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

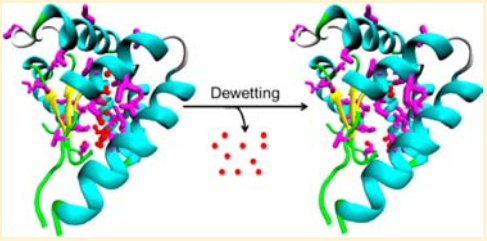
Interplay between Drying and Stability of a TIM Barrel Protein: A Combined Simulation–Experimental Study

Payel Das,[†] Divya Kapoor,[‡] Kevin T. Halloran,[‡] Ruhong Zhou,*[,†,§] and C. Robert Matthews*[,‡]

[†]Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, United States
[‡]Department of Biochemistry & Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, United States
[§]Department of Chemistry, Columbia University, New York, New York 10027, United States

Supporting Information

ABSTRACT: Recent molecular dynamics simulations have suggested important roles for nanoscale dewetting in the stability, function, and folding dynamics of proteins. Using a synergistic simulation–experimental approach on the αTS TIM barrel protein, we validated this hypothesis by revealing the occurrence of drying inside hydrophobic amino acid clusters and its manifestation in experimental measures of protein stability and structure. Cavities created within three clusters of branched aliphatic amino acids [isoleucine, leucine, and valine (ILV) clusters] were found to experience strong water density fluctuations or intermittent dewetting transitions in simulations. Individually substituting 10 residues in the large ILV cluster at the N-terminus with less hydrophobic alanines showed a weakening or diminishing effect on dewetting that depended on the site of the mutation. Our simulations also demonstrated that replacement of buried leucines with isosteric, polar asparagines enhanced the wetting of the N- and C-terminal clusters. The experimental results on the stability, secondary structure, and compactness of the native and intermediate states for the asparagine variants are consistent with the preferential drying of the large N-terminal cluster in the intermediate. By contrast, the region encompassing the small C-terminal cluster experiences only partial drying in the intermediate, and its structure and stability are unaffected by the asparagine substitution. Surprisingly, the structural distortions required to accommodate the replacement of leucine by asparagine in the N-terminal cluster revealed the existence of alternative stable folds in the native basin. This combined simulation–experimental study demonstrates the critical role of drying within hydrophobic ILV clusters in the folding and stability of the αTS TIM barrel.



INTRODUCTION

The folding of proteins following synthesis on a ribosome or dilution from a chemically denatured state involves the formation of numerous van der Waals interactions, hydrogen bonds, and electrostatic interactions that stabilize the compact native conformation. It is widely accepted that a necessary structural consequence of the protein-folding reaction is the exclusion of water from the side chains and main chains that become buried in the native state. The thermodynamic consequence of the dehydration reaction reflects the substantial gain in entropy realized by freeing water during folding.

The role of water in protein-folding reactions has been examined by both experimental and computational approaches. Mutational analyses in which nonpolar side chains were replaced with isosteric polar side-chain analogues showed that water is selectively shed prior to the appearance of the native state to enable the formation of critical cores of stability in early intermediates[1] or transition-state ensembles.[2,3] By contrast, time-resolved IR analysis revealed dehydration of the main-chain amides in the final step of folding from the alkaline-denatured states of both α-helical[4] and β-sheet proteins.[5] A third experimental approach for examining the role of water in folding monitored the protection of main-chain amide hydrogens against exchange in deuterated water in partially folded states[6,7] and folding intermediates.[8−10] When hydrogen exchange (HX) techniques were applied to a pair of $(\beta\alpha)_8$ TIM barrel proteins,[10,11] protection against exchange in folding intermediates was found to be selectively associated with clusters of branched aliphatic side chains [isoleucine, leucine, and valine (ILV) clusters]. The molecular rationale for this behavior was ascribed to the preferential partitioning of side-chain analogues of saturated hydrocarbon moieties into the vapor phase relative to their aromatic, sulfur, or polar-containing counterparts that spontaneously dissolve in water.[12] The Branched Aliphatic Side Chain (BASiC) hypothesis was formulated on the basis of these differential solubilities and proposes that clusters of ILV side chains play crucial roles in stabilizing folding intermediates in TIM barrel proteins by selectively excluding water from their interiors.[13,14]

From a computational perspective, nanoscale dewetting transitions[15−17] between hydrophobic surfaces have long been of interest for both physical[16,18−22] and biological systems.[16,19,23−29] Previous molecular dynamics (MD) simulation

Received: October 25, 2012
Published: January 7, 2013

 © 2013 American Chemical Society dx.doi.org/10.1021/ja310544t | J. Am. Chem. Soc. 2013, 135, 1882−1890

studies have identified several proteins or peptides in which a dewetting transition was observed prior to the docking of preformed elements of secondary structure. For example, a remarkable dewetting transition was observed within the nanoscale channel between the four melittin α-helices, each of whose hydrophobic interface comprises three isoleucines, four leucines, one tryptophan, and two valines.[29] A subsequent study on a variety of protein complexes (dimers, tetramers, and two-domain proteins) found that dewetting requires large complementary hydrophobic surfaces with significant contributions from ILVs.[24] In contrast, a marked decrease in water density was not detected at the domain interface in the two-domain 2,3-dihydroxybiphenyl dioxygenase (BphC).[23] The domain interface in BphC is relatively heterogeneous in nonpolar side chains.

Building on the results of the previous experiments and MD simulations, we adopted a combined experimental and computational approach to test the conjecture that large ILV-rich clusters in TIM barrel proteins are prone to undergo dewetting from their interiors. As a target, we chose the α subunit of Trp synthase (αTS), a ~28 kDa TIM barrel $(\beta\alpha)_8$ protein that is a component of the $\alpha_2\beta_2$ tetrameric tryptophan synthase complex. Previously, αTS was observed to offer strong and selective protection against HX in an on-pathway intermediate associated with a large N-terminal ILV cluster.[7] A smaller C-terminal ILV cluster does not offer protection against HX and provides an internal control. As a surrogate for the polarity introduced by water, two buried leucines in the N-terminal cluster and a single leucine in the C-terminal cluster were individually replaced with isosteric, polar asparagines. The effects of these mutations on the water density within the clusters were predicted by MD simulations of artificially displaced versions of their preformed β-sheet and α-helical components. These predictions were then compared with the effects of the mutations on the experimentally determined stabilities and structures of the native state and the folding intermediate. The possibility that wetting could also be enhanced by replacing a buried cysteine adjacent to the N-terminal ILV cluster with an asparagine was also studied. The combined results support the conclusion that the drying of the large N-terminal ILV cluster is crucial to the stability and structure of the native state and of a productive folding intermediate in a TIM barrel protein.

