

Computational protein folding: From lattice to all-atom

by Y. Duan
P. A. Kollman

Understanding the mechanism of protein folding is often referred to as the second half of genetics. Computational approaches have been instrumental in the efforts. Simplified models have been applied to understand the physical principles governing the folding processes and will continue to play important roles in the endeavor. Encouraging results have been obtained from all-atom molecular dynamics simulations of protein folding. A recent microsecond-length molecular dynamics simulation on a small protein, villin headpiece subdomain, with an explicit atomic-level representation of both protein and solvent, has marked the beginning of direct and realistic simulations of the folding processes. With growing computer power and increasingly accurate representations together with the advancement of experimental methods, such approaches will help us to achieve a detailed understanding of protein folding mechanisms.

Proteins support life by carrying out important biological functions, which are determined primarily by their structures. Subjected to evolutionary pressure, only those proteins that are helpful to the survival of living beings have been retained. Though their folding time may not be a subject of active refinement of evolution, proteins are required to be able to adapt well-defined structures soon after being synthesized and transported to their designated locations within cells to perform their functions. Such a requirement sets the upper limit for their folding time and is one of the important aspects of proteins that sets them apart from other polymers, including other nonprotein polypeptides.^{1,2} The astronomically large number of possi-

ble conformations suggests that proteins use some sort of “directed” mechanisms to fold. An elucidation of protein folding mechanisms must address how proteins fold into their well-defined three-dimensional structures within a limited time. We review briefly the history of computational protein folding studies, discuss the recent developments in more detail, and present a perspective of the future.

Under the right physiological conditions, proteins can fold into and subsequently maintain well-defined structures, determined by sequences,³ through delicate balances⁴ of enthalpy and entropy,^{5,6} weak interactions, including van der Waals, electrostatic, and hydrogen-bonding forces, and a balance between protein intramolecular interactions and the interactions with solvent that also play major roles in protein folding.⁷ A major motivation for the mechanistic studies has been the need to understand the roles of these interactions in determining protein structures, since such an understanding can help to improve the accuracy of protein structure prediction. Because of the close association between protein structures and their functions, understanding how protein sequences determine their structures has often been referred to as the second half of genetics. With the explosive growth of genomic sequence data, the need for reliable structural prediction methods that can complement the existing experimental ap-

©Copyright 2001 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

proaches such as X-ray crystallography and NMR (nuclear magnetic resonance) spectroscopy is compelling. In this regard, an appealing aspect of the physically based modeling is its generality. These models use the physical interaction energies as the primary criteria to analyze protein structures. The same set of physical principles that drives protein folding also dictates substrate and ligand binding as well as the induced conformational changes that are often associated with protein functions and are important for a detailed understanding of biochemical processes. Understanding protein folding would inevitably aid in the understanding of these processes. The relatively recent discovery of folding-related diseases⁸⁻¹⁸ reinforces such a need. Despite great progress made using a variety of approaches, it is still difficult to establish detailed descriptions of the protein folding processes and such descriptions are the necessary steps toward the comprehensive understanding of the mechanisms of folding.

Lattice models

Computational studies of protein folding have come of age. Among the early successes was an Ising model simulation on the unfolding and hydrogen exchange of proteins¹⁹ in which a two-state transition was observed, which was not surprising given the three-dimensional nature of the model. Ptitsyn and Rashin²⁰ studied the folding of myoglobin without using a computer by representing the protein at the secondary structure level and treating each α helix as a uniform rigid body cylinder. Using the highly simplified representation, they concluded that the folding was a nucleation process, similar to that of crystal growth. A similar representation²¹ has been applied recently in combination with a Brownian dynamics approach in the study of the folding of a four-helix bundle.

A more detailed representation also appeared²²⁻²⁴ in the late 1970s. Using a combination of Langevin dynamics and energy minimization, Levitt and Warshel studied folding of BPTI (bovine pancreatic trypsin inhibitor)²² and Carp Myogen.²⁴ In these studies, the authors represented each amino acid by two particles. They observed highly complex folding processes in which secondary structures were seen both forming and breaking, and challenged the notion that folding was preceded by forming stable secondary structures first. This pioneering work marked the beginning of physically based models in the studies of protein folding, albeit at a somewhat crude level. The level of approximation, both in the representation and the parameter, naturally implied a

certain level of uncertainty and sometimes even significant error, as pointed out by Hagler and Honig.²⁵

Levitt and Warshel also noted that hydrogen bonds seem to slow down the folding process,²² a finding that has yet to be clarified by further studies. Nevertheless, we should recognize the pioneering nature of the work, which helped to topple the then-popular view that stable secondary structure always forms first in the folding process. The fact that most current structure prediction methods use a similar representation to that of Levitt and Warshel is a strong testament to the power of such an approach. Fifteen years later, using a residue-level lattice model, Skolnick and Kolinski²⁶ have successfully simulated the folding of some small proteins. Interestingly, the parameters were obtained by analyzing protein structures deposited in the Protein Data Bank (PDB),²⁷ similar to the approaches of Miyazawa and Jernigan.²⁸

The advantages of lattice models are clear. The highly simplified models allow efficient sampling of conformational space. This was particularly important at the time when the speed of the most powerful computer was many orders of magnitude slower than a current personal computer. When designed properly, the model can give a well-defined global energy minimum that can be calculated analytically. In fact, one can enumerate all energy states and calculate the corresponding free energies in such models. One can also control other features of the energetic surface. When carefully parameterized, lattice models can be applied to structure prediction and can give encouraging results.²⁶ The lattice model also allows Monte Carlo simulations that give ensemble averages. This was a critical advantage as well, because at the time all experiments were conducted macroscopically and could only give ensemble-averaged results. Single molecule studies were much later developments.^{29,30} This type of model has enjoyed widespread application and has contributed a great deal to our understanding of protein folding mechanisms.

There are two types of lattice model simulations, aimed at two distinct objectives. One, pioneered by Go and coworkers,³¹ was designed to understand the basic physics governing the protein folding process. A key feature of this type of lattice model is its simplicity (the size can range from 3^2 to 5^3 lattice points). A good example of such an approach has been shown by Wolynes and coworkers who, through lattice model simulations, postulated that proteins have a funnel-like energy landscape with a minimally frus-

trated character that “guides” proteins toward their native states.^{32,33} The postulate deviated markedly from the old pathway doctrine and elevated our understanding at the conceptual level.³⁴ Another useful example was done by Dill and coworkers,⁷ who emphasized the importance of hydrophobic interactions. Other examples include studies by Li et al.^{1,2} and by Shakhnovich and coworkers.^{35–38} Some of the work has been reviewed previously.³⁹ Recently, this type of approach has been extended to residue-level off-lattice models.^{40–43} Similar to the approaches of Muñoz et al.,^{40,41} Zhou and Karplus^{42,43} assured the foldability of the model by systematically biasing the energetic surface toward the native state of that particular protein under study in a process consistent with the diffusion-collision model.^{44,45} Because this type of model has not been designed for real proteins, tests on these models have been limited to the studies of general features of protein folding. Nevertheless, a good deal can be learned from these studies. For example, Dill and coworkers^{7,46} have argued that a small set of amino acids (hydrophobic and hydrophilic) can be combined to produce foldable protein-like peptides, a prediction that has been confirmed recently by experiments.⁴⁷

Lattice models by Skolnick and coworkers²⁶ and by Miyazawa and Jernigan^{28,48} belong to the second category. These models are geared toward realistic folding of real proteins and are therefore parameterized using real proteins as templates by statistical sampling of the available structures^{28,48,49} and are often referred to as statistical potentials (or knowledge-based potentials). Works by Crippen,⁵⁰ by Eisenberg and coworkers,^{51,52} and by Sippl and coworkers⁵³ are also good examples in this category that have been reviewed before.^{54–56} Along the same line was the approach by Scheraga and coworkers, who developed a residue-based off-lattice model.^{57–60} Because the residue-level representations are applied to real proteins that have large numbers of energy minima, in contrast to the simplified lattice models described above, their energetic surface can no longer be described exactly, even though exhaustive sampling can be conducted for short sequences (shorter than 100 amino acids).⁶¹ More importantly, the pair-wise discrete neighboring “energy” for the interactions between the nearest neighbors allows only a small number of possible conformations. The lattice coordinates also impose restrictions to the representation, though a high-coordination lattice model has been developed as well.⁶² Given their highly simplified approaches, the successes in predicting protein structure are indeed very encouraging.

