Molecular Dynamics Simulations of the Protein Unfolding/Folding Reaction

VALERIE DAGGETT†

Department of Medicinal Chemistry, University of Washington, Seattle, Washington 98195-7610

Received October 10, 2001

ABSTRACT

All-atom molecular dynamics simulations of proteins in solvent are now able to realistically map the protein-unfolding pathway. The agreement with experiments probing both folding and unfolding suggests that these simulated unfolding events also shed light on folding. The simulations have produced detailed models of protein folding transition, intermediate, and denatured states that are in both qualitative and quantitative agreement with experiment. The various studies presented here highlight how such simulations both complement and extend experiment.

Introduction

How a protein acquires its folded, biologically active conformation remains elusive despite years of intense research activity. To map this folding reaction we need to characterize all states along the way—native (N), transition (TS), intermediate (I), and denatured (D)—as well as the mechanism of conversion between these states. Such characterization has proved difficult because of the dynamic, heterogeneous, and often transient nature of nonnative states. As a result, molecular dynamics (MD) simulations have been performed to complement and extend the information available from experiment to more fully map the folding/unfolding reaction coordinate.

MD is the most realistic simulation technique available, allowing all of the detailed interactions between protein and solvent atoms to be monitored over time. Also MD can be readily applied to elucidate kinetic pathways, which is necessary since sampling is still too limited to reconstruct accurate pathways from energy surfaces.[1] Due to the complexity of the system and the time scale of folding (10 μs−ms for the fastest folding proteins), however, realistic MD simulations of protein folding are still not currently possible, but it is possible to probe unfolding and limited refolding events. The principle of microscopic reversibility dictates that folding and unfolding processes are the same under the same conditions.

The first MD study of thermal protein unfolding focused on the unfolding pathway and intermediate states, i.e., populated, partially folded structures.[2] Studies that followed over the next few years continued with this approach.[3−12] From these studies we can say that there is a connection between intermediate states sampled in the simulations and in experiments. Also, the intermediates have many common properties: They are expanded relative to the native state with disrupted tertiary interactions and hydrophobic cores, they have moderate amounts of dynamic secondary structure, and they have heightened interactions with solvent and in particular have an increased exposure of nonpolar residues.

However, these studies only allowed for glimpses of the early stages of the protein unfolding process. This time limitation has been surmounted more recently due to increased computing power. Nonetheless, such a detailed, atomic-level view began to fill in the missing information about the protein-folding pathway. Also, it became clear that we needed to devise a way to identify transition states, particularly because of the detailed information becoming available via Φ-value analysis,[13−15] and to travel further along the unfolding reaction coordinate to the denatured state.

Unfolding/Refolding Transition States

As in the actual folding/unfolding process, the TS is a high-energy, transiently populated state in MD simulations, which necessitated new approaches for its identification. Previously, the characterization of states along the pathway focused on well-populated conformations that accumulated. This approach clearly will not work for transition states. In addition, the TS of protein folding/unfolding is the state of highest free energy, and unfortunately, free energies cannot be reliably calculated from MD unfolding trajectories with the purpose of mining for conformational states. So, a new method for identifying putative transition states was developed, employing a conformational clustering procedure.[16] This method assumes that the state of highest free energy relates to leaving the first cluster near the native state. This corresponds to a state in which packing interactions are disrupted but the protein has not progressed far enough along the unfolding pathway to allow for a substantial increase in entropy and therefore a decrease in free energy.

In the first application of this approach we focused on the small single-domain, two-state folding protein chymotrypsin inhibitor 2 (CI2, Figure 1a). Fersht and coworkers developed the protein engineering/Φ-values analysis to map more precisely particular regions of a protein transition or intermediate state, and CI2 was the second protein investigated in this way. They made hydrophobic deletion mutants throughout the core and related the destabilization of the TS upon mutation to the effect on stability, or in other words $\Phi = \Delta\Delta G_{TS-D}/\Delta\Delta G_{N-D}$, where Φ-values typically range from 0 (denatured-like extent of structure in the TS) to 1 (nativelike extent of structure).

Valerie Daggett received her B.A. degree in chemistry from Reed College in 1983. She attended the University of California at San Francisco from 1985 to 1990, obtaining her Ph.D. in pharmaceutical chemistry with Peter Kollman and Irwin "Tack" Kuntz. Subsequently, she was a Jane Coffin Childs postdoctoral fellow with Michael Levitt until 1993. She then joined the Department of Medicinal Chemistry at the University of Washington, becoming an Associate Professor in 1998.

† E-mail: Daggett@u.washington.edu.

