Protein Folding−Simulation

Valerie Daggett*

Department of Medicinal Chemistry, Box 357610, University of Washington, Seattle, Washington 98195-7610

Received June 29, 2005

Contents

Valerie Daggett received her B.A. from Reed College in 1983. She then went to graduate school at the University of California, San Francisco, in 1985. She worked with Irwin "Tack" Kuntz and Peter Kollman on the conformational properties of peptides, quantum mechanical calculations of enzymes, free energy perturbation calculations, and electrostatic calculation methods. In 1990, she joined Michael Levitt's laboratory at Stanford University where she concentrated on simulations of protein unfolding, building upon her helix-coil work at UCSF. In January 1993, she moved to the Department of Medicinal Chemistry at the University of Washington as an assistant professor. She is now Professor of Medicinal Chemistry and Adjunct Professor in the Departments of Biomedical and Health Informatics, Biochemistry, and Bioengineering. Her current research focuses on characterization of the process of protein folding and unfolding, conformational changes associated with the early steps in amyloidosis, and dynameomics, the lab's bioinformatics initiative, as well as code, methods, and parameter development.

1. Introduction

1.1. Why Study Protein Folding and Unfolding?

Despite decades of active research, protein folding remains one of the most important unsolved problems in molecular biology, and it represents an important missing link necessary for full utilization of the information becoming available from the mapping of genomic sequences. Structure prediction methods typically rely heavily on information gleaned from native, well-structured proteins, which, unfortunately, has proved insufficient for reliable, high-resolution prediction of structure from sequence,[1] although there has been headway in recent years.[2−6] Thus far, our folding rules are based on the properties of native proteins, but more information about the folding process per se may help to bootstrap our way to better prediction algorithms. Characterization of the unfolding process is equally important, both from the perspective of fully understanding a fundamental biochemical phenomenon and for the light shed on the folding process.[7] An understanding of protein folding/unfolding also has important implications for all biological processes, including protein degradation, protein translocation, aging, and human diseases. In this regard, unfolding is particularly important, as it is now appreciated that it plays a critical role in the growing number of amyloid diseases[8−10] and many cellular processes.[11]

The field of protein folding has seen tremendous advances over the past 15 years due, in large part, to technological advances and communication between theoreticians and experimentalists. The underlying technological breakthroughs have been the following: protein engineering to probe specific portions of the protein; the use of NMR to characterize partially unfolded and denatured states of proteins; fast spectroscopic methods; and improvements in molecular dynamics (MD) procedures coupled with the advent of very fast, inexpensive computers to simulate protein unfolding and limited refolding events, at the atomic level. There is a synergy between these various disciplines: experimental studies need theory so that detailed structural models can be used to interpret and exploit the experimental results, and theory in the absence of experimental verification is of limited utility. Accordingly, theory and experiment are finally

* E-mail: Daggett@u.washington.edu.

10.1021/cr0404242 CCC: $59.00
Published on Web 04/26/2006

becoming truly integrated, building on their strengths, to yield a much richer view of the protein-folding process. Here we focus on MD simulations of proteins in solution to investigate the detailed processes of protein unfolding and refolding.

1.2. Why Use Simulations to Study Protein Unfolding?

To fully map the folding/unfolding process, we need to characterize all conformational ensembles along the way—native, transition, intermediate, and denatured—as well as the mechanism of conversion between these states. Such characterization is experimentally difficult because of the dynamic, heterogeneous, and transient nature of partially folded states. Given that experimental approaches only provide limited information for the structural transitions and interactions occurring during protein folding, theoretical studies can nicely complement and extend experiment. While simplified models of protein folding have enriched our understanding of the fundamental principles of protein folding,[12−16] atomic-level resolution of folding and unfolding events requires MD simulation.

MD is the most realistic simulation technique available, allowing all of the detailed interactions between protein and solvent atoms to be monitored over time. MD can also be readily applied to elucidate kinetic pathways, which is necessary since sampling is generally too limited to reconstruct accurate pathways from pseudo-free energy surfaces.[17] Since the first report of MD simulations of protein unfolding,[18] the technique has become quite popular.[19−28] These studies include many using our approach of high temperature or chemicals to disrupt the native state, as has historically been done experimentally. Also, there is a growing number of "steered", or pulling, unfolding simulations to tie in with recent single-molecule experiments.

There are numerous advantages to studying unfolding rather than folding. Such simulations begin from a well-defined starting point—a crystal or NMR structure, which improves the odds of sampling experimentally relevant regions of conformational space. Simulations from an arbitrary extended structure present too many conformational possibilities, and the search problem becomes insurmountable. The system is less likely to become trapped in a local minimum during unfolding, which is common in attempts to simulate the folding reaction. Fortunately, it is not necessary to have MD sample all of conformational space, as real proteins do not sample all possible conformations in the process of folding and unfolding, but we must ensure that we sample experimentally relevant regions of conformational space. Simulations are currently restricted to time scales of $\sim$1 μs or less, which is far too short for the time scale of greater than a millisecond for the halftime of folding of most proteins. So, simulations are typically performed at high temperature to overcome energetic barriers to unfolding. Another advantage to studying unfolding is that the full reaction coordinate from the native to denatured states can be explored. The principle of microscopic reversibility asserts that the pathways of folding and unfolding are the same under the same conditions. Therefore, the mechanism of folding can, in theory, be probed from both directions, and information obtained from studying unfolding can be used to deduce the mechanism of folding, but given the different conditions, this must be assessed, as addressed below.

It has been essential to benchmark the MD simulations by experiment. First, the potential functions employed are empirical and have approximations. Second, extrapolation from unnaturally high temperatures in silico to experimentally accessible temperatures has been questioned. Third, by using such a high-resolution, CPU-intensive technique sampling is necessarily limited. So far, however, there has been excellent agreement between simulated and measured Φ-values (these provide residue-specific structural information for transient states,[29] as described further below), as well as other comparisons.[19,30−35] Repetitive simulation of the same unfolding reaction[36−42] shows that there are variations in the pathway and transition state (TS) of unfolding, but the different structures form an ensemble that fluctuates around the experimental data, which are themselves the average over a large number of molecules.

1.3. Bridging the Gap in Experimental and Simulation Time Scales

A number of ultrafast folding (1−15 μs) and unfolding (5−10 ns) proteins have been discovered recently[43−49] (although it is debatable whether some of these are peptides or proteins), many using laser induced temperature-jump relaxation methods.[50] These systems are important because they help to bridge the gap in time scales between experiment and simulation. In this way, simulations of unfolding can be performed at experimentally accessible temperatures, thereby avoiding extremely high temperatures. Also, given the increase in computational power and new ultrafast-folding proteins, it is nearly time for direct simulations of protein folding pathways.

There are many labs pursuing folding studies although virtually all of them focus on peptides (see Gnanakaran et al., 2003 for a comprehensive review of recent peptide work).[51] For example, Pande and co-workers are particularly active in this area and have recently claimed success using distributed computing (tens of thousands of short simulations on screensavers around the world) on reproducing the rate constant for folding of a 23-residue designed peptide,[52] although potential problems with this approach have been noted.[53,54] However, it is not enough to estimate a rate in agreement with experiment, it must be demonstrated that the actual process being simulated is correct, as discussed further below.

In a state-of-the art study, Simmerling and co-workers[55] performed simulations of the 20-residue Trp cage designed by Neideigh et al.[46] with a folding time of 4 μs.[56] They were able to correctly, and blindly, predict the structure within 1 Å C_α RMSD of the NMR structure, as well as to correctly predict the side-chain orientations. Later work gave comparable results in most cases.[57−59] Extension of these encouraging studies by performing all-atom simulations of larger, but still tractable, systems with explicit solvent is ongoing.

1.4. The Importance of Solvent in Protein Folding/Unfolding

Small organic molecules in aqueous solution can have profound effects on protein stability, structure, and function. The use of these solutions to stabilize or destabilize proteins in the lab is commonplace. Chemical denaturation, with an agent such as urea, is one of the primary ways to assess protein stability, the effects of mutations on stability, and protein unfolding. Mixed solvents can provide insight into the forces that determine the native structure, and there have been a number of interesting related structural studies.[60−64]

In addition, solvent can override inherent secondary structure tendencies in peptides.[65] There have been a few protein MD simulations using realistic cosolvents: ubiquitin in 60% methanol;[66] barnase in 8 M urea;[67,68] γ-chymotrypsin in hexane;[69] subtilisin in dimethyl formamide;[70] chymotrypsin inhibitor 2 (CI2) in 8 M urea;[71] CI2 in 8 M urea/4 M trimethylamine *N*-oxide (TMAO);[72] ubiquitin and cutinase in hexane;[73] and cytochrome P450 BM-3 in 14% dimethyl sulfoxide.[74] The first of these was able to demonstrate solvent-dependent conformational behavior, yielding a partially unfolded state of ubiquitin consistent with NMR studies under the same solvent conditions. The studies of barnase in urea aimed to address the basis of chemical denaturation, but unfortunately, the simulations were far too short (0.9−2 ns), even given the elevated temperature employed (87 °C, 360 K) to denature the protein. The next two studies were different and were nearly neat organic solvents (hexane and dimethyl formamide), as the authors were addressing protein function in organic media. In the next two studies, urea-induced denaturation was achieved, and the mechanism of action of a chemical that counteracts the effects of urea was delineated.

Here, since our focus is on folding/unfolding, we limit ourselves to recent simulations elucidating the mechanism of action of chemical denaturants and counteracting osmolytes. Despite its widespread use, the molecular basis for urea's ability to denature proteins is just now becoming apparent. Urea has been postulated to exert its effect directly, by binding to the protein, or indirectly, by altering the solvent environment.[75−87] Most versions of the direct interaction model have the urea bind to, and stabilize, the denatured state, thereby favoring unfolding. But this interpretation does not explain how the protein surmounts the kinetic barrier to unfolding. In this regard, urea could bind to the protein and compete with native interactions, thereby actively participating in the unfolding process. Alternatively, it has been proposed that urea acts indirectly by altering the solvent environment, thereby mitigating the hydrophobic effect and facilitating the exposure of core residues. Unfortunately, experimental approaches cannot provide the molecular details of how chemicals denature proteins, so MD simulations are being employed to address this issue.

In addition, mechanisms have evolved in nature for organisms to compensate for, and thrive at, extreme conditions. For example, certain marine animals have adapted to life at high pressures and salinity by using osmolytes to maintain cellular volume and buoyancy;[88,89] many of which are denaturants.[90] Interestingly, these animals contain protective osmolytes, such as betaine and TMAO, to counteract the effect of the denaturant.[88] For example, in organisms that concentrate urea as an osmolyte and buoyancy agent, TMAO has been found in roughly a 2:1 ratio.[91,92] TMAO use is becoming a popular in vitro chemical chaperone because of its ability to restore enzyme structure and function.[92−97] As with denaturants, in some cases the proposed mechanism of action involves direct interactions,[98] while other work suggests that the effect is indirect.[94,95,99]

Osmolytes have also been viewed as crowding agents. In contrast to the dilute conditions used experimentally and in simulations, the protein concentration in vivo is 300−400 mg/mL.[100] Such high concentrations can favor the native folded state: for example, the protein FlgM gains structure in living cells as probed by NMR compared with dilute solution conditions.[101] This protein is a so-called intrinsically disordered protein,[102] yet part of the protein becomes structured upon binding a specific transcription factor and, separately, in the high-concentration environment of a cell. Furthermore, Dedmon et al. showed that the protein also folds in high concentrations (>400 mg/mL) of nonspecific crowding agents (glucose, BSA, and ovalbumin). Hopefully future simulation studies will aid in characterization of the mechanism of action of crowding agents.

2. Molecular Dynamics Simulations of Protein Unfolding

CI2 is the archetypical two-state folding protein.[103] It contains a single "module" of structure, and essentially the entire chain contributes relatively uniform interactions over the entire structure. CI2 represents a basic folding unit and as such serves as a model for folding units in larger multidomain proteins.

Case Study: chymotrypsin inhibitor 2

Given that CI2 is a two-state folding protein, the native, transition, and denatured states must be characterized. The structure of the transition state for folding and unfolding has been studied experimentally by a variety of techniques, including a Φ-value analysis using >100 mutations spanning the length of this 64-residue protein[104,105] and the structures of a large number of truncated mutants and peptide fragments.[106] Φ-Value analysis is critical to the analysis of protein folding/unfolding transition states and for validation of MD simulations of the unfolding process. Φ-Value analysis involves introducing mutations throughout the protein and measuring the effects of the mutation on the energetics of the native state (N), transition state (TS), and denatured state (D) using a combination of traditional kinetic and thermodynamic experiments.[29] Ratios of the resulting free energy changes are referred to as Φ-values:

$$\Phi_F = \frac{\Delta G_{TS-D} - \Delta G'_{TS-D}}{\Delta G_{N-D} - \Delta G'_{N-D}} = \frac{\Delta\Delta G_{TS-D}}{\Delta\Delta G_{N-D}} \qquad (1)$$

where ΔG_{TS-D} and ΔG_{N-D} are the free energies of the transition and native states, respectively, relative to the denatured state for the wild-type protein, and the corresponding terms for the mutant are indicated by a prime. Consequently, $\Delta\Delta G_{N-D}$ and $\Delta\Delta G_{TS-D}$ are the destabilization energies of the native and the transition state, respectively, caused by mutation. Consider a case where, in the transition state of unfolding, the structure of the protein at the site of mutation is the same as in the native state. Then, the protein is immune to the effect of the mutation until after the major transition state, and the transition state is destabilized by exactly the same amount as the native state; that is, $\Delta\Delta G_{TS-D} = \Delta\Delta G_{N-D}$ and $\Phi_F = 1$. Conversely, a Φ_F value of 0 implies that the structure of the transition state at the site of mutation is like the denatured state. Intermediate values represent partial structure in the transition state.