Off-lattice models

A constant driving force in the computational study of protein folding has been the need to develop methods that can reliably differentiate native states from the non-native ones. The most widely used approaches in protein structure prediction have been based on residue-level models (either lattice or off-lattice models) with typically statistical “potentials” obtained from the structural database (PDB). A growing trend in the community has been the development of atomic-level statistical potentials^{63–68} in attempts to improve the accuracy. The application of all-atom representation with physical potentials in structural prediction, on the other hand, has been limited. A typical application would be at the final stage—a minor refinement of the structures using limited energy minimization designed to eliminate the bad contacts. It has been pointed out that the gas phase energy calculated by all-atom molecular mechanics is a poor descriptor of the “quality” of the structures.^{54,69–72} This is not surprising given the critical role that solvent plays in determining protein structures and in fact is reassuring, because gas-phase energy alone should not be able to discriminate good structures from the bad ones. As expected, the accuracy was dramatically improved with the inclusion of the solvent effect.^{69,70,73–77}

An improved level of accuracy has been obtained through a combination of an all-atom representation of protein and a continuum model of solvent. Sung studied the folding of Alanine-based peptides^{78,79} and noted interesting features from the simulations, including the role of electrostatic interactions between the successive amides, which favored extended conformations and caused energy barriers to helix folding, intermediate states, and formation of both 3^{10} and α helices. Karplus and coworkers^{74,80–82} adopted a similar approach and applied it to the studies of the folding free energy of chymotrypsin inhibitor 2⁸³ and that of G-peptide,⁸⁴ using unfolding simulations, and tested this approach on a set of proteins⁷³ and on two peptides.^{81,82} Continuing along this line was the work by Wu and Sung,⁸⁵ who proposed the use of the mean solvation force to represent solvent, and tested this method on alanine-dipeptide. This type of model tries to strike a balance between the accuracy of the representation and the computational cost. Application of the continuum solvent model can significantly reduce the number of particles included in the calculation and, hence, the computational cost, even after considering the overhead due to the added complexity of the

continuum solvent model. It is interesting to note that such development came about two decades after the first residue-level simulation of Warshel and Levitt.²² The new development reflects considerable improvement in the level of sophistication, in addi-

**A powerful extension
of unfolding simulations
is the attempt to reconstruct
the free-energy landscape.**

tion to the improvement due to the differences between all-atom and residue-level models of protein. Compared to the *ad hoc* approach of Warshel and Levitt in parameter generation,²³ the present parameters were based on quantum mechanical calculations and refined against experiments. The solvent model has also been improved substantially from initially simple solvation-free energy approaches²³ to today's solvent model based on macroscopic electrostatics.⁷⁴ As pointed out by many researchers in the field, a common deficiency of the continuum solvent models is that the simulated events can occur at time scales much smaller than those found in experiments,⁸⁶ which, in many cases, can be corrected by taking into account the viscosity of the solvent. A more serious problem can arise when solvent plays a structural role. This becomes an important issue in protein folding, since proteins can have substantial solvent molecules in the interior in some important states, such as molten globule states. Furthermore, studies have suggested that solvent plays a role as the lubricant prior to reaching the native state^{87,88} and ejection of a solvent molecule from the interior may contribute a nontrivial portion to the free-energy barriers.⁸⁹

All-atom models

At an even higher level of sophistication is the all-atom representation of both solvent and protein. A hallmark of models of this type is that their parameters are obtained through high-level quantum mechanical calculations on short peptide fragments. Such an approach has several advantages. It assures the generality and allows further refinement upon the availability of more accurate quantum mechanical methods and upon the need for such an improvement. Such models also allow further extension. For instance, active efforts have been undertaken to pa-

rameterize polarization energy that can be integrated seamlessly into present simulation methods.^{90–92} Some of the earlier developments have been reviewed.⁹³ Because the detailed models require both a large number of particles, typically more than 10000, and a small time step of one to two femtoseconds (10^{-15} seconds), direct simulation of the folding processes, which take place on a microsecond or larger time scale, has been difficult. Therefore, such models have been applied to study the unfolding processes of small proteins that can be accelerated substantially by raising the simulation temperature,^{88,94–100} by changing solvent condition,^{88,101,102} by applying external forces,^{103–105} and by applying pressure.¹⁰² The detailed representation has allowed direct comparisons with experiments and encouraging results have been obtained.^{98,106} Limited refolding simulations were also attempted starting from partially unfolded structures generated from the unfolding simulations and considerable fluctuations were observed.^{97,100} These short-time refolding simulations have also identified the transition states in the vicinity of the native state.^{100,106} Care must be taken, though, because the short-time refolding simulations can only sample the conformational space in the vicinity of the unfolding trajectories. Equilibration of water in this type of short-time refolding simulation is needed to avoid simulating a trivial collapse process of water equilibration when the system is brought to room temperature and to restore faithfully the room-temperature solvent condition that has been distorted significantly due to the entropy-enthalpy imbalance¹⁰⁷ at high temperature. Such an imbalance inherent in the typical unfolding simulations may also be reduced by conducting the unfolding simulations at moderate unfolding temperatures¹⁰⁸ such that both temperature and pressure can be maintained at experimentally relevant conditions.

A powerful extension of unfolding simulations is the attempt to reconstruct the free-energy landscape. Using the weighted histogram method,¹⁰⁹ Brooks and coworkers calculated the free-energy landscapes of folding a three-helix bundle,¹¹⁰ the segment B1 of streptococcal protein G,⁸⁷ and the Betanova⁸⁹ from restrained unfolding simulations. They demonstrated funnel-shaped free-energy landscapes, the existence of multiple folding pathways, and showed that the shapes of the funnels are also dependent on the type of proteins (i.e., α helical or α/β). They also observed that ejection of water from the interior of the intermediate state contributes to the free energy barrier of folding,⁸⁹ suggesting the role that water may play

in the folding process in addition to its role as solvent. Such an observation is only possible with the explicit inclusion of solvent in the simulation. It is noteworthy that the application of the restraint functions is an integral part of the methodology, because it ensures a sufficient number of transitions between neighboring states and hence ensures the reversibility that is absent in the unrestrained unfolding simulations. Nevertheless, the weighted histogram method has also been applied in the analysis of unrestrained unfolding trajectories.⁸³ Free energy profiles (or probability profiles) have also been generated directly from the unfolding trajectories,¹⁰⁰ but it is unclear at what temperature the profiles were generated.

A central question concerning the elucidation of the protein folding mechanism is: how do proteins reach their native state? Therefore, direct simulation of protein folding using an all-atom model has been termed the “holy grail.”¹¹¹ Encouraging developments have been made in the simulations of the folding processes of small peptide fragments with explicit representations of both solvent and peptides.^{112–116} Tobias and Brooks studied the formation of a β -turn in aqueous solution.¹¹² Case and coworkers studied the transitions between two conformations of a β -turn motif and most of the simulated distances agree with NMR data.¹¹³ Daura et al.¹¹⁴ studied the folding and unfolding processes of a short β -peptide in methanol at temperatures below, around, and above T_m of the peptide. They observed reversible formation of secondary structure in a simple two-state manner within 50 nanoseconds. The estimated folding free energies, based on the population of the states, are in qualitative agreement with experimental observations. Chipot et al. studied the formation of undecamer peptides at the water-hexane interface.^{115,116} Wang and coworkers recently developed a method called Self-Guided Molecular Dynamics (SGMD)^{117,118} and applied it to study the folding of a 16-residue peptide.¹¹⁹ These studies provided detailed atomic-level descriptions of the formation of isolated secondary structure motifs in their respective solvent environments.

