Published in final edited form as:

J Am Chem Soc. 2013 February 6; 135(5): 1882–1890. doi:10.1021/ja310544t.

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

Payel Das¹, Divya Kapoor², Kevin T. Halloran², Ruhong Zhou^{1,3,*}, and C. Robert Matthews^{2,*}

¹Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598

²Department of Biochemistry & Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605

³Department of Chemistry, Columbia University, New York, NY 10027

Abstract

Recent molecular dynamics simulations have suggested important roles for nanoscale dewetting on the stability, function, and folding dynamics of proteins. Using a synergistic simulation-experimental approach on the α TS TIM barrel protein, we validate this hypothesis by revealing the occurrence of drying inside hydrophobic amino acid clusters and its manifestation on experimental measures of protein stability and structure. Cavities created within three clusters of branched aliphatic amino acids, isoleucines, leucines and valines (ILV), 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 the less hydrophobic alanine 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 the isosteric and polar asparagine enhanced the wetting of the N- and C-terminal clusters. 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 only experiences 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 in hydrophobic ILV clusters to the folding and stability of the α TS TIM barrel.

Keywords

Dewetting transition; water density fluctuations; ILV clusters; equilibrium unfolding; molecular dynamics

Introduction

The folding of proteins following synthesis on a ribosome or dilution from a chemically-denatured state involves the formation of numerous van der Waal's interactions, hydrogen bonds and electrostatic interactions that stabilize the compact native conformation. It is

*Corresponding Author: C.Robert.Matthews@umassmed.edu, ruhongz@us.ibm.com.

The supplemental information and figures (S1–S5) are available free of charge via the Internet at <http://pubs.acs.org>.

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 are replaced with isosteric polar side chain analogs, have shown that water is selectively shed prior to the appearance of the native state to enable the formation of critical cores of stability in early intermediates² or transition state ensembles^{3,4}. 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 -helical⁵ and β -sheet proteins⁶. A third experimental approach towards examining the role of water in folding monitors the protection of main chain amide hydrogens against exchange in deuterated water in partially-folded states^{7,8} and folding intermediates^{9–11}. When hydrogen exchange (HX) techniques were applied to a pair of $(\beta\alpha)_8$ TIM barrel proteins^{11,12}, protection against exchange in folding intermediates was found to be selectively associated with clusters of branched aliphatic side chains, isoleucine, leucine and valine (ILV). The molecular rationale for this behavior was ascribed to the preferential partitioning of side chain analogs of saturated hydrocarbon moieties into the vapor phase, relative to their aromatic, sulfur or polar-containing counterparts that spontaneously dissolve in water¹³. 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^{14,15}.

From a computational perspective, nanoscale dewetting transitions^{16–18} between hydrophobic surfaces have long been of interest for both physical^{17,19–23} and biological systems^{17,20,24–30}. Previous molecular dynamics (MD) simulation 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 is comprised of 3 isoleucines, 4 leucines, 1 tryptophan and 2 valines³⁰. A subsequent study on a variety of protein complexes (dimers, tetramers and two-domain proteins) found that dewetting required large complementary hydrophobic surfaces with significant contributions from isoleucines, leucines and valines²⁵. In contrast, a marked decrease in water density was not detected at the domain interface in the two-domain 2,3-dihydroxy-biphenyl dioxygenase (BphC)²⁴. 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, α TS, a ~28 kDa TIM barrel $(\beta\alpha)_8$ protein that is a component of the $\alpha_2\beta_2$ tetrameric tryptophan synthase complex. Previously, α TS was observed to offer strong and selective protection against HX in an on-pathway intermediate associated with a large N-terminal ILV cluster⁸. 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 the isosteric and polar asparagine. 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

Molecular Dynamics Simulations

The initial structures of the ILV clusters were taken from the crystal structure deposited in the Protein Data Bank (PDB ID code 1BKS). The clusters were determined using the same protocol described previously¹⁵. Three ILV hydrophobic clusters are found within the α TS 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 for the cavity for all systems was set to be wet, unless otherwise stated. The resulting systems were minimized for 10000 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 (c32b1 parameter set) force field was used and simulations were performed using NAMD2 molecular modeling package with a 2 fs time step. At least 15 different trajectories were run for Cluster 1 and Cluster 2 of the wild type 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 μ s. Additionally, we performed ~100 ns long “folding” simulations, in which both main chain and side chains atoms were free to move.