Using this approach, structure is inferred from energetics, but detailed molecular structures cannot be obtained using this approach. MD simulations, on the other hand, can provide such detailed structural information. This information comes from denaturation simulations and characterization of the transition and intermediate states using a conformational clustering approach[107] (Figure 1). In addition, MD can evaluate the assumption that the mutation is merely a probe of the wild-type unfolding/folding pathway. Combining

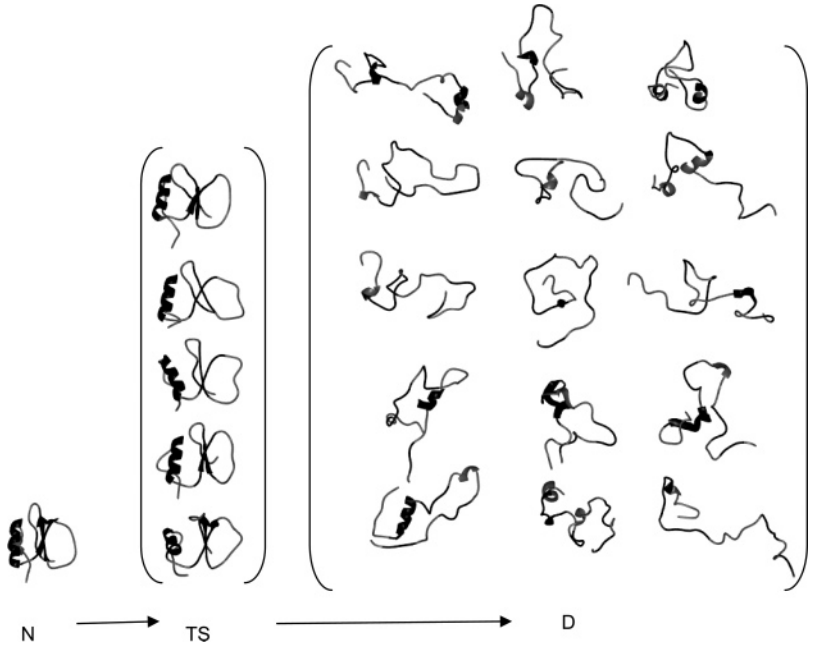

Figure 1. The unfolding of chymotrypsin inhibitor 2. Transition state structures and denatured state snapshots from independent simulations beginning with different members of the NMR ensemble are displayed.

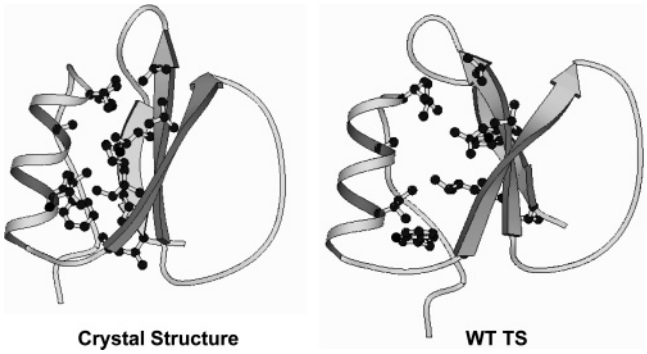

Figure 2. Packing in the hydrophobic core of chymotrypsin inhibitor 2 is disrupted in the transition state.

theory and experiment yields a self-consistent view of the folding/unfolding pathway of CI2. There is a single, common, rate-determining transition state ensemble for folding and unfolding, and the Φ-values for CI2 are independent of whether unfolding or refolding are measured.

The TS of CI2 is quite nativelike with considerable secondary structure and disrupted packing of the side-chains (Figure 2). Experimentally derived Φ-values and the corresponding values that describe the extent of local structure in the MD-generated models (Φ_{MD} or S-values) are in good agreement.[104,105,107,108] The Φ-values for CI2 tend to fall between 0.2 and 0.5. There are some higher values that are found in the α-helix, and in the β-sheet for residues that dock with the helix. In general terms, the transition state for folding resembles a distorted form of the native state, which appears to be increasingly less structured moving out from the helix and where it docks onto the β-sheet. Secondary structure is being consolidated at the same time as long-range interactions.

Multiple unfolding simulations of CI2 (beginning from the crystal structure and different NMR structures) were performed at 498 K, and a transition state was identified from each in the first study to characterize TS ensembles via MD (Figure 2).[36,107] The simulations were done in parallel with the experimental studies in a blind manner; that is, they were performed as predictions, not fits to experiment. The transition states, identified in the simulations by a clustering procedure, were similar overall, and the unfolding pathways only diverged past the transition state as they generate a heterogeneous denatured state. The structures in the TS ensemble have the following characteristics: the hydrophobic core is considerably weakened; the secondary structure, particularly the β-sheet is frayed; and packing of the secondary structure is disrupted considerably (Figure 2).

"Computer mutations" were made to the transition state structures, and the difference in packing contacts between the wild type and the mutant proteins in the transition and native states were evaluated to determine a Φ_{MD} value, which is in very good agreement with experiment for hydrophobic deletion mutants.[104] The best agreement with experiment is when the individual members of the computer-generated transition state ensemble are pooled and averaged, highlighting that the transition state is an ensemble of related structures.[36,108] One of the particularly high Φ-values ($\Phi_F = 1.3$) is for Val 19. Interestingly, this residue makes heightened packing interactions in the TS; that is, new, or nonnative, contacts are made leading to a greater than native extent of structure at that position. Similar unfolding simulations of CI2 by Lazaridis and Karplus,[37] using a different force field, protocols, and program, are consistent with the results described here. In addition, we have always stressed the potential importance of nonnative interactions during protein folding and are skeptical of methods that only consider native interactions. In a recent MD study by Settanni et al.,[109] they also find that nonnative interactions are important and emphasize that misleading results can be obtained if all Φ-values are interpreted in terms of just native contacts.

The agreement between experiment and simulation lends support, on one hand, to the assumption that the protein engineering approach need not dramatically change the folding process and can report on the behavior of the wild-type protein, and, on the other hand, that MD simulations at high temperature provide a credible description of protein unfolding at experimentally accessible temperatures. Further

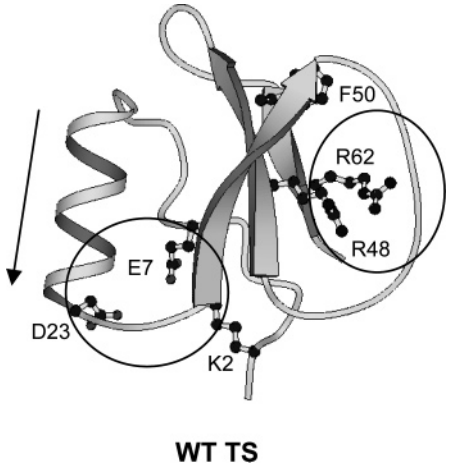

Figure 3. Unfavorable charge interactions in the wild-type transition state of chymotrypsin inhibitor 2. The circle on the left shows the negatively charged Asp 23 at the C-terminus of the alpha-helix. In the active site loop the disposition of the neighboring charged Arg residues is unfavorable (circle on the right).

tests of the simulations were conducted using the simulated MD structures to identify TS-specific mutations that should decrease the energy barrier for folding, thereby increasing the rate.

The models pinpoint a number of unfavorable local interactions at the carboxyl-terminus of the α-helix and in the protease-binding loop region of CI2. So, the prediction is that if unfavorable interactions are removed via mutation, folding will speed up. The first region investigated was the C-terminus of the helix. Asp 23 stabilizes the native protein by making a salt bridge with Lys 2 (Figure 3). But the presence of an Asp at this position in an isolated helix is destabilizing through unfavorable interactions with the carbonyl groups at the end of the helix, which some consider a helix macrodipole. The simulated transition state effectively has an isolated helix when this salt bridge is broken. We predicted that a Asp 23 → Ala mutant should fold faster than wild-type CI2 through stabilization of the transition state (Figure 3). This is found to be the case; the refolding rate constant increases from 56 for wild type to 84 s^{-1} for Ala 23.[110] These increases are especially significant considering that overall destabilization of CI2 generally leads to a decrease in the rate constant for folding.

The second region investigated contains a cluster of positive charges, comprised of Arg 43, Arg 46, Arg 48, and Arg 62 (Figure 3). The hydrophobic side-chains of these residues stack, leading to their guanidinium groups being close to one other and causing electrostatic strain. This strain is partly relieved by a network of hydrogen bonds with the carboxylate of the C-terminal residue, Gly 64, in the native state. This loop region is expanded and more loosely packed in the transition state (Figure 3). The TS models show three or four of the Arg residues in proximity, and the native salt bridges and favorable ionic interactions are not well formed. The removal of some of the unfavorable electrostatic interactions between the positively charged guanidinium groups, and improvement of nonpolar packing in the region, would therefore be expected to stabilize the transition state. The effect was predicted to be a TS-specific effect with only minimal effects on N and D. An Arg 48 → Phe mutation was made, and the rate of folding increases from 56 to 2300 s^{-1} to yield the fastest folding form of CI2 thus far. The mutations described above were designed to yield faster folding versions of CI2 based solely on the MD-generated transition state models. This work shows that MD simulations can aid in the engineering of faster folding proteins.

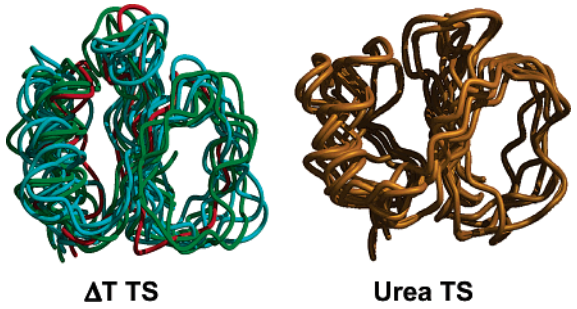

Figure 4. Overlay of transition state structures of chymotrypsin inhibitor 2. Left, structures from multiple 498 K simulations (green) and at a variety of other temperatures (cyan) are superimposed on the crystal structure (red). On the right are transition state structures from 8 M urea simulations at 333 K.

It must also be kept in mind that the transition state is, like any other thermodynamic or kinetically populated state, comprised of an ensemble of conformations. This state is more heterogeneous than that of the native state, but it is constrained relative to intermediate and denatured states. That is, although there might only be a few key contacts in the nucleus of the transition state, this does not mean that the rest of the protein is random or widely divergent. Figure 4 illustrates the heterogeneous nature of the transition-state ensemble and the relative insensitivity of this ensemble to large changes in temperature. From this ensemble, the final steps in folding involve the expulsion of water molecules from the interior and exposed residues, and the fine-tuning of the side-chains, which then leads to the much tighter native-state ensemble.

The unfolding simulations, described above, were continued until the protein unfolded (Figure 1). The denatured state ensemble is expanded with little persistent secondary structure and few tertiary contacts (Figure 1). There is some dynamic, residual native helical structure, but the β-sheet is totally destroyed. There are some dynamic hydrophobic clusters in the denatured ensemble, of which the more persistent ones are found in the center of the protein. The MD-generated view of the denatured state was confirmed through NMR studies. The denatured state of CI2 is largely unstructured as probed by NMR; however, there is some tendency for very weak native helical structure and some weak clustering of hydrophobic residues, particularly near the center of the protein, as revealed by deviations in main-chain and side-chain NMR chemical shifts from random coil values.[111]

The overall picture from the experimental and theoretical studies is that CI2-folds by a nucleation−condensation mechanism, in which the protein collapses around an extended nucleus.[105,112] There is concurrent consolidation of tertiary and secondary structure as the protein collapses around the extended nucleus around the helix. This nucleus is best viewed as a patch of the helix and a portion of the β-sheet. It is dispersed enough, and there is enough degeneracy in the surrounding residues that the nucleus can shift, and, like the overall topology of the TS ensemble, there is some heterogeneity in the nucleus. This nucleation−condensation or nucleation−collapse mechanism has now been found to be quite common.[31,32] As simulated directly by MD, unfolding, expansion, and loss of secondary structure occur concomitantly in unfolding and the nucleation site

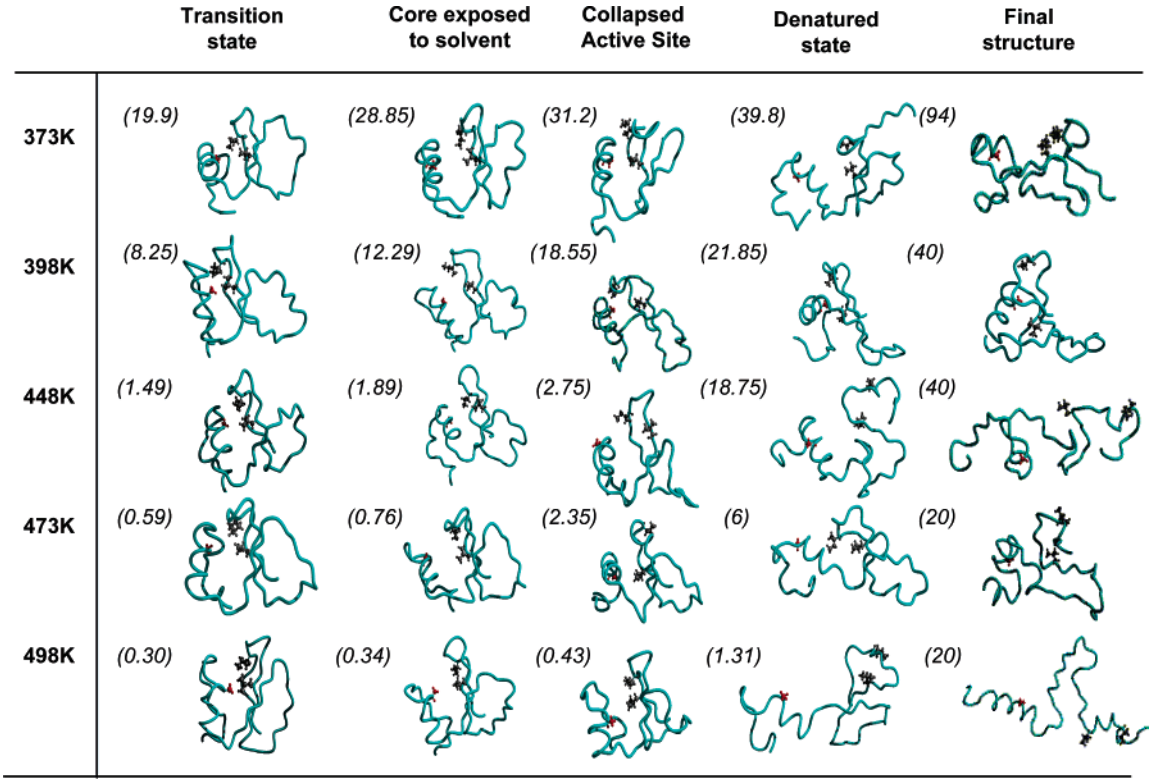

Figure 5. Structures from independent simulations of chymotrypsin inhibitor 2 at different temperatures with times (in ns) of the snapshots in italics.

remains embryonic until sufficient long-range contacts are made[36,107] (Figures 1 and 2). The necessity of involving residues along the entire chain has also been seen in studies of CI2 fragments, particularly involving systematic truncation at the C-terminus.[106] When more than a few residues are removed from the C-terminus, the protein will not fold.