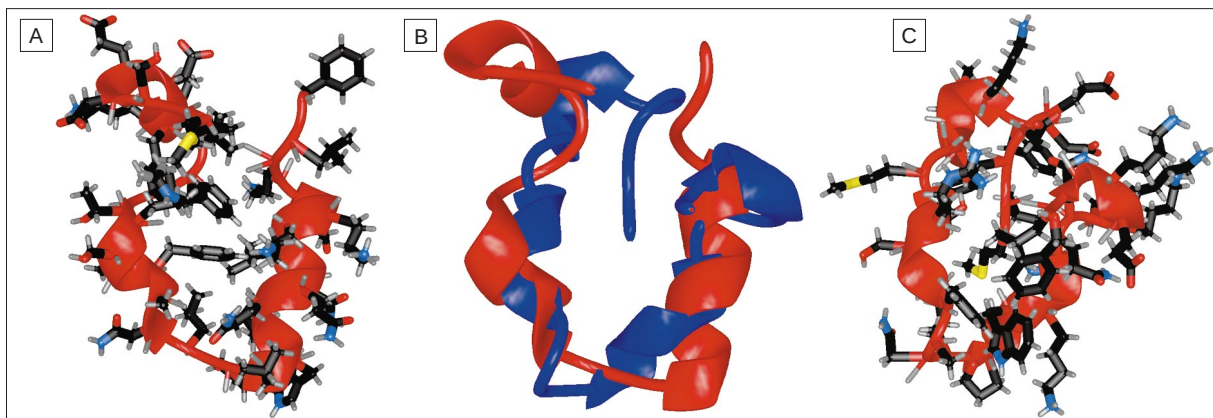
Direct simulations of protein folding with all-atom models

An exciting development was the application of such models in the direct simulations of the early stages of the folding process of small proteins, including a 36-residue villin headpiece subdomain (HP-36)^{120,121} and a zinc-finger-like protein BBA1,¹²² on the micro-

second time scale.^{123,124} Such “top-down” simulations were perceived as impossible for the study of protein folding “in the foreseeable future.”³⁹ Even though both the villin headpiece subdomain and BBA1 are small, they both share the features common to other proteins. The villin headpiece subdomain is a helical protein with well-defined tertiary and secondary structures.¹¹⁹ Its three helices form a unique type of fold with Helix 1 aligned perpendicular to the plane formed by Helices 2 and 3. The three helices are held together by a tightly packed hydrophobic core. Its melting temperature T_m is about 70 degrees Centigrade¹²² and the estimated folding time is about 10 microseconds,^{125,126} making it one of the fastest folding small stable proteins. BBA1 was designed by Imperiali and coworkers¹²² using the zinc finger as the template. It has a three-turn α helix packed against a short β sheet. Unfolding simulations suggested that BBA1 might fold by forming its secondary structures first.¹⁰⁸

The simulation on HP-36 indicated that its initial collapse phase was accompanied by partial formation of the native helices and reduction of hydrophobic surface in a simple downhill process.¹²³ The observed time scale of helix formation, 60 nanoseconds, agrees qualitatively with experimental observations on other proteins.^{86,127,128} The importance of the initial collapse phase has been understated somewhat in the past and has often been termed as the dead time “burst phase,” perhaps due to the fact that experimental studies of these ultrafast processes have been difficult until recently.^{86,127–129} Because early-stage species can often form in the burst phase and they may lead to the subsequent formation of other species and intermediates, the characteristics of these early stage species may affect the folding kinetics, whether or not they themselves are productive intermediates. The observed concomitant formations of both helical domains and hydrophobic clusters in the simulation suggest a way to lower the entropy cost in the subsequent folding processes by reducing the protein internal entropy in the early stages. The reduced protein entropy can be partly compensated by the entropy gained due to releasing water from the surface, as shown in the simulation by the strong correlation between the radius of gyration (R_g) and the solvation free energy (SFE) and by the large decrease of the SFE during the collapse process.¹³⁰ These early stage nascent domains may (1) dissipate, (2) aggregate, or (3) grow. Some of these domains may well be the “nuclei” for the later stage intermediate structures. Formation of native-like domains in the early stage helps the formation of the later-stage interme-

Figure 1 (A) Ribbon representation of the average native NMR structure of the villin headpiece subdomain (HP-36). (B) Superposition of native NMR structure (red) and the intermediate structure obtained in the simulation (blue). Only main chains are shown in ribbon representation. (C) Structure of the intermediate state observed in the simulation.



Reprinted with permission from "Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution," *Science* **282**, No. 5389 (October 23, 1998). © 1998 American Association for the Advancement of Science.

diates structures and perhaps the formation of the native structure. Conversely, the non-native domains will eventually dissipate and those that are retained in the later stage intermediate species will be difficult to dissipate and contribute to the free energy barriers. Correlation calculations indicated that the collapse was driven both by burial of hydrophobic surface and a lowering of the internal energy of the protein.¹³⁰

Considerable fluctuations between compact and extended states were also observed in the simulation, suggesting a shallow free-energy landscape in the vicinity of extended conformations. The residence times of the compact conformations are much longer than those of the extended conformations, suggesting the compact conformations are energetically more favorable than the extended conformations. These indicated that the free-energy landscape is rugged and the existence of intermediate states is likely.

Proteins can fold from fully unfolded states to the native state by going through many intermediate states of varying degrees of stability. In fact, since these intermediate states can be referred to as the "landmarks" of the protein folding free-energy landscape, studying the mechanism of protein folding, in a sense, is to study these intermediate states, the relationship between them, and the relationship between them and the native state. One microsecond, even though it was more than two orders of mag-

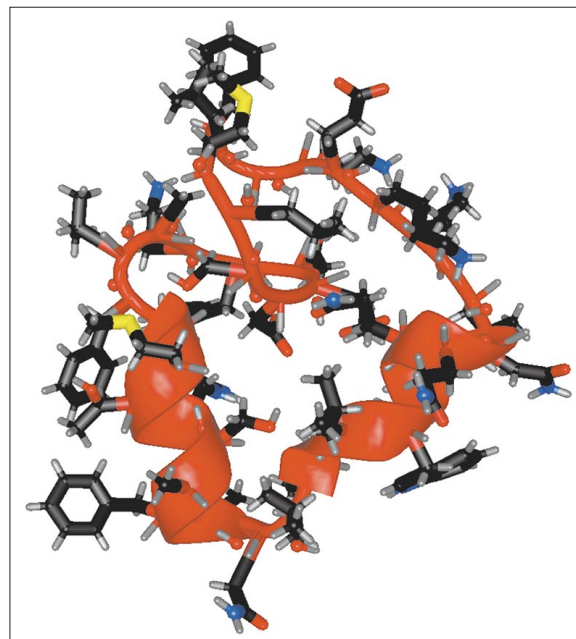
nitude longer than the longest simulation conducted up to 1998 on proteins in water, is still an order of magnitude shorter than the shortest estimated folding time for a protein; thus, it is unrealistic to expect the protein to reach the native state during such a simulation. However, this time scale appears to be sufficient to observe some marginally stable intermediates, if they exist. This was indeed the case in the simulation. A marginally stable intermediate was observed in the simulation and lasted for about 150 nanoseconds. As shown in Figure 1, the main-chain structure of the intermediate was remarkably similar to the native structure, including partial formation of Helices 2 and 3 and a closely packed hydrophobic cluster. The solvation free energy, calculated using the method and parameters developed by Eisenberg and McLachlan,¹³¹ reached a level comparable to that of the native structure. Further analyses using the MM-PB/SA^{132,133} method indicated that the free energy of the intermediate state is the lowest among all the states sampled during the simulation¹³⁴ and both were significantly higher in free energy than the simulation starting from the native structure. This suggests that the reason the simulation did not reach the native state was because of kinetics and not force-field artifacts. Both the solvation free-energy and MM-PB/SA (molecular mechanics-Poisson Boltzmann/surface area) free-energy calculations were independent from the simulation, yet both of their results were consistent with the simulations. Both calculations indicated that the inter-

mediate state was the most favorable one sampled during the simulation, consistent with the long residence time of the state.

It is generally perceived, perhaps even among most specialists in the field, that molecular dynamics simulation is deterministic in a way that is different from stochastic algorithms, such as Monte Carlo, in which the built-in randomness ensures that the simulation will asymptotically approach the ergodic limit. Because of that, it has been argued, molecular dynamics simulation methods are nonergodic. This holds true, however, only to a limited extent. Recent studies by two groups^{135,136} demonstrated that molecular dynamics simulations on complex systems are inherently chaotic. Zhou and Wang¹³⁵ demonstrated that near-identical simulations differing by a root-mean-square deviation (RMSD) of 0.02 Å (angstrom) resulted in two very different trajectories within 1 nanosecond and the resulting structures can differ by an RMSD of as much as 5.0 Å. Yet the RMSD can be reduced substantially to within 2 Å after rigid-body alignment, clearly indicating that the trajectories sampled the same conformational free-energy basin.¹³⁷ Moult and coworkers further demonstrated that despite short-time (<1 picosecond) chaos, which is due to physical interaction (e.g., van der Waals forces), not the algorithmic instability, all trajectories remained close to each other with an RMSD of less than 2 Å,¹³⁶ similar to the typical thermal fluctuation.

