

Transition State Ensemble for the Folding of B Domain of Protein A: A Comparison of Distributed Molecular Dynamics Simulations with Experiments

Shanmei Cheng,[†] Yuedong Yang,[†] Wenrui Wang,[‡] and Haiyan Liu^{*,†}

Key Laboratory of Structural Biology, School of Life Sciences, University of Science and Technology of China (USTC), and Hefei National Laboratory for Physical Sciences at Microscale, Hefei, Anhui, 230026, China, and School of Informational Sciences and Technology, University of Science and Technology of China (USTC), Hefei, Anhui, 230026, China

Received: April 6, 2005; In Final Form: October 8, 2005

Folding pathways of the B domain of *staphylococcal* protein A have been sampled with a distributed computing approach. Starting from an extended structure, the method employs an index measuring topological similarity to the native structure to selectively sample trajectory branches leading to the native fold. Unperturbed and continuous folding trajectories are drawn on a physics-based atomic potential energy surface with an implicit solvent. The sampled folding trajectories demonstrate a similar sequence of events: the earlier stage involves a partial formation of helix 2 and to a less extent of helix 1 at their N terminals, followed by the hydrophobic collapse between residues F14, I17, and L18 on helix 1 and residues R28, F31, and I32 on helix 2, which results in the rigidification of the helix turn from R28 to I32. Helix 2 is then able to extend, allowing for the formation to turn 2. The above description explains one experimental result why a G30A mutant of the protein was observed to be the fastest folder among proteins of its size. And the ensemble of structures right before the final collapse is in good agreement with the transition state ensemble mapped by another recent experiment with Fersht Φ values. We emphasize that because the approach here does not provide quantifications of the free energy landscape, our model of the transition state ensemble emerges from comparisons of simulations and previous experimental results rather than from the simulation results alone. On the other hand, as our approach does not rely on a low-dimensional free energy surface, it can complement methods based on the construction of free energy surfaces.

Introduction

Enormous theoretical and experimental efforts have been devoted to study protein folding. Our understandings of this problem, although still far from completion, have been deepened significantly during the last few decades. Various models of the protein folding process have been proposed, among which the framework and the nucleation–condensation models are quite representative.^{1–3} It can be argued that for different real world proteins or protein folds, there may be an entire spectrum of folding mechanisms between models such as framework and models such as nucleation–condensation. The ultimate determinant of the folding mechanism of a particular protein is its free energy hypersurface. The funnel theory of protein folding tries to resolve the contradiction between the enormous conformational space of a protein and its rapid folding kinetics.⁴ However, it remains not so well-understood what differentiates those tiny fractions of amino acid sequences that possess such funnel-like free energy landscapes from those that do not. Even given the sequence and the native structure of a protein, in most cases it still remains controversial where and how to locate its folding mechanism in the above spectrum.

One closely related problem is to characterize the ensemble of structures corresponding to the transition state ensemble

(TSE) separating the native and the unfolded conformations. If we can achieve this, the natural next step is to identify local (lower contact order) and distant (higher contact order) interactions key to the folding and stability of a protein. Although not directly aimed at the sequence to structure problem, answering such questions is important for fold and design, as they would shed light on the design principles of various protein folds.

For a number of protein domains, remarkable progress has been made to characterize the TSE for folding through joint efforts of experiments and simulations. The progress made and the uncertainties remaining can be highlighted by studies over the folding mechanism of a small protein domain, the B domain of *staphylococcal* protein A (BdpA). Because of its small size and fast folding kinetics,^{5,6} this protein domain has been widely used as a model system in theoretical and experimental studies. It is a three-helix bundle with the following amino acid sequence: TADNKFNKEQQNAFYIEILHLPNLNEEQRNGFIQSLKDDPSQSANLLAEAKKLNDQAAPKA. In the following text, helix 1 (residues 12–18), turn 1 (residues 19–24), helix 2 (residues 25–37), turn 2 (residues 38–41), and helix 3 (residues 42–53) will be noted as H1, T1, H3, T2, and H3, respectively. Besides the internal stabilizing interactions within individual helices, the inter-helix side chain contacts are also important to stabilize the native conformation.

A number of experiments studying the folding/unfolding of BdpA have been reported. Bottomley et al. probed the unfolding of BdpA using fluorescence spectroscopy,⁷ which suggested that H1 unfolded first, followed by the unfolding of H2 and H3 together. Bai et al. studied peptide fragments corresponding to

* Address correspondence to this author. E-mail: hylu@ustc.edu.cn.

[†] Key Laboratory of Structural Biology, School of Life Sciences, University of Science and Technology of China (USTC), and Hefei National Laboratory for Physical Sciences at Microscale.

[‡] School of Informational Sciences and Technology, University of Science and Technology of China (USTC).

individual helices and their partial combinations by circular dichroism, which suggested some marginal stability of the H2/T2/H3 hairpin.⁸ Myers and Oas used NMR to probe the folding/unfolding kinetics of BdpA as functions of denaturant concentration.⁶ Their data indicated an extremely rapid two-state folding kinetics of this protein. On the basis of the dependence of the folding/unfolding rate on denaturant conditions they estimated that the transition state buried nearly 80% of the solvent accessible area that the native state did. Dimitriadis et al. recently reported the folding kinetics of a F14W G30A mutant of BdpA probed by nanosecond laser-induced temperature jumping experiments.⁹ They identified this mutant as the fastest folding protein of its size studied so far. And the observed dependence of the folding rate on denaturant also indicated a two-state folding kinetics. Vu et al. reported temperature-jump experiments using both fluorescence (Tyr15 on H1) and IR absorbance as probes.¹⁰ The IR probed both ~ 70 ns and ~ 5.3 μ s relaxation processes, which were interpreted as the initial helix-coil transition and the subsequent assembling of helices, respectively, while the fluorescence probe captured only the slower process. The authors tended to interpret this as an indication for a later formation of H1 during folding.

Sato et al. recently preformed extensive analyses to map the transition state ensemble for the folding of BdpA¹¹ using Fersht Φ values.¹² On the basis of Φ values for Ala to Gly mutations at surface-exposed positions, they summarized that at the transition state region, H2 was virtually native (Φ values of 0.8–0.9), H3 only had weakly formed secondary structures at its N-terminal (Φ values of 0.5 decreasing to 0), T1 was unstructured, and T2 might have some structures (two probes gave Φ values of 0.2–0.5). The measurable Φ values on H1 were low (Φ values of around 0.3). From Φ values for the tertiary structure, they found that contacts between H1 and H2 were important for stabilizing the TSE, while the N-terminal of H3 made some weaker contributions.

A variety of theoretical models^{13–23} have been applied to simulate the folding/unfolding process of BdpA. These include off-lattice¹³ and all-atom¹⁴ G \ddot{o} model simulations of the folding process, all-atom force field simulations of the unfolding processes,¹⁵ all-atom force field simulations with umbrella sampling,¹⁶ and replica-exchange molecular dynamics¹⁷ to explore the free energy landscape. A recent review by Wolynes has given insightful analyses of these studies from several aspects.²⁴ As summarized in his review, predictions on important global features of the folding process such as the extent of formations of secondary and tertiary structures and the coupling between formations of secondary and tertiary structure were grossly consistent among different simulations and between simulations and experimental Φ value analyses. The details of the folding processes produced by simulations fell into two classes. Both the studies of Brooks and co-workers employing umbrella sampling¹⁶ and of Garcia and Onuchic employing replica exchange molecular dynamics¹⁷ aimed at constructing the free energy surface as functions of structural reaction coordinates. They suggested that there were substantial contacts formed between H1 and H2 at the transition state. Other simulations emphasized the roles of formations of H3 and the contacts between H2 and H3 for the TSE.

