triangle, in each contact map, is for structures in the top panel. The contact map is colored by the probability of contact formation. The extent of contact formation is given by the color scale on the right that goes from zero to unity. The analysis of the structures of the TSE in this cluster is given in Table 2. The superposition of structures in the second dominant cluster for the three conditions is given below the contact maps. The corresponding contact maps are shown in the lower right triangle in the middle panel. Both the structures and the contact maps show fraying in β_3 and expansion of the core of the protein formed from β_2 and β_3.

between the TSE in the 1PIN and hYAP WW domains are not surprising based on the chemical sequence entropy $S_{CSE}(n)$.[30] It was shown that $S_{CSE}(n)$, which measures the conservation of the chemical character of the nth residue, is nearly perfectly conserved ($S_{CSE}(n)$ values are nearly zero) for a number of residues in the β_2 and β_3 strands. The present computations, experiments,[42] and calculations based on $S_{CSE}(n)$ suggest that the topology is an indicator of the TSE structures especially in the bulk. The good agreement between the simulations and experiments in the bulk[42] also validates the use of C_α−SCM in obtaining the structures of the TSE for the WW domain.

Transition State Ensemble Is Not Dramatically Altered by Crowding and Confinement for Model A. Using the PVC algorithm, we determined the transition state structures in the presence of crowding agents and in a spherical pore using model A. Just as in the bulk, the TSE structures predominantly partition into two dominant clusters. In the most dominant cluster, the overall topology of the WW domain under crowding and confinement is similar to that in the bulk (see the upper panel and the contact map in all three parts of Figure 3). Maps of the probability of contact formation and the associated structures show that on average the structures in the second dominant cluster are more open with considerable disorder in β_3 (see lower panels in Figure 3). The TSE structures show (Figure 3) the formation of short-ranged contacts that are close to the diagonal region in the contact maps. Interestingly, when the size of the pore decreases ($R_c = 3.05R_g^N$) or the volume fraction of crowders increases ($\Phi_c = 0.15$), the rmsd of the first dominant cluster decreases (Table 2), indicating that the width of the TSE decreases. We conclude that the global features of the transition state the folds are unperturbed at $\Phi_c \leq 0.15$ and with a pore radius as small as $R_s = 3.05R_g^N$. The observation that structures of the TSE do not change significantly at these volume fractions validates the assumption[10] that crowding due to excluded volume predominantly affects the denatured state ensemble.

Close inspection of the structural characteristics of the TSE shows changes in the overall shapes in the presence of crowding agents and in spherical cavities. The radius of gyration of the TSE is about $\approx(1.1-1.2)R_g^N$ which reflects the similarity in the topology of the TSE structures in the bulk and in the presence of crowding agents. We find that R_g is not a good measure of the structures of the TSE. In contrast, there are substantial differences in the shapes of the WW domain between the bulk case and due to crowding and confinement (see Table 2). Following our previous study, we use the asphericity parameter

TABLE 2: Structural Features of the TSE in the First Dominant Cluster of the WW Domain Using Model A

model	percentage[a]	rmsd[b]	$\langle R_g\rangle^c/R_g^N$	$\langle\Delta\rangle^d/\Delta^N$	$\langle S\rangle^e/S^N$
(a) bulk	34	0.92 ± 0.02	1.22	0.97	0.04
(b) $R_s = 3.5R_g^N$	30	0.81 ± 0.01	1.16	0.60	0.02
(c) $R_s = 3.05R_g^N$	36	0.79 ± 0.01	1.17	0.55	0.05
(d) $\Phi_c = 10\%$	37	0.98 ± 0.01	1.21	1.02	0.17
(e) $\Phi_c = 15\%$	45	0.72 ± 0.02	1.13	0.55	0.13

[a] Denotes the percentage of structures in the most dominant cluster. [b] rmsd (in units of σ) is computed with respect to the average structure in the cluster. [c] $\langle R_g\rangle$ is the average value of the radius of gyration, and $R_g^N = 10.1$ Å is the value in the native state. [d] $\langle\Delta\rangle$ is the average asphericity parameter, and $\Delta_N = 0.16$ is the value in the native state. [e] $\langle S\rangle$ is the average shape parameter, and $S_N = 0.10$ is the value in the native state.