2.1. The Effect of Temperature on Protein Unfolding

Molecular dynamics simulations of protein unfolding generally employ high temperature to accelerate the unfolding process so that the accompanying conformational transitions can be viewed on the ns time scale. For example, extrapolation of the experimentally determined temperature dependence of the unfolding rate constant of CI2 in the temperature range 290−313 K[103] indicates that the average time necessary for unfolding is 9×10^{-2} s at 360 K and 3×10^{-8} s (30 ns) at 498 K. If experimentalists were limited to the nanosecond regime, they too would have to use very extreme conditions. At 498 K, high pressure is required to maintain water as a liquid. At 498 K, the density of water is 0.829 g/cm^3, which corresponds to liquid water at a pressure of ∼26 atm.[113] Under these conditions, the structural and dynamical properties of the water model used in the simulations are in agreement with corresponding experiments.[38,114] Nonetheless, as early as the first MD simulations of a protein unfolding in solution, temperatures of 423 and 498 K were used, and similar behavior was observed at the two temperatures, suggesting that raising the temperature speeds up the process without changing the pathway.[18]

Understandably, the use of such high temperatures has its detractors. It has been claimed that high temperature distorts the energy landscape and that high-temperature simulations are thus irrelevant. This is usually given as a statement of fact. We do not, however, know that the energy landscape is changed dramatically. It may be that high temperature merely affects the rate of unfolding, as it would for a traditional activated process. Given our experience with many different protein systems in which very good, and even quantitative agreement, is obtained between low-temperature experiments and high-temperature simulations, it seems unlikely that the high-temperature energy landscape is grossly different from that probed by experimentalists at lower temperatures. Furthermore, in MD unfolding simulations at a variety of temperatures the overall pathway of unfolding is conserved, while detailed interactions may differ given enthalpic and entropic responses to changing temperature. In any case, the relative insensitivity of the unfolding/folding pathway to temperature has been tested directly for a variety of proteins, including CI2,[38] the WW domain,[115] the engrailed homeodomain,[45,116,117] cmyb,[41,118] FF,[119,120] and the designed protein α_3D.[121] The findings for two of these systems are described briefly below. First, we focus on CI2.

Systematic investigation of the unfolding of CI2 as a function of temperature in water at seven different temperatures ranging from 298 to 498 K (25−225 °C), with the 348 K simulation falling near the experimental T_m indicates that unfolding is essentially an activated process.[38] That is, the pathway is not substantially changed across the 200 degree range of temperatures (Figure 5). At all temperatures, the protein unfolds by expanding slightly with a corresponding disruption of core packing. This initial expansion of the protein is followed by fraying of the sheet. Then, the protein reaches the unfolding transition state, which has a weakened hydrophobic core and some loss of secondary structure. Once it passes through the transition state, the protein core is disrupted and becomes more fully solvated. The active site loop is highly distorted. The protein reaches a denatured state with virtually no native structure, although there is fluctuating

secondary and nonnative tertiary structure. The denatured state is more expanded at the highest temperatures, but the local structure, or lack thereof, is similar at all temperatures.

The primary effect of lowering the temperature is merely to increase the time it takes to reach the transition and denatured states. For example, it takes 20 ns to reach the transition state at 373 K and only 0.3 ns at 498 K. The transition state structures sampled at the different temperatures are similar, but it is a heterogeneous ensemble with a Cα RMSD spread across the different temperatures of 4−5 Å (Figure 4). The heterogeneity for transition state structures across this 125 K range in temperature is comparable to the heterogeneity observed for multiple simulations at a particular temperature.

No single force emerges as a dominant contributor to the thermal behavior of the protein. The order of loss of specific native contacts was not conserved across these temperatures. The total number of contacts formed in the transition state ensemble and order of global unfolding events is, however, essentially the same at all temperatures. These observations suggest that the thermal denaturation of proteins is an activated process taking place on an energy landscape that is not grossly changed by elevated temperatures. The barriers to unfolding on this energy landscape can be thought of as the sum of the interaction energies of each contact. While the precise order in which these contacts are broken changes from one simulation to the next, the protein crosses the lower barriers before higher ones, regardless of temperature, and the overall unfolding pathway is conserved.

Similar results were obtained for other proteins (references provided above), but particularly interesting is the engrailed homeodomain (EnHD) because it is an ultrafast unfolding and folding protein, allowing the time scale of the process by MD to be probed directly at experimentally accessible temperatures. In temperature-jump experiments, En-HD folds to an intermediate state in ≤1.5 μs, and the transition from the intermediate to native state takes ∼15 μs. The relaxation kinetics were followed to 338 K and extrapolated to 373 K (Figure 6). Simulations were performed in the same temperature range, at 348 and 373 K, thereby requiring only minor extrapolation compared with past studies. The time taken to reach the transition state in these simulations is in agreement with the unfolding times determined experimentally (Figure 6).

The TS of EnHD contains nativelike secondary structure and a partially packed hydrophobic core, which is consistent with a framework mechanism of folding. The calculated and experimental Φ-values for the TS are in good agreement.[116,118,122] As with CI2, the simulated unfolding process is independent of temperature, and essentially the same transition states are obtained at 348, 373, and 498 K (Figure 6).

From the transition state, reorientation of the helices, expansion, and disruption of the helix docking leads to the intermediate state (Figure 7). This intermediate has a high helical content and few tertiary contacts. Continuation of the simulation shows that the unfolded state of En-HD contains little residual secondary structure, is expanded, and very dynamic (Figure 7). The denatured state contains a low amount of fluctuating helical structure, both native and nonnative. This unfolded state is not populated appreciably under conditions that favor folding ("physiological conditions").[116] Instead, the denatured state under "physiological conditions" is the folding intermediate (Figure 7). The

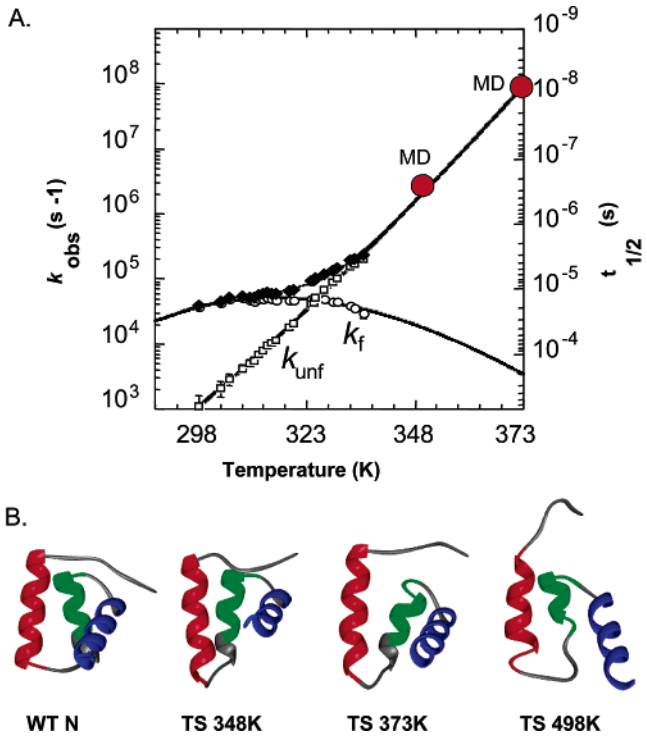

Figure 6. Quantitative agreement in unfolding times by experiment and simulation and invariance in unfolding pathway with temperature. (A) Kinetics of folding and unfolding the engrailed homeodomain in laser T-jump experiments and the times from MD simulations in red. (B) The transition state structure is independent of temperature. All have roughly nativelike structure with helix III (blue) pulled away from the core.

denatured states of many proteins under physiological conditions may be best described as folding intermediates, and highly unfolded denatured states are not usually obtained except under more extreme conditions.

One interesting thing about the unfolding of proteins is that regions of the protein that fluctuate greatly in the native state tend to be among the earliest regions to unfold. This was first shown for bovine pancreatic trypsin inhibitor.[18] The turns and loops of the protein experienced heightened dynamics relative to the rest of the protein in the native state, and they were lost early during unfolding, which is a common theme now that multiple protein unfolding transition states have been determined. Roccatano et al.[123] explored this idea in depth using essential dynamics analysis to determine the preferred motions during the thermal unfolding trajectories of cytochrome *c*. They found a correlation between the deformation motions in the early stages of unfolding and the essential, dominant motions typifying the 300 K state of the protein. One loop region in particular stands out and essential dynamics sampling along that collective mode (i.e., in effect filtering out other motions) leads to rapid unfolding. The authors conclude that thermal denaturation involves the selective excitation of one of a few specific collective motions.

2.2. The Effect of Solvent on Protein Unfolding

Protein structure and function are critically dependent on the solvent environment. Experimental studies suggest that the mobility of proteins decrease in organic solvent,[124−128] which can certainly affect function. Organic solvents are, of course, a common approach for modulating the conformational behavior of proteins, in the most extreme case by

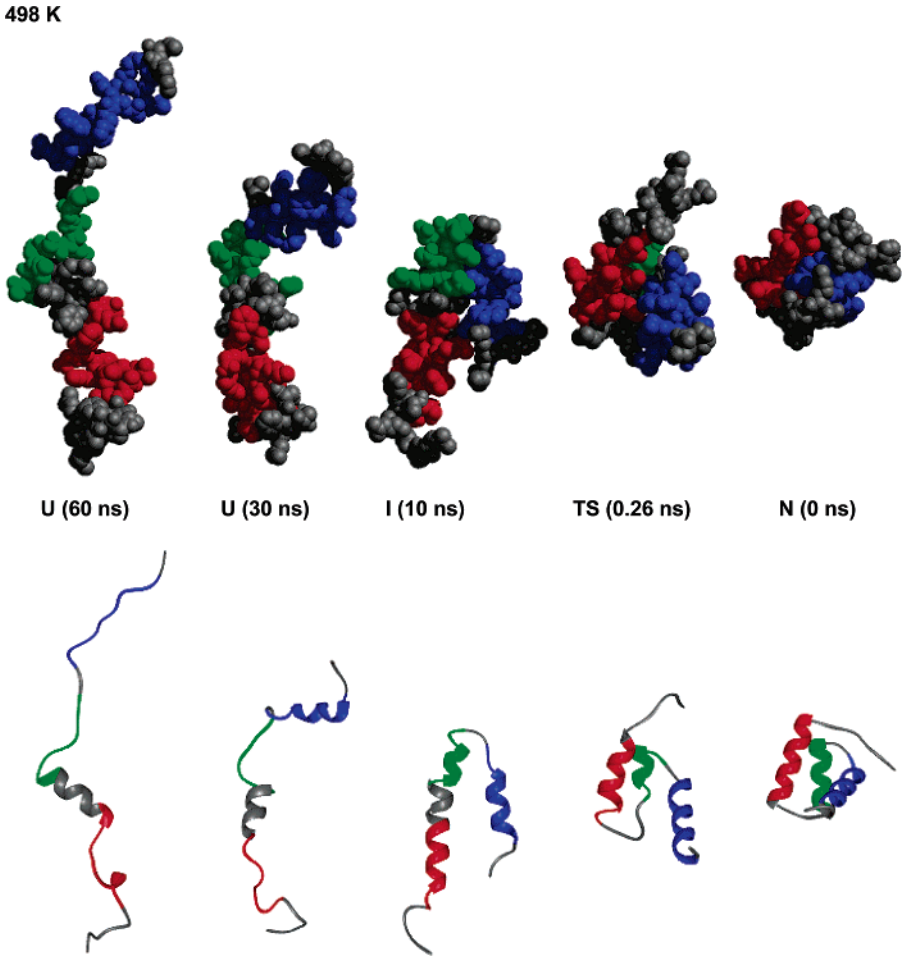

Figure 7. The unfolding pathway (shown in reverse) of the engrailed homeodomain at 498 K. The collapse of the molecule is evident in the space-filling representation at the top. The main-chain traces below show the fluctuating helical structure, which was both native and nonnative in origin. In all cases, the structures are aligned at the end of helix II (in green). The large motion of the chain to bring helices II and III into rough proximity, followed by rotation and collapse of helix I are highlighted by the arrows. The helices are colored in the order red, green, blue.