MATERIALS AND METHODS

MD Simulations. The initial structures of the ILV clusters were taken from the crystal structure deposited in the Protein Data Bank (PDB entry 1BKS). The clusters were determined using the protocol described previously.[14] Three hydrophobic ILV clusters are found within the αTS native structure: (i) a large external-to-the-barrel cluster spanning the N- and the C-termini (cluster 1), (ii) a second external-to-the-barrel cluster in the C-terminal region (cluster 2), and (iii) an internal-to-the barrel cluster (cluster 3) (Figure 1a). The helical parts of the clusters were pulled 4−6 Å away from the β-sheet region to create the cavity for investigating dewetting. The system was solvated in a box of TIP3P water. The initial state of the cavity for all systems was set to be wet, unless otherwise stated. The resulting systems were minimized for 10 000 steps followed by a 16 ns MD simulation at 310 K and 1 atm. During this simulation, the protein heavy atoms were constrained, whereas water molecules were free to move. The particle-mesh Ewald (PME) method was used for the long-range electrostatic interactions, while the van der Waals interactions were treated with a cutoff distance of 12 Å. The CHARMM force field (c32b1 parameter set) was used, and the simulations were performed using the NAMD2 molecular modeling package with a 2 fs time step.

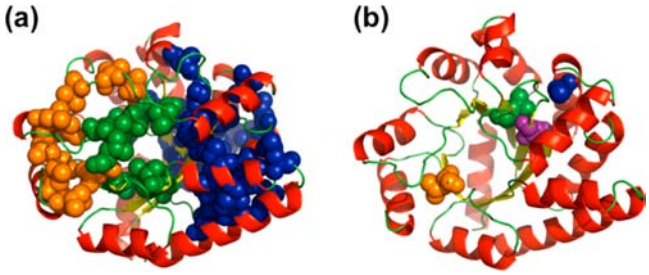


Figure 1. Ribbon diagrams of αTS (a) highlighting the three hydrophobic clusters formed by the ILV residues [cluster 1 (blue), cluster 2 (orange) and cluster 3 (green)] obtained using a 4.2 Å cutoff distance between pairs of ILV side chains and (b) showing the location of hydration mutations in the crystal structure [Leu50 in β2 (purple), Cys81 in α2 (blue), Leu99 in β3 (green), and Leu176 in β6 (orange)]. Coordinates of αTS from *Salmonella typhimurium* were used to generate the figure from a refined version of PDB entry 1BKS using PyMOL version 1.3.[30]

At least 15 different trajectories were run for cluster 1 and cluster 2 of the wild-type (WT) protein and its C81, L99, and L176 mutants (in silico variants). For all other systems, at least three different trajectories were run. The total aggregate simulation time was about 5 μs. Additionally, we performed ~100 ns long "folding" simulations in which both the main-chain and side-chain atoms were free to move.

Site-Directed Mutagenesis. The codon-optimized wild-type αTS gene was synthesized by Genscript in pUC 57 and recloned into a modified pGS-21a vector with an N-terminal His_6 tag and tobacco etch virus (TEV) protease site using EcoRV and *Bam*HI restriction sites. Various hydration mutations were made with mutagenic oligonucleotides purchased from Integrated DNA Technologies using the Stratagene QuikChange site-directed mutagenesis kit, and the mutations were confirmed using DNA sequencing [a full list of mutants is shown in Table S1 in the Supporting Information (SI)]. The pGS-21a plasmid DNA was transformed into BL21 (DE3) pLysS cells for protein expression and purification (see the SI for protein purification details).

Equilibrium and Kinetic Unfolding Experiments. The thermodynamic properties of both wild-type αTS and various hydration mutants were determined by titrations with urea as the denaturant on a Jasco J-810 spectropolarimeter. Samples at varying urea concentrations were prepared using a Hamilton 540B automatic titrator and incubated overnight at 25 °C for complete equilibration. Data were collected using a quartz cuvette with a 2 mm path length at a bandwidth of 2.5 nm. The spectra were recorded every 1 nm over the wavelength range from 215 to 260 nm at a scan speed of 50 nm min^{-1} with an averaging time of 8 s. The denaturant dependences of the ellipticities for αTS and its variants were fit to a three-state model using Savuka, an in-house nonlinear least-squares program, assuming a linear dependence of the free energy of unfolding on the denaturant concentration.[31] These fits provided the free energy differences among the three thermodynamic states, the denaturant dependences of these free energy differences, and the Z parameter required to estimate the ellipticity of the intermediate.[32]

The manual-mixing kinetic unfolding jumps began in the absence of denaturant and ended at urea concentrations of 4.0−6.0 M, with the final protein concentration ranging from 3 to 5 μM. Data were collected at 222 nm and 25 °C in a cuvette with a path length of 1 cm. The relaxation times were obtained by fitting the kinetic traces to a single-exponential function in Savuka.[31]

RESULTS

A ribbon diagram of αTS showing the location of its three ILV clusters is presented in Figure 1a. Cluster 1, which contains 31 ILVs, forms the interface between the exterior of the β-barrel and the interior of the α-helical shell in $(\beta\alpha)1-4$. Cluster 2, which contains 12 ILVs, is found at the interface between the β-

dx.doi.org/10.1021/ja310544t | J. Am. Chem. Soc. 2013, 135, 1882−1890

barrel and the α-helical shell in $(\beta\alpha)5-6$. Cluster 3 is located in the interior of the cylindrical barrel and is formed from eight ILVs individually contributed by seven of the eight β-strands and α-helix 0 at the N-terminus.

MD Simulations of Hydration in ILV Clusters. *Hydration in Cluster 1.* A cavity inside cluster 1, with an estimated volume of ~1300 Å^3, was created by pulling the α1 and α2 helices away from the β1, β2, and β3 strands by a separation distance d varying from 4 to 6 Å (Figure 2a) and filled with water molecules. Previous reports by us and others have described nanoscopic dewetting transitions in proteins with cavity volumes on a similar order.[24,33] During the 16 ns simulation time, the water molecules were free to move, but the protein heavy atoms remained constrained. Figure 2b shows plots of water density as a function of simulation time for cluster 1 at d = 4 and 6 Å. The cavity underwent intermittent transitions between wet and dry states at a separation distance of 4 Å, but no drying transition was observed at d = 6 Å. These drying transitions typically occur in 200−300 ps. To check the convergence of our results, we ran simulations starting from two different initial states with d = 4 Å: the "wet" state and a "dry" state in which all of the initial water molecules were removed manually to create a dry cavity (Figure 2c; longer time scale simulations can be found in Figure S1 in the SI). Within the first 1−4 ns, the cavity underwent wetting/dewetting transitions in which the normalized water density inside the cavity switched between a maximum of 0.8 (wet) and a minimum of 0.2 (dry) from both initial states. The normalized water density was obtained by dividing the number of water molecules by the maximum number of water molecules inside the cavity. Snapshots of the cavity in the wet and dry states (Figure S2 in the SI) suggested that the water density was lowest near the center of the cavity, as a vapor bubble was frequently formed in this region and was stable for several nanoseconds. The two termini of the β-strand triplet remained relatively wet, the N-terminus being drier than the C-terminus. The latter results are consistent with the stronger protection against amide hydrogen exchange with solvent in this region observed in native-state HX experiments.[11]