(Δ) and the shape parameter (S) to characterize the overall shape of the TSE structures. The shape parameters Δ and S are computed using the inertial tensor.[43] For a sphere $\Delta = 0$ and $S > 0$ (<0) corresponds to a prolate (oblate) ellipsoid. The values of Δ in the bulk and at $\Phi_c = 0.10$ are similar to that in the native state. However, the TSE structures are markedly aspherical when confined to a pore or when Φ_c is increased to 0.15. Comparison of the $\langle S\rangle$ values shows that the TSE in all cases is more oblate than in the native state. The distribution of the S values (data not shown) shows that there are a large number of TSE structures that are oblate ($S \leq 0$) while the WW domain in the native state is prolate. This implies that in the transition to the native state there has to be a substantial change in the shape especially in the presence of crowding agents. These results show that crowding and confinement induce substantial perturbations in the shapes of TSE structures compared to the bulk. The present analysis shows that changes in the shapes of polypeptide chains may be the slow degrees of freedom in the transition from the TSE to the native state, and could serve as a suitable low dimensional reaction coordinate.

Non-native Side-Chain Contacts and Interaction with the Pore Dramatically Alters the Nature of the TSE. In many situations, such as proteins encapsulated in the GroEL cavity or a polypeptide chain in the exit tunnel of the ribozome, there are water-mediated interactions between the polypeptide chain and the walls of the cavity. We have used model B to mimic such interactions. The strength of the interactions is given by the dimensionless parameter λ. We had shown in a previous study[30] that at an optimal value of λ (=1) not only is there a modest enhancement in folding rate compared to the bulk but also an increase in the yield of the native state. The superposition of the structures in the first dominant cluster of the TSE for model B in the bulk and in a spherical cavity ($R_s = 5R_g^N$) at various values of λ, including $\lambda = 0$, is shown in Figure 4. While the outlines of the global topology of the WW domain may be discerned in Figure 4, it is clear that the structures are much more disordered compared to those shown in Figure 3. The TS region is, in all likelihood, much broader when non-native interactions are taken into account. These results suggest that as long as λ is not too large inclusion of non-native interactions can help in the folding process itself.[28,44] A more detailed analysis of the TSE structures (see Table 3) of the dominant cluster shows that on average the TSE structures are compact. Inclusion of non-native interactions or addition of pore−WW domain interaction (λ is nonzero) results in aspherical and prolate ellipsoids for the overall shapes of the TSE structures (Table 3).

In order to further illustrate the role of λ in modulating the TSE structures, we have compared, for model B, the TSE for