While having brought about significant insights, each of the theoretical simulations is subjected to their own limitations because of the model employed. For example, simulations with the G \ddot{o} -type models allow for sufficient sampling of possible folding pathways to obtain proper statistics of the free energy landscape, but at the price that the funnel type of energy

landscape employed by such models has been set up artificially, i.e., extracted from native contacts. Simulations with the all-atom force field employed more physical descriptions for the energy landscape. This, however, limits the sampling: simulations usually do not directly sample continuous folding pathways. Instead, a low-dimensional free energy surface is constructed based on a series of simulations, each sampling different regions along one or two heuristically chosen coordinates. When the definition of a transition state region is based on a free energy surface of only a few dimensions, conformation samples drawn in this region may fall into many separated minima on the higher dimensions. Thus sampling near the top of a barrier on a low-dimensional free energy surface may simply produce conformations within these minima, which do not necessarily correspond to the true TSE. For this important reason, comparisons of these conformations with the experimental mapping of the TSE are questionable without additional proof that the ensemble of structures sampled around the barrier top corresponds to the true TSE of folding.

In this work, we apply a new approach to simulate the folding process of BdpA. The method employs distributed computing with an accelerated sampling technique, and is able to sample continuous folding molecular dynamics trajectories from the extended state to the native conformational state on an essentially unperturbed physics-based atomic potential energy surface. While a major disadvantage of the approach presented here is that a quantification of the change of free energy introduced by the branch selection procedure is not available, the results complement free energy surface-based studies as no pre-chosen reaction coordinates or order parameters are employed to identify the transition state region. In the following sections, we first introduce the method, then discuss in detail the folding mechanism of BdpA revealed by comparing our simulations with previous experimental and theoretical studies, emphasizing the model of the TSE and the sequence of events that lead to TSE, which shed light on specific amino acid residues/sites that may have dictated the folding of BdpA and the structural characteristics of its TSE.

Method

Sample Folding Trajectories That Use Distributed Computing Guided by Topological Similarity to the Native Fold.

Processes such as protein folding are rare events at a molecular time scale, which means that an actual folding trajectory may span a time scale sufficiently short to be covered by a single molecular dynamics trajectory. Such trajectories are, however, rare in the sense that if we use an initial state randomly drawn from an ensemble of unfolded state, the chance that the trajectory will end up folding within the simulation time is very small.

Distributed computing can be employed to handle this problem. Projects such as fold@home²⁵ make use of the computing power shared by hundreds of thousands of desktop computers around the world to simulate a vast number of trajectories for a single protein, from which folding trajectories can be extracted and studied in detail. In such context, synchronization between different desktop computers is neither feasible nor required. A slightly more sophisticated but much more efficient approach is possible when synchronization can be easily realized. One strategy is to start M simulations from the same local conformational state in parallel, and once one of the simulations escapes from the local state (i.e., a synchronization criterion is satisfied), re-synchronization of the simulations is carried out by stopping the remaining $M - 1$ simulations and restarting another $M - 1$ from the new local state produced

by the leading trajectory. It has been proven that if the transitions between local states are Markovian, such a strategy accelerates the diffusive kinetics between different local states by a factor of M .²⁶ The Markovian requirement is easily fulfilled as long as the definition of local states is sufficiently coarse-grained compared with the level at which the dynamics is described.

Different criteria have been employed to set the synchronization points, including changes of energy in parallel replica dynamics²⁷ and changes of heat capacity in ensemble dynamics.²⁸ We consider the following factors in choosing our criteria. First it has to be sufficiently coarse-grained so that any chain of synchronization events is indeed Markovian. Second, given that the accuracy of the current atomic force field is still insufficient for ab initio folding, the synchronization condition should bias the sampling toward identifying folding trajectories. We note that in the case where we have an ideal potential energy model, that is, the native fold is its global free energy minimum, we may not need to consider the second factor. By considering the second factor we may be able to sample straightforward folding trajectories in a very computationally efficient manner, but at the price of losing the capability of estimating the folding rate directly from the probability distributions.

In our work, we define the following parameter that we call the index of topological similarity (I_{tops}), which measures the similarity in topology between the native and the current fold:

$$I_{\text{tops}} = \frac{1}{2} \sum_{i \neq j} \theta_{ij} S_{ij}$$

$$\theta_{ij} = \begin{cases} 0, & \text{if } i \text{ and } j \text{ do not contact in the current fold, otherwise} \\ 1, & \text{if } i \text{ and } j \text{ also form a contact in the native fold, and} \\ -1, & \text{if } i \text{ and } j \text{ do not form a contact in the native fold} \end{cases}$$
(1)

S_{ij} is the sequence separation (contact order) between residues i and j , i.e., $|i - j|$. Two residues with a sequence separation of at least 3 are treated as in contact when the distance between their C α atoms is shorter than a cutoff distance (7 Å in this work).

The synchronization condition in our distributed computing is the increase in I_{tops} . That is, if any of the M simulations, all starting from the same conformation but with different random initial velocities, samples a conformation with a higher I_{tops} than the starting conformation, the other $M - 1$ simulations are stopped and then restart from this conformation. This process is continued until the native fold is reached, or none of the simulations could result in any increase in I_{tops} within a given length of simulation time. For those processes ending in the native folds, one of the M simulations between each pair of consecutive synchronization points contributes a segment to a complete and continuous folding trajectory. As the bias is only applied to filtering the trajectories at sparse time points but not to restraining the dynamics, the resulting folding trajectory is drawn from an unperturbed physics-based potential energy surface.

Compared with the commonly used Q value (the fraction of formed native contacts), our choice of the topological similarity as an order parameter to measure the degree of folding is based on an important earlier concept of the relative contact order and experimental observations on its correlations with the folding rate.²⁹ They implied that bottlenecks for folding must involve formations of higher order native contacts. Consistent with this implication, trajectories resulting in higher order native

contacts should be spawned and explored further, which is exactly the reasoning behind our choice of the order parameter. Another important concept is the effective contact order, which measures the “effective” separation in sequence between long-range contacts to be formed after more local contacts have been formed.³⁰ Both the relative contact order and effective contact order are not order parameters and cannot be used directly for our purpose. An additional ingredient in our definition is the inclusion of negative contributions from non-native contacts: non-native contacts in the current structure make the current structure less similar to the native.

We have also tested a number of alternative criteria other than I_{tops} , including the number of native contacts formed (Q value) and the root-mean-square deviation from the targeted native structure. They did not result in the sampling of successful folding trajectories. A brief comparison of different criteria will be given in the Results and Discussion section.

We emphasize that the above pathway-sampling scheme differs fundamentally from the “steered” simulation approaches. There is no ad hoc “steering” potential added on the top of the physical potential. The synchronization criterion affects only the selection of trajectory branches at a few time points. And the system is not forced to satisfy an earlier synchronization condition in subsequent simulations.

To minimize the perturbation introduced by the trajectory-branch selection procedure, the synchronization condition is only checked after the new simulations have produced conformations significantly deviated from each starting conformation. So if a nativelike feature has not been persistently stable already at a particular folding stage, it would disappear in the succeeding rounds of simulations. Thus although the selection process does introduce some unquantified bias, we expect the bias to be very weak, as demonstrated by the overall downhill potential energy changes along the sampled trajectories (see below).