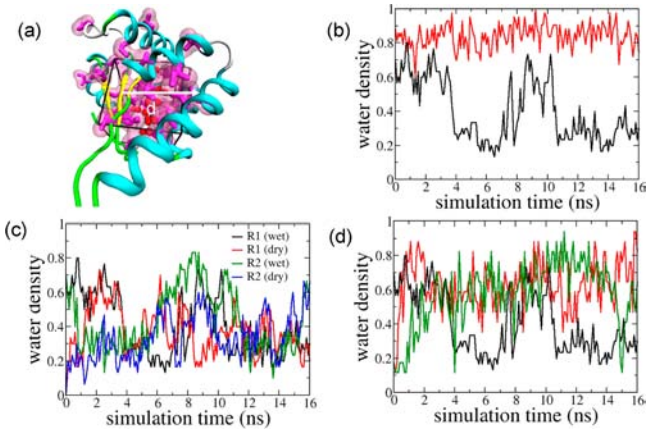


Figure 2. (a) Ribbon representations of ILV cluster 1, in which the helices (shown in cyan) are manually separated from the β-strands (shown in yellow) by a separation distance d to create a hydrophobic cavity. The cavity was initially filled with water molecules (red spheres). The observation volume of ~1300 Å^3 is shown with black lines. Branched aliphatic side chains (heavy atoms only) are shown as sticks and as molecular surfaces. The cluster is oriented in such a way that the bottom of the *figure* is the N-terminus of the β-strands. (b) Plots of normalized water density as a function of simulation time of ILV cluster 1 for d = 4 Å (black) and 6 Å (red). The normalized water density was obtained by dividing the number of water molecules by the maximum number of water molecules inside the cavity ($N_{w,max}$ = 30). (c) Plot of water density as a function of simulation time for four different trajectories, two starting from the "wet" initial state and two other starting from the "dry" initial state, with d = 4 Å. (d) Plots of normalized water density as a function of simulation time for the three ILV clusters of αTS (cluster 1 in black, cluster 2 in red, and cluster 3 in green).

To provide insight into the role of individual ILV side chains in the dehydration observed in cluster 1, 10 of its constituent members were individually substituted with alanines, and the simulations were performed on these alanine mutants. The residues selected, V23, L25, I37, I41, L48, L50, L85, I95, I97, and L99 (Figure 3a), have previously been shown to eliminate an early kinetic trap in folding when replaced by alanine, and all but the I41A and L85A mutations significantly destabilize the on-pathway equilibrium intermediate.[14] The minimal effects of the I41A and L85A variants are thought to reflect their location in helices α1 and α2, which are self-contained elements of secondary structure on the surface of the protein that can more readily mitigate the effect of mutations on stability than their β-barrel counterparts. Figure 3b shows histograms of the water density within the cavity for cluster 1 of the wild-type protein and all 10 alanine mutants. The histogram for the wild-type protein shows a bimodal distribution, confirming that the cavity underwent transitions between a dry state (water density ≈ 0.3) and a wet state (water density ≈ 0.65), the dry state being more

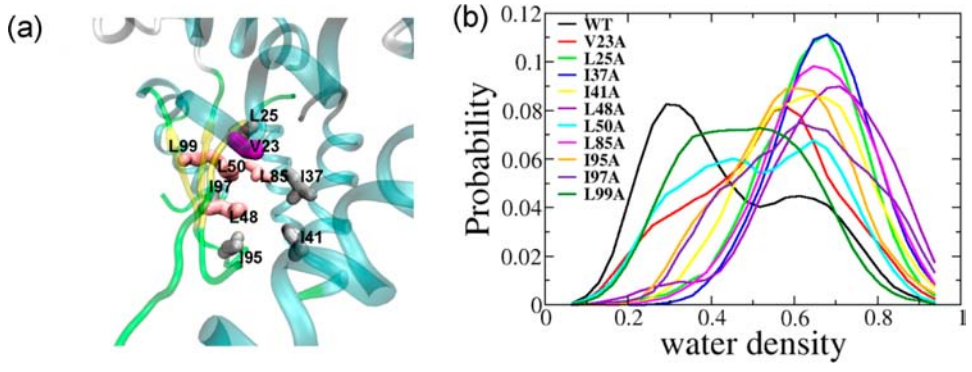


Figure 3. (a) Ribbon diagram of the interior of cluster 1, with the ILV side chains selected for alanine-scanning mutagenesis portrayed in space-filling format. (b) Histograms of water density inside cluster 1 for the wild-type protein and the 10 alanine variants.

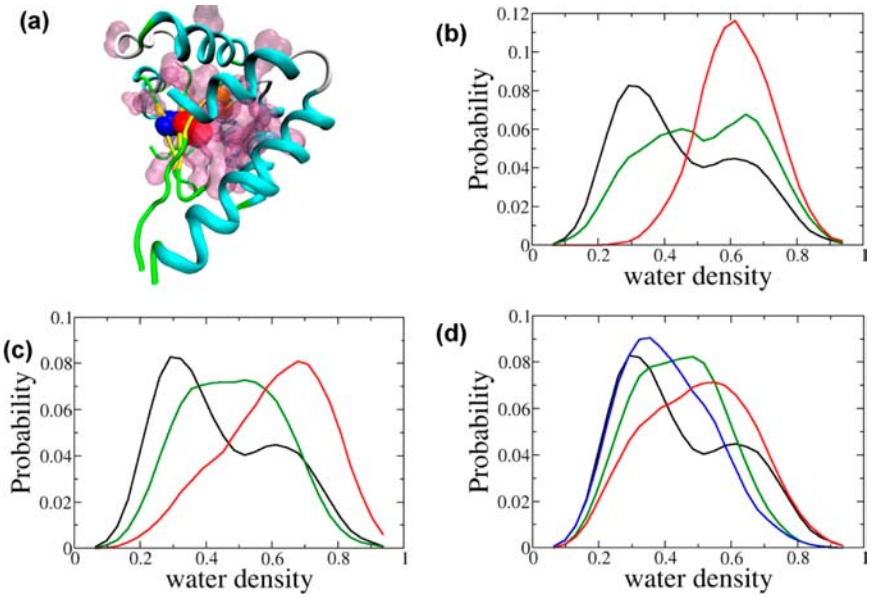


Figure 4. (a) Position of Leu50 (red), Cys81 (orange), and Leu99 (blue) within ILV cluster 1. Mutation sites are shown as van der Waals spheres. (b–d) Histograms of water density inside the cavity of cluster 1: (b) L50 with WT in black, L50A in green, and L50N in red; (c) L99 with WT in black, L99A in green, and L99N in red; (d) C81 with WT in black, C81I in green, C81V in blue, and C81N in red.

probable than the wet state. With the exception of the V23A, L50A, and L99A mutants, the cavity in the mutated proteins experienced complete or nearly complete loss of dewetting (Figure 3b). The histograms of water density for all of these mutants are Gaussian in nature with peaks centered at a water density of 0.6–0.7, showing that the cavity largely wets upon alanine substitution. These findings show that cluster 1 in the wild-type protein is fairly "optimized" in terms of drying and stability. The water fluctuation within the cavity, $\sigma_{N_w}{}^2/\langle N_w\rangle^2$, which can be used as an indicator of the dewetting transition propensity of a system,[34] also shows a similar trend (Table S2 in the SI). Subtle changes in the surface topography and chemistry (e.g., single mutation I/L → A) can tip the balance of the cavity to a more wet state, potentially lowering the stability of the cluster and protein. In contrast, replacement of residues V23, L50, and L99 with alanine resulted in partial loss of dewetting, with the L99A mutation being most resistant to wetting. The histograms of these three mutants show a considerable population of the low water density states (Figure 3b). These β-strand residues are centrally located in the cluster, facing the helical shell, and are surrounded by neighboring ILV residues (Figure 3a). To "wet" the cavity inside cluster 1 further, we performed a more radical perturbation of the hydrophobic surface by substituting L50 and L99 with their isosteric, polar counterpart, asparagine. The water density histograms for the L50N and L99N mutant proteins show that the introduction of a polar side chain at positions 50 and 99 resulted in complete or significant loss of dewetting (Figure 4b,c), respectively. The histograms for those two asparagine mutants have a water density peak centered at 0.6–0.7; the shoulder near 0.4 for L99N shows a limited propensity for dewetting. As was later confirmed experimentally, these results led to the expectation that native states and the folding intermediates for the L50N and L99N variants of αTS would be substantially destabilized relative to their wild-type counterpart (see the experimental results described below).