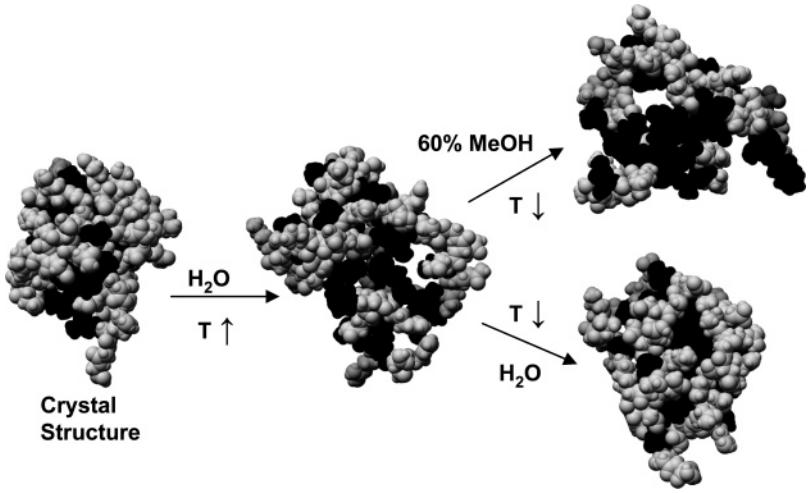

Figure 8. Solvent-dependent conformational behavior of ubiquitin. A structure from a high-temperature thermal unfolding simulation at 498 K adopts an expanded intermediate state in 60% methanol and collapses in pure water at 333 K.

denaturing them, which we focus on below. In the first example of MD simulations in an aqueous organic solvent, solvent-dependent conformational behavior was demonstrated.[66] In this case, the thermal unfolding pathway of ubiquitin and the effect of 60% methanol on intermediate structures quenched from the high-temperature trajectory were investigated (Figure 8). An intermediate was obtained in methanol that was partially structured with a β-hairpin, helix, and the rest of the protein unfolded but with some tendency toward local helical conformations. This intermediate agreed with the available experimental evidence, including hydrogen exchange data and NOEs, which helps to explain discrepancies between the two experimental approaches. This intermediate was not stable in a control simulation in pure water, in agreement with experiment. This control simulation in water instead showed the protein

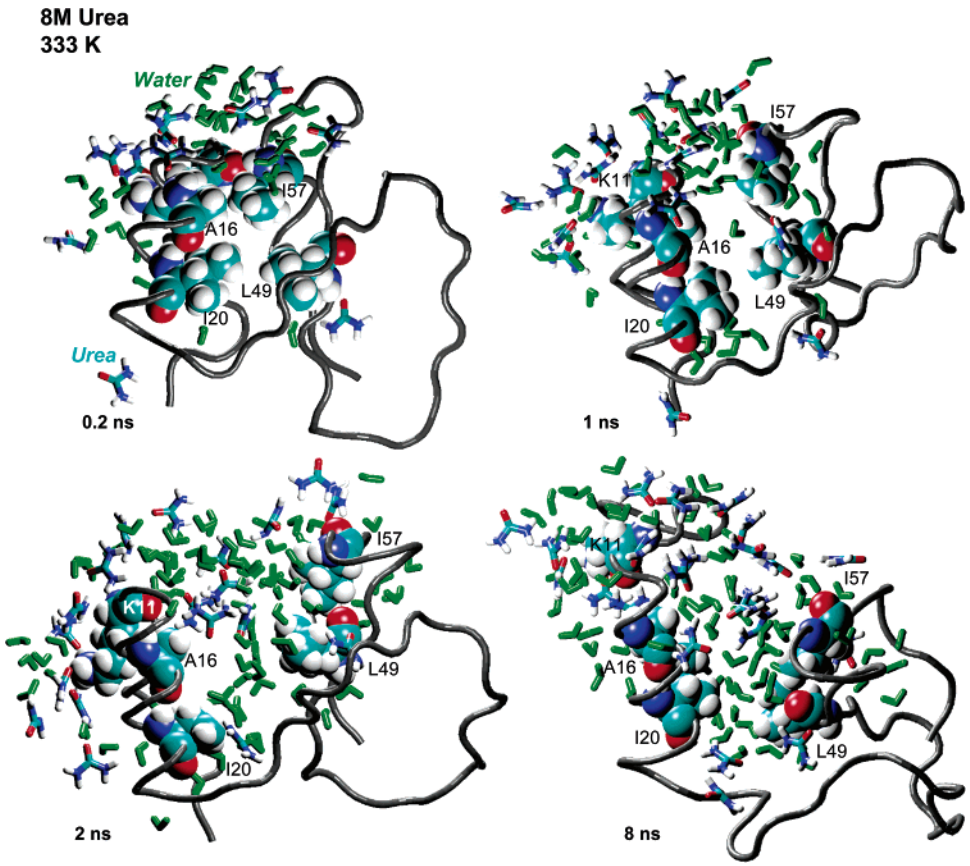

Figure 9. Urea-induced denaturation of chymotrypsin inhibitor 2. Structures from the simulation show the invasion of the active site by water (in green), which is later followed by the introduction of urea.

collapsing and becoming more nativelike, which turned out to be the first simulation of hydrophobic collapse.[66,129] Thus, MD simulations are sufficiently robust that they can provide atomic-resolution information for solvent-dependent conformational behavior.

2.2.1. The Effect of Chemical Denaturants on Protein Unfolding

In preparation for simulations of proteins in aqueous urea solutions, the effect of urea on water structure and dynamics was determined, particularly around alkanes[130] and cyclic dipeptides.[99] Urea has little effect on water structure with respect to radial distribution functions.[130] But the number of hydrogen bonds per water molecule is a useful metric to distinguish the solutions and the strength of the hydrogen bond,[131] as it can reveal subtle differences not seen in other ensemble-averaged properties. The number of hydrogen bonds per water is lower in the hydration shell of nonpolar molecules, such as octane, as these waters are restricted due to their efforts to maximize interactions with neighboring waters while minimizing interactions with hydrocarbon.[130]

Urea leads to a similar decrease in water−water interactions and hydrogen bond strength, as well as local ordering of water around urea's polar atoms, thereby lowering the penalty for exposure of nonpolar groups to solvent relative to pure water.[99] Interestingly, the number of hydrogen bonds per water fails to converge to a constant value until a radius of 6 Å from hydrocarbon or urea is reached. Thus, projection of the urea upon the surrounding aqueous environment results in a 6 Å "radius of influence". The sphere of influence is notable; at concentrations as low as 4 M, 76% of the water is in contact with at least one urea molecule. Thus, urea appears to better solubilize hydrophobic solutes by perturbing water and preloading the solvent to accept nonpolar groups by subtle disruption of water's preferred structure. At higher concentrations, there are more direct interactions between the urea and the solute. These studies suggest that urea acts both directly and indirectly on the solute; then the question is how does it affect proteins.

Simulations of CI2 in 8 M urea at 333 K to match experimental conditions necessary to denature the protein have been described[71] (Figure 9). Urea catalyzes denaturation of CI2 by a combination of direct and indirect mechanisms. First, water structure and dynamics are perturbed by urea, yielding weakened water−water interactions and decreased water diffusion. The water diffusion constants in 8 M urea at 333 K are similar to those of pure water at 298 K, which is unexpected. Many of the early explanations of urea-dependent denaturation relied on chaotropic arguments: urea disorders water structure so that hydrophobic molecules are more easily solvated. At high concentration, this effect on water is magnified and contributes to the "chaotropic" properties of urea and allows for easier solvation of nonpolar residues. Also, the hydrophobic effect increases with temperature,[132,133] reflecting the increasing disparity between perturbed waters forced to align themselves around nonpolar groups and those free to interact with bulk solvent. In 8 M urea, the hydrophobic effect is mitigated by the decrease in water dynamics at 333 K. In effect, the solvent environment is better able to solvate hydrophobic groups and the exclusion of nonpolar side-chains from solvent offers little advantage; thus the indirect effects of urea act to stabilize nonnative states of the protein, including the transition state, thereby accelerating protein unfolding.

Direct interactions between solvent and CI2 are also clearly evident in the simulations. Disruption of water structure by urea diminishes the cohesion of the water, freeing it to be the primary denaturant early in unfolding, thereby providing a link between the direct and indirect effects. The number of water−protein hydrogen bonds increases ~50% from N → TS, with less of a change in urea interactions. However, urea also interacts directly with the protein, particularly after disruption of the secondary structure. The number of urea hydrogen bonds with the peptide backbone increases from TS → D, while water hydrogen bonds remained relatively constant. The increase in "bound" urea upon unfolding is in agreement with estimates based on experiment.[79] The extent of interaction of the main-chain with urea in the transition state in the simulations (30−35% of the residues) is also in good agreement with estimates based on experiment (36%).[81]

Solvent not only participates in specific electrostatic interactions with the protein, it also screens hydrophobic interactions. For example, residues comprising the edge of the hydrophobic core interact with urea and water until stabilizing hydrophobic and hydrogen-bonding interactions are broken, which lead to the influx of water and then urea to the hydrophobic core (Figure 9). As the protein exposes more backbone atoms to solvent, urea interacts preferentially with these atoms and excludes water in the process, which is consistent with previous calorimetric,[99,134] solubility,[86] solvation studies.[76,77,135−137] Relative enrichment of urea around the protein has been observed in previous simulations.[67,68,78]

Simulations of CI2 in 8 M urea indicate that urea promotes unfolding by both indirect and direct mechanisms. Direct urea interactions consist of hydrogen bonding to the polar moieties of the protein, particularly peptide groups, leading to screening of intramolecular hydrogen bonds. Solvation of the hydrophobic core proceeds via the influx of water molecules, then urea. Urea also promotes protein unfolding in an indirect manner by altering water structure and dynamics, as also occurs upon the introduction of nonpolar groups to water, thereby diminishing the hydrophobic effect and facilitating the exposure of the hydrophobic core residues. Overall, urea-induced effects on water contribute indirectly to unfolding by encouraging hydrophobic solvation, while direct interactions provide the pathway.

Then the question is whether urea affects the structure of the transition and denatured states. With respect to the transition state, the TS structures identified in the 8 M urea simulations are very similar, with Cα RMSDs to the TS ensembles obtained with high temperature of approximately 4 Å, which is comparable to the spread seen in the TS ensemble for multiple simulations at a particular temperature (Figure 4).

Urea has a more pronounced effect on the denatured state. Overall, like the thermally denatured state, the protein contains residual structure in the form of dynamic, native helical structure and hydrophobic side-chain clusters (Figure 9), in agreement with experiment.[111] But, the presence of urea shifts the conformational ensemble to a higher population of polyproline II structure relative to the thermally denatured state (Figure 10). A recent CD study of various peptides by Whittington et al.[138] shows that the population of polyproline II increases with increasing urea concentration. Hence, urea alters the conformational properties of the chain in the near absence of tertiary contacts.

While most theoreticians and experimentalists alike tend to consider studies of chemical and thermal denaturation as "natural" ways to perturb the protein, in fact such situations are relatively rare in vivo. So, it is also of interest to probe the effect of force on proteins, for which there are many examples in nature, as described in section 2.3.

2.2.2. The Effect of Chemical Chaperones on Protein Unfolding

With respect to chemicals that modulate protein conformational behavior, osmolytes, or chemical chaperones that stabilize proteins and reverse the effects of chemical denaturation, are very interesting. MD studies in this area are just beginning. Recently simulations of CI2 in mixtures of urea and TMAO, specifically 8 M urea/4 M TMAO at 333 K were performed to address whether TMAO can counteract the effect of urea.[72] These results are compared with the protein's unfolding behavior in 8 M urea at 333 K. The effect of adding 4 M TMAO to CI2 in 8 M urea is dramatic. In 8 M urea, the protein begins to unfold within a few nanoseconds, while the protein remains very structured when 4 M TMAO is included (Figure 11). Thus, the simulations capture the protective effects of TMAO.

Like urea, TMAO acts both directly and indirectly. A small proportion of the total TMAO molecules interact specifically with Lys and Arg residues. The solution is very crowded with a density of 1.13 gm/ml, which hinders motion and large-scale conformational changes. But the more prevalent effect is that TMAO molecules order water and decrease water−water hydrogen bond lengths in both the hydration layer and the bulk solvent. In effect, TMAO ties up the water and discourages it from attacking the protein; in 8 M urea, water is the first denaturant, and urea moves in after the water. In the presence of TMAO, urea interactions with protein decrease in the hydration shell. The net effect is that increased exposure of the protein in TMAO is discouraged, and experimental studies have also reported that proteins favor compact conformations in sucrose, another osmolyte.[139]

2.3. The Effect of Force on Protein Unfolding

Proteins, such as the giant muscle protein titin, may unfold and then refold as part of the process of extension and contraction in their biological function. The extension may be mimicked in experiments using atomic force microscopy (AFM) to provide an extending force to the ends of the molecule and in simulations by applying a pulling/pushing force. Since AFM allows for single-molecule experiments, this is a particularly exciting area for the comparison of MD simulations and experiment. In addition, single-molecule AFM experiments open up the possibility of comparisons between force-, chemical-, and temperature-induced perturbations of protein structure.

