PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

All-atom replica exchange molecular simulation of protein BBL

Jian Zhang,[1,2] Wenfei Li,[1] Jun Wang,[1] Meng Qin,[1] and Wei Wang[1,2]*

[1] Department of Physics, National Laboratory of Solid State Microstructure, Institute of Biophysics, Nanjing University, Nanjing, China

[2] State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, Nanjing, China

ABSTRACT

Downhill folding is one of the most important predictions of energy landscape theory. Recently, the Escherichia coli 2-oxoglutarate dehydrogenase PSBD was described as a first example of global downhill folding (Garcia-Mira et al., Science 2002;298:2191), classification that has been later subject of significant controversy. To help resolve this problem, by using intensive all-atom simulation with explicit water model and the replica exchange method, we sample the phase space of protein BBL and depict the free energy landscape. We give an estimate of the free energy barrier height of 1–2 k_BT, dependent on the way the energy landscape is projected. We also study the conformational distribution of the transition region and find that the three helices generally take the similar positions as that in the native states whereas their spatial orientations show large variability. We further detect the inconsistency between different signals by individually fitting the thermal denaturation curves of five structural features using two-state model, which gives a wide spread melting temperature of 19 K. All of these features are consistent with a picture of folding with very low cooperativities. Compared with the experimental data (Sadqi et al., Nature 2006; 442:317), our results indicate that the Naf-BBL (pH5.3) may have an even lower barrier height and cooperativity.

Proteins 2008; 72:1038–1047.

Key words: global downhill folding; protein BBL; all-atom simulation; replica exchange molecular dynamics.

INTRODUCTION

One important trend in protein folding researches is to pursue ultra-fast folders and predict folding speed limit.[1–11] The obvious significance of this effort is that it makes the direct comparison of experiments and simulations possible, which will give much of what one would like to know about the mechanism of folding. Secondly, the vanishing barrier of ultra-fast folders allows the exploration of the molecular timescale of folding, and the roughness of the energy landscape thus provides a good chance to test protein theories.[7,8,12,13] Thirdly, these fast folders are potential candidates of "global downhill" folding, where the free energy barrier separating native and denatured states disappears. Unlike cooperative folding, global downhill folding has only one thermodynamic state that changes gradually from folded-like to unfolded-like as its stability decreases. This picture is also distinct from the original Bryngelson model, where the barrier separating the native and denatured states is eliminated only under extreme conditions where there is a strong energetic bias toward or against the native state; after the energetic bias is reduced the barrier reemerges.[12] The downhill folding allows direct observation of the intermediate structures on the pathway from the unfolded to the folded state. The study of downhill folding will also enrich our understanding of the folding cooperativities and the underlying conformational distributions.[14,15] There are several excellent reviews on the theoretical frames and experimental advances on fast folding,[2–4] on the summary of most fast folders and estimation of folding limit,[4,6] on the downhill folding and its biological significance,[6,8,16] as well as its relevance with folding cooperativities and conformational distributions.[15]

In 2002, the *Escherichia coli* 2-oxoglutarate dehydrogenase PSBD (Naf-BBL, pH5.3) was suggested as a global downhill folder based on the statistical mechanical analysis of experimental results with a variety of techniques.[17] This declaration has elicited great interest as well as debates.[17–25] Later, Ferguson *et al.* examined a slightly longer sequence of BBL under a different solvent condition (QNND-BBL, pH7.0); they found that the thermal denaturation of wild-type unlabelled BBL satisfies standard criteria

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: National Natural Science Foundation of China; Grant numbers: 90403120, 10504012, 10021001, 10704033; Grant sponsor: National Basic Research Program of China; Grant number: 2006CB910302.

*Correspondence to: Wei Wang, Department of Physics, Nanjing University, Nanjing 210093, China.
E-mail: wangwei@nju.edu.cn

Received 5 September 2007; Revised 25 December 2007; Accepted 31 December 2007
Published online 4 March 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22001

for apparent two-state transitions and concluded the folding is cooperative.[18] This study was criticized later by Naganathan *et al.* on the inappropriate choice of wavelength, ignorance of unphysical DSC baselines, and attribution of the inconsistency between different probes to experimental errors.[19] As a response, Ferguson *et al.* reported the raw data of 15 thermal denaturation curves of wild-type BBL and performed a careful error analysis; they showed that the entire dataset could be fitted to common thermodynamic values, consistent with cooperative denaturation.[20] Interestingly, they also presented a chevron plot of BBL H166W mutant and found it has a low slope and the values of m_{D-N} and ΔG_{D-N} determined by T-jump experiment were substantially lower than that measured in equilibrium experiments, in contrast to its two homologs. They attributed this behavior to a multistate, barrier-limited refolding through a populated intermediate, or the variable two-state model proposed by Dill.[26] More recently, Sadqi *et al.* measured the thermal unfolding of 158 backbone and side-chain protons in NMR experiment and reported a broad range of midpoint temperatures spanning about 60 K, which is suggested to be a strong evidence for the global downhill picture of this protein.[22] In short, the controversy has made people reexamine several issues intrinsic to conventional folding experiments, including the effect of the extrinsic fluorophores, the requirement of different probes, the experimental conditions, the baseline subtraction and normalization of the signals, and the estimation of fitting errors. Furthermore, in serial papers,[27–30] Muñoz *et al.* questioned the validity of the conventional chemical models used to interpret experimental data. They suggested that such models may not be appropriate for the ultra-fast folders, since the separation of states implies high free energy barriers thus conflicts with the ultra-fast folding rates. Obviously, the problem of BBL is by no means only a matter of folding behavior of protein BBL, but reflects the conflicts between two different methodologies used to interpret experimental data and two different viewpoints, chemical and physical.[19] It is of increasing importance nowadays because more and more ultra-fast protein folders are discovered.