Probing Non-ILV Positions for Effects on Hydration of Cluster 1. Inspection of the simulations for wild-type αTS found residual water density near C81, which is adjacent to cluster 1 in helix α2 and near the C-termini of strands β3 and β4 (Figure 4a). To discover the effect of side chains proximal to cluster 1 on hydration, C81 was substituted with isoleucine, valine, and asparagine. The simulated water density distributions for the C81I and C81V mutants (Figure 4d) show a cavity that fluctuated between wet and dry states, with C81V making the cavity noticeably drier in comparison with the wild-type protein. These results suggest that the C81V mutation would be the best candidate for further dewetting of cluster 1. Unfortunately, the larger steric bulk of valine versus cysteine precluded an unambiguous experimental test of this conjecture. In contrast, the cavity in the C81N variant favored the wet state, with the maximum in the water density probability located around 0.6 (Figure 4d). These findings illustrate the sensitivity of the water probability inside the cavity to the local environment. In an effort to design a drier cavity, the effect of isoleucine substitutions on the water population within the cavity was also tested for a number of residues, such as T24, S33, A47, G51, P78, R89, and G98 (see Figure S3 in the SI). T24, A47, G51, and G98 are positioned in the helices of cluster 1, whereas the other three belong to the β-strands. As Figure S3 shows, none of those isoleucine substitutions resulted in a drier cavity.

*Hydration in Clusters **2** and **3**.* We also compared the water density fluctuations of the two other ILV clusters of αTS for a separation distance of 4 Å (Figure 2d). The ~500 Å^3 cavity for cluster 2, created by displacing the α5 and α6 helices and the β5, β6, and β7 strands, experienced strong fluctuations in water density during the 16 ns simulation time. However, this cluster did not experience extended periods of dehydration at d = 4 Å. The ~700 Å^3 cavity in cluster 3 was created by pulling the α0 helix away from the β1 and β8 strands. This cluster showed hydration even at d = 4 Å for most of the simulation time (Figure 2d); thus, cluster 3 was not considered further. In contrast to the behavior of cluster 1, the cavity in cluster 2 primarily remained in the wet state even at the small separation distance of 4 Å. Closer inspection showed that drying is more

dx.doi.org/10.1021/ja310544t | J. Am. Chem. Soc. 2013, 135, 1882–1890

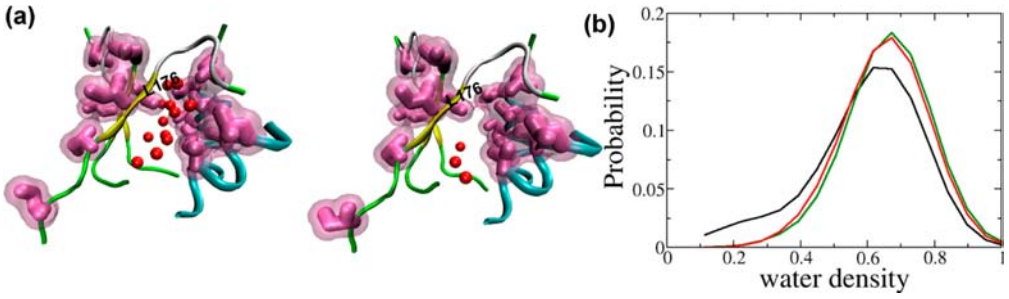


Figure 5. Dewetting of ILV cluster 2. (a) Snapshots of typical wet (left) and dry (right) states of cluster 2 populated during the simulation. The cluster is oriented in such a way that the bottom of the *figure* is the N-terminus of the β-strands. The separation distance was 4 Å. The maximum number of water molecules within the observation volume of ~500 Å^3 was 17. (b) Histograms of water density inside ILV cluster 2 for the wild-type protein (black) and its two mutants, L176A (green) and L176N (red).

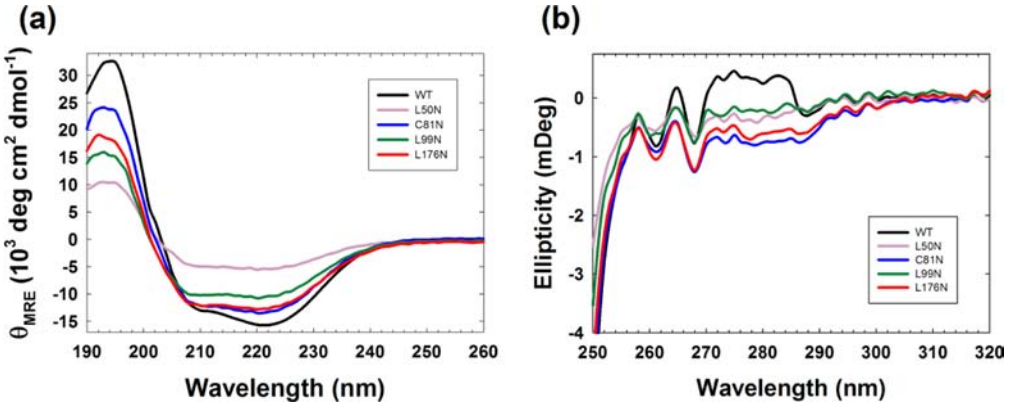


Figure 6. (a) Far-UV CD spectra of wild-type αTS and the L50N, C81N, L99N, and L176N variants from 190 to 260 nm with protein concentrations ranging from 3 to 7 μM. (b) Near-UV spectra of wild-type αTS and the L50N, C81N, L99N, and L176N variants from 250 to 320 nm.

favored toward the C-terminus of β-strands. In particular, residue L176 protrudes from β6 toward the α5 and α6 helices, acting as a roadblock inside the cavity (Figure 5a). To check the sensitivity of this position to mutation in terms of dewetting, we performed two in silico mutations, L176A and L176N. The water density distribution for the wild-type protein indicates that the cavity in cluster 2 preferred the wet state; however, there was a small but not insignificant probability for drying (Figure 5b). Both mutations resulted in enhanced wetting of the cavity, as the water density distributions shifted to the high-density side with a peak around 0.7 (Figure 5b). The distributions of the two mutants appear to be almost identical, suggesting that alanine substitution at position 176 is sufficient to wet the cavity further. This behavior contrasts with that for positions L50 and L99 in cluster 1, where a much stronger perturbation, such as a mutation to asparagine, was needed.