Potential Energy Function and Simulation Method. We used the GROMOS96 force field for the protein,³¹ and an implicit generalized Born/solvent accessible surface area (GB/SA) model³² to describe the solvation. The GB/SA model had been parametrized with the GROMOS96 force field to reproduce molecular dynamics behaviors of globular proteins in an explicit solvent.³³ Implicit solvent models such as GB/SA treat equilibrium solvation in an efficient manner, and seem to be able to reproduce results of more computationally demanding explicit solvent simulations in various contexts. For folding problems, for example, recently Jorgensen and co-workers have successfully reproduced the native folds of a number of peptides based on a GB/SA model.³⁴

To further accelerate large-scale motions in the folding process, we used the amplified-collective-motion (ACM) method in individual simulations.³⁵ Details of the method have been given elsewhere. In summary, the method couples a few lowest frequency collective coordinates to a higher temperature bath with the remaining coordinates coupled to an ambient temperature bath. The lower frequency subspace is determined locally on the conformational space with use of an coarse-grained anisotropic elastic network (ANM) model,³⁶ and updated frequently as the conformation evolves. It has been demonstrated that the method could amplify interdomain motions in lysozyme in an explicit solvent with little influence on the intradomain dynamics, as well as accelerate the refolding of a small helical peptide in an implicit solvent.³⁵ The use of ACM is meant to increase the sampling efficiency of individual simulations. An alternative is to use standard molecular dynamics and increase the allowed duration of individual simulations. This is in

principle possible, but a higher failure ratio is expected if this duration is not long enough, as ACM introduces higher frequencies to escape local conformational traps than standard MD. The use of ACM is reasonable for the exploration of folding mechanisms as only a few low-frequency modes are coupled to a higher temperature, and the overall temperature of our system always stays close to ambient and far below the melting temperature of this protein.

Computational Details. The native structure is the NMR averaged structure of BdpA (pdb code: 1BDD).³⁷ The C- and N-terminal loops were removed and residues 10–55 were kept. After energy minimization with the steepest descent method and a 50 ps equilibration simulation with positions of all non-hydrogen atoms restrained, a 5 ns molecular dynamic simulation was performed to sample an ensemble of structures. The ensemble quickly equilibrated around 2 Å of root-mean-square deviations from the NMR structure. One of the equilibrated conformations was used to define the native fold. Secondary structure and native contacts were calculated according to the native fold. There was no essential difference in the definitions of native contacts and native secondary structures if they were defined by using averages over the stabilized section of the simulation.

The starting conformation for the folding simulations was a fully extended conformation, constructed with use of Tinker.³⁸ The simulations were performed distributively on eight processors, each starting a simulation with initial velocities randomly drawn from a Maxwell–Boltzmann distribution after receiving a starting conformation from a server processor. Each simulation lasted at most for 1 ns and after that a new simulation restarted on the same processor with the same starting conformation but with a new set of velocities. A simulation ended earlier if the given synchronization condition was met (to fulfill the Markovian requirement, the condition was only checked after a simulation had deviated significantly (2 Å in root-mean-square deviation) from its starting conformation so that higher frequency fluctuations in the synchronization parameter were filtered out). In this case the processor communicated the fulfilled conformation to the server, which synchronized all processors to restart M simulations from the new conformation. Thus M , the number of simulations branched from each synchronization point, was not fixed. The simulations were ended either when the native fold had been reached, or each processor had repeatedly performed 1 ns simulation 2–3 times (so M was around 20) with none of them having sampled a conformation satisfying the synchronization condition. The individual simulations employed the amplified-collective-motion scheme, in which four lowest frequency modes had been coupled to 900 K and the other modes stayed at 300 K, with temperature relaxation times of 0.01 and 0.1 ps, respectively. The short and long cutoffs in the anisotropic network model (ANM)³⁶ were 0.7 and 1.4 nm, respectively. The ANM modes were updated every 50 time steps (100 fs). The short coupling time for the slow modes was required because these modes were delocalized and their energies dissipated rapidly. As the number of higher temperature modes was so few, the temperature of the overall system was always close to that of the ambient temperature bath.

Eight groups of simulations were performed, each starting from the extended conformation and employing eight processors to carry out the distributed computing as described above. The trajectories reported in the results were recorded every 10 ps and composed from segments generated from different processors. The computations were performed with GROMOS96 programs,³¹ with in-house modifications to include the GB/SA

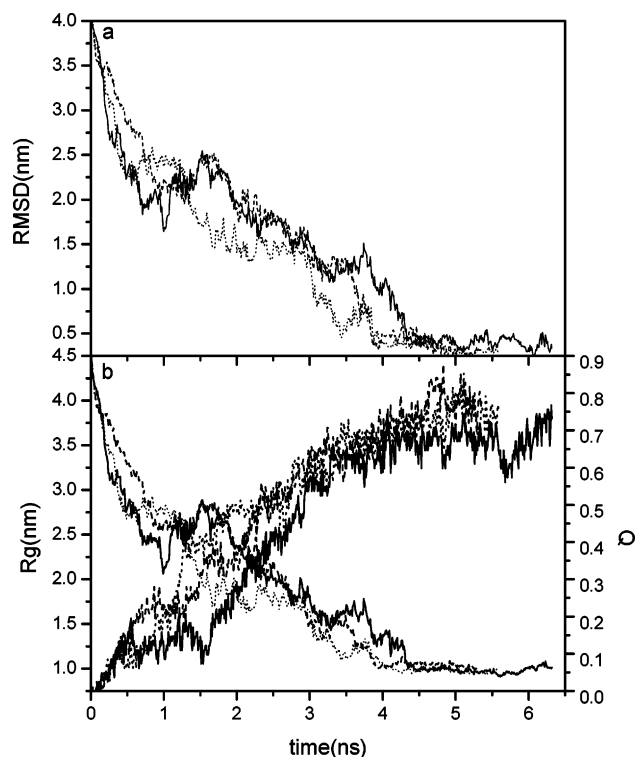


Figure 1. (a) Root-mean-square deviations of C α positions from the native fold and (b) the radius of gyration and Q values as functions of simulation time in the three sampled folding trajectories.

model, the amplified-collective-motion method, and the distributed computing. Some of the analyses were performed with tools contained in GROMACS³⁹ and MOLMOL.⁴⁰

Results and Discussion

Folding Trajectories. Among eight groups of the simulations, three have successfully produced continuous folding trajectories ending with the native fold. The highest Q values for these three paths are 0.85, 0.87, and 0.77, respectively, and their final all-C α -atom root-mean-square deviations from the native fold generated by an equilibrium simulation (see Computational Details below) are all below 3.0 Å, with the lowest being 2.3 Å (Figure 1), and from the averaged NMR model below 3.5 Å, with the lowest being 2.5 Å.

The remaining five groups were unable to produce successful folding trajectories. By definition, these nonfolding groups ended in conformations from which around 20 new simulations were spawned and none of them produced conformations closer to the native fold judged by the given similarity measure within a limited period of 1 ns each. These groups of simulations have been aborted at rather earlier stages of folding: conformations at their respective last synchronization points all have a radius of gyration larger than 1.25 nm, with little formation of any nonlocal native contacts. The existence of these aborted groups of simulations reflects partially the enormous conformational space that a protein polypeptide chain has to explore before folding correctly, partially the incompleteness of the exploration in the conformational space by the finite number and limited duration of spawned simulations, and partially the inaccuracy of the potential energy model.