According to the works mentioned above, it is clear that the determination of the exact folding picture of ultra-fast folders like BBL is very difficult. This is largely attributed to their very marginal stability and the concomitant difficulty of determining baselines accurately, which hinders the ultimate detection of the inconsistency between different structural features. Further, even if the disagreements between signals could be confirmed, there is still possibility that it is due to cooperative folding via intermediates or structured denatured states.[18,20] On the contrary, the global downhill system can also appear to undergo cooperative thermal denaturation if the changes in enthalpy and specific heat are linearly related to the change in spectral signal that is used to monitor the system, according to Fersht and coworkers[31] This is a somewhat ironic case and indicates that the equilibrium experiment may be not suitable to give a conclusive picture. In a very recent issue of PNAS, Fersht and coworkers discussed this difficulty and suggested that the feature of chevron plot is a better criterion to distinguish two mechanisms.[31] However, the chevron plot they presented is for the homologue of BBL but not the BBL itself, although it provides important insight into the folding mechanism of similar ultra-fast folders, direct measurement on BBL is needed to give an ultimate conclusion.

There are a few theoretical works that directly attack the folding problem of BBL. By using an off-lattice Go-like model, Zuo *et al.* showed that BBL folds in an exact non-cooperative manner.[32] And from a statistics of 17 proteins, they suggested that the number of non-local contacts per residue could be a good parameter to predict from topology whether a protein folds in downhill manner. In a recent issue of this journal, Knott and Chan discussed the relevance of downhill folding with cooperativities and the underlying conformation distributions.[15] By using a coarse-grained model, they evaluated the effectiveness of the sharpness of transitions, the profile of heat capacity, the chevron rollovers, and the relaxation behaviors at defining the underlying distribution and folding picture. They then suggested that, if the distribution of single-molecule radius of gyration were available, it would permit discrimination between unimodal and bimodal underlying distributions. Although these previous works have given many insights into the features of downhill folding, they did not attack the folding problem of BBL directly by giving the free energy landscape and an estimate of the barrier height, due to the limitation of the coarse-grained model.

As an important complement of experiments, all-atom simulations have been used extensively to study the folding process of small proteins and help to reveal much about the complexity therein.[33–36] In case of the folding problem of BBL, all-atom simulation can help clarify the underlying mechanism because it is free of many complicated factors in experiments and can provide direct information on both energy landscape and structural features. Further, no presumed model is needed to interpret the data, which is of particular importance in this issue.

In this work, by using an intensive all-atom simulation with explicit water model and the replica exchange molecular dynamics (REMD), we sample the phase space of protein BBL and depict its free energy landscape. Based on the energy landscape, we give an estimation of the free energy barrier separating the denatured and native states. We also investigate the underlying conformational distribution of the transition states. Further, we detect possible inconsistency between different structural features by following the similar proce-

dures as in experiments. At last, we compare our results with experiments and give brief conclusions based on our simulation.

METHODS

The sequence used in our simulation is LSPAIRRLLAEHNLDASAIKGTGVGGRLTREDVEKHL. This sequence corresponds to the PDB entry 1BBL and is shorter than those used by both Fersht and Muñoz laboratories. However, we believe this is not a big drawback because the stability of this sequence (2.3 kcal/mol at 296 K, according to Table II that will be presented later) is comparable with that of QNND-BBL,[18] indicating that the boundary effect is not so important for stability. In our simulation, the residue histidine is modeled as a neutral species with hydrogen at the delta position. The protein is solvated in a TIP3P water box with about 3500 water molecules. The water density in the simulation is kept at $\sim$1.0 g/cm^3. The AMBER parm03 force field is used. Electrostatic interactions are calculated by using PME method with a cutoff of 8.0 Å, and the same cutoff is used in the calculation of nonbonded list.