Experimental Analysis of Structure and Stability for Hydration Mutations in ILV Clusters. The effects of the asparagine hydration mutations on the structural properties of αTS were determined by circular dichroism (CD) spectroscopy. The far-UV CD spectra of the cluster 1 variants L50N, C81N and L99N and the cluster 2 variant L176N all display a broad negative minimum between 222 and 208 nm and a positive band at 195 nm (Figure 6a), indicative of α-helix and β-sheet contributions. However, the reductions in the ellipticities at 195 and 222 nm for the variants show that the introduction of a polar side chain at all four positions disrupted the secondary structure to varying degrees. The C81N and L176N mutations decreased the ellipticity at 222 nm by 20%. Surprisingly, the L50N and L99N mutations had a more dramatic effect, decreasing the ellipticity by 70 and 40%, respectively. The near-UV CD spectra, which provide insight into the chiral packing of aromatic side chains, reveal that all of the hydration variants had altered tertiary structures (Figure 6b). The positive band observed for tyrosines between 270 and 285 nm for wild-type αTS became negative for C81N and L176N and was eliminated for L50N and L99N. The phenylalanine bands between 255 and 270 nm were present for all of the variants, but the bands at 265 nm were comparably reduced in magnitude relative to those of wild-type αTS for L50N and L99N.

The effects of the mutations on the thermodynamic properties were determined by monitoring the far-UV CD spectra as the proteins were denatured with urea. The equilibrium unfolding transitions, as illustrated by the changes in ellipticity at 222 nm, are shown in Figure 7. All of the variants display a nearly urea-independent baseline indicative of a thermodynamically stable state in the absence of denaturant. As previously observed for wild-type αTS,[14] the equilibrium unfolding reactions of the four hydration variants are well-described by a three-state model, native (N) $\rightleftarrows$ intermediate (I) $\rightleftarrows$ unfolded (U), to fit the CD data. For the L176N variant with a limited native baseline, the stability of the N $\rightleftarrows$ I transition was determined by measuring the amplitude of the rate-limiting N $\rightarrow$ I unfolding phase as a function of the initial urea concentration while jumping to the same final urea concentration. Because the amplitude is proportional to the fraction of the native state at the initial urea concentration, the

dx.doi.org/10.1021/ja310544t | J. Am. Chem. Soc. 2013, 135, 1882−1890

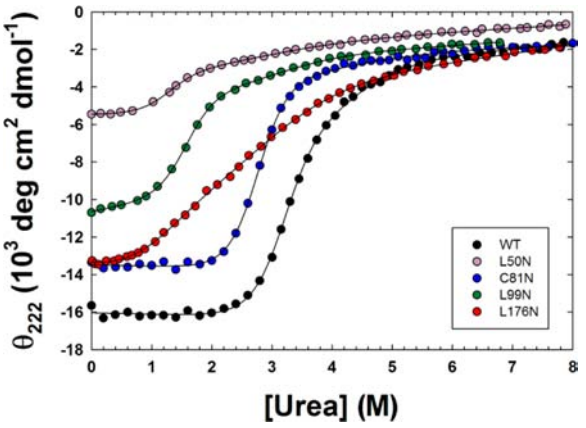


Figure 7. Urea-induced equilibrium unfolding profiles for wild-type αTS and the hydration variants L50N, C81N, L99N, and L176N. The continuous lines represent fits of the data to a three-state model.

fit of the amplitude to a 2-state model yields the desired thermodynamic parameters (Figure S4 in the SI).

The free energy differences for the N ⇄ I and I ⇄ U transitions for the variants are shown in Figure 8A and Table 1.

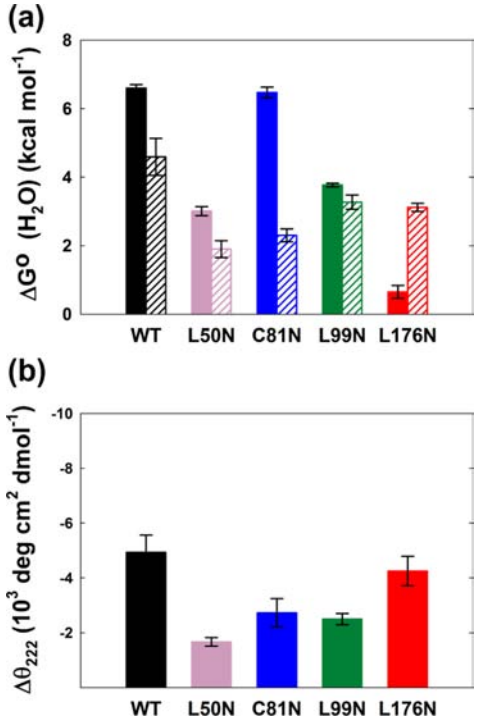


Figure 8. (a) Bar graph showing the free energy differences for the N ⇄ I (solid bars) and I ⇄ U (hatched bars) transitions obtained by fitting the data for the hydration variants to a three-state model. (b) Bar graph showing the differences between the mean residue ellipticities of the intermediate (I) and unfolded (U) states at 222 nm (ΔMRE) for wild-type αTS and the L50N, C81N, L99N, and L176N variants.

The stability of the N state relative to the I state was substantially decreased for the L50N, L99N, and L176N mutations, but the C81N mutation left the stability virtually unchanged. In addition to the stability, the fits also provided values of m, a measure of the sensitivity of the folding free energy to the denaturant concentration that is proportional to the change in buried surface area.[35] The average of the m values of the N ⇄ I transition for the L50N, C81N, and L99N variants, $m = -2.32 \pm 0.04$ kcal mol^{-1} M^{-1}, was larger than the value for the wild-type αTS, $m = -2.05 \pm 0.03$ kcal mol^{-1} M^{-1}, suggesting that all three were less well-folded in the I state (Table 1). The smaller m value for the N ⇄ I transition for L176N, -0.72 ± 0.11 kcal mol^{-1} M^{-1}, could reflect a less compact folded state or the existence of additional intermediates in the conversion of N to I. If present, the additional species would lead to an overestimation of the perturbation of the stability for the native state.

The stability of the I state relative to the U state is reduced for all four αTS variants, indicating that both clusters are sensitive to the state of hydration of the mutated side chains. The reductions in the m values for the I ⇄ U transitions for L50N and C81N are very similar in magnitude to the increases seen for the N ⇄ I transition (Table 1), again suggesting a less well folded I state. The L99N variant, however, does not display this behavior, and with a total m value of -3.39 ± 0.05 kcal mol^{-1} M^{-1} versus an average of -2.94 ± 0.07 kcal mol^{-1} M^{-1} for the wild-type, L50N, and C81N variants, L99N may experience a disruption of residual nonpolar structure in the U state. The unfolding of the intermediate for the L176N variant exposes an amount of buried surface area comparable to that for the wild-type protein.