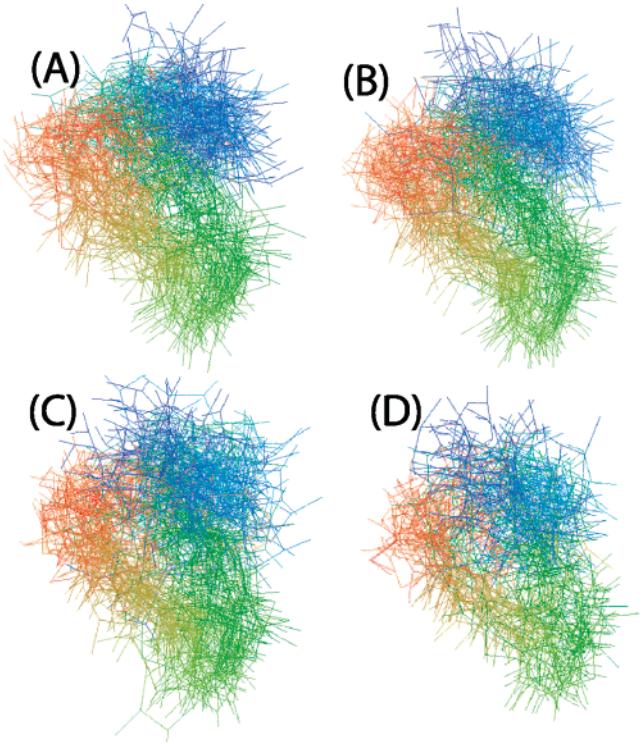

Figure 4. Superposition of TSE structures from the first dominant cluster for model B of the WW domain. TSE structures are for (A) the bulk case, (B) the purely spherical confinement with $R_s = 5R_g^N$ and $\lambda = 0$, (C) the spherical confinement with $R_s = 5R_g^N$ and $\lambda = 1$, (D) the spherical confinement with $R_s = 5R_g^N$ and $\lambda = 3$. R_s is the size of the spherical nanopore, and λ is given in eq 4. The analysis of the structural characteristics is further given in Table 3. R_s is the size of the spherical nanopore, and λ is the strength of the pore−protein interaction.

TABLE 3: TSE Structure Characteristics from the First Dominant Cluster of Model B of the WW Domain [The Columns Have the Same Meaning as in Table 2; λ Gives the Strength of Interaction between the WW Domain and the Nanopore (see eq 4)]

model	percentage	rmsd	$\langle R_g\rangle/R_g$	$\langle\Delta\rangle/\Delta^N$	$\langle S\rangle/S^N$
(a) bulk	36	1.42 ± 0.02	1.05	0.87	0.73
(b) $R_s = 5R_g^N$, $\lambda = 0$	44	1.47 ± 0.02	1.04	0.80	0.59
(c) $R_s = 5R_g^N$, $\lambda = 1$	49	1.60 ± 0.02	1.04	0.82	0.62
(d) $R_s = 5R_g^N$, $\lambda = 3$	31	1.62 ± 0.02	1.03	0.90	0.85

$\lambda = 0$ and $\lambda = 3$ with respect to structures when $\lambda = 1$. The probabilities of native and non-native contact formation in the TS for both native contact pairs are computed for the TSE in the first dominant cluster for the WW domain confined to a sphere with different wall−protein interactions (λ). The differences of these contact maps with respect to the contact probability at $\lambda = 1$ are shown in Figure 5. We chose $\lambda = 1$ as a reference because in the previous[30] study the rate of folding at this value was higher than either $\lambda = 0$ and $\lambda = 3$. More generally, it has been shown that rate enhancement is observed at optimal values of pore−peptide interactions.[30,45,46]

Figure 5A shows the difference contact map between $R_s = 5R_g^N$, $\lambda = 0$, and $R_s = 5R_g^N$, $\lambda = 1$. Data in the upper triangle represent native side-chain contacts, and those in the lower triangle are non-native ones. The TSE structures from the pure confinement case ($\lambda = 0$) favor the formation of short-range native as well as non-native contacts (shown in blue boxes in Figure 5A). Inclusion of optimal protein−wall attractions ($\lambda = 1$) helps rescue kinetically trapped structures that are mostly dominated by the formation of short-range non-native contacts.