10.1021/ar0100834 CCC: $22.00
Published on Web 03/22/2002

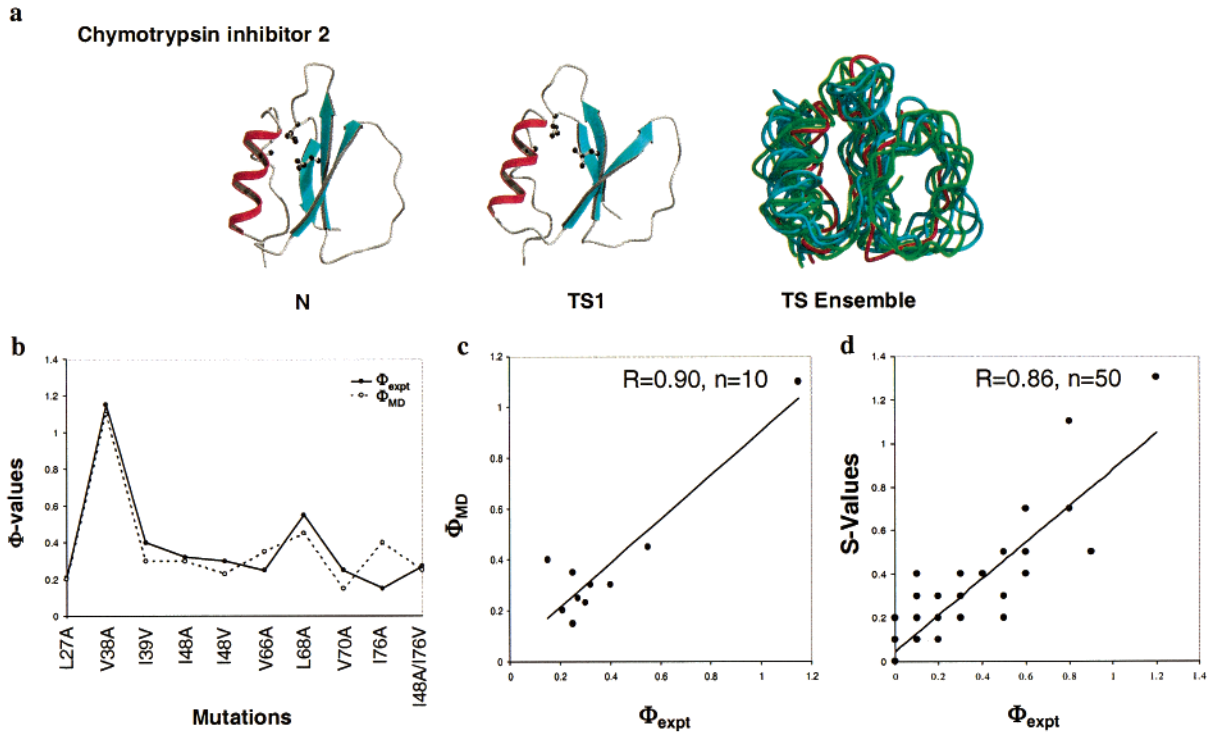

FIGURE 1. Transition state of folding/unfolding of chymotrypsin inhibitor 2. (A) Ribbon representations of the main chain fold of the protein are shown for the native and the average transition state structure from one of the thermal denaturation simulations (TS1 from MD1) on the left. The images on the right show the conformational heterogeneity of the transition state ensemble relative to the native state. The blue transition state structures are from four independent high-temperature simulations (498 K), and those in green are for transition states found in simulations ranging from 348 to 498 K (348, 373, 398, 423, 448, and 498 K). The structure in red is the crystal structure. (b) The experimental Φ-values and the calculated Φ_{MD}-values for hydrophobic deletion mutants are shown as a function of position along the sequence. (Note that the "old" numbering system beginning with residue 19 instead of 1 is used.) (c) A scatter plot of the points shown in panel b is presented. (d) A scatter plot comparing all of the experimental Φ-values and the calculated average *S*-values (the product of the local secondary structure based on ϕ,φ angles and the percent of native-state contacts) for the 4 independent 498 K simulations shown in blue in panel a is presented.

Complementary calculated Φ-values (Φ_{MD}) were predicted while the experiments were being carried out. These calculated values were based on the packing interactions lost upon mutation in the TS model. When both studies were complete, the results were compared and found to be in not only qualitative but also quantitative agreement (Figure 1b,c).[16,17] However, due to complications with calculating Φ_{MD}-values for polar and surface residues in the TS models, a more general, but approximate, approach to reflect the degree of structure in the TS was introduced to keep up with the new mutants being produced in the Fersht laboratory.[18−20] This new structure index, or *S*-value, is the product of the local secondary structure and packing interactions of the residue of interest. The *S*-values are also in decent agreement with experiment (Figure 1d). The TS ensemble does not change appreciably when the denaturation temperature varies between 348 and 498 K (compare the green and cyan structures in Figure 1a). The TS models have been further verified by using them to design faster folding CI2 variants[21] and by performing free energy perturbation calculations to directly compare with the experimentally determined free energy changes.[22]

Further independent support for the overall simulated unfolding events comes from work by Lazaridis and Karplus.[23] They performed 24 independent high-temperature unfolding simulations using an implicit solvent. While the details of any one simulation often differed from that of another, an overall average pathway of unfolding was evident, and it is consistent with the simulations employing explicit solvent described above. The TS ensembles in the different studies are also similar, although that of Lazaridis and Karplus is broader. Their more diverse TS ensemble could be due to the improved sampling (24 versus 10 simulations), the different solvent environments, or just that they did not explicitly identify TS ensembles. Nevertheless, the overall results are in agreement, which is particularly encouraging given the different solvent treatments, force fields, simulation programs, and protocols employed by the two groups.

The nature of the barrier to refolding can be surmised by inspection of the differences between the transition and native states. The side chains are not tightly interdigitated in the simulated TS structures. Hence, they are more mobile than in the native state, which provides for a difference in entropy, but the side chains are still constrained relative to more unfolded structures occurring after the TS. In an attempt to investigate the determinants of the barrier to refolding in more depth, we have also performed temperature-quenched simulations of structures before and after the TS.[24] Structures taken before the TS and simulated under quasi-native conditions became more nativelike on the basis of Cα root-mean-

square deviation (rmsd), specific contacts, and hydrogen bond formation, among other criteria. In contrast, structures taken after the TS did not refold and instead they either unfolded further or remained unchanged. This is striking given that the starting structures are very similar. These results demonstrate that topology is not everything. Structures before and after the TS clearly have the same topology, yet their behavior when placed under folding conditions was quite different. Instead, it is the detailed interactions within the protein and between the protein and solvent that determine whether a structure refolds or not.