Because the source of chaotic behavior is the non-linearity of the physical interactions, it is not surprising that earlier tests on simple harmonic systems failed to reveal such an important aspect. This has important implications to our simulations of protein folding. Because of the presence of chaos in the system, which is a source of randomness, the simulations have a stochastic character in the long time scale, hence can asymptotically approach the ergodicity limit, as well as have deterministic behavior in the short time scale. Simulations of similar conditions are expected to sample the areas that are close to each other in the phase space and produce similar trajectories differing in detail. Due to the randomness, which is a source of uncertainty, one should focus on the qualitative behavior in the analyses of individual trajectories, such as the time scale of the events and general trend.¹²³ The randomness also contributes to the level of fluctuations exhibited in the folding simulations.¹²³ When one wants to focus on the detail, such as the role of individual hydro-

Figure 2 The structure of an intermediate state (at 350 nanoseconds) in one of the 500-nanosecond simulations



gen bonds and contacts, multiple simulations are needed for statistically meaningful results.

We therefore conducted two additional simulations on HP-36, each to 0.5 microseconds starting from different states.¹³⁸ In both simulations, HP-36 reached compact states within about 50 nanoseconds with the radius of gyration comparable to that of the native state. More importantly, substantial secondary structures were formed during the initial collapse phase. Both the time scale and the concomitant occurrence of the hydrophobic collapse and the secondary structures were consistent with our earlier simulation^{123,130} and with experiments.^{86,127,128} One of these two trajectories also reached a state with structure resembling that of the intermediate state found earlier, including similarities of both topology and formation of a tightly packed hydrophobic cluster, as shown in Figure 2.

Perspective

Understanding the mechanisms of protein folding has been an evolving field. Increasingly detailed studies, both at structural level and time scale, have high-

lighted recent developments. For instance, earlier studies would try to address questions like the cooperativity, high rate of folding, ability to converge on the native state from so many starting conformations

**The quality of knowledge-based
structure prediction methods
depends on the available
experimental structures.**

in the unfolded state, and secondary structures. Many current studies try to address questions on transition states, the relationship between protein structures and the folding processes, and characterization of intermediate and molten globule states. Many of these developments have been catalyzed by the advancements of experimental methods, which have been summarized recently.¹³⁹ Mutageneses have allowed extensive and detailed characterization of the transition states for a few small proteins.^{140–142} Hydrogen exchange experiments have been applied to study the dynamics of proteins.^{143,144} An atomic force microscope (AFM) can probe the stability of proteins¹⁴⁵ that can be compared directly with simulations.¹⁴⁶ Solution structure techniques can be used to study some of the intermediate states, including molten globule,^{147,148} partially folded,¹⁴⁹ and unfolded^{150,151} states. Ultrafast kinetic experiments can provide increasingly detailed information on the early-stage folding processes.^{40,86,127,128} Single-molecule techniques have made it possible to monitor the dynamics and folding and unfolding processes of individual molecules.³⁰

Encouraging progress has been made recently and we are now in an era of active application of molecular dynamics simulations to study the folding process. Because of the vital importance of water in protein folding and in the cell, the explicit representation of both solvent and proteins is expected to play an increasingly important role in our understanding of protein folding mechanisms. With the promises of even more accurate models and considerably faster computer speed, together with the advancement of experimental approaches, the goal of understanding the folding of small proteins should be achievable in the future.

Besides the understanding of protein folding mechanisms, a major goal is the prediction of structures

from sequences. It has become increasingly clear that structural prediction methods have made considerable progress,^{152–154} along with the development of protein design,^{122,155–157} despite our lack of comprehensive understanding of the folding mechanisms. The quality of knowledge-based structure prediction methods depends on the available experimental structures. They can reach reasonable accuracy for proteins with sequence homologies to proteins with known structures at above 50 percent. But the quality decreases substantially for lower sequence homologies.¹⁵⁴ Other related issues include the need to predict the binding affinities of small molecules to proteins when the flexibility of either ligands or proteins plays a role and the need to understand conformational changes using simulation methods. These reinforce the need for further understanding of protein folding mechanisms.

The subject of protein folding is perhaps one of the most challenging areas in biophysical chemistry. Given the complexity of protein structures, diverse folding processes should be expected, including the possible role of certain folding-assisting domains within large proteins.¹⁵⁸ Furthermore, the complexity of *in vivo* folding processes is, as yet, another challenge.^{159–163} Therefore, full understanding of protein folding mechanisms is indeed a daunting task. Despite this, recent developments have made us believe that an eventual solution may lie ahead. From a theoretical perspective, an immediate objective is to accurately replicate the complete folding process of small fast-folding proteins on computers, including atomic details. Such simulation results would provide the data for developing abstract models at a conceptual level that describe general and unambiguous features of protein folding mechanisms. The success of such simulations would itself be a strong testament to the accuracy of the method and parameters. The diversity of protein structures and the complexity of the *in vivo* folding process can, in principle, be dealt with by a combination of experiments and further simulations. We may also be able to answer questions such as whether or not a particular part of a protein is designed to assist the folding of the rest of the protein, as found in “intramolecular chaperons,” and how this assistance occurs. The mechanism of chaperon-assisted folding processes can also be better understood. But without the basic understanding of the folding process of single-domain, small, fast-folding proteins, understanding of more complex folding processes will be more difficult to achieve.

Acknowledgment

We are in debt to Ken Dill, Michael Levitt, and Tack Kuntz for helpful discussions. Computer time was provided by the Pittsburgh Supercomputer Center and by Silicon Graphics, Inc. This work has been supported by the National Institutes of Health (Grant GM-29072) and a University of California Biotechnology Star grant from Amgen Inc. (to Dr. Kollman).