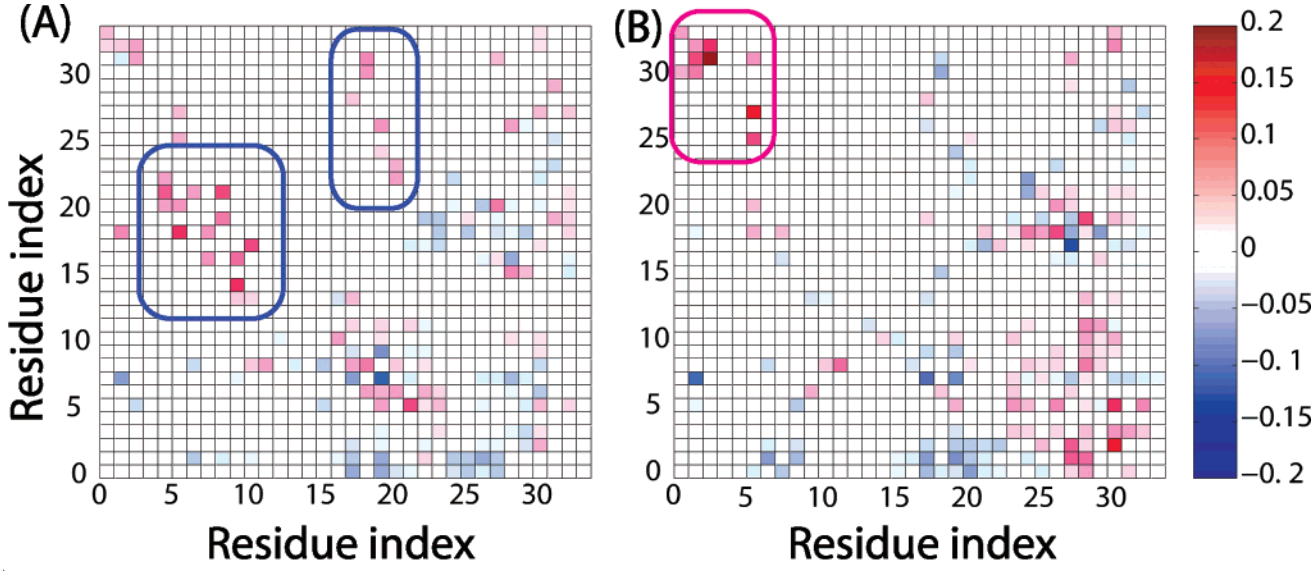

Figure 5. Differences in the probability of contact formation in TSE structures from the first dominant cluster in the form of a contact map for model B of the 1PIN WW domain. (A) Difference contact map between $R_s = 5R_g^N$, $\lambda = 0$, and $R_s = 5R_g^N$, $\lambda = 1$. (B) Same as part A except for $R_s = 5R_g^N$, $\lambda = 3$, and $R_s = 5R_g^N$, $\lambda = 1$. The data in the upper triangle are for native contact pairs, and those in the lower triangle are for non-native contact pairs. Both of the difference contact maps were computed with respect to the TSE structures from $R_s = 5R_g^N$, $\lambda = 1$, to demonstrate that a change in the wall−protein interactions can alter the TSE structures. The maximum change occurs in the regions as circled in blue (A) and in red (B). The extent of the difference is quantified using the color scale on the right. The largest changes occur in residues in β_1 (Figure 1).

Figure 5B shows the differences between contact map $R_s = 5R_g^N$, $\lambda = 3$, and $R_s = 5R_g^N$, $\lambda = 1$. The distributions of the TSE structures are sensitive to the sequence detail when wall−protein interactions are strong. Contacts associated with tryptophan residues (residues 5 and 28) are most susceptible to this interaction. Figure 5B shows that a larger wall−protein interaction results in the long-range contact formation associated with 5 and/or 28 (shown in red box), while the probability of short-ranged contacts is much less than $\lambda = 1$. However, these interactions also lead to long-range non-native contacts associated with these large hydrophobes. In both parts A and B of Figure 5, the distribution of TSE structures is highly sensitive to the protein−wall interactions. These calculations show that there is a subtle interplay between non-native contacts in the protein and pore−protein interactions. At modest values of λ, incorrect long-range contacts can be resolved, whereas, as λ increases, such contacts are enthalpically stabilized in the TSE, resulting in small refolding rates and highly dispersed transition state structures.

Conclusions

Folding rates of proteins in the presence of crowding or in confined spaces are typically, but not always, larger compared to the bulk. The rate enhancement is readily explained in terms of entropic destabilization of the denatured states.[11,47] More generically, we can envision two extreme scenarios: (i) One is based on the notion of the entropic stabilization which is applicable when excluded volume interactions between the protein and the crowding agents dominate or when the protein is encapsulated in an inert pore. (ii) The other extreme case corresponds to the enthalpic limit in which nonspecific interaction between the protein and the pore can profoundly influence folding rates and stability. In the entropic stabilization limit, we expect that the TSE should be similar to that in bulk as long as the native state is not altered. On the other hand, the TSE changes cannot be easily predicted for the case when enthalpic contribution affects stability and folding rates.