In a pioneering study of the wild-type titin domain T1 I27,[140] correspondence between the activation energies, or rates, of unfolding by force and by conventional bulk solution measurements were reported. On the basis of this finding, the authors asserted that the pathways under the different conditions were the same. However, with respect to comparison of force-induced unfolding with bulk solution experiments using chemical denaturants, it is not enough to get correspondence of rates. Evidence is required that the rate constants are for the same processes. More in-depth study of the pathways of unfolding by simulation,[141,142] mutation,[143] and AFM studies on mutants[144,145] indicates that the protein

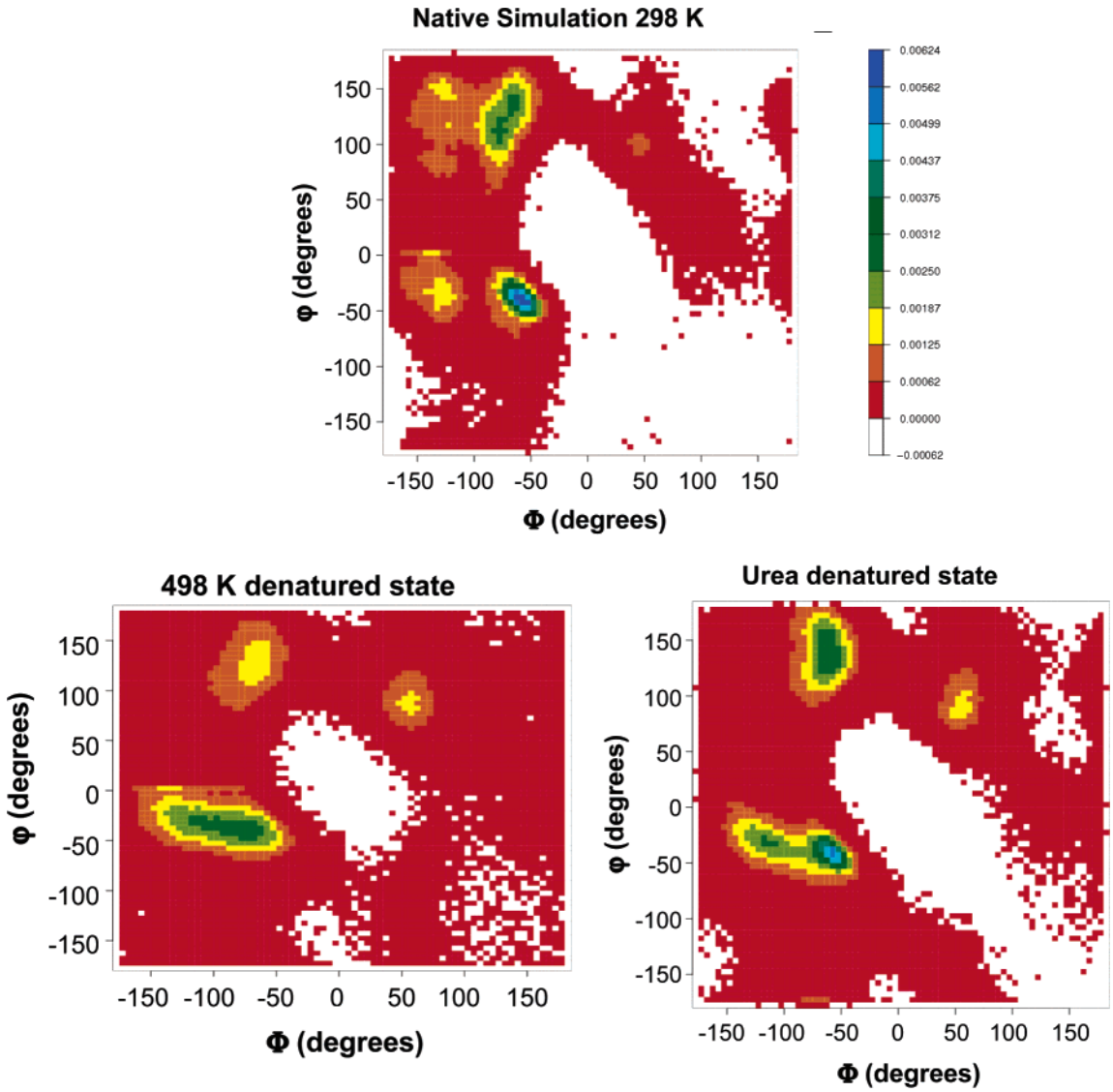

Figure 10. Ramachandran maps for the native (298 K MD), thermally denatured (498 K), and chemically denatured (8 M urea) states of chymotrypsin inhibitor 2.

unfolds by different pathways under the different conditions. The ensemble experiments involving chemical denaturation suggest that the transition state is expanded but nativelike, while the pulling experiments predict a TS that is essentially the native state with one β-strand pulled off. While the pathways of unfolding may not be the same, in this case one can learn about properties of the protein relevant to its biological function when pulling muscle proteins. However, then one wonders how "normal" globular proteins respond to force.

To address this issue, Clarke and co-workers undertook AFM studies of barnase.[146] Barnase is a good model system for evaluating these effects, as its folding/unfolding pathway in solution has been mapped in detail by combining experimental studies with simulation.[147−154] High-temperature simulations provide a TS ensemble with nativelike structure with some loss of secondary structure and disruption of packing interactions (Figure 12). The MD-generated ensemble is in good agreement with experimental Φ-values.[150] Further unfolding leads to a more expanded structure with retention of some secondary structure and contacts in the main hydrophobic core (Figure 12). As with the transition state, this intermediate is in agreement with the experimental Φ-values.[151] Finally, further unfolding yields a denatured state with considerable residual structure, which has also been validated by experiment.[153] This residual structure helps to set up loose, native topology in the denatured state (Figure 12).

The picture of how barnase unfolds is substantially different when force is applied. When barnase is pulled, the N-terminal helix is pulled away from the protein, the final strand peels off, and then the rest of the structure progressively pulls away (Figure 12).[146] The force required is much lower than that required to unfold the muscle proteins. The unfolding pathway involves the unraveling of the protein from the termini, with much more nativelike secondary and tertiary structure being retained in the transition state than is observed in simulations of thermal unfolding or experimentally, using chemical denaturant. These results suggest that proteins that are not selected for tensile strength may not resist force in the same way as those that are and that proteins with similar unfolding rates in solution need not have comparable unfolding properties under force. While force-induced unraveling of the termini may not be relevant to the biological function of most globular proteins, it may be relevant to how proteins are unfolded for transport across membranes.[11]

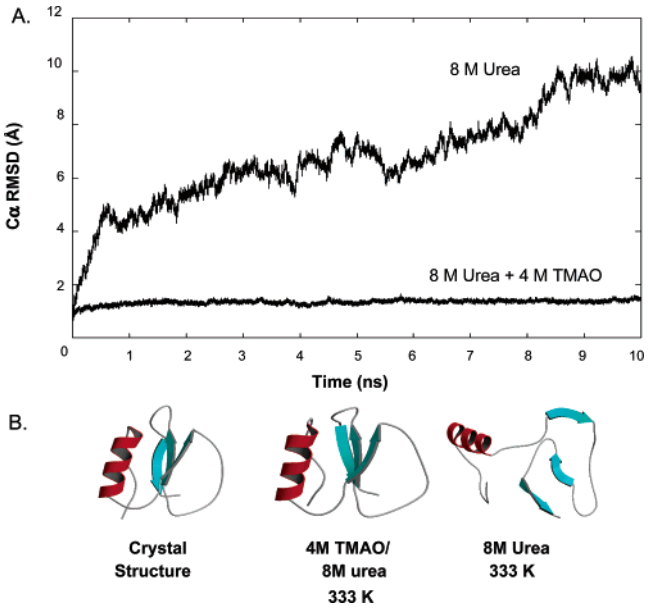

Figure 11. Prevention of urea denaturation in a ternary solvent containing the chemical chaperone TMAO. (A) Ca RMSD as a function of simulation time shows that the protein deviates greatly from the crystal structure upon addition of 8 M urea, while it remains in the native state when TMAO is added to the solution. (B) Final structures (10 ns) from the simulations are displayed.

2.4. Sampling: How Many Simulations Are Required and How Representative Is Any Given Simulation?

A reasonable criticism of all-atom MD simulations of proteins is that the small number of simulations one can routinely perform may not be able to reproduce the average behavior of the $10^{15}-10^{18}$ molecules in an experimental sample. A corresponding criticism of most experiments involving proteins is that they only tell us the average behavior of an ensemble of molecules and do not give us information for individual members of the ensemble. Only recently have experiments been devised that measure the behavior of individual molecules and the distribution of signal about a mean.[155] These experiments show definitively that the properties of a protein observed in experiment are due to broad ensembles of conformations.[156]

MD simulations are the classic single-molecule "experiments", providing atomic-resolution structural and dynamic information. However, the single-molecule nature of the technique has also been its shortcoming, with frequent criticisms of sampling inadequacies and questions regarding the ensemble behavior of large numbers of molecules. Given the increase in computer power, we can now address this issue by performing a large number of simulations and comparing individual and ensemble properties. A recent study describes 100 independent MD simulations of CI2 that were carried out for 20 ns each at 498 K in water to more fully describe the potentially diverse routes of protein unfolding and investigate how representative a single trajectory can be.[42] Rapid unfolding was observed in all cases with the trajectories distributed about an average "ensemble" path in which secondary and tertiary structure was lost concomitantly, with tertiary structure loss occurring slightly faster. Individual trajectories sample conformations far from the average path with very heterogeneous time-dependent properties (Figure 13). Nevertheless, all of the simulations but one follow the average "ensemble" pathway, such that a

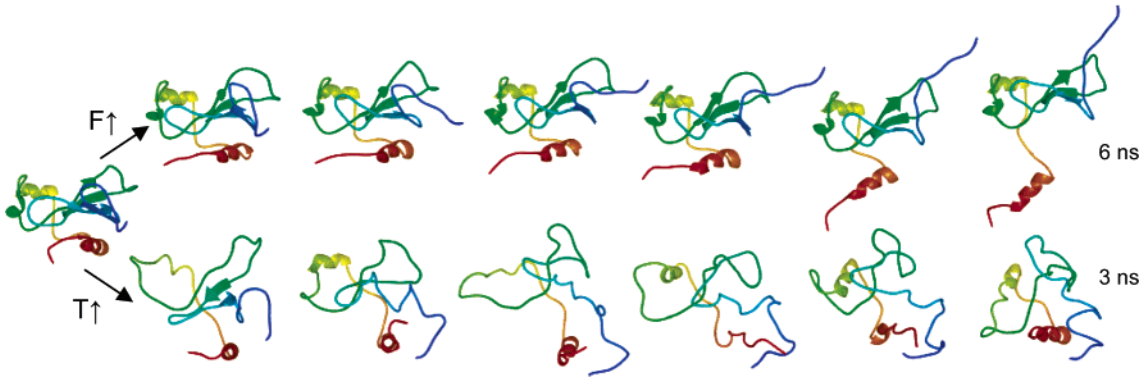

Figure 12. Unfolding of barnase through AFM-like pulling from the ends of the structure and at high temperature (498 K). The pathway of unfolding is different depending on the mode of disrupting the structure. Structures are provided every nanosecond for the force-induced unfolding and every 0.5 ns for the high-temperature simulation.

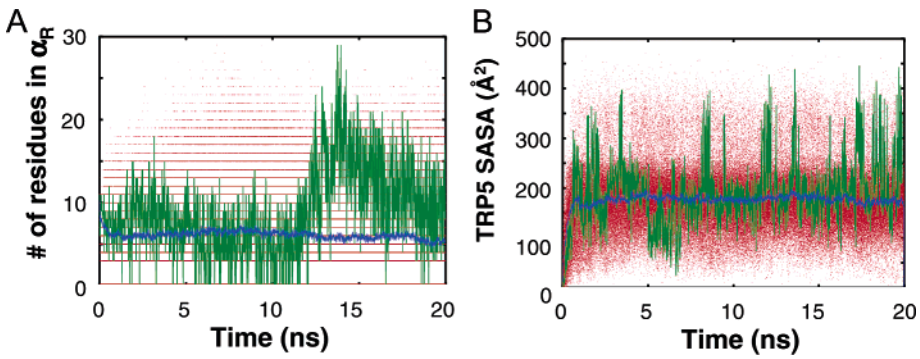

Figure 13. Average properties for 100 unfolding simulations of chymotrypsin inhibitor 2 (blue) at 498 K and all instantaneous values (in red) and a single trajectory (in green) for the (A) alpha-helix content and (B) solvent accessible surface area of Trp 5.

small number of simulations (5−10) is sufficient to capture the average properties of these states and the unfolding pathway.

That a good description of unfolded state properties can be obtained with so few simulations is not due to all of the simulations sampling similar conformations. The same ensemble averages can be determined from quite different ensembles. In many cases, none of the structures sampled late in an individual simulation are seen in any of the other members of a set of five simulations, but the average properties of two independent sets of five simulations will be similar. Additionally, the correlation between small sets of simulations and the full set of 100 increases rapidly as the size of the set increases. While widely different conformations can have similar overall properties, the unfolded per residue properties from individual simulations generally correlate very poorly with the averaged per residue properties. Thus, different simulations are sampling widely different areas of conformational space, but averaging yields a common set of properties.

In contrast, the structure of the transition state from individual simulations is relatively similar to the average structure. That is, the properties of individual molecules are similar to the ensemble properties. While there are some outliers, most of the transition state structural properties have correlation coefficients to the averaged transition state between 0.7 and 0.95. As the transition state is chosen to be the point of exit from the structurally homogeneous native state, these similarities are not surprising. The overall pathway of unfolding is quite well conserved, however, and involves early loss of structure in β-sheet and loops, followed by complete core opening and loss of structure in the α-helix. Only one simulation deviates substantially from this overall description with early unfolding of the α-helix. The small number of simulations required to accurately describe the average properties of the transition state is partly attributable to this relative homogeneity.