Similar results were found earlier for ubiquitin. In that study many structures were extracted from high-temperature simulations and then simulated under quasi-native conditions. The behavior of the quenched structures fell into three groups: nativelike structures near the TS (≤5 Å Cα rmsd, from the crystal structure); partially unfolded structures with some kernel of secondary structure (5−10 Å Cα rmsd); unfolded structures (Cα rmsd > 10 Å). The results obtained with the first group are very similar to those described above for the structures on the native side of the transition of CI2—they essentially refolded. The intermediate, partially unfolded structures collapsed and acquired some native secondary and tertiary structure, but they did not refold. Instead, these structures seem to accumulate near the TS, but they do not pass over it and refold. Thinking that this stumbling block was just due to insufficient simulation time, Duan and Kollman[25] were motivated to perform a much more impressive μs simulation of the villin headpiece beginning with a similarly unfolded, intermediate structure. Their results are consistent with those described above; the protein collapsed and formed some native secondary structure and some tertiary contacts, but it failed to completely refold.

With respect to ubiquitin, the final group of structures underwent successive cycles of collapse and expansion in the search for more productive conformations, and with each collapse, the burial of nonpolar surface area improved. However, these structures did not refold and did not become more nativelike on the whole. One of the more important findings from this study was that contact order seemed to be very important in determining whether a structure would go on to become more nativelike or just cycle through collapsed and expanded states. That is, those structures that moved toward the native state, some of which had Cα rmsd's of 10 Å, had low contact order: the protein first made local interactions and then brought more distant portions of the protein together. In the other case of unproductive collapse, the protein made high contact order, very nonlocal interactions first and these effectively trapped the molecule and prevented productive folding. This idea was picked up and popularized to explain folding rates for a large variety of proteins,[26] but it is not robust.[27] Also, as mentioned above for the CI2 quenching simulations, topology is not the sole determinant of folding.

The focus here has been mostly on CI2, but similar studies have now been performed on a variety of small

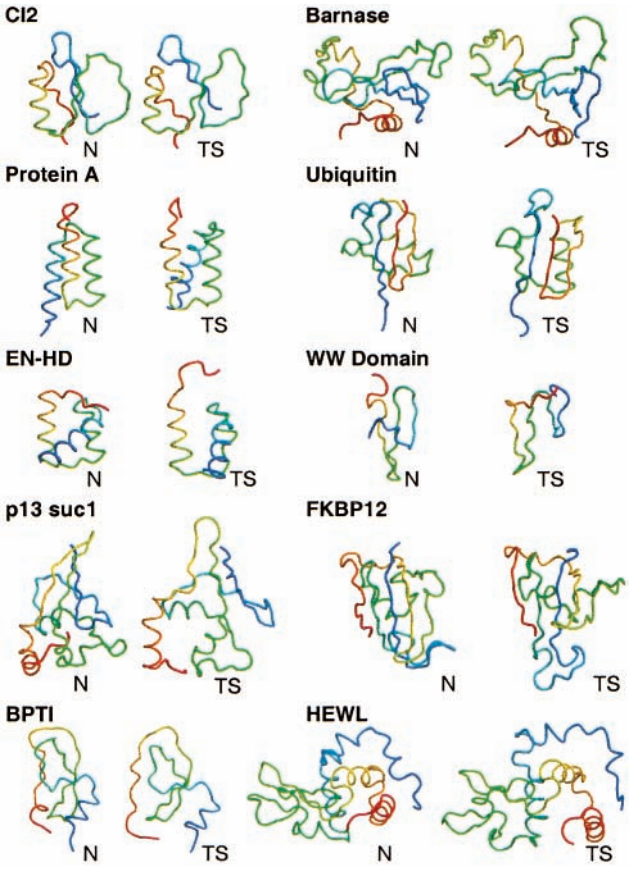

FIGURE 2. Native and transition state structures from thermal denaturation simulations of various small proteins. For each protein the native structure, either the crystal or NMR structure, and a representative transition state structure are provided, showing the overall fold colored from red at the N-terminus to blue at the C-terminus.

proteins (Figure 2). All of the transition states characterized thus far are distorted forms of the native state, and all of these, except barnase, were done as predictions. The predictions have been borne out through comparisons with Φ-values in the cases of CI2,[16−20] p13 suc1,[28,29] FKBP12,[30] and the WW domain.[31] The barnase TS models are also in agreement with experiment,[32] but most of the experimental Φ-values were available before the simulations were performed. Less quantitative experimental results concerning the TS of folding/unfolding of hen egg white lysozyme (HEWL),[33] protein A,[1] ubiquitin,[34] the engrailed homeodomain (EN-HD),[35] and bovine pancreatic trypsin inhibitor (BPTI)[36] support the TS assignments, but their verification awaits further experimental studies.

The various TS structures in Figure 2 share a number of properties. They are all relatively compact and nativelike. Their secondary structure is partially to largely intact with obvious fraying around the edges, although bear in mind that only one structure is displayed for each TS ensemble. Most of the loops and turns are severely altered in the TS. In all cases, the side chain packing is disrupted. In most cases the packing interactions are strongest near the central core of the protein and are weakest near the perimeter. There are some exceptions, however, which tend to be polar interactions involving specific partners

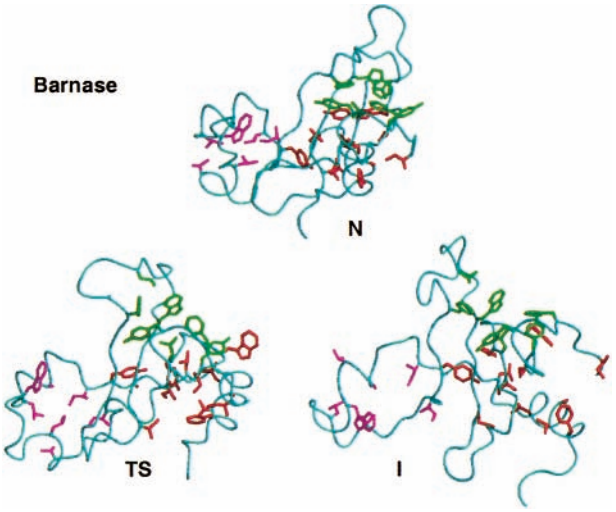

FIGURE 3. Disintegration of packing interactions in the intermediate state of barnase. Side chains in the three hydrophobic cores of the protein are highlighted in the native, transition, and intermediate state structures.