In order to explore both of the limits, we used a fully automated algorithm to investigate the changes in the structures of the TSE in the presence of macromolecular crowding effects and upon confinement to spherical pores. For illustrative purposes, we used the WW protein that has been well-characterized in the bulk[42] and whose folding has been previously described using C_α Go models.[48] The TSE structures are obtained from the folding trajectories without assuming any model for the reaction coordinate. The TSE structures in the bulk partition into two distinct clusters. In both of the clusters, the structures are, on average, compact, while they are more structured in the dominant cluster than in the subdominant (less populated) cluster. For the modified C_α−SCM Go-like models (model A), whose folding in the presence of crowding agents is well explained by the entropic stabilization scenario,[10] the TSE structures due to crowding and confinement effects are similar to the bulk. The TSE structures are compact and native-like in which the contacts in the β_2 and β_3 strands are formed with substantial probability. However, when the volume fraction increases or the pore size decreases, the distribution of TSE in the first dominant cluster is narrower. The most significant changes in the transition from the unfolded state to the folded state occur in the shape of the WW domain. The WW domain samples oblate conformations in the transition state even though the native conformation is prolate.

When enthalpic interactions become relevant, the nature of the TSE changes greatly. In this case, the TSE structures are much less ordered and the TS region may itself be broad. The inherent plasticity in these structures suggests that modest non-native interactions or pore−protein interactions can lead to escape from kinetic traps or in the annealing of TS structures that have long-range non-native contacts. When the pore−protein interactions are large, the formation of long-range contact TS structures can impede the folding rates.

Although we have used the WW domain as an example, the conclusions are fairly general. We predict that the nature of the TSE does not change significantly when folding occurs in inert

pores or in the presence of crowding agents when excluded volume interactions dominate. On the other hand, when there are interactions between crowding agents and the protein, the changes in the TSE are harder to predict. However, we propose that at modest values of λ the potential long-range non-native contacts, that can more easily form in restricted spaces or when crowding agents are present, can be more easily resolved. Our results also show that shape parameters (Δ and S) which change substantially during the folding reaction are appropriate low dimensional folding reaction coordinates.

Acknowledgment. M.S.C. is grateful to A. P. Sloan foundation for a postdoctoral fellowship and the new faculty start-up fund from the University of Houston. The computations were performed on the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputing Center and Teragrid (TG-MCB050048N). This work was supported in part by a grant from the National Institute of Health through Grant no. (1R01GM067851-01) and the National Science Foundation (CHE-05-14056).