Overall, these simulations show that a small number of trajectories, typically, 5−10, are adequate to describe the average properties of protein conformational states. While this appeared to be the case earlier given the good agreement with experiment in studies employing a small number of simulations, now proof has been obtained. But, of course, one could always say that thousands, millions, or more simulations are required, and until they are performed, we do not know that a small number of simulations is adequate. On the basis of our findings over the past decade with many proteins and those from many other labs, this assertion does not appear to be reasonable, but unfortunately it is untestable at present. So we take a pragmatic approach: small numbers of simulations provide reasonable results and provide insight into protein folding. More extensive sampling may, and probably will, provide more insight, but in my opinion it is better to perform a smaller number of simulations and sample new protein folds, say, than to perform hundreds of simulations of the same fold and learn little beyond what can be garnered from the smaller set. This is the basis of a new endeavor we are calling dynameomics in which we are simulating a representative of each nonredundant protein fold.[157]

2.5. Is Unfolding the Reverse of Folding?

Unfolding and folding have been used somewhat interchangeably here, which may not be correct. Given the agreement between the results of unfolding simulations with experiments probing *both* folding and unfolding, we believe that unfolding simulations represent plausible folding pathways when viewed in reverse. But we recognize that sampling is limited, and there is no a priori guarantee that the forward and reverse pathways are the same. Microscopic reversibility dictates that the mechanism of folding can be probed in both the directions of unfolding and refolding, at least under the same conditions. In practice, such conditions are very difficult to achieve for such a complicated process, but Fersht et al.[29] have shown that extrapolating folding and unfolding experiments to obtain results corresponding to the same experimental conditions is valid. In addition, Dinner and Karplus[158] have shown that protein unfolding is the reverse of folding in lattice simulations. We believe that to be the case for MD simulations, as well, but to better address the "reversibility", temperature-quenched simulations to obtain glimpses of refolding events have been undertaken.

Refolding of small peptides can be accomplished using simplified models with better sampling.[159,160] Although we cannot yet fully simulate the refolding process using all-atom MD simulations with inclusion of explicit solvent models, the combination of faster computers and faster proteins means that it should be possible soon. In the meantime, it is possible to simulate portions of the folding reaction coordinate and piece them together for a more complete description of folding.

The first simulations of protein collapse and partial refolding began as control simulations for 60% methanol mixed-solvent simulations of ubiquitin.[66,129] As these represented the first mixed solvent simulations of a protein, simulations in pure water were necessary to control for the effect of extracting partially unfolded conformations from a thermal denaturation simulation. In this case, 11 quenched simulations in water under quasi-native conditions were performed (elevated temperature, 333 K, but below the T_m of the protein), spanning almost native to fully denatured structures. The different simulations addressing different aspects of the refolding pathway were patched together in the end to create a fuller description of the folding process.

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); and unfolded structures (Cα RMSD > 10 Å) (Figure 8). The results obtained with the first group are very similar to those described below for the structures on the native side of the transition of CI2—they essentially refold. The intermediate, partially unfolded structures collapse and acquire some native secondary and tertiary structure, but they do 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[161] were motivated to perform a much longer microsecond simulation of the villin headpiece beginning with a similarly unfolded, intermediate structure. Their results are consistent with those described for ubiquitin; the protein collapses, forms some native secondary structure and some tertiary contacts, but it fails to completely refold.

With respect to ubiquitin, the final group of very unfolded structures undergo successive cycles of collapse and expansion in the search for more productive conformations. With each collapse, the burial of nonpolar surface area improves. However, these structures do not refold and do not become

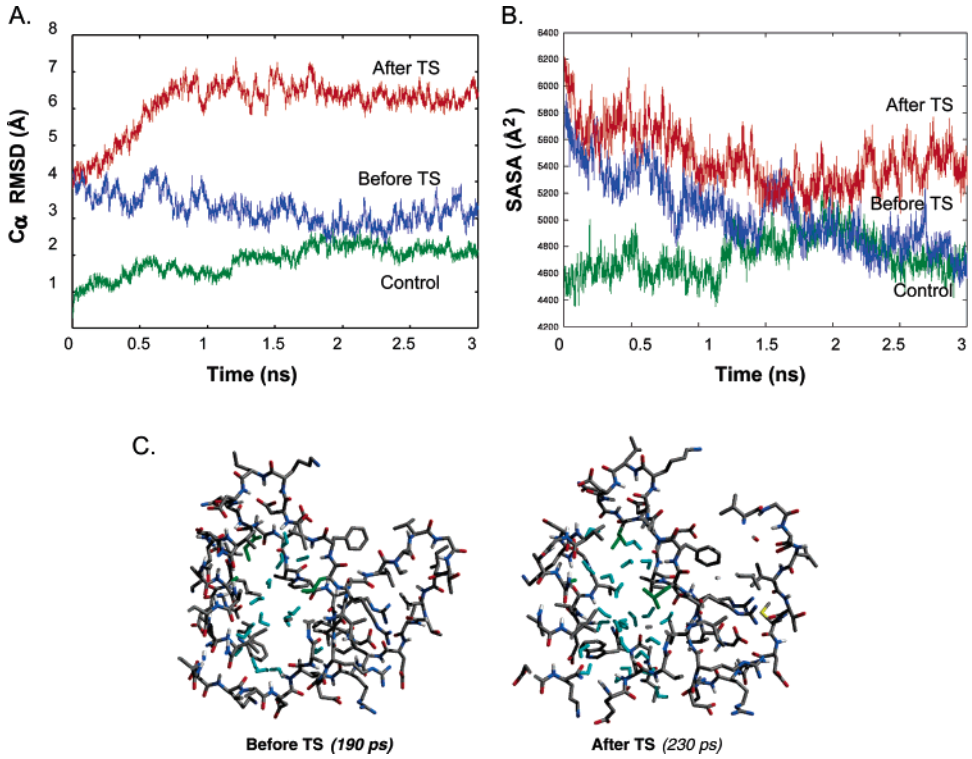

Figure 14. Thermal quenches of chymotrypsin inhibitor 2 of structures around the TS. (A, B) The protein collapses, refolds, and approaches the control with respect to distance from the crystal structure (Cα RMSD) and solvent accessible surface area. Water plays a role in the process, and hydration of the core is shown in panel C. The structure before the TS has few waters in the hydrophobic core, and they are primarily self-associated, which facilitates their expulsion upon refolding. In contrast, more extensive hydration of the core occurs after the TS, and the waters interact avidly with the protein interior.

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. Structures that move toward the native state, some of which have Cα RMSDs of 10 Å, have low contact order: that is, the protein first makes local interactions and then brings more distant portions of the protein together. In the other case of unproductive collapse, the protein makes high contact order, very nonlocal interactions first, and these effectively trap the molecule and prevent fast productive folding. These observations were generalized to explain folding rates for a large variety of proteins.[162]

Similar results were obtained later in temperature-quenched simulations of structures before and after the TS of CI2.[163] Nine structures within ~35 ps and ~3 Å C_α RMSD of the transition state ensemble were extracted and simulated under quasi-native conditions (elevated temperature of 333 K but below the T_m of the protein). All of the structures undergo hydrophobically driven collapse in response to the drop in temperature. Structures less denatured than the transition state became structurally more nativelike, while structures that are more denatured than the TS tend to show additional loss of native structure (Figure 14). The structures in the immediate region of the transition state fluctuate between becoming more and less nativelike. All of the starting structures have the same nativelike topology and are quite similar (within 3.5 Å C_α RMSD). That the structures all shared nativelike topology, yet diverge into either more or less nativelike structures depending on which side of the transition state they occupied on the unfolding trajectory, indicates that topology alone does not dictate protein folding. Instead, our results suggest that a detailed interplay of packing interactions and interactions with water determine whether a partially denatured protein will become more nativelike under refolding conditions. Similar results were obtained regarding the relative importance of topology and detailed packing interactions in an interesting study of two different SH3 domains with the same fold and different sequences and two circular permutants.[39]

All the CI2 structures contained internalized waters that are expelled during collapse due to the drop in temperature and concomitant increase in solvent density. Structures before and after the TS are very similar and 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.

While there are some differences between the starting structures before and after the transition state, they appear to be relatively minor. Consequently, the solvation of these structures was also investigated (Figure 14B). For example, the 190 ps structure before the TS contains 474 hydrating waters. In contrast, the 230 ps starting structure after the TS contains 491 hydrating waters. During the simulations, the proteins collapse in response to the lower temperature, and the number of hydration waters drops by roughly the same amount in the 190 and 230 ps structures. This loss of hydrating waters is due to the expulsion of many internalized waters: in the case of the 190 ps starting structure they were fully extruded, while the 230 ps structure was unable to fully expel the waters, at least in these short simulations. Therefore, the later structure must not only improve its packing interactions to become more nativelike, it must also extrude more water. But it is not expulsion of waters per se that is problematic. Instead, what distinguishes the pre- and post-transition-state structures and determines whether collapse is quickly productive or not from a folding perspective

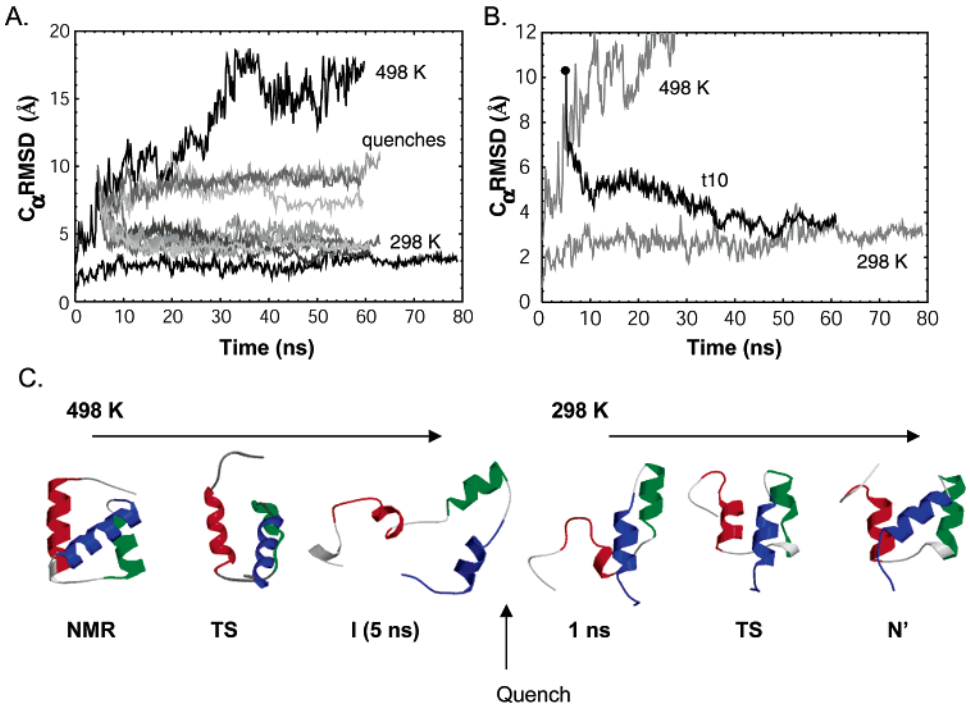

Figure 15. Unfolding and refolding of the engrailed homeodomain. (A) Temperature quenched simulations of the protein from 498 to 298 K show that the protein is approaching the native state in some simulations. (B) Blow-up of the *y*-axis in panel A for one particular target simulations, t10. (C) The thermal denaturation pathway and structures after the thermal quench of t10 show the refolding and docking of the helices, as well as the similarity between the TS ensembles for unfolding and refolding. The coloring is as described in Figure 7.

involves the extent of intermolecular, water−protein hydrogen bonds. While the 190 ps starting structure contains internalized waters in the hydrophobic core between the helix and sheet, they do not interact substantially with the protein's main-chain hydrogen bonding groups (Figure 14). That is, the waters self-associate. In contrast, the 230 ps starting structure contains more waters in the hydrophobic core, and over half of these form hydrogen bonds with the protein main-chain. Many of these waters are involved in multiple hydrogen bonds with the protein. This difference allows the 190 ps structure to expel the water molecules without the need to break a large number of intermolecular hydrogen bonds, as is the case with the 230 ps starting structure.

In more recent studies with the engrailed homeodomain, instead of performing single simulations of multiple targets, multiple simulations of a single target were performed.[164] A post TS structure (5 ns, approximately the intermediate) was extracted from a 498 K thermal unfolding run of En-HD and used to seed 12 independent temperature quench/refolding simulations at 298 K (Figure 15). The 5 ns starting structure is 10.5 Å C_α RMSD from the crystal structure. This intermediate (see the first I structure in Figure 15) is nonnative with very few tertiary contacts, each helix lacks several turns, and the N-terminus contains a nonnative helical segment. The 12 quenches were prepared identically except for the random number seed, used for the assignment of the initial velocities to the atoms. From elementary statistical mechanics, the probability of seeing refolding events early in a quench simulation is enhanced by performing multiple simulations.[53,165]

Figure 15A shows the C_α RMSD from the crystal structure as a function of simulation time for the 12 quench simulations (shades of gray) and the 298 K native state and the 498 K thermal unfolding simulation. The C_α RMSD to the crystal structure ranges from 2 to 4 Å in the control simulation at 298 K, with a final value of 3.6 Å. While the C_α RMSDs may seem large, this protein is only marginally stable at room temperature ($\Delta G = 2.5$ kcal/mol[116]). At 498 K, the RMSD rapidly diverges from the range of values experienced by the native ensemble to a value of 18.6 Å at 60 ns. The refolding simulation starts from the 5 ns, 10.5 Å, unfolding intermediate. The final structure of target 10 (t10, Figure 15B) of the folding simulation after 55 ns at 298 K has an RMSD of 3.6 Å, and the lowest value during the simulation was 2.6 Å. The total solvent accessible surface area for the final nanosecond of the control and t10 refolding simulations overlap, while the accessibility of the 498 K simulation does not. The total number of side-chain contacts for the "refolding" and control simulations are also very similar. Overall, the protein in the quenched, refolding simulation becomes very nativelike.