Cited references

1. H. Li, R. Helling, C. Tang, and N. Wingreen, "Emergence of Preferred Structures in a Simple Model of Protein Folding," *Science* **273**, No. 5275, 666–669 (1996).
2. H. Li, C. Tang, and N. S. Wingreen, "Are Protein Folds Atypical?" *Proceedings of the National Academy of Sciences (USA)* **95**, No. 9, 4987–4990 (April 28, 1998).
3. C. B. Anfinsen, "Principles That Govern the Folding of Protein Chains," *Science* **181**, No. 96, 223–230 (July 20, 1973).
4. B. Honig and A.-S. Yang, "Free Energy Balance in Protein Folding," *Advances in Protein Chemistry* **46**, 27–58 (1995).
5. P. L. Privalov, "Stability of Proteins: Small Globular Proteins," *Advances in Protein Chemistry* **33**, 167–241 (1979).
6. G. I. Makhatadze and P. L. Privalov, "Energetics of Protein Structure," *Advances in Protein Chemistry* **47**, 307–425 (1995).
7. K. A. Dill, S. Bromberg, K. Z. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, "Principles of Protein Folding: A Perspective from Simple Exact Models," *Protein Science* **4**, No. 4, 561–602 (1995).
8. P. J. Thomas, B. H. Qu, and P. L. Pedersen, "Defective Protein Folding as a Basis of Human Disease," *Trends in Biochemical Sciences* **20**, No. 11, 456–459 (1995).
9. V. E. Bychkova and O. B. Ptitsyn, "Folding Intermediates Are Involved in Genetic Diseases?," *FEBS Letters* **359**, 6–8 (1995).
10. R. N. Sifers, "Defective Protein Folding as a Cause of Disease," *Nature Structural Biology* **2**, 355–357 (1995).
11. P. Choudhury, Y. Liu, and R. N. Sifers, "Quality Control of Protein Folding: Participation in Human Disease," *News in Physiological Sciences* **12**, 162–166 (1997).
12. B. H. Qu, E. Strickland, and P. J. Thomas, "Cystic Fibrosis: A Disease of Altered Protein Folding," *Journal of Bioenergetics and Biomembranes* **29**, No. 5, 483–490 (1997).
13. S. B. Prusiner, "Prion Diseases and the BSE Crisis," *Science* **278**, No. 5336, 245–251 (1997).
14. G. Kuznetsov and S. K. Nigam, "Mechanisms of Disease: Folding of Secretory and Membrane Proteins," *New England Journal of Medicine* **339**, No. 23, 1688–1695 (December 3, 1998).
15. J. Baum and B. Brodsky, "Folding of Peptide Models of Collagen and Misfolding in Disease," *Current Opinion in Structural Biology* **9**, 122–128 (1999).
16. C. M. Dobson, "Protein Misfolding, Evolution and Disease," *Trends in Biochemical Sciences* **24**, No. 9, 329–332 (1999).
17. S. E. Radford and C. M. Dobson, "From Computer Simulations to Human Disease: Emerging Themes in Protein Folding," *Cell* **97**, No. 3, 291–298 (1999).
18. J. C. Rochet and P. T. Lansbury, "Amyloid Fibrillogenesis: Themes and Variations," *Current Opinion in Structural Biology* **10**, No. 1, 60–68 (2000).
19. J. J. Hermans, D. Lohr, and D. Ferro, "Unfolding and Hydrogen Exchange of Proteins: The Three-Dimensional Ising Lattice as a Model," *Nature* **224**, No. 215, 175–177 (1969).
20. O. B. Ptitsyn and A. A. Rashin, "A Model of Myoglobin Self-Organization," *Biophysical Chemistry* **3**, No. 1, 1–20 (1975).
21. A. Rojnuckarin, S. Kim, and S. Subramaniam, "Brownian Dynamics Simulations of Protein Folding: Access to Milliseconds Time Scale and Beyond," *Proceedings of the National Academy of Sciences (USA)* **95**, No. 8, 4288–4292 (1998).
22. M. Levitt and A. Warshel, "Computer Simulation of Protein Folding," *Nature* **253**, 694–698 (1975).
23. M. Levitt, "A Simplified Representation of Protein Conformations for Rapid Simulation of Protein Folding," *Journal of Molecular Biology* **104**, 57–116 (1976).
24. A. Warshel and M. Levitt, "Folding and Stability of Helical Proteins: Carp Myogen," *Journal of Molecular Biology* **106**, No. 2, 421–437 (1976).
25. A. T. Hagler and B. Honig, "On the Formation of Protein Tertiary Structure on a Computer," *Proceedings of the National Academy of Sciences (USA)* **75**, No. 2, 554–558 (1978).
26. J. Skolnick and A. Kolinski, "Simulations of the Folding of a Globular Protein," *Science* **250**, 1121–1125 (1990).
27. F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, "The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures," *Journal of Molecular Biology* **112**, 535 (1977).
28. S. Miyazawa and R. L. Jernigan, "Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures: Quasi-Chemical Approximation," *Macromolecules* **18**, 534–552 (1985).
29. H. P. Lu, L. Y. Xun, and X. S. Xie, "Single-Molecule Enzymatic Dynamics," *Science* **282**, No. 5395, 1877–1882 (1998).
30. A. D. Mehta, M. Rief, J. A. Spudich, D. A. Smith, and R. M. Simmons, "Single-Molecule Biomechanics with Optical Methods," *Science* **283**, No. 5408, 1689–1695 (1999).
31. N. Go, "Theoretical Studies of Protein Folding," *Annual Review of Biophysics and Bioengineering* **12**, 183–210 (1983).
32. P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, "Navigating the Folding Routes," *Science* **267**, 1619 (March 17, 1995).
33. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, "Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis," *Proteins: Structure, Function, and Genetics* **21**, No. 3, 167–195 (1995).
34. K. A. Dill and H. S. Chan, "From Levinthal to Pathways to Funnels," *Nature Structural Biology* **4**, No. 1, 10–19 (1997).
35. A. Sali, E. Shakhnovich, and M. Karplus, "How Does a Protein Fold?" *Nature* **369**, No. 6477, 248–251 (1994).
36. A. Dinner, A. Sali, M. Karplus, and E. Shakhnovich, "Phase Diagram of a Model Protein Derived by Exhaustive Enumeration of the Conformations," *Journal of Chemical Physics* **101**, No. 2, 1444–1451 (1994).
37. V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, "Free-Energy Landscape for Protein-Folding Kinetics: Intermediates, Traps, and Multiple Pathways in Theory and Lattice Model Simulations," *Journal of Chemical Physics* **101**, No. 7, 6052–6062 (1994).
38. A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich, "Is Burst Hydrophobic Collapse Necessary for Protein Folding?" *Biochemistry* **34**, No. 9, 3066–3076 (1995).
39. E. I. Shakhnovich, "Theoretical Studies of Protein-Folding Thermodynamics and Kinetics," *Current Opinion in Structural Biology* **7**, No. 1, 29–40 (1997).
40. V. Muñoz, P. A. Thompson, J. Hofrichter, and W. Eaton,

- "Folding Dynamics and Mechanism of Beta-Hairpin Formation," *Nature* **390**, No. 6656, 196–199 (1997).
41. V. Muñoz and W. A. Eaton, "A Simple Model for Calculating the Kinetics of Protein Folding from Three-Dimensional Structures," *Proceedings of the National Academy of Sciences (USA)* **96**, No. 20, 11311–11316 (1999).
 42. Y. Q. Zhou and M. Karplus, "Interpreting the Folding Kinetics of Helical Proteins," *Nature* **401**, 400–403 (1999).
 43. Y. Q. Zhou and M. Karplus, "Folding of a Model Three-Helix Bundle Protein: A Thermodynamic and Kinetic Analysis," *Journal of Molecular Biology* **293**, No. 4, 917–951 (1999).
 44. D. Bashford, D. L. Weaver, and M. Karplus, "Diffusion-Collision Model for the Folding Kinetics of the Lambda-Repressor Operator-Binding Domain," *Journal of Biomolecular Structure and Dynamics* **1**, 1243–1255 (1984).
 45. M. Karplus and D. L. Weaver, "Protein-Folding Dynamics: The Diffusion-Collision Model and Experimental Data," *Protein Sciences* **3**, No. 4, 650–668 (1994).
 46. K. A. Dill, "Dominant Forces in Protein Folding," *Biochemistry* **29**, 7133–7155 (1990).
 47. D. S. Riddle, J. V. Santiago, S. T. Brayhall, N. Doshi, V. P. Grantcharova, Q. Yi, and D. Baker, "Functional Rapidly Folding Proteins from Simplified Amino Acid Sequences," *Nature Structural Biology* **4**, No. 10, 805–809 (1997).
 48. S. Miyazawa and R. L. Jernigan, "Residue-Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term for Simulation and Threading," *Journal of Molecular Biology* **256**, No. 3, 623–644 (1996).
 49. I. Bahar and R. L. Jernigan, "Inter-Residue Potentials in Globular Proteins and the Dominance of Highly Specific Hydrophilic Interactions at Close Separation," *Journal of Molecular Biology* **266**, No. 1, 195–214 (1997).
 50. G. M. Crippen, "Easily Searched Protein Folding Potentials," *Journal of Molecular Biology* **260**, No. 3, 467–475 (1996).
 51. J. U. Bowie, R. Luthy, and D. Eisenberg, "A Method to Identify Protein Sequences that Fold into a Known 3-Dimensional Structure," *Science* **253**, No. 5016, 164–170 (1991).
 52. J. U. Bowie and D. Eisenberg, "An Evolutionary Approach to Folding Small Alpha-Helical Proteins That Uses Sequence Information and an Empirical Guiding Fitness Function," *Proceedings of the National Academy of Sciences (USA)* **91**, No. 10, 4436–4440 (1994).
 53. M. Hendlich, P. Lackner, S. Weitkus, H. Floeckner, R. Froschauer, K. Gottsbacher, G. Casari, and M. J. Sippl, "Identification of Native Protein Folds Amongst a Large Number of Incorrect Models: The Calculation of Low Energy Conformations from Potentials of Mean Force," *Journal of Molecular Biology* **216**, No. 1, 167–180 (1990).
 54. J. Moult, "Comparison of Database Potentials and Molecular Mechanics Force Fields," *Current Opinion in Structural Biology* **7**, No. 2, 194–199 (1997).
 55. M. J. Sippl, "Knowledge-Based Potentials for Proteins," *Current Opinion in Structural Biology* **5**, No. 2, 229–235 (1995).
 56. J. P. A. Kocher, M. J. Rooman, and S. J. Wodak, "Factors Influencing the Ability of Knowledge-Based Potentials to Identify Native Sequence-Structure Matches," *Journal of Molecular Biology* **235**, No. 5, 1598–1613 (1994).
 57. A. Liwo, S. Oldziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga, "A United-Residue Force Field for Off-Lattice Protein-Structure Simulations: I. Functional Forms and Parameters of Long-Range Side-Chain Interaction Potentials from Protein Crystal Data," *Journal of Computational Chemistry* **18**, No. 7, 849–873 (1997).
 58. A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, S. Oldziej, and H. A. Scheraga, "A United-Residue Force Field for Off-Lattice Protein-Structure Simulations: II. Parameterization of Short-Range Interactions and Determination of Weights of Energy Terms by Z-Score Optimization," *Journal of Computational Chemistry* **18**, No. 7, 874–887 (1997).
 59. A. Liwo, R. Kazmierkiewicz, C. Czaplewski, M. Groth, S. Oldziej, R. J. Wawak, S. Rackovsky, M. R. Pincus, and H. A. Scheraga, "United-Residue Force Field for Off-Lattice Protein-Structure Simulations: III. Origin of Backbone Hydrogen-Bonding Cooperativity in United-Residue Potentials," *Journal of Computational Chemistry* **19**, No. 3, 259–276 (1998).
 60. M. H. Hao and H. A. Scheraga, "Designing Potential Energy Functions for Protein Folding," *Current Opinion in Structural Biology* **9**, No. 2, 184–188 (1999).
 61. G. M. Crippen and Y. Z. Ohkubo, "Statistical Mechanics of Protein Folding by Exhaustive Enumeration," *Proteins: Structure, Function, and Genetics* **32**, No. 4, 425–437 (1998).
 62. J. Skolnick and A. Kolinski, "Dynamic Monte Carlo Simulations of a New Lattice Model of Globular Protein Folding, Structure, and Dynamics," *Journal of Molecular Biology* **221**, No. 2, 499–531 (1991).
 63. S. Subramaniam, D. K. Tchong, and J. M. Fenton, "A Knowledge-Based Method for Protein Structure Refinement and Prediction," *Proceedings, Fourth International Conference on Intelligent Systems for Molecular Biology*, St. Louis, MO (June 12–18, 1996), pp. 218–229.
 64. S. E. DeBolt and J. Skolnick, "Evaluation of Atomic Level Mean Force Potentials via Inverse Folding and Inverse Refinement of Protein Structures: Atomic Burial Position and Pairwise Non-Bonded Interactions," *Protein Engineering* **9**, No. 8, 637–655 (1996).
 65. M. J. Sippl, M. Ortner, M. Jaritz, P. Lackner, and H. Flockner, "Helmholtz Free Energies of Atom Pair Interactions in Proteins," *Folding and Design* **1**, No. 4, 289–298 (1998).
 66. R. Samudrala and J. Moult, "An All-Atom Distance-Dependent Conditional Probability Discriminatory Function for Protein Structure Prediction," *Journal of Molecular Biology* **275**, No. 5, 895–916 (1998).
 67. A. Rojnuckarin and S. Subramaniam, "Knowledge-Based Interaction Potentials for Proteins," *Proteins: Structure, Function, and Genetics* **36**, No. 1, 54–67 (1999).
 68. M. Levitt, personal communication.
 69. Y. H. Wang, H. Zhang, W. Li, and R. A. Scott, "Discriminating Compact Nonnative Structures from the Native Structure of Globular-Proteins," *Proceedings of the National Academy of Sciences (USA)* **92**, No. 3, 709–713 (1995).
 70. Y. H. Wang, H. Zhang, and R. A. Scott, "A New Computational Model for Protein Folding Based on Atomic Solvation," *Protein Sciences* **4**, No. 7, 1402–1411 (1995).
 71. B. Park and M. Levitt, "Energy Functions That Discriminate X-Ray and Near-Native Folds from Well-Constructed Decoys," *Journal of Molecular Biology* **258**, No. 2, 367–392 (1996).
 72. B. H. Park, E. S. Huang, and M. Levitt, "Factors Affecting the Ability of Energy Functions to Discriminate Correct from Incorrect Folds," *Journal of Molecular Biology* **266**, No. 4, 831–846 (1997).
 73. T. Lazaridis and M. Karplus, "Discrimination of the Native from Misfolded Protein Models with an Energy Function Including Implicit Solvation," *Journal of Molecular Biology* **288**, No. 3, 477–487 (1999).
 74. T. Lazaridis and M. Karplus, "Effective Energy Function for Proteins in Solution," *Proteins: Structure, Function, and Genetics* **35**, No. 2, 133–152 (1999).
 75. A. S. Yang and B. Honig, "Free Energy Determinants of