References and Notes

(1) Ellis, R. J.; Minton, A. P. *Nature* **2003**, *425*, 27−28.
(2) Ellis, R. J. *Trends Biochem. Sci*. **2001**, *26*, 597−604.
(3) Asakura, S.; Oosawa, F. *J. Chem. Phys*. **1954**, *22*, 1255−1256.
(4) Zimmerman, S. B.; Minton, A. P. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 27−65.
(5) Minton, A. P. *Biophys. J.* **2005**, *88*, 971−985.
(6) Spencer, D. S.; Xu, K.; Logan, T. M.; Zhou, H. X. *J. Mol. Biol*. **2005**, *351*, 219−232.
(7) van den Berg, B.; Ellis, R. J.; Dobson, C. M. *EMBO J.* **1999**, *18*, 6927−6933.
(8) Sasahara, K.; Mcphie, P.; Minton, A. P. *J. Mol. Biol*. **2003**, *326*, 1227−1237.
(9) Dedmon, M. M.; Patel, C. N.; Young, G. B.; Pielak, G. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12681−12684.
(10) Cheung, M. S.; Klimov, D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4753−4758.
(11) Klimov, D. K.; Newfield, D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8019−8024.
(12) Thirumalai, D.; Lorimer, G. H. *Ann. Rev. Biophys. Biomol. Struct*. **2003**, *30*, 245−269.
(13) Woolhead, C. A.; McCormick, P. J.; Johnson, A. E. *Cell* **2004**, *116*, 725−736.
(14) García, A. E. *Phys. Rev. Lett.* **1992**, *68*, 2696−2699.
(15) Onuchic, J.; Socci, N.; Luthey-Schulten, Z.; Wolynes, P. *Fold. Des.* **1996**, *1*, 441−450.
(16) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. *J. Chem. Phys.* **1998**, *108*, 334−350.
(17) Shoemaker, B. A.; Wang, J.; Wolynes, P. G. *J. Mol. Biol*. **1999**, *287*, 675−694.
(18) Nymeyer, H.; Socci, N. D.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 634−639.
(19) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11−27.
(20) Shea, J.-E.; Brooks, C. L.; Onuchic, J. N. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499−535.
(21) Portman, J. J.; Takada, S.; Wolynes, P. G. *J. Phys. Chem*. **2001**, *114*, 5082−5096.
(22) Li, L.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13014−13018.
(23) Hummer, G. *J. Chem. Phys*. **2004**, *120*, 516−523.
(24) Cho, S.; Levy, Y.; Wolynes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 586−591.
(25) Klimov, D. K.; Thirumalai, D. *Chem. Phys.* **2004**, *99*, 8019−8024.
(26) Deng, F.; Guo, W.; Dokholyan, N. V.; Shakhnovich, E. I.; Shea, J.-E. *J. Mol. Biol.* **2005**, *350*, 1035−1050.
(27) Guo, Z.; Thirumalai, D. *Fold. Des*. **1997**, *2*, 377−391.
(28) Klimov, D. K.; Thirumalai, D. *Proteins* **2001**, *43*, 465−475.
(29) Klimov, D. K.; Thirumalai, D. *J. Mol. Biol*. **2005**, *353*, 1171−1186.
(30) Cheung, M. S.; Thirumalai, D. *J. Mol. Biol*. **2006**, *357*, 632−643.
(31) Cheung, M. S.; Finke, J. M.; Callahan, B.; Onuchic, J. N. *J. Phys. Chem. B* **2003**, *107*, 11193−11200.
(32) Sololev, V.; Wade, R.; Vriend, G.; Edelman, M. *Proteins* **1996**, *25*, 120−129.
(33) Betancourt, M. R.; Thirumalai, D. *Protein Sci.* **1999**, *8*, 361−369.
(34) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.
(35) Sanbonmatsu, K. Y.; García, A. E. *Proteins* **2002**, *46*, 225−234.
(36) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(37) Veitshans, T.; Klimov, D. K.; Thirumalai, D. *Fold. Des.* **1997**, *2*, 1−22.
(38) Honeycutt, J. D.; Thirumalai, D. *Biopolymers* **1992**, *32*, 695−709.
(39) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem*. **1992**, *13* (8), 1011−1021.
(40) Klimov, D. K.; Thirumalai, D. *J. Mol. Biol*. **1998**, *282* (2), 471−492.
(41) Jager, M.; Nguyen, H.; Crane, J. C.; Kelley, J. W.; Gruebele, M. *J. Mol. Biol*. **2001**, *311*, 373−393.
(42) Crane, J. C.; Koepf, E. K.; Kelley, J. W.; Gruebele, M. *J. Mol. Biol.* **2000**, *298*, 283−292.
(43) Dima, R. I.; Thirumalai, D. *J. Phys. Chem. B* **2004**, *108*, 6564−6570.
(44) Clementi, C.; Plotkin, S. S. *Protein Sci.* **2003**, *13*, 1750−1766.
(45) Betancourt, M. R.; Thirumalai, D. *J. Mol. Biol*. **1999**, *287*, 627−644.
(46) Jewett, A. I.; Baumketner, A.; Shea, J.-E. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *101*, 13192−13197.
(47) Takagi, F.; Koga, N.; Takada, S. *Proc. Natl. Acad. U.S.A.* **2003**, *100*, 11367−11372.
(48) Karanicolas, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3954−3959.