Site-directed Mutagenesis

The codon-optimized α TS WT gene was synthesized by Genscript in pUC 57 and re-cloned into a modified pGS-21a vector with an N-terminal 6X His tag and TEV protease site using EcoRV and BamHI restriction sites. Various hydration mutations were made using mutagenic oligonucleotides purchased from Integrated DNA Technologies using the Stratagene Quick-change site-directed mutagenesis kit and mutations were confirmed using DNA sequencing (a full list of mutants is shown in Table S1). The pGS-21a plasmid DNA was transformed into BL21 (DE3) pLysS cells for protein expression and purification (See Supplemental Information for protein purification details).

Equilibrium and Kinetic Unfolding Experiments

The thermodynamic properties of both wild-type α TS and various hydration mutants were determined by urea titrations on a Jasco J-810 spectropolarimeter. Samples at varying urea concentrations were prepared using a Hamilton 540B automatic titrator and were incubated overnight at 25 °C for complete equilibration. Data were collected using a 2 mm pathlength quartz cuvette and a 2.5 nm bandwidth. The spectra were recorded at every 1 nm in the wavelength range from 215 nm to 260 nm with a scan speed of 50 nm min⁻¹ and an eight second averaging time. The denaturant dependence of the ellipticities for α TS and its variants was fit to a three-state model using Savuka, an in-house nonlinear least squares program, and assuming a linear dependence of the free energy of unfolding on the denaturant concentration³¹. These fits provided the free energy differences between the three

thermodynamic states, the denaturant dependences of these free energy differences and the Z parameter required to estimate the ellipticity of the intermediate³².

The manual-mixing kinetic unfolding jumps began in the absence of denaturant and ended between 4.0 M to 6.0 M urea, with the final protein concentration ranging from 3–5 μ M. Data were collected at 222 nm and at 25 °C in a 1 cm pathlength cuvette. The relaxation times were obtained by fitting the kinetic traces to a single exponential function in Savuka³¹.

Results

A ribbon diagram of α TS and the location of its three ILV clusters are shown in Figure 1a. Cluster 1, containing 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, containing 12 ILVs, is found at the interface between the β -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 8 ILVs individually contributed by 7 of the 8 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 $\sim 1300 \text{ \AA}^3$, was created by pulling the $\alpha 1$ and $\alpha 2$ helices away from the $\beta 1$, $\beta 2$ and $\beta 3$ strands by a separation distance d varying from 4 $\text{\AA}$ to 6 $\text{\AA}$ (Figure 2a) and filled with water molecules. Previous work by us and others have reported nanoscopic dewetting transitions in proteins with cavity volumes of a similar order^{25,33}. During the 16 ns simulation time, the water molecules were free to move, but the protein heavy atoms remained constrained. Figure 2b shows the water density plots as a function of simulation time for Cluster 1 at two separation distances d , 4 $\text{\AA}$ and 6 $\text{\AA}$. The cavity undergoes intermittent transitions between wet and dry states at a separation distance of 4 $\text{\AA}$, however, no drying transition was observed at $d = 6 \text{ \AA}$. These drying transitions typically occur in 200–300 ps. To check the convergence of our results, we started the simulations from two different initial states with $d = 4 \text{ \AA}$: one from the ‘wet’ state and a second starting from a ‘dry’ state, in which all of the initial water molecules were removed manually to create a dry cavity (Figure 2c, longer timescale simulations can be found in Figure S1). 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 is obtained by dividing number of water molecules with maximum number of water molecules inside the cavity. Snapshots of the cavity in the wet and dry states (Figure S2) 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 (HX) with solvent in this region observed in native-state HX experiments¹².