The REMD method is used to overcome the multiple-minimum problem because the replicas at low temperature can be exchanged to high temperatures and get high probabilities there to transfer to new regions of phase space. The standard REMD algorithm implemented in AMBER 8.0 is used to perform the simulations.[37,38] The simulation is carried out with 64 replicas, with the temperatures ranging from 273 to 560 K. All of the temperatures used are detailed in the supplemental materials. The strategy used to choose temperature is as follows. First, a short REMD simulation with trial temperatures is run, and then the temperature dependence of the average energy is calculated and fitted with a five-order polynomial function, after that the temperatures that guarantee desired exchange rates can be analytically calculated with the metropolis-like exchange criterion. The goodness of this strategy can be further verified by the final exchange rates of all replicas. It is found that they range from 23 to 29% and keep constant during the entire time of our simulation, suggesting that this strategy works well and the temperatures are appropriately distributed (the details are given in the supplemental materials).

The simulation is started from an energy-minimized native structure. From this structure, a 2-ns simulation is performed at 300 K at constant pressure to achieve proper size of water box and equilibrium of water molecules. The final configuration of this run is further subjected to 64 simulations at constant temperature in the temperature range from 273 to 560 K, lasting 2 ns without exchange between replicas. The final configurations of these 64 simulations are used as the initial conformations of the REMD simulations. Of the 64 initial conformations, 38 have an RMSD $<$ 4 Å, 12 have 4 Å $<$ RMSD $<$ 6 Å, 8 have 6 Å $<$ RMSD $<$ 8 Å, 2 have 8 Å $<$ RMSD $<$ 10 Å, 2 have 10 Å $<$ RMSD $<$ 12 Å, and 2 have 12 Å $<$ RMSD $<$ 14 Å. These initial conformations span a large conformational range from native state to unfolded states, which is comparable with that in the REMD simulation of protein A.[33] The temperature coupling constant is set to 0.1 ps; the replica exchanges are attempted every 1 ps. Each replica is run for 100 ns, resulting in a total simulation time of $\sim$6.4 μs, which is comparable with the folding time ($\sim$7 μs at 310 K[20]). Additional technical details about REMD simulation are referred to our previous work.[39]

We use totally nine order parameter (OP) to characterize the free energy landscape, including the fraction of native contact Q, the fraction of native contact in the secondary structure Q_2, that in the tertiary structure Q_3, the average length of helix-1 (measured by number of residues), that of helix-3, the radius of gyration R_g, the radius of gyration of the hydrophobic core residues R_g^{core}, the all-atom cRMSD with respect to the native state, and the number of native backbone hydrogen bonds nHB. In our calculations, a native contact is defined if the distance between any pair of heavy atoms (non-hydrogen atoms) of two residues is within 5 Å. A native contact is assumed a tertiary contact if the sequence separation between two residues is larger than 6. A helical residue is defined if it occupies the α-helical region of the $\phi\psi$ map ($\phi = -60^\circ \pm 35^\circ$ and $\psi = -40^\circ \pm 30^\circ$). The helix-1 is assumed extending from residue I5 to E11, the helix-3 from R30 to H36, and the middle 3_{10} helix from A16 to A18. The hydrophobic core residues used to calculate R_g^{core} include I5, L8, L9, L14, A16, I19, L28, V33, and L37. A hydrogen bond is counted if the distance between donor and acceptor is less than 3.5 Å and the angle N—H$\cdots$O is larger than 120°. In the entire paper, the RMSD always refers to the all-atom cRMSD with respected to the X-ray native structure, unless indicated otherwise. The WHAM method is used to calculate thermodynamic quantities from simulations.[40,41] In the WHAM calculation, the bin size of the discrete parameters, including Q, Q_2, and Q_3, nHB and length of helix, is chosen so that the neighboring bins differ by only one number of contacts, hydrogen bonds or residues. For the other parameters, the bin sizes are 0.16 Å, 0.19 Å and 0.2 Å for R_g, R_g^{core} and RMSD, respectively.

We also calculate the minimal pathway along free energy surface in high dimensional space to alleviate the reaction coordinate (RC) problem, because it is hard to determine which order parameter corresponds to the real RC and how many of them are need to characterize the free energy landscape. In detail, we construct a four-dimensional (4D) space using the following four order parameters, Q, Q_3, R_g^{core}, and RMSD. These order parameters are chosen because they reflect mostly tertiary structures hence are better to characterize the fold-

ing process, taking account of there are already significant helical structure in the denatured ensemble. We calculate the free energy surface in this 4D space and then the minimal pathway along it as a function of Q. The minimal pathway is defined by the combination of (Q, Q_3, R_g^{core}, RMSD) parameters for which the free energy is minimal for each value of Q, following the method used by Garcia *et al.*[33] Note that in our implementation a distance constraint is used to avoid unreasonable long jump between basins in the free energy surface.