Insights into the impact of the mutations on the secondary structure of the intermediate can be calculated from the Z parameter employed in the three-state fits to the equilibrium unfolding data (see Materials and Methods). The Z parameter reflects the normalized change in ellipticity of the intermediate state relative to the unfolded state and is defined as $Z = (\theta_I - \theta_N)/(\theta_U - \theta_N)$. This equation was used to extract θ_I from the Z values obtained from the fits. The ellipticities of the intermediates relative to their respective unfolded states for all three polar replacements in or near cluster 1 were found to decrease by 2−3-fold (Figure 8b). By contrast, the ellipticity of the intermediate for the L176N variant is very similar to that for its wild-type counterpart.

Although all of the asparagine replacements retained a three-state unfolding profile, the disruption of as much as 70% of the ellipticity in the L50N variant raised the possibility that their folded states no longer reside in the native basin for wild-type αTS. To explore this issue, denaturant jumps from the native state to the unfolded state were employed to monitor the rate limiting N → I unfolding reaction. All of the variants displayed a slow unfolding reaction whose relaxation times differed from that for the wild-type protein by less than a factor of 5 and, similar to that for wild-type αTS, decreased exponentially with increasing denaturant concentration (Figure S5 in the SI). The denaturant dependence of the observed relaxation times for the L50N, C81N, and L99N variants varied from that for wild-type αTS by less than 10% (Figure S5). The minimal perturbation of the N → I unfolding dynamics for these L → N variants, all in or near cluster 1, relative to the wild-type protein implies that the mutations have a very similar effect on the energies of the native state and the transition-state ensemble (TSE). The similar denaturant dependences for these variants imply the exposure of a comparable amount of buried surface to access the TSE.[35] By contrast, the L176N variant exhibited a 50% reduction in the denaturant dependence for its unfolding reaction (Figure S5). In view of the more than 50% decrease in the m value for the N → I reaction at equilibrium (Table 1), the folded state of L176N must be less compact than that of wild-

dx.doi.org/10.1021/ja310544t | J. Am. Chem. Soc. 2013, 135, 1882−1890

Table 1. Thermodynamic Parameters for Urea-Induced Unfolding of Wild-Type and Hydration Variants of αTS[a,b,c]

	$\Delta G^{\circ}_{NI}(H_2O)$	$-m_{NI}$	Cm_{NI}	$\Delta G^{\circ}_{IU}(H_2O)$	$-m_{IU}$	Cm_{IU}	Z	$\Delta G^{\circ}_{total}(H_2O)$	$-m_{total}$
WT	6.60 ± 0.10	2.05 ± 0.03	3.22 ± 0.06	4.59 ± 0.54	1.09 ± 0.09	4.21 ± 0.60	0.68 ± 0.05	11.19 ± 0.55	3.14 ± 0.10
L50N	3.01 ± 0.13	2.30 ± 0.09	1.31 ± 0.07	1.90 ± 0.25	0.72 ± 0.06	2.64 ± 0.41	0.58 ± 0.05	4.91 ± 0.28	3.02 ± 0.11
C81N	6.47 ± 0.16	2.29 ± 0.06	2.83 ± 0.10	2.30 ± 0.19	0.57 ± 0.05	4.04 ± 0.48	0.80 ± 0.02	8.77 ± 0.25	2.86 ± 0.07
L99N	3.77 ± 0.05	2.36 ± 0.03	1.59 ± 0.03	3.27 ± 0.21	1.03 ± 0.05	3.17 ± 0.26	0.75 ± 0.02	7.04 ± 0.22	3.39 ± 0.05
L176N	0.65 ± 0.19[d]	0.72 ± 0.11[d]	0.90 ± 0.30	3.13 ± 0.12[e]	0.94 ± 0.03[e]	3.33 ± 0.15	0.86 ± 0.05[e]	3.78 ± 0.22	1.66 ± 0.11

[a]The equilibrium unfolding data were fit to a three-state model, N ⇄ I ⇄ U. $\Delta G^{\circ}(H_2O)$, m, and C_m represent the free energy of unfolding in the absence of urea, the urea dependence of the free energy of unfolding, and the concentration of urea at the midpoint of the transition, respectively. The subscripts NI and IU denote the N ⇄ I and I ⇄ U transitions, respectively. [b]Units are as follows: $\Delta G^{\circ}(H_2O)$ in kcal mol^{-1}, m in kcal mol^{-1} M^{-1}, and C_m in M, where M denotes mol of urea L^{-1}. [c]Errors for $\Delta G^{\circ}(H_2O)$ and m are standard errors from the fits. Errors in C_m, $\Delta G^{\circ}_{total}(H_2O)$, and $-m_{total}$ were obtained by standard error propagation based on the equations $C_m = \Delta G^{\circ}(H_2O)/m$, $\Delta G^{\circ}_{total}(H_2O) = \Delta G^{\circ}_{NI}(H_2O) + \Delta G^{\circ}_{IU}(H_2O)$, and $m_{total} = m_{NI} + m_{IU}$, respectively. [d]Value was obtained by fitting the amplitude of the N → I kinetic unfolding reaction to a two-state model. [e]Value was obtained from equilibrium data with the N ⇄ I transition constrained by the values obtained from the kinetic unfolding experiment.

type αTS. The retention of the three-state unfolding model, the similar degrees of compaction implied by the totals of the m values for the two transitions (with the exception of the L176N variant) (Table 1), and the similar barriers to unfolding all argue that the variants occupy the same native basin as wild-type αTS.

DISCUSSION

We have characterized the relationship between nanoscale dewetting transitions and the stability and structure of αTS, a TIM barrel protein, using a combination of MD simulations and experimental studies. The simulations revealed that cavities created inside the two large hydrophobic ILV clusters of αTS undergo either intermittent or strong water density fluctuations, depending on the size and composition of the cluster. The largest ILV cluster (cluster 1) was found to be optimized in terms of dehydration, but it proved to be difficult to design a drier cavity. In silico isoleucine substitutions at several residues (e.g, T24, S33, A47, G51, P78, R89, and G98) were attempted in order to design a drier cavity; however, none of these attempts showed significant success (see Figure S3 in the SI). Substituting selected ILV residues with alanine was found to weaken or completely diminish dewetting, which strongly depends on the local environment of the mutation site. Our simulations also showed that the replacement of buried leucines in both clusters with asparagines is sufficient to wet their cavities completely.

The experiments performed in this study suggested a folding mechanism in which segments of large ILV clusters adopt folded-like conformations prior to final collapse and expulsion of water. In our simulations, we introduced a cavity by separating the α-helices from the β-strands of the ILV clusters in their native structures to capture how single amino acid substitutions of the critical ILV residues affect this final stage of folding. Additionally, we did not observe any significant conformational changes upon mutation in ∼100 ns long "folding" simulations of the wild-type protein and the L50N and L99N mutants, which was not unexpected because the conversions of the U state to the I state and the I state to the N state occur on the millisecond time scale.[31]

As found in our simulations, ILV cluster 1 experiences frequent transitions between a wet and a dry state, whereas cluster 2 sits on the wet side. Previous studies have shown that small perturbations, such as single amino acid substitutions, can shift the protein from a dry state to a wet state.[29] For example, a single I2A or I2V mutation can tip the tetramer channel of the protein melittin from dry to wet.[29] Along this line, recent simulation studies by Garde and co-workers also showed that water near protein surfaces can be sensitive to subtle changes in surface conformation, topology, and chemistry and that small changes can tip the balance from dry to wet or vice versa. That is, the protein can be "sitting at the edge" of the dewetting transition.[34] For example, melittin sits on the dry side of a dewetting transition, while another protein, BphC, sits on the wet side. It is possible to tip the balance to the other side for both melittin and BphC by introducing additional perturbations (e.g., point mutations). Taken together, the findings by Garde and co-workers and our current results suggest that biomolecules often sit at the edge of dewetting transitions and are sensitive to perturbations.[34] We have further shown that such sensitivity to perturbations can be readily manipulated by protein engineering, which allows the TIM barrel protein to fine-tune its stability and folding.