- Secondary Structure Formation: I. α -Helices," *Journal of Molecular Biology* **252**, No. 3, 351–365 (1995).
76. A.-S. Yang and B. Honig, "Free Energy Determinant of Secondary Structure Formation: II. Antiparallel β -Sheets," *Journal of Molecular Biology* **252**, No. 3, 366–376 (1995).
 77. A.-S. Yang, B. Hitz, and B. Honig, "Free Energy Determinants of Secondary Structure Formation: III. β -Turns and Their Role in Protein Folding," *Journal of Molecular Biology* **259**, No. 4, 873–882 (1996).
 78. S. S. Sung, "Helix Folding Simulations with Various Initial Conformations," *Biophysical Journal* **66**, No. 6, 1796–1803 (1994).
 79. S. S. Sung, "Folding Simulations of Alanine-Based Peptides with Lysine Residues," *Biophysical Journal* **68**, No. 3, 826–834 (1995).
 80. M. Schaefer and M. Karplus, "A Comprehensive Analytical Treatment of Continuum Electrostatics," *The Journal of Physical Chemistry* **100**, 1578–1599 (1996).
 81. M. Schaefer, C. Bartels, and M. Karplus, "Solution Conformations and Thermodynamics of Structured Peptides: Molecular Dynamics Simulation with an Implicit Solvation Model," *Journal of Molecular Biology* **284**, No. 3, 835–848 (1998).
 82. M. Schaefer, C. Bartels, and M. Karplus, "Solution Conformations of Structured Peptides: Continuum Electrostatics versus Distance-Dependent Dielectric Functions," *Theoretical Chemistry Accounts* **101**, No. 1–3, 194–204 (1999).
 83. T. Lazaridis and M. Karplus, "'New View' of Protein Folding Reconciled with the Old Through Multiple Unfolding Simulations," *Science* **278**, No. 5345, 1928–1931 (1997).
 84. A. R. Dinner, T. Lazaridis, and M. Karplus, "Understanding Beta-Hairpin Formation," *Proceedings of the National Academy of Sciences (USA)* **96**, No. 16, 9068–9073 (1999).
 85. X. W. Wu and S. S. Sung, "Simulation of Peptide Folding with Explicit Water: A Mean Solvation Method," *Proteins: Structure, Function, and Genetics* **34**, No. 3, 295–302 (1999).
 86. P. A. Thompson, W. A. Eaton, and J. Hofrichter, "Laser Temperature Jump Study of the Helix $\rightleftharpoons$ Coil Kinetics of an Alanine Peptide Interpreted with a 'Kinetic Zipper' Model," *Biochemistry* **36**, No. 30, 9200–9210 (1997).
 87. F. B. Sheinerman and C. L. Brooks, "Calculations on Folding of Segment B1 of Streptococcal Protein G," *Journal of Molecular Biology* **278**, No. 2, 439–456 (1998).
 88. A. Cafisch and M. Karplus, "Acid and Thermal-Denaturation of Barnase Investigated by Molecular-Dynamics Simulations," *Journal of Molecular Biology* **252**, No. 5, 672–708 (1995).
 89. B. D. Bursulaya and C. L. Brooks, "Folding Free Energy Surface of a Three-Stranded Beta-Sheet Protein, *Journal of the American Chemical Society* **121**, No. 43, 9947–9951 (1999).
 90. L. X. Dang, J. E. Rice, J. Caldwell, and P. A. Kollman, "Ion Solvation in Polarizable Water-Molecular Dynamics Simulations," *Journal of the American Chemical Society* **113**, No. 7, 2481–2486 (1991).
 91. W. D. Cornell, J. W. Caldwell, and P. A. Kollman, "Calculation of the Phi-Psi Maps for Alanine and Glycyl Dipeptides with Different Additive and Non-Additive Molecular Mechanical Models," *Journal de Chimie Physique et de Physico-Chimie de Biologie* **94**, 1417–1435 (1997).
 92. D. N. Bernardo, Y. B. Ding, K. Kroghjerspersen, and R. M. Levy, "An Anisotropic Polarizable Water Model: Incorporation of All-Atom Polarizabilities into Molecular Mechanics Force Fields," *The Journal of Physical Chemistry* **98**, 4180–4187 (1994).
 93. T. A. Halgren, "Potential Energy Functions," *Current Opinion in Structural Biology* **5**, No. 2, 205–210 (1995).
 94. J. Tirado-Rives and W. L. Jorgensen, "Molecular Dynamics Simulations of the Unfolding of an Alpha-Helical Analogue of Ribonuclease A S-Peptide in Water," *Biochemistry* **30**, No. 16, 3864–3871 (1991).
 95. V. Daggett and M. Levitt, "A Model of the Molten Globule State from Molecular Dynamics Simulations," *Proceedings of the National Academy of Sciences (USA)* **89**, No. 11, 5142–5146 (1992).
 96. V. Daggett and M. Levitt, "Protein Folding Unfolding Dynamics," *Current Opinion in Structural Biology* **4**, No. 2, 291–295 (1994).
 97. D. O. V. Alonso and V. Daggett, "Molecular Dynamics Simulations of Protein Unfolding and Limited Refolding: Characterization of Partially Unfolded States of Ubiquitin in 60% Methanol and in Water," *Journal of Molecular Biology* **247**, No. 3, 501–520 (1995).
 98. V. Daggett, A. J. Li, and A. R. Fersht, "Combined Molecular Dynamics and Phi-Value Analysis of Structure-Reactivity Relationships in the Transition State and Unfolding Pathway of Barnase: Structural Basis of Hammond and Anti-Hammond Effects," *Journal of the American Chemical Society* **120**, No. 49, 12740–12754 (1998).
 99. A. M. J. J. Bonvin and W. F. van Gunsteren, "Beta-Hairpin Stability and Folding: Molecular Dynamics Studies of the First Beta-Hairpin of Tendamistat," *Journal of Molecular Biology* **296**, No. 1, 255–268 (2000).
 100. V. S. Pande and D. S. Rokhsar, "Molecular Dynamics Simulations of Unfolding and Refolding of a β -Hairpin Fragment of Protein G," *Proceedings of the National Academy of Sciences (USA)* **96**, No. 16, 9062–9067 (1999).
 101. C. A. Schiffer, V. Dötsch, K. Wüthrich, and W. F. van Gunsteren, "Exploring the Role of the Solvent in the Denaturation of a Protein: A Molecular Dynamics Study of the DNA Binding Domain of the 434 Repressor," *Biochemistry* **34**, No. 46, 15057–15067 (1995).
 102. P. H. Hunenberger, A. E. Mark, and W. F. van Gunsteren, "Computational Approaches to Study Protein Unfolding: Hen Egg-White Lysozyme as a Case-Study," *Proteins: Structure, Function, and Genetics* **21**, No. 3, 196–213 (1995).
 103. H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, "Unfolding of Titin Immunoglobulin Domains by Steered Molecular Dynamics Simulation," *Biophysical Journal* **75**, No. 2, 662–671 (1998).
 104. A. Krammer, H. Lu, B. Isralewitz, K. Schulten, and V. Vogel, "Forced Unfolding of the Fibronectin Type III Module Reveals a Tensile Molecular Recognition Switch," *Proceedings of the National Academy of Sciences (USA)* **96**, No. 4, 1351–1356 (1999).
 105. E. Paci and M. Karplus, "Forced Unfolding of Fibronectin Type 3 Modules: An Analysis by Biased Molecular Dynamics Simulations," *Journal of Molecular Biology* **288**, No. 3, 441–459 (1999).
 106. A. Li and V. Daggett, "Identification and Characterization of the Unfolding Transition State of Chymotrypsin Inhibitor 2 by Molecular Dynamics Simulations," *Journal of Molecular Biology* **257**, No. 2, 412–429 (1996).
 107. A. V. Finkelstein, "Can Protein Unfolding Simulate Protein Folding?" *Protein Engineering* **10**, No. 8, 843–845 (1997).
 108. L. Wang, Y. Duan, R. Shortle, B. Imperiali, and P. A. Kollman, "Study of the Stability and Unfolding Mechanism of BBA1 by Molecular Dynamics Simulations at Different Temperatures," *Protein Sciences* **8**, No. 6, 1292–1304 (1999).
 109. E. M. Boczek and C. L. Brooks, "Constant-Temperature