RESULTS AND DISCUSSIONS

Convergence diagnostics

To test the convergence of our simulation, we calculated the fraction of folded as a function of time for replicas at low temperatures, as shown in Figure 1(a). The curves at high temperatures are not given since they converge faster. As a quantity relevant to the number of structures, the fraction of folded is more sensitive to the convergence than other quantities such as the average Q or average radius of gyration. The structure is counted as folded if its RMSD is less than 5 Å. The fraction of folded is calculated by averaging on the data collected in a 20-ns window; for example, the last data point at 90 ns is calculated from the structures collected from 80 to 100 ns. According to Figure 1(a), the calculation becomes convergent at about 66 ns. To check this judgment further, the free energy as a function of Q is calculated at 330 K and shown in Figure 1(b), which is a much stricter criterion since it tests the relative population between many states rather than just two [as Fig. 1(a) does]. The temperature 330 K is chosen as a compromise between the convergent quality and the interested temperature, that is, the melting temperature (QNND-BBL, 328.30 K,[18] Naf-BBL, 321 K[19]). The convergence at low temperatures below 320 K is not as good as at 330 K as shown in Figure 1(b).

As shown in Figure 1(b), the free energy curves calculated from 60 to 80 ns and 80 to 100 ns suggest different conformational distributions along Q, but have comparable magnitudes of ruggedness. It may be because that the 20-ns sampling is not long enough to cover the available phase space due to the correlation between motions and the vast volume of phase space. This argument is supported by the free energy curves calculated from larger windows. The free energy curves calculated from 66 to 90 ns and 66 to 100 ns agree with each other very well, except at the low Q region ($Q < 0.3$), which is obviously due to the huge number of denatured states there. However, we believe it will not affect the major goal of this work because (1) we are mainly interested in the transition region ($0.3 < Q < 0.8$), which has been sampled adequately [Fig. 1(b)]; 2) the ensemble average, such as $\langle Q \rangle$ or fraction of folded, converges much fast than the energy landscape itself, which allows us to study the temperature dependent properties.

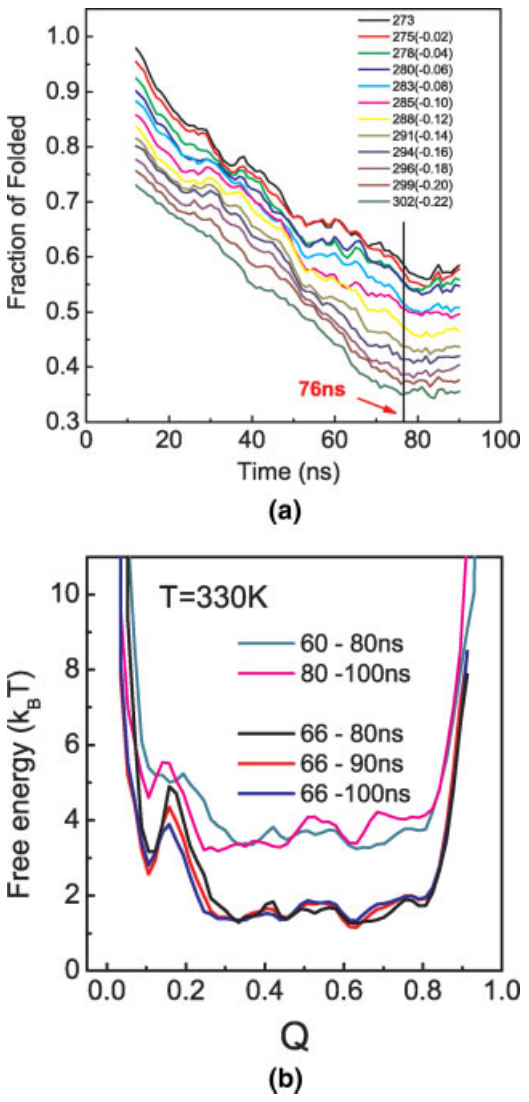

Figure 1

*(**a**) The fraction of folded for replicas at low temperatures. The data points are obtained by averaging on conformations collected within 20-ns window. Each curve has been shifted 0.02 downward with respect to its lower temperature neighbor for clear representation; the shifted values are labeled in the corresponding legends. (**b**) The free energy as a function of Q at temperature 330 K, calculated from different running windows. The curves have also been shifted along the vertical axis due to the same reason. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

We also checked the time serials of Q and temperature of all replicas. The time serials of Q and temperature both span a broad region, indicating a good sampling quality. We also found many unfolding events and several refolding events. The relevant figures are given in the supplemental materials.

Figure 1(a) deserves further investigation. It can be seen that the fraction of folded at all temperatures keeps decreasing before the convergence is reached, indicating that the general trend of the simulation is "unfolding." However, it does not preclude us from investigating the geometry of free energy landscape and the folding pictures, because the REMD is a "sampling" algorithm in nature, thus the key point is to ensure that the interested regions in phase space are adequately sampled and the relative populations between different regions are correct, which has been tested in Figure 1. The unfolding trend is irrelevant to the replica exchange algorithm itself, but only due to the initial conformations, which only affects the convergence speed. However, the REMD algorithm does preclude us from discussing the real kinetics of folding/unfolding because the trajectories at fixed temperatures are discontinuous. Therefore, it should be noted that, (1) when the terms folding/unfolding are used in this article, they only refer to the progression along pre-chosen order parameters and do not necessarily correspond to the real kinetic folding/unfolding processes[42,43]; (2) the kinetics relevant conclusions drawn in this article are only correct in an approximate manner, based on the assumption that the kinetics follows the thermodynamically calculated energy landscape; however, we believe this assumption can be largely fulfilled by a cautious choice of order parameter.