To provide insight into the role of individual ILV side chains to 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 (Figure 3a), V23, L25, I37, I41, L48, L50, L85, I95, I97 and L99, have previously been shown to eliminate an early kinetic trap in folding when replaced by alanine and all but I41A and L85A significantly destabilize the on-pathway equilibrium intermediate¹⁵. The minimal effects of the I41 and L85 variants are thought to reflect their location in helices $\alpha 1$ and $\alpha 2$, 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 plots the histograms showing the probability of the water density within of the cavity for Cluster 1 of the wild-type protein and all ten alanine mutants. The histogram for the wild type protein shows a bimodal distribution, confirming that the cavity undergoes transitions between a dry (water density ~ 0.3) and a wet state (water density ~ 0.65), the dry state being more probable over the wet state. With the exception of V23A, L50A and L99A, 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 the Cluster 1 in wild-type protein is fairly

“optimized” in terms of drying and stability. The water fluctuation, $\frac{\sigma_{N_w}^2}{\langle N_w \rangle^2}$, within the cavity, which can be used as an indicator of dewetting transition propensity of a system³⁴, also shows a similar trend (Table S2). Subtle changes in the surface topography and chemistry (e.g. single mutation I/L $\rightarrow$ A) can tip the balance of the cavity to a more wet state, potentially lowering the stability of the cluster and protein. In contrast, alanine replacement to residues V23, L50, and L99 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). In order to further ‘wet’ the cavity inside Cluster 1, we performed a more radical perturbation on the hydrophobic surface by substituting L50 and L99 with their isosteric and polar counterpart, asparagine. The water density histograms of L50N and L99N mutant proteins illustrate that the introduction of a polar side chain at positions 50 and 99 results in a 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 will be confirmed experimentally, these results led to the expectation that native states and the folding intermediates for the L50N and L99N variants of α TS would be substantially destabilized relative to their wild-type counterpart (see experimental results).

Probing Non-ILV Positions for Effects on Hydration of Cluster 1

Inspection of the simulations for wild-type α TS found residual water density near C81, which is adjacent to Cluster 1 in helix $\alpha 2$ and near the C-termini of strands $\beta 3$ and $\beta 4$ (Figure 4a). To discover the effect of side chains proximal to Cluster 1 on hydration, C81 was substituted with isoleucine, valine, and asparagine, respectively. The simulated water density distributions of C81I and C81V mutants (Figure 4d) showed a cavity that fluctuates between wet and dry states, C81V making the cavity noticeably drier compared to the wild-type protein. These results suggest that C81V mutation would be the best candidate to further dewet Cluster 2. Unfortunately, the larger steric bulk of valine versus cysteine precludes an unambiguous experimental test of this conjecture. In contrast, the cavity in the C81N variant favors the wet state with the maximum of water density probability around 0.6 (Figure 4d). These findings illustrate the sensitivity of 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, G98 (see Figure S3). 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 α TS for the separation distance of 4 Å (Figure 2d). The ~ 500 Å³ cavities for Cluster 2, created by

displacing the $\alpha 5$ and $\alpha 6$ helices and the $\beta 5$, $\beta 6$ and $\beta 7$ strands, experienced strong fluctuations in water density during the 16 ns simulation time. However, this cluster did not experience extended periods of dehydration at a separation distance of 4 Å. The ~ 700 Å³ cavities in Cluster 3 was created by pulling the $\alpha 0$ helix away from the $\beta 1$ and $\beta 8$ strands. This cluster showed hydration even at a separation of 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 a small separation distance of 4 Å. Closer inspection showed that drying is more favored toward the C-terminus of β -strands. In particular, residue L176 protrudes from $\beta 6$ toward the $\alpha 5$ and $\alpha 6$ 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 of the wild type protein indicated that the cavity in Cluster 2 prefers 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 shift to the high-density side with a peak around 0.7 (Figure 5b). The distributions of the two mutants appear almost identical, suggesting that an alanine substitution is sufficient at position 176 to further wet the cavity. This behavior contrasts with that for positions L50 or L99 in Cluster 1, where a much stronger perturbation, such as mutation to asparagine, is 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 α TS were determined by CD spectroscopy. The far-UV CD spectra of the Cluster 1 variants L50N, C81N, L99N, and the Cluster 2 variant L176N all display a broad negative minimum between 222 nm 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 the introduction of a polar side at all four positions disrupts the secondary structure to varying degrees. The C81N and L176N mutations decrease the ellipticity at 222 nm by 20%. Surprisingly, the L50N and L99N mutations have 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 have altered tertiary structures (Figure 6b). The positive band observed for tyrosines between 270 and 285 nm for wild-type α TS becomes negative for C81N and L176N and is eliminated for L50N and L99N. The phenylalanine bands between 255 and 270 nm are present for all of the variants, however, the bands at 265 nm are comparably reduced in magnitude vs. wild-type α TS for L50N and L99N.