that are maintained in the TS. Interestingly, in almost all cases there are a small number of residues with relatively well-maintained, nonlocal packing interactions. These residues constitute the folding nucleus, although note that in all cases there is some degeneracy built into the nucleus, with neighbors of the primary residues in positions to help or act as surrogates in the case of mutation.

Another complementary approach to evaluating the importance of particular residues in the TS was introduced recently by Vendruscolo et al.[37] The procedure is similar in spirit to how native protein structures are constructed from NMR data. In this case, they created models in which the strength of the contacts was determined by the magnitude of the experimentally determined Φ-values. Interestingly, they found that, of 24 residues used to generate the structures, only 3 were necessary to define the fold of the molecule. Furthermore, it was proposed that these residues constitute the folding nucleus.

Folding/Unfolding Intermediates

In the case of barnase, after the TS, the protein unfolds further to yield an intermediate state. In contrast, a well-populated intermediate is not observed with CI2. The major intermediate of barnase has the same overall features as the intermediates described above: it is expanded, it has a disrupted core and largely disordered turns and loops, and the secondary structure is frayed and in some cases largely disrupted (Figure 3). But, unlike the early simulations addressing partially folded intermediates, Φ-values are available for this state of barnase,[14,38−39] which allows us to better evaluate the validity of the MD-generated intermediate state ensemble.[40] As with the TS of barnase, the agreement between experimental Φ-values and calculated *S*-values is good with a correlation coefficient of 0.82 for 37 sites of the protein. The comparison improves considerably ($R = 0.88$) if we exclude 3 particular problematic residues that retain their side chain interactions in the intermediate but with altered backbones. The overall results presented here for barnase are similar to those obtained by Caflisch and Karplus.[7,8] While they did not calculate Φ-values, it seems likely that their models would also give rise to *S*-values in good agreement with experiment.

Hen egg white lysozyme has also been studied extensively and populates a well-characterized intermediate during folding.[41] Early unfolding simulations of hen egg-white lysozyme resulted in intermediate structures with an unfolded α-domain and a structured β-domain.[3] In contrast, the β-domain unfolded first in a more recent study by Kazmirski and Daggett,[33] leaving a structured α-domain (Figure 4), which is what is observed experimentally.[41] Various intermediates were then populated that differed with respect to the packing of the helices and the elements of nonnative structure adopted (Figure 4). Overall, the pooled ensemble of these intermediates is in agreement with the experimental data for the major kinetic intermediate,[41] suggesting that the kinetic intermediate consists of distinct, but rapidly interconverting, partially folded conformations distinguished primarily by differences in helix packing. The nonnative structure observed in the simulations is critical for explaining many of the experimental observations. Also, it is worth noting that the pooled, simulated intermediates did a better job of reproducing the experimental data than did any individual intermediate conformational ensemble.

Denatured State

Recent studies have shown that residual structure in denatured states is quite common. Previously it was assumed that denatured proteins were random coils, an idea originating from theoretical considerations.[42] The term random coil refers to a "freely jointed chain" lacking restrictions of the bond and dihedral angles, as well as excluded volume effects. Rarely is such a state reached with proteins except under very extreme conditions.

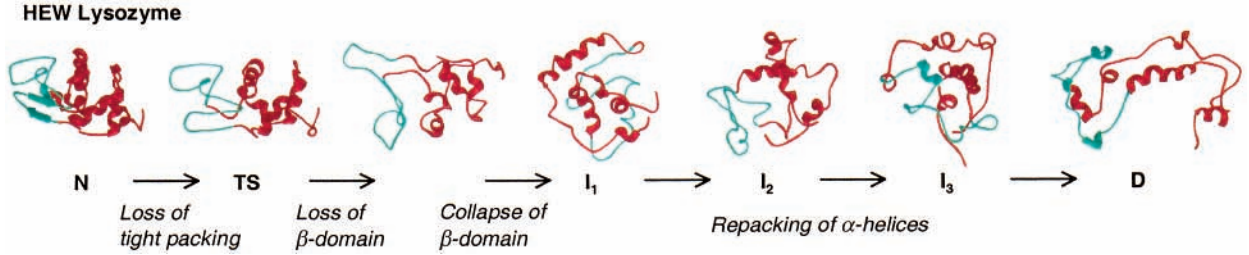

FIGURE 4. Lysozyme unfolding through a variety of intermediates that differ with respect to the packing of the helices.