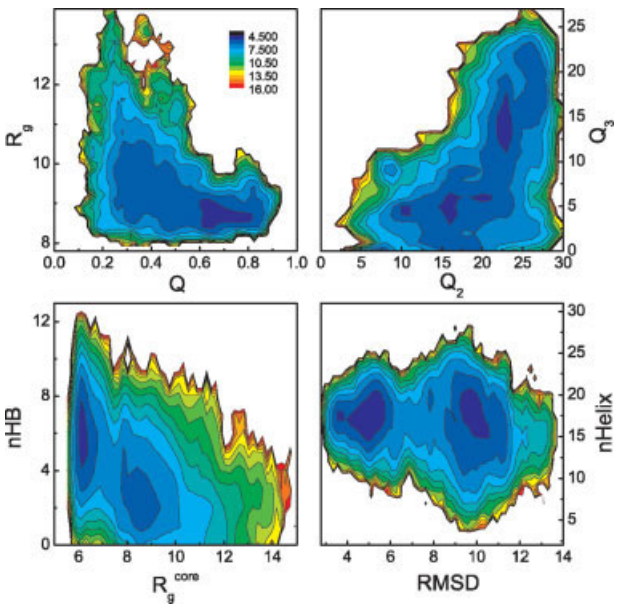

Figure 2

The 2D free energy landscape calculated at temperature 330 K. The figure is calculated from the conformations collected from 66 to 100 ns. The height between neighboring contour lines is $1k_BT$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The free energy landscape and barrier height

Figure 2 shows the two-dimensional (2D) projection of free energy landscape at 330 K, slightly higher than experimental melting temperature (QNND-BBL, 328.30 K,[18] Naf-BBL, 321 K[19]). The figure is calculated from the conformations collected from 66 to 100 ns. If not indicated otherwise, all of the following figures are calculated likewise. To characterize the energy landscape faithfully, a total of four combinations of eight OPs are used to project the free energy surface. Presumably, eight OPs are enough due to the simple native topology of protein BBL. The free energy landscape is rather complex: Figure 2(a,b) shows only one broad populated states, whereas Figure 2(c,d) shows two separated states. Therefore, to determine what the true picture is, one has to determine which OP corresponds to the true reaction coordinate. This is of course not a trivial matter. However, even for Figure 2(c,d) the free energy barrier separating two basins is only $\sim 2k_BT$, indicating that the folding cooperativity is rather low. Figure 2 also suggests that among all OPs, R_g^{core} and RMSD could be good parameters to detect possible separation of states, although it is not known whether the thermodynamic surface projected on these OPs reflects faithfully the underlying kinetics and barrier height. Besides, it should be noted that Figure 2 shows no evidence of existing of intermediate states, ruling out the multi-state picture to some extent.

Because the energy landscape is dependent on the choice of OPs, we try to circumvent this problem by constructing it in high-dimensional space. The idea is that, (1) presumably, more OPs together can characterize the free energy landscape better; (2) although highly cooperative folding (indicating high free energy barrier) only needs one reaction coordinate to describe the reaction, weakly or non-cooperative folding (indicating flat free energy surface) may need more to describe the motions along the surface. After constructing the energy landscape in 4D space, we calculate the minimal pathway leading from the denatured state to the native state as a function of Q. The curves calculated at temperatures 300, 310, 320, 330, and 340 K are shown in Figure 3. It can be seen that along these pathways, the free energy barriers are very low. For example, it is only $1k_BT$ at 330 K and 1.3 k_BT at 300 K.

By combing the pictures from Figures 2 and 3, it can be concluded that the free energy barrier is very low, 1–$2k_BT$ near the experimental melting temperature, depending on the way the energy landscape is projected. Moreover, taking consideration of the slower convergent speed at the barrier region, the barrier height may decrease slightly further with the elongate of simulations. Therefore 1–$2k_BT$ could be an upper limit of the barrier height near the melting temperature.

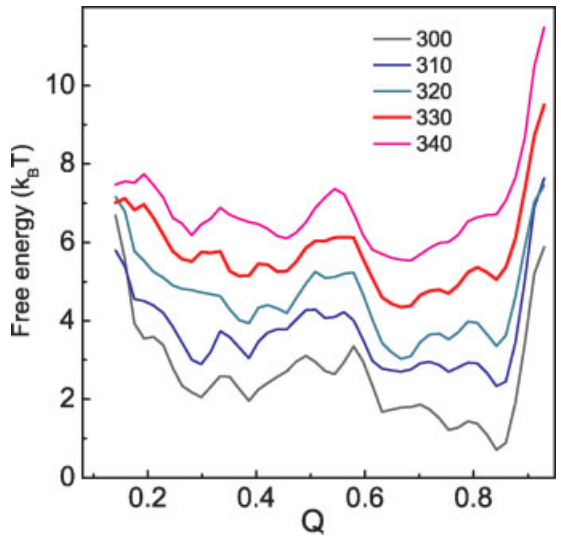

Figure 3

The minimal pathway along the 4D free energy surface as a function of Q, *calculated at temperatures indicated by the legends. The curves have been shifted along the vertical axis for clear representation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