The effects of the mutations on the thermodynamic properties were determined by monitoring the far-UV CD spectrum 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 α TS¹⁵, the equilibrium unfolding reactions of the four hydration variants are well-described by a 3-state model, $N \rightleftharpoons I \rightleftharpoons U$, to fit the CD data. For the L176N variant with a limited native baseline, the stability of the $N \rightleftharpoons I$ 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 fit of the amplitude to a 2-state model yields the desired thermodynamic parameters (Figure S4).

The free energy differences for the $N \rightleftharpoons I$ and $I \rightleftharpoons U$ transitions for the variants are shown in Figure 8A and Table 1. The stability of the N state vs. the I state is substantially decreased for the L50N, L99N and L176N mutations, however, the C81N mutation leaves the stability

virtually unchanged. In addition to the stability, the fits also provide the m -value, a measure of the sensitivity of the folding free energy to the denaturant concentration that is proportional to the change in buried surface area³⁵. The average of the m -values of the $N \rightleftharpoons I$ transition for the L50N, C81N and L99N variants, $\langle m \rangle = -2.32 \pm 0.04 \text{ kcal mol}^{-1} \text{ M}^{-1}$, is larger than for the wild-type α TS, $-2.05 \pm 0.03 \text{ kcal mol}^{-1} \text{ M}^{-1}$, suggesting that all three are less well-folded in the I state (Table 1). The smaller m -value for the $N \rightleftharpoons I$ transition for L176N, $-0.72 \pm 0.11 \text{ kcal mol}^{-1} \text{ M}^{-1}$, 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 in stability for the native state.

The stability of the I state vs. the U state is reduced for all four α TS 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 $I \rightleftharpoons U$ transition for L50N and C81N are very similar in magnitude to the increases seen for the $N \rightleftharpoons I$ 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 $-3.39 \pm 0.05 \text{ kcal mol}^{-1} \text{ M}^{-1}$ versus an average of $-2.94 \pm 0.07 \text{ kcal mol}^{-1} \text{ M}^{-1}$ for the wild-type, L50N and C81N variants, may experience a disruption of residual nonpolar structure in the U state. The unfolding of the intermediate for the L176N variant exposes a comparable amount of buried surface area as 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 3-state fits to the equilibrium unfolding data (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 $Z = (\theta_I - \theta_N)/(\theta_U - \theta_N)$. Rearranging to extract θ_I , the ellipticities of the intermediates vs. 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 its wild-type counterpart.

Although all of the asparagine replacements retain a 3-state unfolding profile, the disruption of as much as 70% of the ellipticity in the L50N variant raises the possibility that their folded states no longer reside in the native basin for wild-type α TS. To explore this issue, denaturant jumps from the native state to the unfolded state were employed to monitor the rate limiting, $N \rightarrow I$, unfolding reaction. All of the variants display a slow unfolding reaction whose relaxation times differ less than a factor of 5 from that for the wild-type protein and, similar to wild-type α TS, decrease exponentially with increasing denaturant concentration (Figure S5). The denaturant dependence of the observed relaxation times for the L50N, C81N and L99N variants varies less than 10% from that for wild-type α TS (Figure S5). The minimal perturbation of the $N \rightarrow I$ unfolding dynamics for these L $\rightarrow$ N variants, all in or near Cluster 1, vs. 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 TSE³⁵. By contrast, the L176N variant has a 50% reduction in the denaturant dependence for its unfolding reaction (Figure S5). When considered with the more than 50% decrease in the m -value for the $N \rightarrow I$ reaction at equilibrium (Table 1), the folded state of L176N must be less compact than for wild-type α TS. The retention of the 3-state unfolding model, a similar degree of compaction implied by the total of the m -values for the two transitions (with the exception of the L176N variant) (Table 1) and the same barrier to unfolding all argue that the variants occupy the same native basin as wild-type α TS.