Figure 1a shows root-mean-square deviations from the native fold, and Figure 1b shows the radius of gyration (R_g) and Q values (number of native contacts in a conformation divided by the total number of contacts in the native fold), as functions of time for the three folding trajectories. Although different from

each other in many details, the three folding trajectories share some common global features, which may be the keys to the folding mechanism.

The initial coiling of the extended chain results in decreases of R_g from 4.5 to 2.5–3.0 nm. The Q values after this initial coiling are below 0.3 in all three trajectories. This initial collapse is because of the use of a fully extended starting structure. Being of little relevance to the “real” physics of folding, it will not be discussed later.

Subsequently, the peptide chain condenses further and there is a plateau with R_g around 1.75 nm. We will focus our discussions on the trajectory sections from R_g around 1.75 nm and down to 1.0 nm. Between, there are two additional plateaus separated by rapid collapses: one with R_g around 1.5 nm and the other with R_g around 1.25 nm.

The root-mean-square deviations from the native fold behave quite similarly to R_g : the switching between the rapid decreases and the plateau phases in the root-mean-square deviations are clearly synchronized with the changes in R_g .

The separations between the collapsing phases and the plateau phases are unclear in the Q values versus time plots probably because at larger Q values the fluctuations in short-range contacts blur the increasing trend in Q . However, it can still be roughly spoken that configurations at plateaus with R_g around 1.75, 1.5, and 1.25 nm mostly correspond to Q values of about 0.45–0.55, 0.55–0.65, and 0.65–0.7, respectively. Because of this and also because of the lack of a better reaction coordinate or order parameter than the Q value for the folding process, we will still use the Q value as a coordinate to unite the three folding trajectories in part of the discussions below, despite the fluctuations in Q values. This is done by dividing the Q value into bins of width of 0.05, and for each bin, configurations are collected from the three folding trajectories and analyzed.

Changes in the Overall Energies, Hydrophobic Surface Areas, and Secondary Structures as Functions of Q . Figure 2a indicates that, coarse-grainedly, the effective energy which is the sum of the intramolecular energy of the peptide chain and the solvation free energy computed by the GB/SA model decreases monotonically with the increase in Q . This has two implications:

First, it justifies the physical relevance of the folding pathways identified by using our approach. That is, unlike other guided sampling approaches to simulate pathways such as the targeted⁴¹ or steered molecular dynamics simulations,⁴² the pathway sampling method employed here does not force the protein to climb over potential energy barriers. Rather, through repeatedly exploring different branches of the trajectories stemming from the same conformation and selecting the folding branch with the given synchronization condition, physically plausible trajectories, i.e., trajectories toward the native fold along valleys on the potential energy surface, have been selectively sampled.

Second, this monotonically downhill energy profile is in agreement with the observed two-state folding mechanism of BdpA.^{5,6} The increase in Q is accompanied by the decrease in conformational entropy of the peptide chain. This loss in entropy cannot be fully compensated for by the decrease in the effective energy until the transition state is reached, resulting in only a single barrier along the folding pathway at a coarse-grained level.

The total effective energy was broken into components (see Figure 2a, plots in the subwindow). We observe decreases in the intramolecular interaction energy are partly compensated for by increases in the solvation free energy. This mutual compensation between intramolecular and solvation contribu-

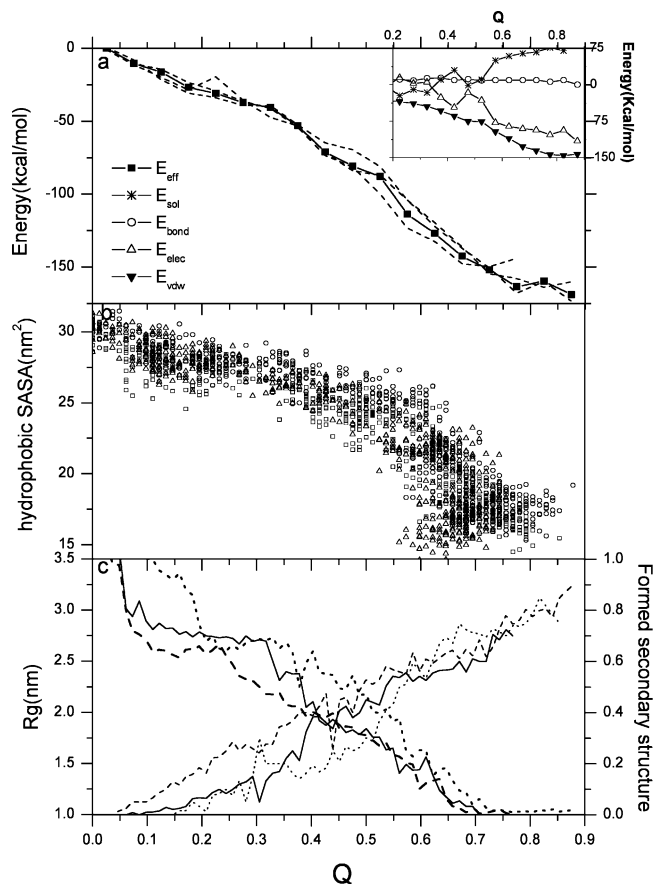


Figure 2. (a) Effective energies averaged over conformations within each Q bin. If not specified, the three folding trajectories have been combined in the analyses. E_{eff} : the total effective energy, which is the sum of the total solute energy and the solvation free energy. E_{sol} : solvation free energy. E_{bond} : bonded energy of the solute. E_{elec} : electrostatic energy of the solute. E_{vdw} : van der Waals energy of the solute. Dashed lines around E_{eff} are total effective energies from individual trajectories. (b) The total hydrophobic solvent accessible surface areas (SASA) for conformations sampled along the three different trajectories (squares, circles, and triangles represent different folding trajectories) and (c) the radius of gyration and the portion of formed secondary structures averaged over conformations within each Q bin (solid, dashed, and dotted lines represent different folding trajectories).

tions has been observed before.⁴³ The increase in the solvation free energy mainly comes from the desolvation of the polar protein groups, as they become intramolecular interaction partners. Thus we observe a strong anticorrelation between the steep decrease in the intramolecular electrostatic energy and the sharp increase in the solvation free energy for Q bins between 0.45 and 0.55. Starting from the Q bin between 0.55 and 0.6, however, the increasing trend in the solvation free energy is flattened because of the starting of the hydrophobic collapse process (see below). Inconsistent with the latter, the intramolecular van der Waals interaction energies decrease more rapidly.