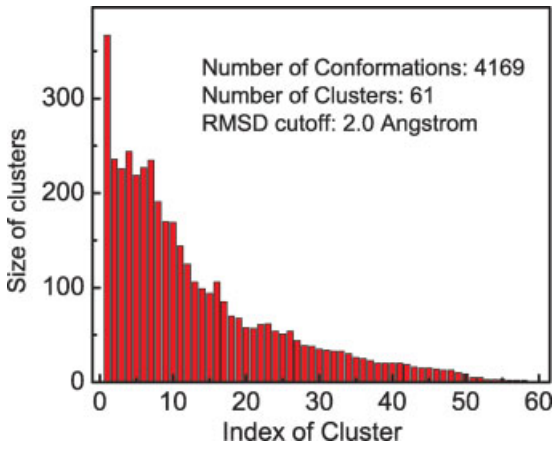

Figure 4

The occupation of the conformational clusters. Totally 4169 conformations are collected from the transition region. The clustering algorithm uses a cutoff of 2 Å and gives 61 clusters. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Conformational distribution at the transition region

One important difference between cooperative and downhill folding is the conformational distribution of the transition states (TS).[15,22] It is well established that for perfect cooperative folding, the transition states will resemble the structure of native state, leading to a small conformational heterogeneity. This picture is also consistent with the fact that the transition state has a low population and acts as an entropic barrier. In contrast, in downhill folding, the conformational distribution shows a broad unimodal peak centering at the middle region of reaction coordinate at the melting temperature, reflecting a large heterogeneity in the conformations.[15,22,32] To characterize the conformational distribution of the TS, we collect all the conformations that satisfy 6.6 Å $< R_g^{core} <$ 7.9 Å and 6 Å $<$ RMSD $<$ 8.5 Å in the last 30-ns run at 330 K. This criterion is based on the free energy landscape in Figure 2. Note that these conformations may not be the transition states in a strict sense since it is hard to prove the OPs used in Figure 2 correspond to the real RC.[39] However, they may provide a good approximation because we use many OPs simultaneously to characterize the energy landscape. Totally we obtain 4169 conformations. We then cluster these conformations by following the algorithm proposed by Snow *et al.* using an RMSD cutoff of 2.0 Å.[44] This procedure gives 61 clusters, with their occupations shown in Figure 4. The large number of clusters indicates the underlying conformational heterogeneities at the transition region. The overall structures of these conformations are shown in Figure 5, which are obtained by taking the representative structure of each cluster and then rotating it to achieve maximal superimposition with the native structure. It can be seen that, roughly speaking, the relative positions of three helices are similar to that in the native state thus similar in all conformations [Fig. 5(a)]. However, the spatial orientations of the helices show large variability,

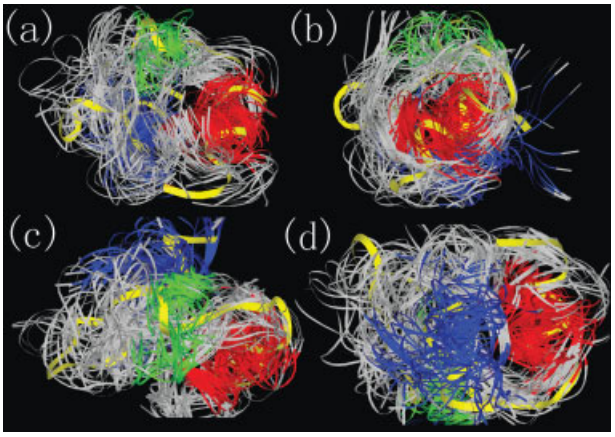

Figure 5

*The superimposition of the native structure and the 61 representative structures of TS, taken from each cluster. The native structure is shown by thick yellow ribbon, with the first residue located at the middle bottom. The others are shown by narrow gray ribbons with the first helix, the middle 3_{10} helix and the third helix colored red, green and blue, respectively. (**a**) the top view, showing the relative positions of three helices; (**b–d**) the side view focusing on the first, middle and third helix, respectively.*