Discussion

We have characterized the relationship between nanoscale dewetting transitions and the stability and structure of α TS, a TIM barrel protein, using a combined molecular dynamics simulations and experimental approach. Simulations reveal that cavities created inside the two large hydrophobic ILV clusters of α TS 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, such as T24, S33, A47, G51, P78, R89, G98 were attempted in order to design a drier cavity; however, none of these attempts show significant success (see Figure S3). 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 completely wet their cavities.

The experiments performed in this study suggest 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 α -helices from β -strands of the ILV clusters in their native structures to capture how single amino acid substitutions to the critical ILV residues affect this final stage of folding. Additionally, we did not observe any significant conformational changes upon mutation from ~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 scales³¹.

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 state³⁰. For example, a single I2A or I2V mutation can tip the protein melittin tetramer channel from dry to wet³⁰. Along this line, recent simulation studies by Garde and coworkers also show that water near protein surfaces can be sensitive to subtle changes in surface conformation, topology, and chemistry, and 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 transition³⁴. For example, melittin sits on the dry side of a dewetting transition, while another protein BphC on the wet side. It is possible to tip the balance to the other side for both melittin and BphC proteins by introducing additional perturbations, e.g. point mutations. Taken together, these findings by Garde and coworkers and our current results suggest that biomolecules often sit at the edge of dewetting transitions and are sensitive to perturbations³⁴. We further show 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 α TS not only demonstrated dehydration in both native and intermediate states but also revealed that the introduction of polarity substantially decreased the stabilities and had 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. The 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, there appears to be sufficient driving force from the need to sequester the remaining 30 aliphatic side chains from solvent to populate these alternative states. The substantial decrease in the CD signal at 222 nm and 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 α 2 and adjacent to Cluster 1 had a lesser effect on the secondary structure and left the stability of the native versus the intermediate state virtually unchanged (Table I). Although the midpoint of the urea-induced transition, 2.83 M, is lower than wild-type, 3.22M, the larger m-value for C81N results in a stability that is coincidentally the same as 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 α 2, reflecting the conformational adaptability of a surface helix¹⁵. 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 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 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, leading to only marginal drying of the cavity in the simulations (Figure 5). The limited effect of the L176N mutation on the stability and secondary structure of the intermediate state, in contrast to the substantial effects on the native state, suggests 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 globule-like structure^{36,37}.

It was surprising to observe that the MD simulations for the entire set of 10 ILV $\rightarrow$ A mutations in Cluster 1 resulted in the 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 α TS¹⁵. The simulations suggest that the enhanced propensity of the alanine variants in Cluster 1 to wet, i.e., 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 the wild-type α TS would mitigate any enhanced hydration from alanine replacements and minimize the perturbation on 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 structure and the apparent loss of tight packing around the 7 tyrosines with a compactness comparable to the

wild-type protein. Although further experiments are required to rule out the coincidental cancellation of positive and negative bands for the tyrosines and confirm their putative dynamic properties, native-like compactness with mobile side chains are characteristics of the “dry molten globule”³⁸. The dry molten globule was initially proposed as a model for folding transition states or to arise in a membrane environment³⁹ 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 α TS, however, do not unlock all of the phenylalanines and have a substantially altered secondary structure compared to the canonical TIM barrel. Further studies are required to determine if the folded states of the L50N and L99N variants are indeed dry molten globules, as envisioned by Shakhnovich and Finkelstein³⁹, 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 sequences⁴³.

Conclusion

The results of this combined experimental-simulation study on α TS 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 the Rossmann-fold families⁴⁴, and they also define coiled coils^{45,46}, repeat-sequence proteins^{47–49}, β -sandwich motifs^{50,51}, and anti-parallel β -sheet arrays found in amyloidogenic peptides⁵². Thus, our findings may provide useful insights into the link between hydrophobicity, dewetting, and stability of a large number of protein motifs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Xuhui Huang for the help with the simulation system setup at the beginning of this project. We also thank Silvia Cavagnero, Bruce Berne, Vijay Pande, Zhen Xia and Sagar Kathuria for many fruitful discussions. RZ acknowledges the financial support from the IBM Blue Gene Science Program. CRM acknowledges funding from the National Institutes of Health through grant GM 23303.