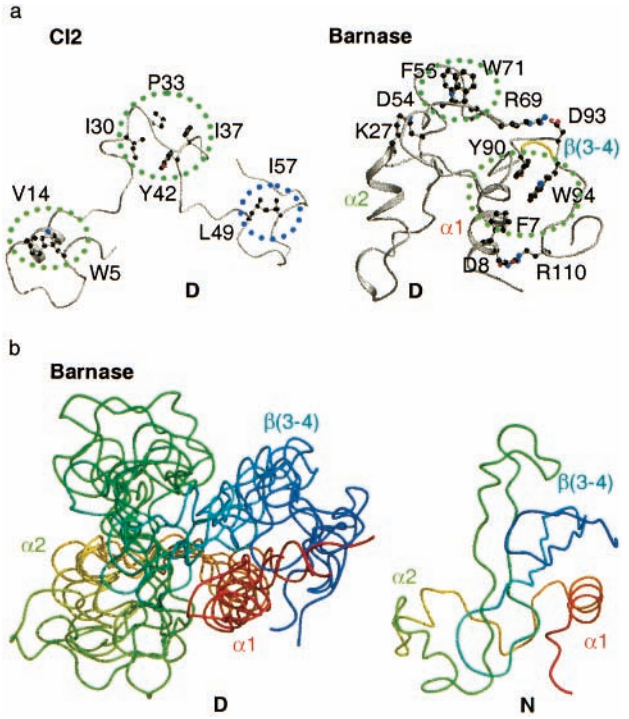

FIGURE 5. Structure of the denatured state: (a) residual structure found in the form of local and nonlocal hydrophobic clusters and both native and nonnative secondary structure; (b) overall chain topology of the native state preserved in the denatured state. Representative barnase structures are provided and are colored from red (N-terminus) to blue.

Instead, 'unfolded' proteins under denaturing conditions tend to contain some degree of residual structure.

Increased computer power has now made it possible to continue denaturation simulations to the point where the denatured state can be sampled and characterized. The first such study focused on barnase.[43] Experiments and simulations show that barnase contains a considerable amount of residual structure comprised of fluctuating native helical structure and both native and nonnative tertiary interactions. The residual tertiary structure is primarily the result of dynamic clustering of hydrophobic side chains. In particular, residues within the $\beta(3-4)$ hairpin participate in such clusters to maintain, or as a result of, the turn between the two strands. In addition, these residues interact with other portions of the protein, specifically with residues in the helix $\alpha1$ (Figure 5). These interactions trigger and then participate in the refolding of the helix, which was termed contact-assisted helix formation. Interestingly, the $\beta(3-4)$ region became disrupted to aid in re-formation of the helix, and it did not form the necessary tertiary interactions with the side chains of the helix when in a native conformation. The average Cα RMSD between the denatured state structures and the native state is 10−15 Å. Despite the dramatic changes in the structure upon denaturation, the overall topology of the protein was retained (Figure 5), which was somewhat unexpected.[43] The retention of gross topology may narrow the conformational search problem associated with protein folding. Recent experimental studies suggest that this phenomenon is not limited to barnase.[44] More expanded investigation of the denatured state of barnase incorporating additional simulations and NMR studies of the denatured protein support the overall findings described above and in addition validate the residual structure observed in the simulations.[45]

Two other studies focusing on the denatured state have been reported, that of BPTI[36] and CI2.[46] In the case of BPTI, the simulated reduced, denatured state is highly dynamic, but it contains some residual hydrophobic clustering in the β-sheet and as well as fluctuating native helical structure. The simulated structures agree with 2D NMR[47] and fluorescence energy transfer experiments,[48] which suggest that there is residual structure. Furthermore, the loss of nativelike secondary structure and increased solvent accessible surface area of the aromatic groups in the simulations is consistent with previous experiments suggestive of random coil behavior. Thus, the simulated denatured state of BPTI is in agreement with both the experiments taken as evidence for random coil and for nonrandom, residual structure. This observation illustrates how ensembles of fluctuating structures can give rise to experimental observables that are seemingly at odds, especially when comparing the results from methods with different inherent resolutions.

An even more extreme view of the denatured state is provided by CI2. The overall conclusion from a combined NMR and MD study is that the denatured state of CI2 is highly unfolded.[46] The chemical shift deviations from random coil suggest that there are only two small regions (residues 19−21 and 30−36) with residual structure (Figure 5). Three of five independent simulations had residual native, helical structure in the denatured state localized between residues 17−21 (Figure 5). The other area of residual structure in the MD simulations was the turn stabilized by hydrophobic interactions around residues 29−37 (Figure 5). The residual structure was transient and weakly populated such that it was not evident in the coupling constants, average Φ−values, or NMR relaxation results. These regions of structure not only limit the conformational space sampled by the denatured state but they also may bias the conformational search toward the native state during folding and help guide the protein along the folding pathway, thereby overcoming the Levinthal paradox.

Putting It All Together and Running the MD Simulations in Reverse: How Does a Protein Fold?

To address how a protein folds we have to cheat and view our simulations in reverse, although we can supplement this picture with the results from simulations addressing limited aspects of refolding (such as the formation of secondary structure and hydrophobic collapse). To begin, we consider the general features of the various denatured states investigated to date. Proteins in the denatured state are dynamic, searching for favorable local interactions and attempting to avoid unfavorable interactions (Figures 6 and 7). They make some nonlocal interactions with