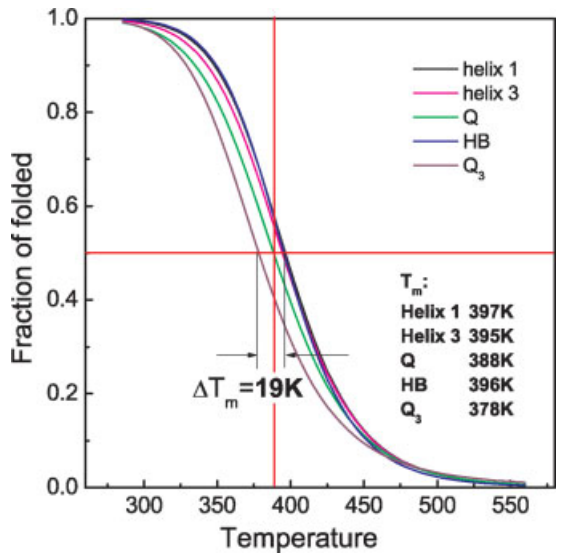

Figure 6

The fraction of folded as a function of temperature, monitored by five different structural features. The curves are obtained by fitting the raw simulation data of five structural features individually with two-state model. Different structural signals give different T_ms, which span a range of 19 K. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

especially the second and the third one [Fig. 5(c,d)]. Besides, the long loop region between the last two helices is completely disordered. By comparing Figure 5 with the TS structure obtained from a coarse-grained model [Fig. 7(b,e), Knott and Chan[15]], it can be seen that the conformations in Figure 5 show a much higher degree of variability than that of protein NTL9, whereas a less degree than that predicted by the coarse-grained model. The comparison suggests that our construct indeed exhibits some cooperativity; however, it is very weak due to the large conformational variability at the transition region.

Detect conformational variability from thermal unfolding curves

On the experimental front, major emphasis has been placed on checking for inconsistency in T_m values for different probes. Here we follow the same procedure as in experiments by monitoring the temperature dependence of the signals of five different structural features. The structural probes include the fraction of native contacts Q, the fraction of native tertiary contacts Q_3, the average lengths of helix-1 and helix-3, and the number of native hydrogen bonds nHB.

Figure 6 shows the thermal unfolding curves calculated for five different structural features. The curves have been baseline subtracted with a two-state analysis using the Gibbs–Helmholtz equation. Note that the fit is carried out for each curve individually. The raw data, baselines and fitted parameters are given in Figure 7 and Table I. It can be seen that the curves do not coincide with each other: the structural features that mostly define the tertiary structure melt at lower temperatures (For example, the Q_3 curve), whereas secondary structure sustains much higher temperatures. The melting temperatures range from 378 to 397 K, spanning a width of 19 K. A similar value, 17 K, is obtained if the T_m is calculated from direct derivative of the unfolding curves, which procedure is free of possible baseline problem. The large spread of T_m suggests that the five structural features unfold in a weakly or non-cooperative manner, consistent with the small free energy barrier shown in Figures 2 and 3.

The spread of 19 K is significantly smaller than that of Naf-BBL reported by Sadqi *et al.* (60 K).[22] One possible reason may be due to the different pH values used; a neutral solvent is modeled in our simulation whereas pH 5.3 is used in that experiment. Physically, the low pH will change the electrostatic distribution of protein thus affect the enthalpy–entropy interplay in folding process, leading to a different free energy barrier and different cooperativity. If we naively relate the spread of T_m with the barrier height, since both of them are relevant to the folding cooperativity, the comparison of the spread of T_m indicates that the Naf-BBL at pH 5.3 may have an even lower barrier height than $1–2k_BT$. In this way our simulation supports the global downhill picture of BBL proposed by Muñoz *et al.*

We also tried to detect the inconsistency between thermal unfolding curves by global fit, using the same fixed

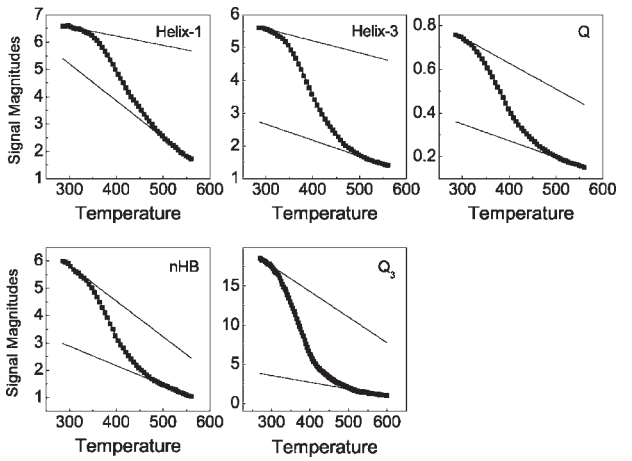

Figure 7

The global fit of the temperature dependence of five structural features. The squares represent the raw simulation data. The fitted curves are shown by the solid lines, which are almost indistinguishable from the raw data. The baselines are shown as straight lines, which are the exactly same as that used in the individual fit.