Experimental analysis of leucine to asparagine mutations in the N-terminal ILV cluster in αTS not only demonstrated dehydration in both the native and intermediate states but also revealed that the introduction of polarity substantially decreases the stabilities and has a dramatic effect on the structures of both states. The substitution of an acetamide group for an isobutyl group at L50 and L99 reduced the secondary structure of the native state by 40−70% and appeared to mobilize the tyrosine side chains. The secondary structures of the corresponding intermediate states were also greatly diminished for these variants. These results are consistent with the prediction that the interior of this cluster strongly prefers to dewet in a TIM barrel configuration and the conclusion that this configuration also exists for the intermediate state. What is very surprising, however, is that these mutations do not simply destabilize the TIM barrel fold or its folding intermediate. Rather, the presence of the polar side chain leads to distinct high-energy thermodynamic states in the native basin on the TIM barrel folding free energy surface. In the case of cluster 1, the need to sequester the remaining 30 aliphatic side chains from solvent appears to provide sufficient driving force to populate these alternative states. The substantial decrease in the CD signal at 222 nm and the loss of signal at 280 nm could reflect a highly dynamic α-helical shell that enables the partial exposure of the asparagine side chains at positions 50 and 99 to water while retaining buried surface area.

Interestingly, the C81N mutation in helix α2 and adjacent to cluster 1 had a lesser effect on the secondary structure and left the stability of the native relative to the intermediate state virtually unchanged (Table 1). Although the urea concentration at the midpoint of the urea-induced transition (2.83 M) was

lower than that for the wild-type (3.22 M), the larger m value for C81N resulted in a stability that was coincidentally the same as that for the wild-type. As noted above, the increased m value reflects a less well folded I state. The ready adaptation to the polar side chain at position 81 is similar to the previously described response of the L85A mutation, also in helix α2, reflecting the conformational adaptability of a surface helix.[14] The stability of the I state versus the U state and the m value, however, were markedly reduced. The lower inherent stability of the I state apparently does not provide a sufficient driving force to accommodate the asparagine side chain and maintain the secondary structure and compactness for the C81N variant.

Although an asparagine mutation in cluster 2 had a lesser effect on the secondary structure in the native state and little or no effect on the intermediate, the L176N variant could not achieve the same stability or degree of compactness in the native state as the wild-type protein. The distinct changes in the near-UV CD spectrum might reflect perturbations in the packing of the adjacent Y173 and Y175 residues inside the β-barrel as well as more global effects accompanying the decreased packing efficiency. In contradiction to the predictions of the simulations, the wet state favored for the interior of this cluster was not capable of supporting the presence of the polar side chain in the native conformation. The contradiction may reflect the smaller size of cluster 2, which led to only marginal drying of the cavity in the simulations (Figure 5). The limited effects of the L176N mutation on the stability and secondary structure of the intermediate state, in contrast to the substantial effects on the native state, suggest that the side chain is only partially dehydrated at this stage of folding. All of these findings are consistent with the previous conjecture that the region encompassing cluster 1 is well-packed in the intermediate state while cluster 2 is best described as a loosely folded, molten, globulelike structure.[36,37]

It was surprising to observe that the MD simulations for the entire set of 10 ILV → A mutations in cluster 1 resulted in wetting of the cavity. One might have expected that the removal of 2−3 carbons from a cluster of 31 branched aliphatic side chains would have little effect on the propensity of water to occupy the exposed nonpolar volume. However, the sensitivity of drying to the composition and/or structure of the cavity may be the explanation for the previous experimental observation that these same alanine replacements substantially reduce the stability of the intermediate in αTS.[14] The simulations suggested that the enhanced propensity of the alanine variants in cluster 1 to wet (i.e., to favor a less well-folded state) is, along with the loss of packing interactions, a mechanism for destabilizing the intermediate. The tendency of the cavity in cluster 2 to wet in wild-type αTS would mitigate any enhanced hydration from alanine replacements and minimize the perturbation of the stability of the intermediate, as observed.

An unanticipated outcome of creating the L50N and L99N variants was the discovery of discrete thermodynamic states that have substantially disrupted secondary structures and the apparent loss of tight packing around the seven tyrosines with a compactness comparable to that in the wild-type protein. Although further experiments are required in order to rule out the coincidental cancellation of positive and negative bands for the tyrosines and confirm their putative dynamic properties, nativelike compactness with mobile side chains are characteristics of the "dry molten globule",[38] which was initially proposed as a model for folding transition states or to arise in a membrane environment[39] and, subsequently, as a discrete state in the native basin.[40−42] The putative dry molten globule states for the L50N and L99N variants of αTS, however, do not unlock all of the phenylalanines and have a substantially altered secondary structure compared with the canonical TIM barrel. Further studies are required in order to determine whether the folded states of the L50N and L99N variants are indeed dry molten globules, as envisioned by Shakhnovich and Finkelstein,[39] or represent related high-energy states in the native basin. Intriguingly, the existence of such states might provide a path for the evolution of the sequence to produce TIM barrels with alternative locations for their ILV sequences.[43]

CONCLUSION

The results of this combined simulation−experimental study of αTS demonstrate the critical role of dehydration in a large hydrophobic ILV cluster in determining the stability and structure of a TIM barrel fold and a critical folding intermediate. ILV clusters are common in the other $(\beta\alpha)$-repeat motifs, such as the flavodoxin fold and Rossmann fold families,[44] and they also define coiled coils,[45,46] repeat-sequence proteins,[47−49] β-sandwich motifs,[50,51] and antiparallel β-sheet arrays found in amyloidogenic peptides.[52] Thus, our findings may provide useful insights into the links among hydrophobicity, dewetting, and stability of a large number of protein motifs.

ASSOCIATED CONTENT

Supporting Information

Supplementary Materials and Methods, Tables S1 and S2, and Figures S1−S5. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

ruhongz@us.ibm.com; C.Robert.Matthews@umassmed.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Xuhui Huang for help with setting up the simulation system at the beginning of this project and Silvia Cavagnero, Bruce Berne, Vijay Pande, Zhen Xia, and Sagar Kathuria for many fruitful discussions. R.Z. acknowledges financial support from the IBM Blue Gene Science Program. C.R.M. acknowledges funding from the National Institutes of Health through Grant GM 23303.