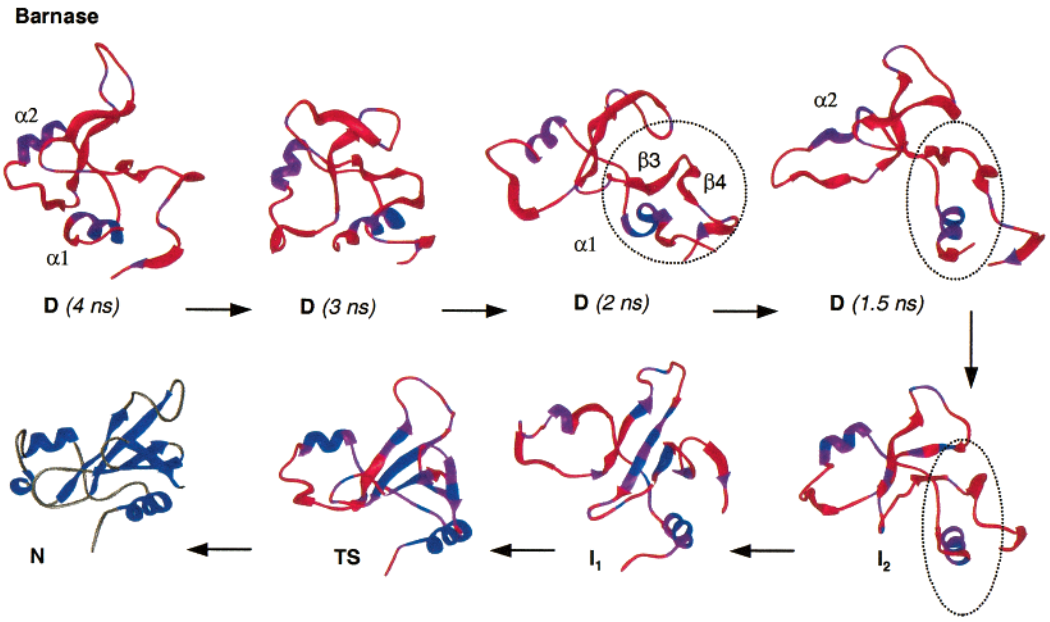

FIGURE 6. Running a barnase unfolding simulation in reverse to illustrate a possible folding pathway.

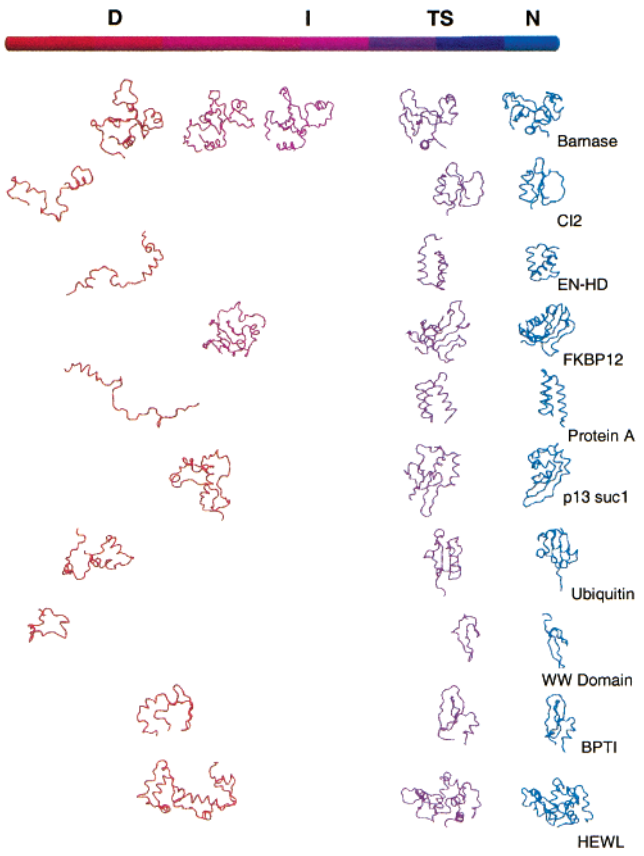

FIGURE 7. Schematic representation of the folding reaction coordinate. Representative structures from the unfolding simulations of a variety of proteins are displayed below the reaction coordinate at their approximate positions along the folding pathway.

residues distant in sequence. Since they are not restricted to just one conformation, a relatively high number of dynamic contacts can be formed without overly compromising their favorable entropy. The overall size and degree of interactions in the denatured state vary greatly among the different proteins studied. The nonnative structure populated in the denatured state tends to be less stable than the native counterpart. But, both native and nonnative structure can be important as folding initiation sites. For example, in the case of barnase, the refolding of the first α-helix is catalyzed by interactions with side chains in the β(3−4) hairpin region.[43,45] These side chains reach out and interact with side chains in the helix and aid in its winding into the helical state. This phenomenon has also been observed in simulations of protein A.[1] The interactions between the different portions of the proteins leading to this tertiary contact-assisted formation of helical structure are both native and nonnative, but it appears that the secondary structure of the β-hairpin from barnase must be disrupted for its side chains to participate in this refolding event.

In addition, after the contact-assisted helix formation in barnase (which is observed directly and not just in reverse), α1 and β(3−4) remain near one another and act as a scaffold for the neighboring β-strands (Figure 6). The residues of these neighboring strands build upon the scaffold to yield a loose, relatively unstructured β-sheet and the associated α1 helix, which is the major intermediate. The sheet is almost planar. This is a common feature of our simulated intermediates with β-structure.[10,34] The β-structure does not acquire its nativelike twist until the packing interactions improve near the TS.

If one moves further along the folding reaction coordinate, the packing interactions in the intermediate improve, the β-sheet twists, and the secondary structure continues to build upon the scaffold and tightens up (Figure 6). Getting to this state involves expulsion of many waters and a considerable loss in both side chain and main chain entropy. In the TS ensemble, the folding nucleus forms, bringing together distant regions of the sequence loosely into place (Figures 2 and 7). Consolidation about this nucleus occurs, leading to the tight interdigitation of side chains. Tightening of the secondary

structure and the hydrophobic core(s) occurs concomitantly with the consolidation about the nucleus. Water-solvating surface residues destined to become more buried during the TS to native state process are expelled into the bulk phase, and any remaining internalized waters are squeezed out.

Concluding Remarks

A prevalent criticism of thermal denaturation simulations is that the high temperature necessary to observe the unfolding process on the computationally accessible nanosecond time scale distorts the energy landscape and does not reflect the process under milder conditions. The agreement between the simulations and experiment, even quantitative agreement, for approximately a dozen proteins with different architectures and spanning transition to denatured states shows that this blanket criticism is not valid. Also, it is now possible to perform much longer simulations such that the unfolding process can be investigated at lower, experimentally accessible temperatures. So far this has been done for CI2,[49] the engrailed homeodomain,[35] and the WW domain,[31] and the results demonstrate that the overall pathway of unfolding is independent of temperature and that the process is merely accelerated by increasing temperature.