References

1. PyMOL; Version 1.3. Schrödinger, LLC;
2. Bartlett AI, Radford SE. *J Mol Biol.* 2010; 396:1329. [PubMed: 20053361]
3. Brun L, Isom DG, Velu P, Garcia-Moreno B, Royer CA. *Biochemistry.* 2006; 45:3473. [PubMed: 16533028]
4. Fernandez-Escamilla AM, Cheung MS, Vega MC, Wilmanns M, Onuchic JN, Serrano L. *Proc Natl Acad Sci U S A.* 2004; 101:2834. [PubMed: 14978284]
5. Nishiguchi S, Goto Y, Takahashi S. *J Mol Biol.* 2007; 373:491. [PubMed: 17850819]
6. 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. [PubMed: 18757727]
7. Englander SW. *Annu Rev Biophys Biomol Struct.* 2000; 29:213. [PubMed: 10940248]
8. Wintrodde PL, Rojsajakul T, Vadrevu R, Matthews CR, Smith DL. *J Mol Biol.* 2005; 347:911. [PubMed: 15784252]
9. Miranker A, Robinson CV, Radford SE, Aplin RT, Dobson CM. *Science.* 1993; 262:896. [PubMed: 8235611]
10. Jones BE, Matthews CR. *Protein Sci.* 1995; 4:167. [PubMed: 7757007]

11. Gu Z, Rao MK, Forsyth WR, Finke JM, Matthews CR. *J Mol Biol.* 2007; 374:528. [PubMed: 17942114]
12. Vadrevu R, Wu Y, Matthews CR. *J Mol Biol.* 2008; 377:294. [PubMed: 18234216]
13. Radzicka A, Wolfenden R. *Biochemistry.* 1988; 27:1664.
14. Kathuria SV, Day IJ, Wallace LA, Matthews CR. *J Mol Biol.* 2008; 382:467. [PubMed: 18619461]
15. Wu Y, Vadrevu R, Kathuria S, Yang X, Matthews CR. *J Mol Biol.* 2007; 366:1624. [PubMed: 17222865]
16. Lum K, Chandler D, Weeks JD. *The Journal of Physical Chemistry B.* 1999; 103:4570.
17. Huang X, Margulis CJ, Berne BJ. *Proc Natl Acad Sci U S A.* 2003; 100:11953. [PubMed: 14507993]
18. ten Wolde PR, Chandler D. *Proc Natl Acad Sci U S A.* 2002; 99:6539. [PubMed: 11983853]
19. 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. *Angewandte Chemie-International Edition.* 2006; 45:3611.
20. Hua L, Zangi R, Berne BJ. *J Phys Chem C.* 2009; 113:5244.
21. Hummer G, Rasaiah JC, Noworyta JP. *Nature.* 2001; 414:188. [PubMed: 11700553]
22. Li J, Liu T, Li X, Ye L, Chen H, Fang H, Wu Z, Zhou R. *J Phys Chem B.* 2005; 109:13639. [PubMed: 16852709]
23. Li X, Li J, Eleftheriou M, Zhou R. *J Am Chem Soc.* 2006; 128:12439. [PubMed: 16984193]
24. Zhou R, Huang X, Margulis CJ, Berne BJ. *Science.* 2004; 305:1605. [PubMed: 15361621]