Panels b and c in Figure 2 show the total hydrophobic solvent accessible surface area and the radius of gyration of the protein domain as functions of Q , respectively. The hydrophobic solvent accessible surface area tends to decrease monotonically with the increase in Q (Figure 2b). Interestingly, the relation is piecewise linear, with the first slower decreasing phase turning into the second more rapid decreasing phase at the Q value of about 0.55. This turning point corresponds to where R_g starts decreasing from around 1.75 nm in the folding trajectories

(Figure 2c, see also Figure 1b). Thus both the decrease in R_g and the increased reduction in the total hydrophobic surface area with Q are indications of the starting of the same hydrophobic collapse process. As a result, the increase in the solvation free energy with Q slows down at the Q bin between 0.55 and 0.6 (see above and Figure 2a). There the favorable contributions from burying the hydrophobic surface areas start to compensate for the unfavorable desolvation contributions of polar groups accompanying the increase in Q . At the $R_g \sim 1.25$ nm plateaus in Figure 1b, most conformations have Q values of around 0.65 to 0.7 (Figure 2c). There about two-thirds of the hydrophobic collapse process has been finished if measured by the total hydrophobic surface area to be buried: the hydrophobic surface area is ca. 24 nm² when the collapse starts, the final area is around 16 nm², and at a Q value of around 0.65–0.7 the surface area is around 19 nm². If the comparison is made against the fully extended state (hydrophobic areas of about 30 nm²), about 80% of the surface-burying process has been finished at the $R_g \sim 1.25$ nm plateaus.

In Figure 2c the portion of formed secondary structures has been plotted as a function of Q for each of the three trajectories (a value of 1.0 corresponds to a complete formation of the native secondary structures). The secondary structure contents were calculated as the number of formed helical hydrogen bonds (H-bonds) versus the total number of helical H-bonds in the three helices (the criterion for the helical H-bond is that the distance between the carbonyl oxygen of residue i and the amide hydrogen of residue $i + 4$ is less than 2.5 Å, and the angle between the corresponding C=O and H–N vectors is below 60°). In all three folding trajectories the overall secondary structure contents increase almost linearly with Q . At $Q = 0.55$, which mainly corresponds to conformations before the collapse from R_g around 1.75 nm, around 50% to 60% of overall secondary structures have been formed. At the $R_g \sim 1.25$ nm plateaus, this ratio is increased by about 10%.

From the above overall features of the folding trajectories, our simulations display a hybridization (or cooperation) of the framework and nucleation–condensation models in BdpA folding. The secondary structures are partially formed before significant condensation of the protein in the folding trajectories. Further formation of the secondary structures is at least not necessary for the starting of the hydrophobic collapse process. The hydrophobic collapse is, however, necessarily accompanied by the further increasing in secondary structure contents. In other words, the hydrophobic collapse and the increase in regular secondary structure contents from around 55% to around 65% represent highly cooperative steps in folding.

Sequence of Events during the Folding Process. Figure 3 shows the portions of formed helices within each of the individual helices as functions of time in the three folding trajectories. It indicates that there are some variations in the formations of helices among the three trajectories. We note that these curves are influenced by secondary structure fluctuations near the exact criteria employed to compute the helical contents. Despite these, it is quite clear that H2 is always the best-formed secondary structure element and its helical content is well maintained before and throughout the collapsing phases. A significant portion of H1 has also been formed before and well maintained during the collapsing phases. There are, however, large fluctuations in the computed helical content for H1 after collapsing. H3 is the worst formed among the three helices both before and during the collapsing phases.

To identify the shared sequence of events by the three folding trajectories, we segmented each trajectory based on the R_g

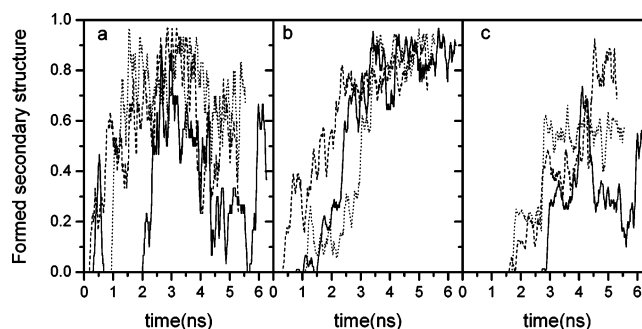


Figure 3. Portions of formed secondary structures within each of the three helices as functions of time for each of the three folding trajectories. Plots a–c correspond to helices 1 to 3, respectively. A value of 1.0 represents a complete formation of the corresponding helix segment.

versus time plots, and grouped corresponding segments (with similar R_g) from the three trajectories. This produced five ensembles of conformations, identified by the corresponding approximate R_g values: the $R_g \sim 2.75$ nm ensemble (combination of 0.49 to 1.28 ns, 0.85 to 1.90 ns, and 1.10 to 1.91 ns segments of the three folding trajectories, respectively), $R_g \sim 1.75$ nm ensemble (combination of the 1.65 to 3.00 ns, 2.75 to 2.94 ns, and 2.77 to 3.01 ns segments), $R_g \sim 1.50$ nm ensemble (combination of the 3.10 to 3.31 ns, 3.31 to 3.60 ns, and 3.40 to 3.77 ns segments), $R_g \sim 1.25$ nm ensemble (combination of the 3.58 to 3.84 ns, 3.64 to 3.80 ns, and 3.86 to 4.31 ns segments), and $R_g \sim 1.0$ nm ensemble (combination of the 4.27 to 5.60 ns, 3.90 to 5.35 ns, and 4.32 to 6.33 ns segments). In Figures 4a,b variations of each backbone dihedral angle within three of the structure ensembles at different R_g plateaus are compared. These give a more detailed picture of how the helical structures have evolved. Among the $R_g \sim 1.75$ nm ensemble, there are already observable sections of the peptide chain with obviously higher backbone rigidity than the rest of the chain. These correspond to the middle sections (closer to the N-terminals) of H1 and H2. Besides these regions, the middle section of H3 is also somewhat slightly less flexible than the remaining chain, but does not really show any formation of helix.

One most significant change in backbone torsion angle variations from the $R_g \sim 1.75$ nm ensemble to the $R_g \sim 1.5$ nm ensemble is the rigidification of the backbone from residue 29 to 32. The helical section between residue 29 and 33 within H2 is consistently formed and stabilized within the $R_g \sim 1.5$ nm ensemble but not within the $R_g \sim 1.75$ nm ensemble. Another change in backbone rigidity between the two ensembles is that the backbone torsion angles of the last two residues on H1 show higher variations within the $R_g \sim 1.5$ nm ensemble (probably to accommodate the docking of H1 onto H2).

The most significant change from the $R_g \sim 1.5$ nm ensemble to the $R_g \sim 1.25$ nm ensemble shown in Figure 4a,b is that the rigidified backbone sections within H2 extend further to the C terminal. In the mean time, some backbone torsion angles of H3 are also rigidified.

The averaged structures of the various ensembles are shown in Figure 5. We note that because of the deformation generated by averaging rather different structures, this figure does not provide as detailed information about the local structures as Figure 4a,b. Despite this, the average of the $R_g \sim 1.75$ nm ensemble still shows consistent helical conformations within H1 and H2. And the average over the $R_g \sim 1.5$ nm ensemble shows significant lengthening in H2 compared with the $R_g \sim 1.75$ nm ensemble, while the average over the $R_g \sim 1.25$ nm