Another common criticism of the approach outlined here is that the process of unfolding is unrelated to folding. We can appeal to the principle of microscopic reversibility, but critics will point out that the conditions for the two processes are different and that the simulations monitor nonequilibrium events. However, the temperature-quenched simulations of ubiquitin, CI2, and the villin headpiece are able to directly provide information about some aspects of refolding. The refolding events observed are the opposite of what is observed during the higher temperature unfolding simulations. Also, again, we can bring up the agreement with experiment and the predictive success of the models. These simulated unfolding pathways are able to provide good models for TS ensembles (which are probed in both the folding and unfolding directions experimentally), intermediates (which are usually observed in refolding experiments and are not populated appreciably during unfolding), and the denatured state. Thus, we do have some check of the ability of unfolding simulations to address folding events, and we contend that viewing an unfolding pathway in reverse does in fact reveal a folding pathway. However, even if one can be convinced that the unfolding and refolding pathways are linked and that the unfolding processes simulated represent plausible folding pathways, they are certainly not the only folding pathways. This issue has also been addressed by Dinner and Karplus[50] using a 125-residue lattice protein model at both low and high temperature. They found that the folding and unfolding pathways are indeed similar, but they point out that the high-temperature simulations are probably best applied to "fast track" folding. In conclusion, MD appears to very effectively capture the dominant features of protein folding/unfolding pathways, and as our sampling improves, we can more fully characterize the conformational ensembles populated during folding and the diversity of paths between these states. In addition, through a concerted "dynameomics" effort we are increasing our coverage of sequence and structure space to characterize the unfolding of representatives from all fold classes in an attempt to discern the general rules for folding.

I am grateful for financial support provided by the Office of Naval Research (Grant N00014-95-1-0484) and the National Institutes of Health (Grant GM 50789). UCSF MidasPlus[51] and Molscript[52] were used for the protein displays.

References

(1) Alonso, D. O. V.; Daggett, V. Staphylococcal Protein A: Unfolding Pathways, Unfolded States, and Differences between the B and E Domains. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 133−138.
(2) Daggett, V.; Levitt, M. A model of the molten globule state from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5142−5146.
(3) Mark, A. E.; van Gunsteren, W. F. Simulation of the thermal denaturation of hen egg white lysozyme: trapping the molten globule state. *Biochemistry* **1992**, *31*, 7745−7748.
(4) Daggett, V. A Model for the Molten Globule State of CTF Generated Using Molecular Dynamics. In *Techniques in Protein Chemistry IV*; Angeletti, R. H., Ed.; Academic Press: San Diego, CA, 1993; pp 525−532.
(5) Daggett, V.; Levitt, M. Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* **1993**, *232*, 600−619.
(6) Tirado-Rives, J.; Jorgensen, W. L. Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry* **1993**, *32*, 4175−4184.
(7) Caflisch, A.; Karplus, M. Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1746−1750.
(8) Caflisch, A.; Karplus, M. Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *J. Mol. Biol.* **1995**, *252*, 672−708.
(9) Schiffer, C. A.; Dotsch, V.; Wüthrich, K.; van Gunsteren, W. F. Exploring the role of the solvent in the denaturation of a protein: a molecular dynamics study of the DNA binding domain of the 434 repressor. *Biochemistry* **1995**, *34*, 15057−10567.
(10) Alonso, D.; Daggett, V. Molecular dynamics simulations of protein unfolding and limited refolding: Characterization of partially unfolded states of ubiquitin in 60% methanol and in water. *J. Mol. Biol.* **1995**, *247*, 501−520.
(11) Storch, E.; Daggett, V. The Structural Consequences of Heme Removal: Molecular Dynamics Simulations of Rat and Bovine Apocytochrome b_5. *Biochemistry* **1996**, *35*, 11596−11604.
(12) Laidig, K. E.; Daggett, V. Molecular dynamics simulations of apocytochrome b_{562}: A "highly ordered" molten globule. *Folding Des.* **1996**, *1*, 353−364.
(13) Matouschek, A.; Kellis, J. T., Jr; Serrano, L.; Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **1989**, *340*, 122−126.
(14) Serrano, L.; Matouschek, A.; Fersht, A. R. The folding of an enzyme III. Structure of the transition state for unfolding of barnase analyzed by a protein engineering procedure. *J. Mol. Biol.* **1992**, *224*, 805−818.
(15) Fersht, A. R.; Matouschek, A.; Serrano, L. The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **1992**, *224*, 771−782.
(16) Li, A.; Daggett, V. Characterization of the transition state of protein unfolding: Chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10430−10434.
(17) Otzen, D. E.; Itzhaki, L. S.; elMasry, N. F.; Jackson, S. E.; Fersht, A. R. Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10422−10425.
(18) Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. Detailed structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering. *J. Mol. Biol.* **1995**, *254*, 260−288.