25. Hua L, Huang X, Liu P, Zhou R, Berne BJ. *J Phys Chem B.* 2007; 111:9069. [PubMed: 17608515]
26. Young T, Hua L, Huang XH, Abel R, Friesner R, Berne BJ. *Proteins-Structure Function and Bioinformatics.* 78:1856.
27. Krone MG, Hua L, Soto P, Zhou RH, Berne BJ, Shea JE. *J Am Chem Soc.* 2008; 130:11066. [PubMed: 18661994]
28. Liu XG, Zhang Y, Goswami DK, Okasinski JS, Salaita K, Sun P, Bedzyk MJ, Mirkin CA. *Science.* 2005; 307:1763. [PubMed: 15774755]
29. Berne BJ, Weeks JD, Zhou R. *Annual review of physical chemistry.* 2009; 60:85.
30. Liu P, Huang X, Zhou R, Berne BJ. *Nature.* 2005; 437:159. [PubMed: 16136146]
31. Bilsel O, Zitzewitz JA, Bowers KE, Matthews CR. *Biochemistry.* 1999; 38:1018. [PubMed: 9893998]
32. Gualfetti PJ, Bilsel O, Matthews CR. *Protein Sci.* 1999; 8:1623. [PubMed: 10452606]
33. Yu N, Hagan MF. *Biophysical journal.* 2012; 103:1363. [PubMed: 22995509]
34. Patel AJ, Varilly P, Jamadagni SN, Hagan MF, Chandler D, Garde S. *J Phys Chem B.* 2012; 116:2498. [PubMed: 22235927]
35. Myers JK, Pace CN, Scholtz JM. *Protein Sci.* 1995; 4:2138. [PubMed: 8535251]
36. Wu Y, Vadrevu R, Yang X, Matthews CR. *J Mol Biol.* 2005; 351:445. [PubMed: 16023136]
37. Wu Y, Kondrashkina E, Kayatekin C, Matthews CR, Bilsel O. *Proc Natl Acad Sci U S A.* 2008; 105:13367. [PubMed: 18757725]
38. Baldwin RL, Frieden C, Rose GD. *Proteins.* 2010; 78:2725. [PubMed: 20635344]
39. Shakhnovich EI, Finkelstein AV. *Biopolymers.* 1989; 28:1667. [PubMed: 2597723]
40. Kiefhaber T, Labhardt AM, Baldwin RL. *Nature.* 1995; 375:513. [PubMed: 7777063]
41. Kiefhaber T, Baldwin RL. *Proc Natl Acad Sci U S A.* 1995; 92:2657. [PubMed: 7708700]
42. Jha SK, Udgaonkar JB. *Proc Natl Acad Sci U S A.* 2009; 106:12289. [PubMed: 19617531]
43. Forsyth WR, Bilsel O, Gu Z, Matthews CR. *J Mol Biol.* 2007; 372:236. [PubMed: 17619021]
44. Bueno M, Campos LA, Estrada J, Sancho J. *Protein Sci.* 2006; 15:1858. [PubMed: 16877708]
45. Munson M, Balasubramanian S, Fleming KG, Nagi AD, O'Brien R, Sturtevant JM, Regan L. *Protein Sci.* 1996; 5:1584. [PubMed: 8844848]
46. Dalal S, Canet D, Kaiser SE, Dobson CM, Regan L. *Protein Eng Des Sel.* 2008; 21:197. [PubMed: 18299293]
47. Yoder MD, Lietzke SE, Jurnak F. *Structure.* 1993; 1:241. [PubMed: 8081738]