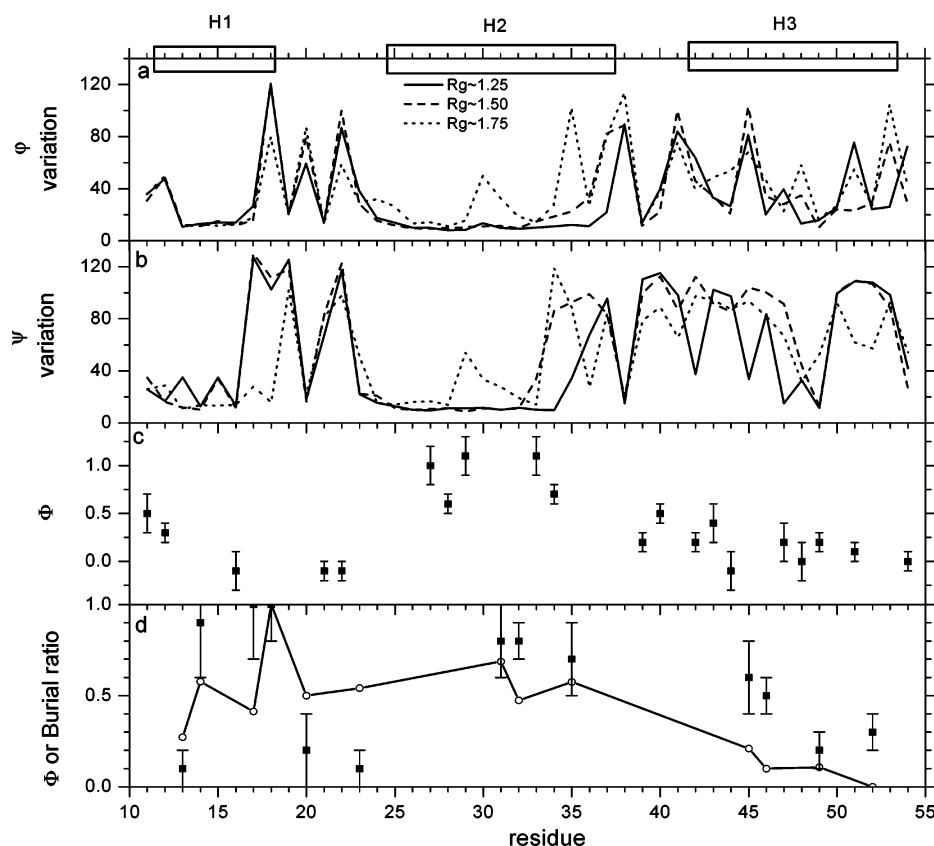


Figure 4. (a) Peptide backbone ϕ and (b) ψ dihedral angle variations of each residue within different structure ensembles at different R_g . (c) Experimental Φ values for secondary structure reported by Sato et al.¹¹ (d) Experimental Φ values for tertiary structure reported by Sato et al., and the relative burial ratios at the transition state ensemble of these residues with experimentally measured Φ values. The relative burial ratio for a residue is computed as $(S_T - S_D)/(S_F - S_D)$, where S stands for the averaged hydrophobic surface area of a residue. The subscripts refer to the ensemble for averaging: T for TSE, modeled by the $R_g \sim 1.25$ nm ensemble; D for denatured, modeled by combinations of segments 0.49 to 3.31 ns, 0.85 to 3.60 ns, 1.10 to 3.77 ns of the three folding trajectories; and F for folded, modeled by the 5 ns simulation of the native structure.

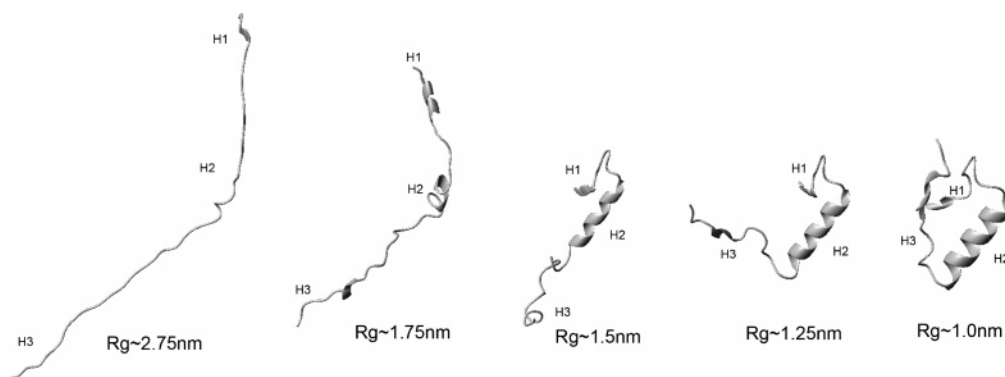


Figure 5. Structures averaged over different R_g ensembles.

ensemble shows further significant lengthening in H2 toward the C terminal, as well as some helical formation of the middle section of H3.

Figure 5 completes the picture more by providing changes in the global structure of the protein domain. Within the $R_g \sim 1.75$ nm ensemble, the three helices are well separated. Within the $R_g \sim 1.5$ nm ensemble, H1 has been docked onto H2, with H3 still extending away from the rest of the chain. Within the $R_g \sim 1.25$ nm ensemble, T2 connecting H2 and H3 starts taking shape. It, however, takes another collapse in R_g from around 1.25 nm to 1.0 nm for H3 to be packed with H2 and H1.

It is interesting to correlate the above cooperativity between the collapse in R_g and the rigidification of secondary structures with the amino acid sequence of the protein. The flexibility of the helix turn from R28 to I32 may be related to the presence

of G30. Our results indicate that without nonlocal contact between H1 and H2, this particular helix turn is not stable. After the collapse, the nonpolar part of R28 on H2 forms contacts with I17 and L18 on H1, and F31 and I32 on H2 form contacts with F14 on H1. As the separations in sequence between the respective residues on each helix are both one helix turn, these contacts will stabilize the helical conformations of both the segment 14–18 on H1 and the segment 28–32 on H2. In return, the contacts freeze the otherwise fluctuating segments within H2 into the helical conformation. On the basis of the well-known cooperativity in the helical formation (i.e., the helical extension is easier than its initialization), it is reasonable to argue that this extension and stabilization of H2 from R28 to I32 is a prerequisite for the next phase of folding, which is another demonstration of cooperativeness: both the extension of H2

into a full length at its C-terminal and the taking shape of T2 take place within a single burst step (the collapse of Rg from 1.5 to 1.25 nm) in the three folding trajectories.

TSE for Folding: Comparisons with Experiments and Earlier Simulations. Because of the prementioned limitation of our approach, the aim of our simulations does not include locating free energy barriers along the path. From the simulation alone, we cannot identify which of the structural ensembles along the folding trajectories corresponds to the TSE. We note that for lattice models a pattern recognition algorithm has been developed to extract folding nuclei from folding trajectories.^{44,45} The method is not applicable in our case because we do not have a large enough number of trajectories for pattern recognition.

Comparisons between the three independently sampled folding trajectories and experimental results indicate that the Rg \sim 1.25 nm ensemble may be the most likely candidate for a TSE. In all three trajectories, it collapses into the native fold without any further intermediate plateau structures. It possesses most of the characteristics of the TSE mapped by experiments (see below). Within this ensemble, the overall helical content is between 60% and 70%, comparing well with the estimated upper limit.²⁴ About 80% of the total buried hydrophobic surface area has been buried, also in agreement with the experimental estimation.⁶

The pathways sampled in our simulations also show the importance of contacts between H1 and H2. This is in agreement with the works of Brooks and colleagues¹⁶ and of Garcia and Onuchic.¹⁷ Our results, however, generate some unique features, which reconstruct into a coherent picture of the folding process. First, structural ordering of the T1 region seems to play an unimportant role for the formation of TSE. Rather, the formations of contacts between H1 and H2 are mainly driven by hydrophobic forces between residues 14, 17, 18 on H1 and residues 28, 31, 32 on H2. For this to take place, some increases in disorder of the C-terminal of H1 and the T1 region are necessary. Second, these residues are separated exactly by one helix turn on the respective helices. Before the collapse, the turn on H1 is reasonably well formed, but the other on H2 is not so well structured. Instead, H2 is well formed from the N-terminal only up to residue 28. Thus the collapse essentially freezes residues 28–32 on H2 into the helical conformation. Third, this may facilitate the further extension of H2, the formation of T2, and the initiation of the docking of H3 onto the well-formed H2.