- Free-Energy Surfaces for Physical and Chemical Processes," *The Journal of Physical Chemistry* **97**, No. 17, 4509–4513 (1993).
110. E. M. Boczek and C. L. Brooks, "First-Principles Calculation of the Folding Free-Energy of a 3-Helix Bundle Protein," *Science* **269**, No. 5222, 393–396 (1995).
111. H. J. C. Berendsen, "Protein Folding: A Glimpse of the Holy Grail?" *Science* **282**, No. 5389, 642–643 (1998).
112. D. J. Tobias, J. E. Mertz, and C. L. Brooks, "Nanosecond Time Scale Folding Dynamics of a Pentapeptide in Water," *Biochemistry* **30**, No. 24, 6054–6058 (1991).
113. E. Demchuk, D. Bashford, and D. A. Case, "Dynamics of a Type VI Reverse Turn in a Linear Peptide in Aqueous Solution," *Folding and Design* **2**, No. 1, 35–46 (1997).
114. X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark, "Reversible Peptide Folding in Solution by Molecular Dynamics Simulation," *Journal of Molecular Biology* **280**, 925–932 (1998).
115. C. Chipot and A. Pohorille, "Structure and Dynamics of Small Peptides at Aqueous Interfaces: A Multi-Nanosecond Molecular Dynamics Study," *Journal of Molecular Structure: THEOCHEM* **398**, No. 399, 529–535 (1997).
116. C. Chipot and A. Pohorille, "Folding and Translocation of the Undecamer of Poly-L-Leucine Across the Water-Hexane Interface: A Molecular Dynamics Study," *Journal of the American Chemical Society* **120**, No. 46, 11912–11924 (1998).
117. X. W. Wu and S. M. Wang, "Enhancing Systematic Motion in Molecular Dynamics Simulation," *Journal of Chemical Physics* **110**, No. 19, 9401–9410 (1999).
118. X. W. Wu and S. M. Wang, "Self-Guided Molecular Dynamics Simulation for Efficient Conformational Search," *The Journal of Physical Chemistry B* **102**, No. 37, 7238–7250 (1998).
119. Y. Pak and S. M. Wang, "Folding of a 16-Residue Helical Peptide Using Molecular Dynamics Simulation with Tsallis Effective Potential," *Journal of Chemical Physics* **111**, No. 10, 4359–4361 (1999).
120. C. J. McKnight, D. S. Doering, P. T. Matsudaira, and P. S. Kim, "A Thermostable 35-Residue Subdomain Within Villin Headpiece," *Journal of Molecular Biology* **260**, No. 2, 126–134 (1996).
121. C. J. McKnight, P. T. Matsudaira, and P. S. Kim, "NMR Structure of the 35-Residue Villin Headpiece Subdomain," *Nature Structural Biology* **4**, No. 3, 180–184 (1997).
122. M. D. Struthers, R. P. Cheng, and B. Imperiali, "Design of a Monomeric 23-Residue Polypeptide with Defined Tertiary Structure," *Science* **271**, No. 5247, 342–345 (1996).
123. Y. Duan and P. A. Kollman, "Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution," *Science* **282**, No. 5389, 740–744 (1998).
124. Y. Duan and P. A. Kollman, "Early State Folding of a Small α/β Protein BBA1 in Aqueous Solution Observed in Molecular Dynamics Simulations," in preparation.
125. K. W. Plaxco, K. T. Simons, and D. Baker, "Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins," *Journal of Molecular Biology* **277**, No. 4, 985–994 (1998).
126. K. W. Plaxco, personal communication (1998).
127. S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer, "Fast Events in Protein Folding: Helix Melting and Formation in a Small Peptide," *Biochemistry* **35**, No. 3, 691–697 (1996).
128. R. Gilmanshin, S. Williams, R. H. Callender, W. H. Woodruff, and R. B. Dyer, "Fast Events in Protein Folding: Relaxation Dynamics of Secondary and Tertiary Structure in Native Apomyoglobin," *Proceedings of the National Academy of Sciences (USA)* **94**, No. 8, 3709–3713 (1997).
129. R. M. Ballew, J. Sabelko, and M. Gruebele, "Direct Observation of Fast Protein Folding: The Initial Collapse of Apomyoglobin," *Proceedings of the National Academy of Sciences (USA)* **93**, No. 12, 5759–5764 (1996).
130. Y. Duan, L. Wang, and P. A. Kollman, "The Early Stage of Folding of Villin Headpiece Subdomain Observed in a 200-Nanosecond Fully Solvated Molecular Dynamics Simulation," *Proceedings of the National Academy of Sciences (USA)* **95**, No. 7, 9897–9902 (1998).
131. D. Eisenberg and A. D. McLachlan, "Solvation Energy in Protein Folding and Binding," *Nature* **319**, No. 3086, 199–203 (1986).
132. J. Srinivasan, T. E. I. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case, "Continuum Solvent Studies of the Stability of DNA, RNA and Phosphoramidate-DNA Helices," *Journal of the American Chemical Society* **120**, No. 37, 9401–9409 (1998).
133. T. E. I. Cheatham, J. Srinivasan, D. A. Case, and P. A. Kollman, "Molecular Dynamics and Continuum Solvent Studies of the Stability of PolyG-PolyC and PolyA-PolyT DNA Duplexes in Solution," *Journal of Biomolecular Structure and Dynamics* **16**, No. 2, 265–280 (1998).
134. M. Lee, Y. Duan, and P. A. Kollman, "Use of MM-PB/SA in Estimating the Free Energies of Proteins: Application to Native, Intermediates, and Unfolded Villin Headpiece," *Proteins: Structure, Function, and Genetics* **39**, 309–316 (2000).
135. H. B. Zhou and L. Wang, "Chaos in Biomolecular Dynamics," *The Journal of Physical Chemistry* **100**, No. 20, 8101–8105 (1996).
136. M. Braxenthaler, R. Unger, D. Auerbach, J. A. Given, and J. Moult, "Chaos in Protein Dynamics," *Proteins: Structure, Function, and Genetics* **29**, No. 4, 417–425 (1997).
137. L. Wang, personal communication.
138. Y. Duan and P. A. Kollman, "Highly Compact Intermediate States Observed in 500-Nanosecond Molecular Dynamics Simulations in Aqueous Solution," in preparation.
139. D. J. Brockwell, D. A. Smith, and S. E. Radford, "Protein Folding Mechanisms: New Methods and Emerging Ideas," *Current Opinion in Structural Biology* **10**, No. 1, 16–25 (2000).
140. A. R. Fersht, "Characterizing Transition States in Protein Folding: An Essential Step in the Puzzle," *Current Opinion in Structural Biology* **5**, 79–84 (1995).
141. L. S. Itzhaki, D. E. Otzen, and A. R. Fersht, "The Structure of the Transition State for Folding of Chymotrypsin Inhibitor 2 Analysed by Protein Engineering Methods: Evidence for a Nucleation-Condensation Mechanism for Protein Folding," *Journal of Molecular Biology* **254**, No. 2, 260–288 (1995).
142. L. S. Itzhaki, J. L. Neira, J. Ruizsanz, G. D. Gay, and A. R. Fersht, "Search for Nucleation Sites in Smaller Fragments of Chymotrypsin Inhibitor 2," *Journal of Molecular Biology* **254**, No. 2, 289–304 (1995).
143. J. Hollien and S. Marqusee, "Structural Distribution of Stability in a Thermophilic Enzyme," *Proceedings of the National Academy of Sciences (USA)* **96**, No. 24, 13674–13678 (1999).
144. J. J. Englander, G. Louie, R. E. McKinnin, and S. W. Englander, "Energetic Components of the Allosteric Machinery in Hemoglobin Measured by Hydrogen Exchange," *Journal of Molecular Biology* **284**, No. 5, 1695–1706 (1998).
145. T. E. Fisher, A. F. Oberhauser, M. Carrion-Vazquez, P. E. Marszalek, and J. M. Fernandez, "The Study of Protein Mechanics with the Atomic Force Microscope," *Transactions in Biochemical Sciences* **24**, No. 10, 379–384 (1999).