Refolding of t10 occurs very much as the reverse of denaturation: after quenching at 298 K, transient nonnative helical segments are lost, and most of the native helical structure returns (Figure 15C). Subsequently, the helix scaffold forms, and the swing arm of helix III begins to move toward the core. Helices I and III are still missing a turn of helix at their N- and C-terminal ends, respectively (Figure 15). Fraying (and recovery) in these positions is also observed in the control simulation. These preliminary results indicate that refolding at 298 K is the reverse of unfolding at 498 K, and they suggest that the potential function and procedures can lead to correctly folded structures in solution provided there is adequate sampling either by performing extremely long simulations or hedging your bets and performing multiple shorter simulations with the hope that one will make it.

We have another way to evaluate whether the process of folding is the reverse of unfolding. As mentioned above, simulations of CI2 at its T_m lead to unfolding and refolding. Under these conditions, microscopic reversibility can be tested directly because the conditions are the same for the

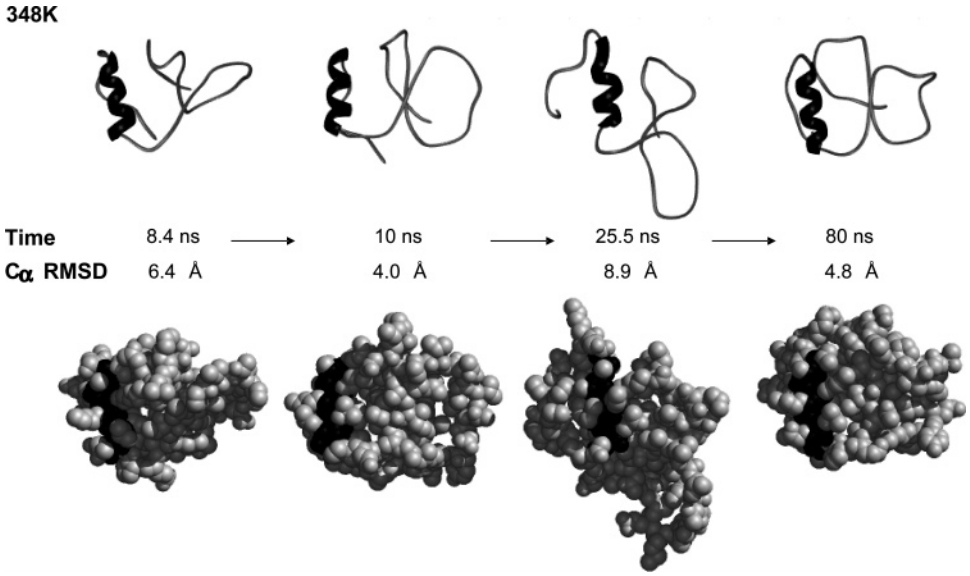

Figure 16. The conformational behavior of chymotrypsin inhibitor 2 at its T_m. Main-chain and space-filling renditions of the structure are provided to illustrate the changes in shape and secondary structure upon unfolding and refolding.

folding and unfolding processes. In particular, we focus on a single simulation of CI2 at 348 K in which the protein initially unfolds to a highly distorted structure, before refolding to a stable nativelike conformation. While the protein does not reach a fully unfolded conformation at this temperature, it has a maximal Cα RMSD of ~9 Å from the crystal structure (Figure 16).[166] This simulation allows direct comparison of the early unfolding process to the refolding process. As the unfolding and refolding processes take place in a single simulation, true differences between the unfolding and refolding pathways can be separated from temperature effects.

The unfolding from N and refolding to N′ (the native state at elevated temperature) proceeds through a series of minimally stable conformations before settling into the final, stable conformation (Figure 16). The unfolding is similar to early stages of the unfolding pathway described previously.[38] There is initially some movement in the termini and loops, as well as a loss of well-defined β-sheet structure. The core becomes partially solvated as the N- and C-termini separate. The N-terminal end of the α-helix pulls away from the β-sheet. Unlike previous unfolding simulations where the separation of the helix and sheet is followed by unfolding of the helix, the C-terminal turn of the α-helix is preserved in the 348 K simulation, as are its hydrophobic contacts with residues from strands 1 and 2 from the β-sheet (Figure 16). By 25.6 ns, this core of structure is essentially all that holds the protein together and the Cα-RMSD to the crystal structure is 8.9 Å. The N- and C-termini are separated from one another, and the core of the protein and the active site loop is highly deformed.

The unfolding pathway up to this point is most similar to an unfolding trajectory at 398 K[38] (Figure 5). The conformations in the 348 K simulation from 8 to 36 ns represent the unfolding pathway at 348 K (Figure 16). The Cα-RMSD between structures in the two simulations continues to increase as the distance to the native starting structure increases, but the partially unfolded structures are always more similar to one another than to N or N′ in the 348 K simulation or fully unfolded conformations in the 398 K simulation.

Rather than continuing to completely unfold, the protein begins to refold from the nucleus of structure formed between the C-terminus of the α-helix and the N-termini of strands 1 and 2 (Figure 16). In the crystal structure, Leu 21 α-helix is on the surface of the protein, and I20 is packed against Ile 29, Val 31, Val 47, and Leu 49 on strands 1 and 2 in the core of the protein. As the protein unfolds and Trp 5 is exposed to solvent, the helix twists and unfolds, bringing Leu 21 into the core. Leu 21 maintains contact with Ile 29 and Leu 49, forming a nucleus for refolding. Refolding begins when contacts between the strands 1 and 2 recover their native register, bringing Leu 21 into contact with Ile 29 and Val 47. This pulls the active site loop back into a more nativelike conformation. The α-helix reforms and twists back near its native conformation, bringing Ile 20 back into contact with Ile 29, Val 31, Val 47, and Leu 49 on strands 1 and 2. Some contacts between the N- and C-termini are reformed, including burial of Trp 5, but they do not regain their precise tight native packing with the helix and strands 1 and 2. The active site loop remains fairly distorted and does not adopt the precise conformation of the crystal structure.

It may simply be that the stable structure at 298 K is not stable at 348 K and that the final structure in simulation represents the high-temperature folded state. A distinct high-temperature folded state has been proposed as an explanation for the unfolding behavior of CspA in laser T-jump experiments.[167] In addition, in crystallographic studies of the structural and dynamic behavior of ribonuclease A and metmyoglobin, Petsko and co-workers found that increasing temperature leads to a linear increase in the protein volume.[168,169] The change in volume naturally affects the packing interactions, leading to native states with a shift to longer intermolecular interactions with increasing temperature, consistent with our contention that our refolded conformer represents the native state at 348 K. More recently single-molecule folding studies by Rhoades et al.[170] also observed a shift from the tightly packed native structure to a more loosely packed native state, N′, for folding events at the GndHCl equivalent of T_m. Overall, this work shows that the order in which structure is reformed mirrors the order in

which it is lost, satisfying the principle of microscopic reversibility.

3. Conclusions

Protein unfolding simulations are coming of age. They can be expected to fairly reliably depict protein folding/unfolding transition states, intermediate states, and denatured states, provided explicit solvent and good simulation techniques are employed. Simulations provide a molecular framework for the interpretation of experimental protein folding studies, and they are readily amenable to validation by comparison with experiment. An understanding of these various conformational states has practical implications: it can aid in the design of faster folding proteins[110] as well as more stable proteins.[171−173] In addition, protein unfolding simulations are becoming increasingly important for mapping conformational changes implicated in amyloidosis.[117,174−183]

4. Acknowledgments

The author is grateful for financial support from the National Institutes of Health (GM50789). UCSF MidasPlus was used to make the protein displays.[184]

5. References

(1) Tramontano, A. *Nat. Struct. Biol.* **2003**, *10*, 87.
(2) Chivian, D.; Robertson, T.; Bonneau, R.; Baker, D. *Methods Biochem. Anal.* **2003**, *44*, 547.
(3) Skolnick, J.; Kolinski, A. *Comput. Methods Protein Folding* **2002**, *120*, 131.
(4) Pillardy, J.; Czaplewski, C.; Liwo, A.; Lee, J.; Ripoll, D. R.; Kazmierkiewicz, R.; Oldziej, S.; Wedemeyer, W. J.; Gibson, K. D.; Arnautova, Y. A.; Saunders: J.; Ye, Y. J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2329.
(5) Pellegrini-Calace, M.; Carotti, A.; Jones, D. T. *Proteins* **2003**, *50*, 537.
(6) Jones, D. T. *Proteins* **2001** (Suppl. 5) 127.
(7) Fersht, A. R.; Kellis, J. T.; Matouschek, A. T. E. L.; Serrano, L. *Nature* **1990**, *343*, 602.
(8) Kelly, J. W. *Nat. Struct. Biol.* **2002**, *9*, 323.
(9) Dobson, C. M. *Nature* **2002**, *418*, 729.
(10) Fink, A. L. *Folding Des.* **1998**, *3*, R9.
(11) Prakash, S.; Matouschek, A. *Trends Biochem. Sci.* **2004**, *29*, 593.
(12) Onuchic, J. N.; Nymeyer, H.; Garcia, A. E.; Chahine, J.; Socci, N. D. *Adv. Protein Chem.* **2000**, *53*, 87.
(13) Thirumalai, D.; Klimov, D. K. *Curr. Opin. Struct. Biol.* **1999**, *9*, 197.
(14) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619.
(15) Dill, K. A.; Bromberg, S.; Yue, K.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D.; Chan, H. S. *Protein Sci.* **1995**, *4*, 561.
(16) Godzik, A.; Kolinski, A.; Skolnick, J. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 397.
(17) Alonso, D. O. V.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 133.
(18) Daggett, V.; Levitt, M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5142.
(19) Vendruscolo, M.; Paci, E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 82.
(20) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646.
(21) Brooks, C. L., 3rd. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1099.
(22) Daggett, V. *Acc. Chem. Res.* **2002**, *35*, 422.
(23) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190.
(24) Shea, J. E.; Brooks, C. L., 3rd. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499.
(25) Israelewitz, B.; Gao, M.; Schulten, K. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224.
(26) Doniach, S.; Eastman, P. *Curr. Opin. Struct. Biol.* **1999**, *9*, 157.
(27) Radford, S. E.; Dobson, C. M. *Cell* **1999**, *97*, 291.
(28) Brooks, C. L., 3rd. *Curr. Opin. Struct. Biol.* **1998**, *8*, 222.
(29) Fersht, A. R.; Matouschek, A.; Serrano, L. *J. Mol. Biol.* **1992**, *224*, 771.
(30) Daggett, V. F., A. R. In *Mechanisms of Protein Folding*, 2nd ed.; Pain, R. H., Ed.; Oxford University Press: Oxford UK, 2000; p 175.
(31) Daggett, V.; Fersht, A. R. *Trends Biochem. Sci.* **2003**, *28*, 18.
(32) Daggett, V.; Fersht, A. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497.
(33) Dinner, A. R.; Sali, A.; Smith, L. J.; Dobson, C. M.; Karplus, M. *Trends Biochem. Sci.* **2000**, *25*, 331.
(34) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573.
(35) Dobson, C. M.; Karplus, M. *Curr. Opin. Struct. Biol.* **1999**, *9*, 92.
(36) Li, A.; Daggett, V. *J. Mol. Biol.* **1996**, *257*, 412.
(37) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928.
(38) Day, R.; Bennion, B. J.; Ham, S.; Daggett, V. *J. Mol. Biol.* **2002**, *322*, 189.
(39) Gsponer, J.; Caflisch, A. *J. Mol. Biol.* **2001**, *309*, 285.
(40) Tsai, J.; Levitt, M.; Baker, D. *J. Mol. Biol.* **1999**, *291*, 215.
(41) White, G. W.; Gianni, S.; Grossmann, J. G.; Jemth, P.; Fersht, A. R.; Daggett, V. *J. Mol. Biol.* **2005**, *350*, 757.
(42) Day, R.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13445.
(43) Huang, G. S.; Oas, T. G. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6878.
(44) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
(45) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518.
(46) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425.
(47) Struthers, M.; Ottesen, J. J.; Imperiali, B. *Fold Des.* **1998**, *3*, 95.
(48) Wang, M.; Tang, Y.; Sato, S.; Vugmeyster, L.; McKnight, C. J.; Raleigh, D. P. *J. Am. Chem. Soc.* **2003**, *125*, 6032.
(49) Nguyen, H.; Jager, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3948.
(50) Callender, R.; Dyer, R. B. *Curr. Opin. Struct. Biol.* **2002**, *12*, 628.
(51) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168.
(52) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102.
(53) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14122.
(54) Paci, E.; Clarke, J.; Steward, A.; Vendruscolo, M.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 394.
(55) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258.
(56) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952.
(57) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548.
(58) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711.
(59) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587.
(60) De Young, L. R.; Dill, K. A.; Fink, A. L. *Biochemistry* **1993**, *32*, 3877.
(61) Fan, P.; Bracken, C.; Baum, J. *Biochemistry* **1993**, *32*, 1573.
(62) Yang, Y.; Mayo, K. H. *Biochemistry* **1993**, *32*, 8661.
(63) Buck, M.; Radford, S. E.; Dobson, C. M. *J. Mol. Biol.* **1994**, *237*, 247.
(64) Uversky, V. N.; Ptitsyn, O. B. *Biochemistry* **1994**, *33*, 2782.
(65) Waterhous, D. V.; Johnson, W. C., Jr. *Biochemistry* **1994**, *33*, 2121.
(66) Alonso, D. O. V.; Daggett, V. *J. Mol. Biol.* **1995**, *247*, 501.
(67) Caflisch, A.; Karplus, M. *Structure* **1999**, *7*, 477.
(68) Tirado-Rives, J.; Orozco, M.; Jorgensen, W. L. *Biochemistry* **1997**, *36*, 7313.
(69) Toba, S.; Hartsough, D. S.; Merz, K. M. *J. Am. Chem. Soc.* **1996**, *118*, 6490.
(70) Colombo, G.; Toba, S.; Merz, K. M. *J. Am. Chem. Soc.* **1999**, *121*, 3486.
(71) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5142.
(72) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6433.
(73) Soares, C. M.; Teixeira, V. H.; Baptista, A. M. *Biophys. J.* **2003**, *84*, 1628.
(74) Roccatano, D.; Wong, T. S.; Schwaneberg, U.; Zacharias, M. *Biopolymers* **2005**, *78*, 259.
(75) Schellman, J. A. *Biopolymers* **1978**, *17*, 1305.
(76) Timasheff, S. N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9721.
(77) Timasheff, S. N. *Biochemistry* **2002**, *41*, 13473.
(78) Tobi, D.; Elber, R.; Thirumalai, D. *Biopolymers* **2003**, *68*, 359.
(79) Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1992**, *226*, 491.
(80) Anderson, C. F.; Felitsky, D. J.; Hong, J.; Record, M. T. *Biophys. Chem.* **2002**, *101−102*, 497.
(81) Schellman, J. A. *Biophys. Chem.* **2002**, *96*, 91.
(82) Arakawa, T.; Timasheff, S. N. *Biochemistry* **1984**, *23*, 5912.
(83) Tanford, C. *Adv. Protein Chem.* **1970**, *24*, 1.
(84) Creighton, T. E. *Curr. Opin. Struct. Biol.* **1991**, *1*, 5.
(85) Alonso, D. O. V.; Dill, K. A. *Biochemistry* **1991**, *30*, 5974.
(86) Robinson, D. R.; Jencks, W. P. *J. Am. Chem. Soc.* **1965**, *87*, 2462.
(87) VonHippel, H.; Schleich, T. *Acc. Chem. Res.* **1969**, *2*, 257.