In Figure 4c, the experimentally measured Φ values for secondary structure reported by Sato et al.¹¹ are shown. These Φ values have been measured for 20 sites not involved in nonlocal contacts in the native fold for a specific type of mutation, that is, the Ala to Gly mutations. The experimental values have been reported for two concentrations of detergent and the larger values are plotted. Qualitatively these values should indicate the rigidity of backbone conformation at the respective sites of the TSE relative to the folded state if one assumes different sites have the same rigidity in the unfolded state. These measured Φ values correlate well with the backbone torsion angle variations in the Rg \sim 1.25 nm ensemble in Figure 4, panels a and b: residues with higher measured Φ values for secondary structure possess more rigid backbone conformations within this ensemble.

Figure 4d shows experimental Φ values for tertiary structure reported by Sato et al. For 13 sites forming nonlocal hydrophobic contacts the Φ values have been measured. To compare qualitatively with these values we define the side chain burial

ratio for each residue as the ratio of the averaged buried area within the Rg \sim 1.25 nm structure ensemble over the total buried area in the native fold. As we lack a model for the actual unfolded state, the buried areas are computed as relative to the extended structure. Although the burial ratio may only partially capture the properties of TSE reflected by the experimental Φ value, the data still show significant correlations between the burial ratios and the Φ values. Except for A13 on H1 and L20 and L23 on T1, the experimentally measured Φ values are higher than the burial ratios. For A13 the computed burial ratio is not so meaningful because of the small size of its side chain. The burial ratios for L20 and L23 are lower than most of the other residues considered. And their higher values relative to the experimental Φ values may indicate an underestimation of their buried surface areas for the unfolded state in our model (because we started from an extended chain to reach structures which were used to model the unfolded state in computing relative burial ratios) rather than an overestimation of their importance at the transition state in our model. Overall, the Pearson linear correlation coefficient between the burial ratio and the experimental Φ value is 0.55, and when the above three residues are not considered, the correlation coefficient amounts to 0.82.

A notable fact is the consistently larger Φ values of residues F14, I17, and L18 on H1 and F31 and I32 on H2. These are in agreement with the above coherent picture about the sequence of events during folding. The Φ value for residue R28 is not reported by experiments, but our simulations indicate that the nonpolar parts of its long side chain may play an important role in the docking of H1 onto H2 and rigidifying the helix turn from R28 to I32. Another experimental indication on the importance of the formation of this particular helix turn at the TSE is the recent report that the F14W G30A mutant was the fastest folding protein of its size studied to date.⁹

Some overall features of folding pathways sampled in our simulations and estimated TSE, such as the extensive formation of H2 and the docking of H1 onto H2 before the docking of H3, are in agreement with the all atomic umbrella sampling simulations of Brooks and colleagues and replica-exchange molecular dynamics simulations of Garcia and Onuchic. Both of these recent studies have been based on obtaining free energy surfaces projected onto certain reaction coordinates. Methodologically our approach complements free energy surface approaches, and for BdpA, our results also produce some unique features about its folding mechanism, which reconstruct into a coherent picture of the folding process. First, structural ordering of the T1 region seems to play an unimportant role for the formation of TSE. Intrinsically, the helix turn from R28 to I32 on H2 is not stable, probably because of the presence of G30. The formations of contacts between H1 and H2 are mainly driven by hydrophobic forces between residues 14, 17, 18 on H1 and residues 28, 31, 32 on H2. These residues are arranged in the primary structure in such a way that their collapse onto each other can provide the necessary stabilization effects on H2. This feature of the folding mechanism is highly sequence specific.

We note that the three independently sampled trajectories behave similarly in terms of the above qualitative picture. Higher numbers of folding trajectories may provide better statistics but it seems that they are unlikely to change this overall picture.

Effects of Different Synchronization Conditions. Not being the major focus of the current report, this is only discussed briefly below. Besides the increase in I_{top} , we have tested synchronization conditions such as the decrease in the root-

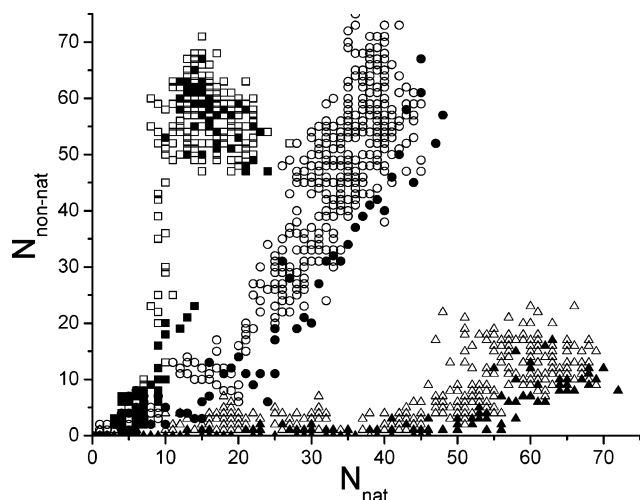


Figure 6. Numbers of non-native contacts ($N_{\text{non-nat}}$) versus numbers of native contacts (N_{nat}) in conformations sampled with different synchronization criteria: root-mean-square deviations from the native fold (squares), Q values (circles), and I_{tops} (triangles). Solid symbols represent conformations that fulfilled the synchronization condition in the corresponding simulations.

mean-square deviation from the native fold and the increase in Q . Only with I_{tops} successful folding trajectories have been sampled.

It seems that the main reason for the failures of using root-mean-square deviation and Q can be linked to the limited frustration of the energy landscape. In Figure 6, the conformations sampled by using different synchronization conditions are plotted as a scattering graph of the number of non-native contacts versus the number of native contacts plot. With I_{tops} as the synchronization condition, the number of non-native contacts is always small, as expected because trajectories branching into conformations with a large number of non-native contacts are not selected. Interestingly, with root-mean-square deviations as the synchronization condition, the sampled trajectories mainly consist of conformations that have more (approximately 3.3 times estimated by the slope) non-native contacts than native contacts. The trajectories generated by using the Q value as the synchronization condition also contain conformations with more (around 1.4 times) non-native contacts than native contacts. It further supports our argument that restrained sampling along Q may not be able to capture the true features of the folding events and the transition state ensemble.

Figure 6 also reflects that our definition of the similarity measure strongly penalizes non-native contacts. Roles of non-native contacts during protein folding processes are in general not well understood. Figure 6 indicates that at least for BdpA most of the non-native contacts may not direct the folding toward the native structure. For proteins with more complicated folding landscapes, which necessitate the temporary formation of non-native contacts during folding, whether the I_{tops} introduced here will be appropriate still remains to be investigated. There can be cases where the problem will not be severe as long as the mediate non-native contacts do not persist during folding, because between the synchronization points the sampling is performed without restraints, thus non-native contacts can be formed and then broken if such events are necessary for the formation of the downstream starting conformation. Arguably, there may also be cases in which there are non-native contacts that are necessarily formed and persist long enough along a folding pathway. Then our definition may not lead to

productive samplings and one may still need to look for better choices for the order parameter to resolve the issue.