REFERENCES

(1) Bartlett, A. I.; Radford, S. E. *J. Mol. Biol.* **2010**, *396*, 1329.
(2) Brun, L.; Isom, D. G.; Velu, P.; Garcia-Moreno, B.; Royer, C. A. *Biochemistry* **2006**, *45*, 3473.
(3) Fernandez-Escamilla, A. M.; Cheung, M. S.; Vega, M. C.; Wilmanns, M.; Onuchic, J. N.; Serrano, L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2834.
(4) Nishiguchi, S.; Goto, Y.; Takahashi, S. *J. Mol. Biol.* **2007**, *373*, 491.
(5) Kimura, T.; Maeda, A.; Nishiguchi, S.; Ishimori, K.; Morishima, I.; Konno, T.; Goto, Y.; Takahashi, S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13391.
(6) Englander, S. W. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 213.
(7) Wintrode, P. L.; Rojsajjakul, T.; Vadrevu, R.; Matthews, C. R.; Smith, D. L. *J. Mol. Biol.* **2005**, *347*, 911.
(8) Miranker, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. T.; Dobson, C. M. *Science* **1993**, *262*, 896.

(9) Jones, B. E.; Matthews, C. R. *Protein Sci.* **1995**, *4*, 167.
(10) Gu, Z.; Rao, M. K.; Forsyth, W. R.; Finke, J. M.; Matthews, C. R. *J. Mol. Biol.* **2007**, *374*, 528.
(11) Vadrevu, R.; Wu, Y.; Matthews, C. R. *J. Mol. Biol.* **2008**, *377*, 294.
(12) Radzicka, A.; Wolfenden, R. *Biochemistry* **1988**, *27*, 1664.
(13) Kathuria, S. V.; Day, I. J.; Wallace, L. A.; Matthews, C. R. *J. Mol. Biol.* **2008**, *382*, 467.
(14) Wu, Y.; Vadrevu, R.; Kathuria, S.; Yang, X.; Matthews, C. R. *J. Mol. Biol.* **2007**, *366*, 1624.
(15) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570.
(16) Huang, X.; Margulis, C. J.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11953.
(17) ten Wolde, P. R.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6539.
(18) Zhang, F.; Du, H.-N.; Zhang, Z.-X.; Ji, L.-N.; Li, H.-T.; Tang, L.; Wang, H.-B.; Fan, C.-H.; Xu, H.-J.; Zhang, Y.; Hu, J.; Hu, H.-Y.; He, J.-H. *Angew. Chem., Int. Ed.* **2006**, *45*, 3611.
(19) Hua, L.; Zangi, R.; Berne, B. J. *J. Phys. Chem. C* **2009**, *113*, 5244.
(20) Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. *Nature* **2001**, *414*, 188.
(21) Li, J.; Liu, T.; Li, X.; Ye, L.; Chen, H.; Fang, H.; Wu, Z.; Zhou, R. *J. Phys. Chem. B* **2005**, *109*, 13639.
(22) Li, X.; Li, J.; Eleftheriou, M.; Zhou, R. *J. Am. Chem. Soc.* **2006**, *128*, 12439.
(23) Zhou, R.; Huang, X.; Margulis, C. J.; Berne, B. J. *Science* **2004**, *305*, 1605.
(24) Hua, L.; Huang, X.; Liu, P.; Zhou, R.; Berne, B. J. *J Phys Chem B* **2007**, *111*, 9069.
(25) Young, T.; Hua, L.; Huang, X. H.; Abel, R.; Friesner, R.; Berne, B. J. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1856.
(26) Krone, M. G.; Hua, L.; Soto, P.; Zhou, R. H.; Berne, B. J.; Shea, J. E. *J. Am. Chem. Soc.* **2008**, *130*, 11066.
(27) Liu, X. G.; Zhang, Y.; Goswami, D. K.; Okasinski, J. S.; Salaita, K.; Sun, P.; Bedzyk, M. J.; Mirkin, C. A. *Science* **2005**, *307*, 1763.
(28) Berne, B. J.; Weeks, J. D.; Zhou, R. *Annu. Rev. Phys. Chem.* **2009**, *60*, 85.
(29) Liu, P.; Huang, X.; Zhou, R.; Berne, B. J. *Nature* **2005**, *437*, 159.
(30) *The PyMOL Molecular Graphics System*, version 1.3; Schrödinger, LLC: Cambridge, MA, 2010.
(31) Bilsel, O.; Zitzewitz, J. A.; Bowers, K. E.; Matthews, C. R. *Biochemistry* **1999**, *38*, 1018.
(32) Gualfetti, P. J.; Bilsel, O.; Matthews, C. R. *Protein Sci.* **1999**, *8*, 1623.
(33) Yu, N.; Hagan, M. F. *Biophys. J.* **2012**, *103*, 1363.
(34) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Hagan, M. F.; Chandler, D.; Garde, S. *J. Phys. Chem. B* **2012**, *116*, 2498.
(35) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138.
(36) Wu, Y.; Vadrevu, R.; Yang, X.; Matthews, C. R. *J. Mol. Biol.* **2005**, *351*, 445.
(37) Wu, Y.; Kondrashkina, E.; Kayatekin, C.; Matthews, C. R.; Bilsel, O. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13367.
(38) Baldwin, R. L.; Frieden, C.; Rose, G. D. *Proteins* **2010**, *78*, 2725.
(39) Shakhnovich, E. I.; Finkelstein, A. V. *Biopolymers* **1989**, *28*, 1667.
(40) Kiefhaber, T.; Labhardt, A. M.; Baldwin, R. L. *Nature* **1995**, *375*, 513.
(41) Kiefhaber, T.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2657.
(42) Jha, S. K.; Udgaonkar, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12289.
(43) Forsyth, W. R.; Bilsel, O.; Gu, Z.; Matthews, C. R. *J. Mol. Biol.* **2007**, *372*, 236.
(44) Bueno, M.; Campos, L. A.; Estrada, J.; Sancho, J. *Protein Sci.* **2006**, *15*, 1858.
(45) Munson, M.; Balasubramanian, S.; Fleming, K. G.; Nagi, A. D.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Sci.* **1996**, *5*, 1584.
(46) Dalal, S.; Canet, D.; Kaiser, S. E.; Dobson, C. M.; Regan, L. *Protein Eng., Des. Sel.* **2008**, *21*, 197.
(47) Yoder, M. D.; Lietzke, S. E.; Jurnak, F. *Structure* **1993**, *1*, 241.
(48) Zitzewitz, J. A.; Bilsel, O.; Luo, J.; Jones, B. E.; Matthews, C. R. *Biochemistry* **1995**, *34*, 12812.
(49) Main, E. R.; Lowe, A. R.; Mochrie, S. G.; Jackson, S. E.; Regan, L. *Curr. Opin. Struct. Biol.* **2005**, *15*, 464.
(50) Cox, A.; Arroyo, M. M.; Mayo, K. H. *Biochem. J.* **2001**, *357*, 739.
(51) Lappalainen, I.; Hurley, M. G.; Clarke, J. *J. Mol. Biol.* **2008**, *375*, 547.
(52) Hills, R. D., Jr.; Brooks, C. L., III. *J. Mol. Biol.* **2007**, *368*, 894.