(88) Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N. *Science* **1982**, *217*, 1214.
(89) Withers, P. C.; Morrison, G.; Guppy, M. *Physiol. Zool.* **1994**, *67*, 693.
(90) Barton, K. N.; Buhr, M. M.; Ballantyne, J. S. *Am. J. Physiol.* **1999**, *276*, R397.
(91) Yancey, P. H.; Somero, G. N. *J. Exp. Zool.* **1980**, *212*, 205.
(92) Yancey, P. H.; Burg, M. B. *Am. J. Physiol.* **1990**, *258*, R198.
(93) Yancey, P. H.; Siebenaller, J. F. *J. Exp. Biol.* **1999**, *202*, 3597.
(94) Lin, T. Y.; Timasheff, S. N. *Protein Sci.* **1996**, *5*, 372.
(95) Wang, A.; Bolen, D. W. *Biophys. J.* **1996**, *71*, 2117.
(96) Baskakov, I.; Bolen, D. W. *Biophys. J.* **1998**, *74*, 2658.
(97) Baskakov, I.; Wang, A.; Bolen, D. W. *Biophys. J.* **1998**, *74*, 2666.
(98) Lin, T. Y.; Timasheff, S. N. *Biochemistry* **1994**, *33*, 12695.
(99) Zou, Q.; Bennion, B. J.; Daggett, V.; Murphy, K. P. *J. Am. Chem. Soc.* **2002**, *124*, 1192.
(100) Minton, A. P. *J. Biol. Chem.* **2001**, *276*, 10577.
(101) Dedmon, M. M.; Patel, C. N.; Young, G. B.; Pielak, G. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12681.
(102) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* **2002**, *41*, 6573.
(103) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10428.
(104) Otzen, D. E.; Itzhaki, L. S.; Elmasry, N. F.; Jackson, S. E.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10422.
(105) Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1995**, *254*, 260.
(106) Gay, G. D.; Ruizsanz, J.; Neira, J. L.; Itzhaki, L. S.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3683.
(107) Li, A.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10430.
(108) Daggett, V.; Li, A.; Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1996**, *257*, 430.
(109) Settanni, G.; Rao, F.; Caflisch, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 628.
(110) Ladurner, A. G.; Itzhaki, L. S.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8473.
(111) Kazmirski, S. L.; Wong, K. B.; Freund, S. M. V.; Tan, Y. J.; Fersht, A. R.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4349.
(112) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10869.
(113) Haar, L. G., J. S.; Kell, G. S. *NBS/NRC Steam Tables*; Hemisphere Publication Corporation: Washington D. C., 1984.
(114) Levitt, M.; Hirshberg, M.; Sharon, R.; Laidig, K. E.; Daggett, V. *J. Phys. Chem. B* **1997**, *101*, 5051.
(115) Ferguson, N.; Pires, J. R.; Toepert, F.; Johnson, C. M.; Pan, Y. P.; Volkmer-Engert, R.; Schneider-Mergener, J.; Daggett, V.; Oschkinat, H.; Fersht, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13008.
(116) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M.; Alonso, D. O.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863.
(117) DeMarco, M. L.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2293.
(118) Gianni, S.; Guydosh, N. R.; Khan, F.; Caldas, T. D.; Mayor, U.; White, G. W.; DeMarco, M. L.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13286.
(119) Jemth, P.; Gianni, S.; Day, R.; Li, B.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sc. U.S.A.* **2004**, *101*, 6450.
(120) Jemth, P.; Day, R.; Gianni, S.; Khan, F.; Allen, M.; Daggett, V.; Fersht, A. R. *J. Mol. Biol.* **2005**, *350*, 363.
(121) Zhu, Y.; Alonso, D. O.; Maki, K.; Huang, C. Y.; Lahr, S. J.; Daggett, V.; Roder, H.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15486.
(122) DeMarco, M. L.; Alonso, D. O.; Daggett, V. *J. Mol. Biol.* **2004**, *341*, 1109.
(123) Roccatano, D.; Daidone, I.; Ceruso, M. A.; Bossa, C.; Nola, A. D. *Biophys. J.* **2003**, *84*, 1876.
(124) Zaks, A.; Klibanov, A. M. *Science* **1984**, *224*, 1249.
(125) Affleck, R.; Haynes, C. A.; Clark, D. S. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5167.
(126) Burke, P. A.; Griffin, R. G.; Klibanov, A. M. *Biotechnol. Bioeng.* **1993**, *42*, 87.
(127) Fitzpatrick, P. A.; Steinmetz, A. C.; Ringe, D.; Klibanov, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8653.
(128) Yennawar, N. H.; Yennawar, H. P.; Farber, G. K. *Biochemistry* **1994**, *33*, 7326.
(129) Alonso, D. O. V.; Daggett, V. *Protein Sci.* **1998**, *7*, 860.
(130) Laidig, K. E.; Daggett, V. *J. Phys. Chem.* **1996**, *100*, 5616.
(131) Muller, N. *Acc. Chem. Res.* **1990**, *23*, 23.
(132) Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1.
(133) Tanford, C.; De, P. K. *J. Biol. Chem.* **1961**, *236*, 1711.
(134) Wang, A.; Bolen, D. W. *Biochemistry* **1997**, *36*, 9101.
(135) Timasheff, S. N. *Biochemistry* **1992**, *31*, 9857.
(136) Timasheff, S. N. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 67.
(137) Timasheff, S. N. *Adv. Protein Chem.* **1998**, *51*, 355.
(138) Whittington, S. J.; Chellgren, B. W.; Hermann, V. M.; Creamer, T. P. *Biochemistry* **2005**, *44*, 6269.
(139) Kim, Y. S.; Jones, L. S.; Dong, A.; Kendrick, B. S.; Chang, B. S.; Manning, M. C.; Randolph, T. W.; Carpenter, J. F. *Protein Sci.* **2003**, *12*, 1252.
(140) Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernandez, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3694.
(141) Lu, H.; Schulten, K. *Biophys. J.* **2000**, *79*, 51.
(142) Paci, E.; Karplus, M. *J. Mol. Biol.* **1999**, *288*, 441.
(143) Fowler, S. B.; Clarke, J. *Structure* **2001**, *9*, 355.
(144) Brockwell, D. J.; Beddard, G. S.; Clarkson, J.; Zinober, R. C.; Blake, A. W.; Trinick, J.; Olmsted, P. D.; Smith, D. A.; Radford, S. E. *Biophys. J.* **2002**, *83*, 458.
(145) Fowler, S. B.; Best, R. B.; Toca Herrera, J. L.; Rutherford, T. J.; Steward, A.; Paci, E.; Karplus, M.; Clarke, J. *J. Mol. Biol.* **2002**, *322*, 841.
(146) Best, R. B.; Li, B.; Steward, A.; Daggett, V.; Clarke, J. *Biophys. J.* **2001**, *81*, 2344.
(147) Serrano, L.; Matouschek, A.; Fersht, A. R. *J. Mol. Biol.* **1992**, *224*, 805.
(148) Matouschek, A.; Kellis, J. T., Jr.; Serrano, L.; Bycroft, M.; Fersht, A. R. *Nature* **1990**, *346*, 440.
(149) Matthews, J. M.; Fersht, A. R. *Biochemistry* **1995**, *34*, 6805.
(150) Daggett, V.; Li, A. J.; Fersht, A. R. *J. Am. Chem. Soc.* **1998**, *120*, 12740.
(151) Li, A.; Daggett, V. *J. Mol. Biol.* **1998**, *275*, 677.
(152) Bond, C. J.; Wong, K. B.; Clarke, J.; Fersht, A. R.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 13409.
(153) Wong, K. B.; Clarke, J.; Bond, C. J.; Neira, J. L.; Freund, S. M.; Fersht, A. R.; Daggett, V. *J. Mol. Biol.* **2000**, *296*, 1257.
(154) Khan, F.; Chuang, J. I.; Gianni, S.; Fersht, A. R. *J. Mol. Biol.* **2003**, *333*, 169.
(155) Zhuang, X.; Rief, M. *Curr. Opin. Struct. Biol.* **2003**, *13*, 88.
(156) Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5179.
(157) Day, R.; Beck, D. A.; Armen, R. S.; Daggett, V. *Protein Sci.* **2003**, *12*, 2150.
(158) Dinner, A. R.; Karplus, M. *J. Mol. Biol.* **1999**, *292*, 403.
(159) Ding, F.; Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. *Biophys. J.* **2002**, *83*, 3525.
(160) Gsponer, J.; Caflisch, A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6719.
(161) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740.
(162) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985.
(163) De Jong, D.; Riley, R.; Alonso, D. O.; Daggett, V. *J. Mol. Biol.* **2002**, *319*, 229.
(164) Beck, D. A.; Daggett, V. *Methods* **2004**, *34*, 112.
(165) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987.
(166) Day, R., Daggett, V. **2006**, submitted for publication.
(167) Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2527.
(168) Frauenfelder, H.; Hartmann, H.; Karplus, M.; Kuntz, I. D., Jr.; Kuriyan, J.; Parak, F.; Petsko, G. A.; Ringe, D.; Tilton, R. F., Jr.; Connolly, M. L.; et al. *Biochemistry* **1987**, *26*, 254−261.
(169) Tilton, R. F., Jr.; Dewan, J. C.; Petsko, G. A. *Biochemistry* **1992**, *31*, 2469−2481.
(170) Rhoades, E.; Gussakovsky, E.; Haran, G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3197−3202.
(171) Yoda, T.; Saito, M.; Arai, M.; Horii, K.; Tsumoto, K.; Matsushima, M.; Kumagai, I.; Kuwajima, K. *Proteins* **2001**, *42*, 49−65.
(172) Pikkemaat, M. G.; Linssen, A. B.; Berendsen, H. J.; Janssen, D. B. *Protein Eng.* **2002**, *15*, 185−192.
(173) Liu, H. L.; Wang, W. C. *Protein Eng.* **2003**, *16*, 19−25.
(174) Kazmirski, S. L.; Alonso, D. O.; Cohen, F. E.; Prusiner, S. B.; Daggett, V. *Chem. Biol.* **1995**, *2*, 305.
(175) Kazmirski, S. L.; Isaacson, R. L.; An, C.; Buckle, A.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Nat. Struct. Biol.* **2002**, *9*, 112.
(176) Alonso, D. O. V.; DeArmond, S. J.; Cohen, F. E.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2985.
(177) Alonso, D. O. V.; An, C.; Daggett, V. *Philos. Trans. R. Soc. London, Ser. A* **2002**, *360*, 1165.
(178) Armen, R. S.; DeMarco, M. L.; Alonso, D. O. V.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11622.
(179) Armen, R. S.; Alonso, D. O. V.; Daggett, V. *Structure* **2004**, *12*, 1847.
(180) Yang, M.; Lei, M.; Huo, S. *Protein Sci.* **2003**, *12*, 1222.
(181) Yang, M.; Lei, M.; Bruschweiler, R.; Huo, S. *Biophys. J.* **2005**, *89*, 433.

(182) Bennion, B. J.; DeMarco, M. L.; Daggett, V. *Biochemistry* **2004**, *43*, 12955.
(183) Gsponer, J.; Ferrara, P.; Caflisch, A. *J. Mol. Graphics Modell.* **2001**, *20*, 169.
(184) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graphics* **1988**, *6*, 13.

CR0404242