48. Zitzewitz JA, Bilsel O, Luo J, Jones BE, Matthews CR. *Biochemistry*. 1995; 34:12812. [PubMed: 7548036]
49. Main ER, Lowe AR, Mochrie SG, Jackson SE, Regan L. *Curr Opin Struct Biol*. 2005; 15:464. [PubMed: 16043339]
50. Cox A, Arroyo MM, Mayo KH. *Biochem J*. 2001; 357:739. [PubMed: 11463344]
51. Lappalainen I, Hurley MG, Clarke J. *J Mol Biol*. 2008; 375:547. [PubMed: 18022190]
52. Hills RD Jr, Brooks CL 3rd. *J Mol Biol*. 2007; 368:894. [PubMed: 17368485]

Figure 1.

Ribbon diagrams of α TS (a) highlighting the three hydrophobic clusters formed by the ILV residues: Cluster 1 (blue), Cluster 2 (orange) and Cluster 3 (green), respectively, 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 β 2 (purple), Cys81 in α 2 (blue); Leu99 in β 3 (green); and Leu176 in β 6 (orange). Coordinates of α TS from *S. typhimurium* were used to generate the figure from a refined version of PDB file 1BKS using PyMOL v. 1.3¹.

Figure 2.

(a) Ribbon representations of ILV Cluster 1, in which the helices (shown in cyan) are manually separated by a separation distance d from the β -strands (shown in yellow) to create a hydrophobic cavity. The cavity is initially filled with water molecules (red spheres). The observation volume of $\sim 1300 \text{ \AA}^3$ is shown with black lines. Branched aliphatic side chains (heavy atoms only) are shown in sticks as well as in molecular surfaces. The cluster is shown such that the bottom of the figure is N-terminus of β -strands. (b) Plots of normalized water density as a function of simulation time (ns) of the ILV Cluster 1 for two separation distances, 4 $\text{\AA}$ (black) and 6 $\text{\AA}$ (red). The normalized water density is obtained by dividing number of water molecules with maximum number of water molecules inside the cavity. The number of maximum water molecules in the cavity is 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. The separation distance is 4 $\text{\AA}$. (d) Plots of normalized water density as a function of simulation time (ns) for the three ILV clusters of α TS (Cluster 1 in black, Cluster 2 in red, and Cluster 3 in green).

Figure 3.

(a) Ribbon diagram of the interior of Cluster 1, portraying the ILV side chains selected for alanine-scanning mutagenesis, in space filling format. (b) Histograms of water density inside Cluster 1 for the wild-type protein and the ten 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.

Figure 5. Dewetting of ILV Cluster 2. (a) Snapshots of typical wet (left) and dry (right) states of Cluster 2 populated during simulation. The cluster is shown such that the bottom of the figure is the N-terminus of the β -strands. The separation distance is 4 Å. The number of maximum water molecules is 17 within the observation volume of $\sim 500 \text{ Å}^3$. (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 α TS and the L50N, C81N, L99N and L176N variants from 190 nm–260 nm with protein concentrations ranging from 3–7 μ M. (b) Near-UV spectra of wild-type α TS and the L50N, C81N, L99N, and L176N variants from 250 nm–320 nm.

Figure 7. Urea-induced equilibrium unfolding profiles for wild-type and the hydration variants, L50N, C81N, L99N and L176N. The continuous lines represent the fit of the data to a 3-state model

Figure 8.

(a) Bar graph of the free energy difference for $N \rightleftharpoons I$ (solid bars) and $I \rightleftharpoons U$ (hatched bars) transitions obtained from fitting the data to a 3-state model for the hydration variants. (b) Bar graph of the difference in the mean residue ellipticity (ΔMRE) at 222 nm of the intermediate state, I, in reference to the unfolded state, U for wild-type αTS and the L50N, C81N, L99N and L176N variants.

Table 1

Thermodynamic parameters for urea-induced unfolding of wild-type and hydration variants of α TS^(a)

	$\Delta G_{\text{NI}}^{\circ}(\text{H}_2\text{O})^{(b,c)}$	$-m_{\text{NI}}$	$\text{Cm}_{(\text{NI})}$	$\Delta G_{\text{IU}}^{\circ}(\text{H}_2\text{O})$	$-m_{\text{IU}}$	$\text{Cm}_{(\text{IU})}$	Z-value	$\Delta G_{(\text{total})}^{\circ}(\text{H}_2\text{O})$	$-m_{\text{total}}$
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, $\text{N} \rightleftharpoons \text{I} \rightleftharpoons \text{U}$. $\Delta G^{\circ}(\text{H}_2\text{O})$, m , and Cm 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 transition, respectively.

^(b) Units are as follows: $\Delta G^{\circ}(\text{H}_2\text{O})$, kcal mol⁻¹; m , kcal mol⁻¹ (M urea)⁻¹; Cm , M (urea).

^(c) Errors for $\Delta G^{\circ}(\text{H}_2\text{O})$ and m are standard errors from the fits. Errors in Cm , $\Delta G_{(\text{total})}^{\circ}(\text{H}_2\text{O})$ and $-m_{\text{total}}$ were obtained by standard error propagation of the equation:

$$\text{C}_m = \Delta G^{\circ}(\text{H}_2\text{O})/m; \Delta G_{(\text{total})}^{\circ}(\text{H}_2\text{O}) = \Delta G_{\text{NI}}^{\circ}(\text{H}_2\text{O}) + \Delta G_{\text{IU}}^{\circ}(\text{H}_2\text{O}); -m_{\text{total}} = -m_{\text{NI}} + -m_{\text{IU}}$$

^(d) Values obtained from fitting the amplitudes of the $\text{N} \rightarrow \text{I}$ kinetic unfolding reaction to a 2-state model.

^(e) Values obtained from equilibrium data with the $\text{N} \rightleftharpoons \text{I}$ transition constrained by the values obtained from the kinetic unfolding experiment.