Conclusions

We successfully applied distributed computing to simulate the folding of BdpA, a small protein domain well studied by recent experiments. The folding trajectories were sampled on an unperturbed physics-based atomic potential energy surface (with an implicit solvent) and they continuously linked the extended states to the native folded state. We emphasize that our analyses should not be viewed as trying to reconstruct a single “folding pathway”, as there seems to be no known consensus among researchers about the level of coarse graining at which protein folding can be discussed in terms of pathways. Instead, we identified common features of independently sampled folding trajectories, and compared them with experimental data when available. Especially, although we only produced three independent folding trajectories, in all of them, the ensemble of structures right before the final collapse into the native fold shared important common features agreeing well with the experimental mapping of the folding TSE through Φ value analysis. Some of them have not been reproduced in previous theoretical studies. Although some of these earlier disagreements might originate from force field models, our results indicated that some of them might stem from locating and sampling a “transition state ensemble” based on lower dimensional free energy surfaces. Besides this, our simulation results pointed to features of specific sequences (amino acid residues at specific sites) that might be the major cause for the cooperative hydrophobic collapse and formation of secondary structure. It should be possible to test these indications by future experiments.

We note that the approach employed here has the limitation that it does not directly quantify the free energy profiles associated with folding. Besides this, there is another important concern whether the trajectory selection process can bias the sampling so much that artifacts are produced in the sampled trajectories. Several reasons lead us to conclude that this is not the case. First, the bias is very gentle, and it is introduced at very few time points. So it cannot drive the system over energy barriers too high to be crossed at ambient temperature. This is reflected by the downhill potential energy curves of the folding trajectories, as well as by the incapability of producing folding trajectories by applying inappropriate criteria for trajectory selection (such as root-mean-square deviations or the Q value). In the latter cases the system is guided into frustrated regions of the energy landscape and the biasing factor does not force it to fold (otherwise arbitrary folding trajectories would have been sampled). Second, between points of selection, the trajectories evolve free of restraints on a physics-based energy surface and the next check for selection is carried out only after the trajectories have diverged significantly from its starting point. Thus, nativelike features that might have contributed to the previous selections should disappear if they are not persistently stable at certain stages of folding. The trajectory would then have to develop other nativelike features to be selected in the next round. Last but not least, we have shown the close agreement between the sampled trajectories and experimental data. If the selection process has produced arbitrary bias, there is no reason that such bias would have been able to lead to the type of agreements shown here in three independently sampled trajectories.

The approach presented here may be employed to dissect key events and their orders in large-scale conformational changes

when a targeted conformation is known. This may provide complements to free energy profile analyses, which suffer greatly from the collapse of a multidimensional surface into one or two heuristically chosen reaction coordinates. One future possibility is to use pathways sampled by the type of approaches employed here to differentiate the true TSE from those structure ensembles that happen to coincide with the “transition state” when projected onto the few reaction coordinates.

Acknowledgment. We thank Professor W. F. van Gunsteren for the GROMOS96 package. This work has been supported by the Chinese National Natural Science Foundation (Grant No. 90103032 and Grant No. 90403120) and by the Chinese Department of Science and Technology (Grant No. 2004-AA235110).

References and Notes

- (1) Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1982**, *51*, 459.
- (2) Daggett, V.; Fersht, A. R. *Trends Biochem. Sci.* **2003**, *28*, 18.
- (3) Guo, Z.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83.
- (4) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545.
- (5) Vu, D. M.; Myers, J. K.; Oas, T. G.; Dyer, R. B. *Biochemistry* **2004**, *43*, 3582.
- (6) Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001**, *8*, 552.
- (7) Bottomley, S. P.; Popplewell, A. G.; Scawen, M.; Wan, T.; Sutton, B. J.; Gore, M. G. *Protein Eng.* **1994**, *7*, 1463.
- (8) Bai, Y.; Karimi, A.; Dyson, H. J.; Wright, P. E. *Protein Sci.* **1997**, *6*, 1449.
- (9) Dimitriadis, G.; Drysdale, A.; Myers, J. K.; Arora, P.; Radford, S. E.; Oas, T. G.; Smith, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3809.
- (10) Vu, D. M.; Peterson, E. S.; Dyer, R. B. *J. Am. Chem. Soc.* **2004**, *126*, 6546.
- (11) Sato, S.; Religa, T. L.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6952.
- (12) Fersht, A. R.; Sato, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7976.
- (13) Zhou, Y.; Karplus, M. *Nature* **1999**, *401*, 400.
- (14) Linhananta, A.; Zhou, Y. *J. Chem. Phys.* **2002**, *117*, 8983.
- (15) Alonso, D. O. V.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 133.
- (16) Guo, Z.; Brooks, C. L., III; Boczek, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10161.
- (17) Garcia, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898.
- (18) Ghosh, A.; Elber, R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10394.
- (19) Shea, J. E.; Onuchic, J. N.; Brooks, C. L., III. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12512.
- (20) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14812.
- (21) Favrin, G.; Irback, A.; Wallin, S. *Proteins* **2002**, *47*, 99.
- (22) Jang, S.; Kim, E.; Shin, S.; Pak, Y. *J. Am. Chem. Soc.* **2003**, *125*, 14841.
- (23) Berriz, G. F.; Shakhnovich, E. I. *J. Mol. Biol.* **2001**, *310*, 673.
- (24) Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6837.
- (25) Shirts, M. R.; Pande, V. S. *Science* **2000**, *290*, 1903.
- (26) Shirts, M. R.; Pande, V. S. *Phys. Rev. Lett.* **2001**, *86*, 4983.
- (27) Voter, A. F. *Phys. Rev. B* **1998**, *57*, 13985.
- (28) Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; Zagrovic, B. *Biopolymers* **2003**, *68*, 91.
- (29) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985.
- (30) Weikl, T. R.; Dill, K. A. *J. Mol. Biol.* **2003**, *329*, 585.
- (31) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *GROMOS96*; Vdf Hochschulverlag: Zurich, Switzerland, 1996.
- (32) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
- (33) Zhu, J.; Shi, Y.; Liu, H. *J. Phys. Chem. B* **2002**, *106*, 4844.
- (34) Ulmschneider, J. P.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2004**, *126*, 1849.
- (35) Zhang, Z.; Shi, Y.; Liu, H. *Biophys. J.* **2003**, *84*, 3583.
- (36) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505.
- (37) Gouda, H.; Torigoe, H.; Saito, A.; Sato, M.; Arata, Y.; Shimada, I. *Biochemistry* **1992**, *31*, 9665.
- (38) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016.
- (39) van der Spoel, D.; van Drunen, R.; Berendsen, H. J. C. *GROMACS*; Nijenborgh: Groningen, The Netherlands, 1999.
- (40) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graph.* **1996**, *14*, 51.
- (41) Schlitter, J.; Engels, M.; Kruger, P. *J. Mol. Graph.* **1994**, *12*, 84.
- (42) Wriggers, W.; Schulten, K. *Proteins* **1999**, *35*, 262.
- (43) Lazaridis, T.; Karplus, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 139.
- (44) Klimov, D. K.; Thirumalai, D. *J. Mol. Biol.* **1998**, *282*, 471.
- (45) Klimov, D. K.; Thirumalai, D. *Proteins* **2001**, *43*, 465.