(19) Li, A.; Daggett, V. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 using molecular dynamics simulations. *J. Mol. Biol.* **1996**, *257*, 412−429.
(20) Daggett, V.; Li, A.; Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* **1996**, *257*, 430−440.
(21) Ladurner, A. G.; Itzhaki, L. S.; Daggett, V.; Fersht, A. R. Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8473−8478.
(22) Pan, Y.; Daggett, V. Direct comparison of experimental and calculated folding free energies for hydrophobic deletion mutants of chymotrypsin inhibitor 2: Free energy perturbation calculations using transition and denatured states from molecular dynamics simulations of unfolding. *Biochemistry* **2001**, *40*, 2723−2731.
(23) Lazaridis, T.; Karplus, M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science* **1997**, *278*, 1928−1931.
(24) DeJong, D.; Riley, R.; Alonso, D. O. V.; Daggett, V. Probing the energy landscape of protein folding/unfolding transition states. *J. Mol. Biol.* **2002**, in press.
(25) Duan, Y.; Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998**, *282*, 740−744.
(26) Plaxco, K. W.; Simons, K. T.; Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **1998**, *277*, 985−994.
(27) Dinner, A. R.; Karplus, M. The roles of stability and contact order in determining protein folding rates. *Nat. Struct. Biol.* **2001**, *8*, 21−22.
(28) Alonso, D. O. V.; Alm, E.; Daggett, V. Characterization of the unfolding pathway of the cell-cycle protein p13suc1 by molecular dynamics simulations: Implications for domain swapping. *Structure* **2000**, *8*, 101−110.
(29) Schymkowitz, J. W. H.; Rousseau, F.; Irvine, L. R.; Itzhaki, L. S. The folding pathway of the cell-cycle regulatory protein p13suc1: clues for the mechanism of domain swapping. *Structure* **2000**, *8*, 89−100.
(30) Fulton, K. F.; Main, E. R. G.; Daggett, V.; Jackson, S. E. Mapping the Interactions Present in the Transition State for Folding/Unfolding of FKBP12. *J. Mol. Biol.* **1999**, *291*, 445−461.
(31) Ferguson, N.; Pires, J. R.; Toepert, F.; Johnson, C. J.; Pan, Y. P.; Volkmer-Engert, R.; Schneider-Mergener, J.; Daggett, V.; Oschkinat, H.; Fersht, A. R. Using flexible loop mimetics to extend Φ-value analysis to secondary structure interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13008−13013.
(32) Daggett, V.; Li, A.; Fersht, A. R. A combined molecular dynamics and ϕ-value analysis of structure−reactivity relationships in the transition state and pathway of barnase: The structural basis of Hammond and anti-Hammond effects. *J. Am. Chem. Soc.* **1998**, *120*, 12540−12554.
(33) Kazmirski, S.; Daggett, V. Non-native interactions in protein folding intermediates: Molecular dynamics simulation of hen lysozyme. *J. Mol. Biol.* **1998**, *284*, 793−806.
(34) Alonso, D. O. V.; Daggett, V. Molecular dynamics simulations of hydrophobic collapse of ubiquitin. *Protein Sci.* **1998**, *7*, 860−874.
(35) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518−13522.
(36) Kazmirski, S.; Daggett, V. Simulations of the structural and dynamical properties of denatured proteins: The "molten coil" state of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **1998**, *277*, 487−506.
(37) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **2001**, *409*, 641−645.
(38) Matouschek, A.; Serrano, L.; Fersht, A. R. The folding of an enzyme IV. Structure of an intermediate in the refolding of barnase analyzed by a protein engineering procedure. *J. Mol. Biol.* **1992**, *224*, 819−835.
(39) Matthews, J. M.; Fersht, A. R. Exploring the energy surface of protein folding by structure−reactivity relationships and engineered proteins: observation of Hammond behavior for the gross structure of the transition state and anti-Hammond behavior for the structural elements for unfolding/folding of barnase. *Biochemistry* **1995**, *34*, 6805−6814.
(40) Li, A.; Daggett, V. The unfolding of barnase: Characterization of the major intermediate. *J. Mol. Biol.* **1998**, *275*, 677−694.
(41) Miranker, A.; Radford, S. E.; Karplus, M.; Dobson, C. M. Demonstration by NMR of folding domains in lysozyme. *Nature* **1991**, *349*, 633−636.
(42) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1953.
(43) Bond, C. J.; Wong, K.; Clarke, J.; Fersht A. R.; Daggett, V. Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: Description of the folding pathway. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 13409−13413.
(44) Shortle, D.; Ackerman, M. S. Persistence of nativelike topology in a denatured protein in 8M urea. *Science* **2001**, *293*, 487−489.
(45) Wong, K. B.; Clarke, J.; Bond, C. J.; Neira, J. L.; Freund, S. M. V.; Fersht, A. R. F.; Daggett, V. Towards a Complete Description of the Structural and Dynamic Properties of the Denatured State of Barnase and the Role of Residual Structure in Folding. *J. Mol. Biol.* **2000**, *296*, 1257−1282.
(46) Kazmirski, S. L.; Wong, K. B.; Freund, S. M. V.; Tan, Y. J.; Fersht, A. R.; Daggett, V. Protein Folding from a Highly Disordered Denatured State: The Folding Pathway of Chymotrypsin Inhibitor 2 at Atomic Resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4349−4354.
(47) Pan, H.; Barbar, E.; Barany, G.; Woodward, C. Extensive nonrandom structure in reduced and unfolded bovine pancreatic trypsin inhibitor. *Biochemistry* **1995**, *34*, 13974−13981.
(48) Ittah, V.; Haas, E. Nonlocal interactions stabilize long-range loops in the initial folding intermediates of reduced bovine pancreatic trypsin inhibitor. *Biochemistry* **1995**, *34*, 4493−4506.
(49) Day, R.; Bennion, B.; Ham, S.; Daggett, V. Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.* **2002**, submitted for publication.
(50) Dinner, A. R.; Karplus, M. Is protein unfolding the reverse of protein folding? A lattice simulation analysis. *J. Mol. Biol.* **1999**, *292*, 403−419.
(51) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. The MIDAS display system. *J. Mol. Graphics* **1988**, *6*, 13−27.
(52) Kraulis, P. J. MolScript: A program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallogr.* **1991**, *24*, 946−950.

AR0100834