146. H. Lu and K. Schulten, "Steered Molecular Dynamics Simulations of Force-Induced Protein Domain Unfolding," *Proteins: Structure, Function, and Genetics* **35**, No. 4, 453–463 (1999).
147. H. J. Dyson and P. E. Wright, "Equilibrium NMR Studies of Unfolded and Partially Folded Proteins," *Nature Structural Biology* **5**, No. 7, 499–503 (1998).
148. D. Eliezer, J. Yao, H. J. Dyson, and P. E. Wright, "Structural and Dynamic Characterization of Partially Folded States of Apomyoglobin and Implications for Protein Folding," *Nature Structural Biology* **5**, No. 2, 148–155 (1998).
149. O. Zhang, L. E. Kay, D. Shortle, and J. D. Forman-Kay, "Comprehensive NOE Characterization of a Partially Folded Large Fragment of Staphylococcal Nuclease Delta131Delta, Using NRM Methods with Improved Resolution," *Journal of Molecular Biology* **272**, No. 1, 9–20 (1997).
150. Y. K. Mok, C. M. Kay, L. E. Kay, and J. Forman-Kay, "NOE Data Demonstrating a Compact Unfolded State for an SH3 Domain Under Non-Denaturing Conditions," *Journal of Molecular Biology* **289**, No. 3, 619–638 (1999).
151. D. W. Yang, Y. K. Mok, D. R. Muhandiram, J. D. Forman-Kay, and L. E. Kay, "H-1-C-13 Dipole-Dipole Cross-Correlated Spin Relaxation as a Probe of Dynamics in Unfolded Proteins: Application to the DrkN SH3 Domain," *Journal of the American Chemical Society* **121**, No. 14, 3555–3556 (1999).
152. A. Sali and J. Kuriyan, "Challenges at the Frontiers of Structural Biology," *Transactions in Biochemical Sciences* **24**, No. 12, M20–M24 (1999).
153. P. Koehl and M. Levitt, "A Brighter Future for Protein Structure Prediction," *Nature Structural Biology* **6**, No. 2, 108–111 (1999).
154. J. Moult, T. Hubbard, K. Fidelis, and J. T. Pedersen, "Critical Assessment of Methods of Protein Structure Prediction (CASP): Round III," *Proteins: Structure, Function, and Genetics* **37**, No. S3, 2–6 (1999).
155. B. I. Dahiyat and S. L. Mayo, "De Novo Protein Design: Fully Automated Sequence Selection," *Science* **278**, No. 5335, 82–87 (1997).
156. T. Kortemme, M. Ramirez-Alvarado, and L. Serrano, "Design of a 20-Amino Acid, Three-Stranded Beta-Sheet Protein," *Science* **281**, No. 5374, 253–256 (1998).
157. S. M. Malakauskas and S. L. Mayo, "Design, Structure and Stability of a Hyperthermophilic Protein Variant," *Nature Structural Biology* **5**, No. 6, 470–475 (1998).
158. J. L. Sohl, S. S. Jaswal, and D. A. Agard, "Unfolded Conformations of Alpha-Lytic Protease Are More Stable Than Its Native State," *Nature* **395**, No. 6704, 817–819 (1998).
159. P. B. Sigler, Z. H. Xu, H. S. Rye, S. G. Burston, W. A. Fenton, and A. L. Horwich, "Structure and Function in GroEL-Mediated Protein Folding," *Annual Review of Biochemistry* **67**, 581–608 (1998).
160. J. P. Hendrick and F. U. Hartl, "Molecular Chaperone Functions of Heat-Shock Proteins," *Annual Review of Biochemistry* **62**, 349–384 (1993).
161. R. Hlodan and F. U. Hartl, "How the Protein Folds in the Cell," *Mechanisms of Protein Folding*, R. H. Pain, Editor, Oxford University Press, New York (1994), pp. 194–228.
162. F. U. Hartl and J. Martin, "Molecular Chaperones in Cellular Protein Folding," *Current Opinion in Structural Biology* **5**, No. 1, 92–102 (1995).
163. D. E. Feldman and J. Frydman, "Protein Folding in Vivo: The Importance of Molecular Chaperones," *Current Opinion in Structural Biology* **10**, No. 1, 26–33 (2000).

Accepted for publication January 2, 2001.

Yong Duan *Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716 (electronic mail: yduan@udel.edu).* Dr. Duan received a Ph.D. degree in 1996 from the University of Pittsburgh and did postdoctoral work at the University of California at San Francisco from 1996 to 2000. He is currently an assistant professor in the Department of Chemistry and Biochemistry at the University of Delaware. His research is in the areas of computational biochemistry and computational biophysics, protein folding, and molecular recognition.

Peter A. Kollman *Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446 (electronic mail: pak@cgl.ucsf.edu).* Dr. Kollman received his B.A. degree from Grinnell College and, in 1970, his Ph.D. degree from Princeton University. After a year as NATO (North Atlantic Treaty Organization) postdoctoral fellow at Cambridge University in England, he joined the faculty of the Department of Pharmaceutical Chemistry at the University of California at San Francisco, where he has worked until the present time. In 1995, he received the Computers in Chemistry award from the American Chemical Society.