Table I
Individual Fit to the Thermal Denaturation Curves Monitored by Different Structural Features

	Helix 1	Helix 3	Q	HB	Q_3
T_m (K)	397.0 ± 0.5	394.6 ± 0.4	388.3 ± 0.4	395.3 ± 0.4	378.3 ± 0.3
ΔH_m (kcal/mol)	12.6 ± 0.2	12.0 ± 0.2	11.1 ± 0.1	13.4 ± 0.2	10.8 ± 0.1
ΔC (kcal/(mol K))	0.03 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.00 ± 0.01	−0.01 ± 0.01
χ^2 ($\times 10^{-3}$)	1.4	1.4	0.01	1.0	15.2
R^2	0.99954	0.99945	0.99968	0.99967	0.99944

baselines as in individual fit. The results are shown in Figure 7 and Table II. According to Figure 7, the fitted curves are almost indistinguishable to the raw data and the fit yields T_m = 380.8 K ± 0.4 K at the 95% confidence level. It is surprising to see that the global fit gives a single T_m with extremely small uncertainty, although the exactly same raw data and baselines as that in individual fit are used. The goodness of the global fit can be further improved slightly if the baselines are permitted to float. Because all the aforementioned features, including the low barrier height, the heterogeneous conformational distribution at the transition region and the large spread given by individual fit are consistently suggesting a very weak folding cooperativity, the result of the global fit shows that it is not sensitive to the inconsistency between signals. This is possibly because it tries to reconcile different unfolding curves, thus introducing pseudo-cooperativities. There also might be potential systematic errors in the global fit. However, the detailed discussion of this issue is beyond of the scope of this article. The reader is referred to several recent communications for interesting discussions.[23–25]

This phenomenon is reminiscent to the similar pattern observed in two recent NMR experiments. In detail, Ferguson *et al.* analyzed their NMR data of 15 selected C^{13} probes by both individual and global two-state fit and found the entire dataset could be fitted to common thermodynamic values; and the global fitting gave a small uncertainty of only ±1 K at the 95% confidence level.[20] In contrast, Sadqi *et al.* measured 158 proton unfolding curves and fitted them individually, finding a broad range of T_m spreading about 60 K.[22] Although the cooperativity suggested by the first experiment cannot be simply attributed to the artifact introduced by the global fit (due to both individual and global fit gave similar result), it should be emphasized that a good global fit with small uncertainty does not necessarily mean cooperative folding. To detect inconsistency between signals, individual fit should always be used, as Sadqi *et al.* did.

Table II
Global Fit to the Thermal Denaturation Curves Monitored by Five Different Structural Features

T_m (K)	ΔH_m (kcal/mol)	ΔC [kcal/(mol K)]	χ^2 ($\times 10^{-3}$)	R^2
380.8 ± 0.4	10.8 ± 0.1	0.01 ± 0.01	26.0	0.99802

Potential problems of the simulation

Our simulation is not without problems. For example, the T_m values calculated by our simulation are much larger than the experimental values (~380 K vs. 321–330 K[18,19]), due to the NTV ensemble used in the simulation and the possibly inaccurate temperature dependence of the modern force fields. The NTV ensemble will cause unreasonable high pressure at high temperatures. To avoid the artifacts arising from this problem, we limit the energy landscape calculation at medium temperatures, a compromise between convergent speed and reasonable pressures. The high pressure at high temperatures may also distort the denaturant baselines and affect the fitting results. However, we argue this does not preclude us from detecting inconsistency between signals because a similar spread of T_m is obtained from direct derivative of the unfolding curves. The latter procedure is only dependent on local slope of the curve hence free of possible baseline problem.

CONCLUSIONS

The experimental determination of folding behaviors of protein BBL is rather difficult and has aroused great debates. The controversy has made people reexamine several issues intrinsic to conventional folding experiments. Recently, the relevant discussions have gone far beyond these issues and led people re-consider the basic theoretical frame that used to interpret experimental data on ultra-fast folders. At this time, it is helpful to study the energy landscape of BBL directly from all-atom simulations, hoping that it can cast some light onto the this issue. The all-atom simulation has the advantage that it can give the energy landscape, the free energy barrier and the conformational distribution directly; no presumed phenomenological model is needed to interpret data. This is of particular importance in determining the exact folding picture of BBL.

We found that the free energy barrier ranges from $1k_BT$ to $2k_BT$, dependent on the way the energy landscape

is projected. We also studied the conformational distribution of the transition states and found that the relative positions of three helices are roughly similar to that in the native state thus similar in all conformations. However, the spatial orientations of three helices show large variability and the long loop region between the last two helices is completely disordered. Furthermore, by individually fitting five denaturation curves with two-state model, we detected a wide spread melting temperature of 19 K, demonstrating the inconsistency between different structural features. All these results are consistent with a picture of folding with very weak cooperativities.

The spread of T_m detected in our construct of BBL is significantly smaller than that of Naf-BBL reported by the experiment.[22] One possible reason of this discrepancy is due to the different pH values used, neutral in our simulation versus pH 5.3 in the experiment. If we naively relate the inconsistency in T_m with the barrier height, as both of them are relevant to folding cooperativity, it is very possible that the Naf-BBL (pH5.3) has an even lower barrier height than $1–2k_BT$. In this way, our simulation supports the downhill picture of BBL. However, to ultimately confirm or falsify this picture, more simulations and experiments have to be done, especially at low pH values. Also, new analyzing methods should be developed further to overcome the difficulties associated with the reaction coordinate problem. These methods are of increasing importance because more and more ultra-fast folders, which imply flat free energy surfaces, are discovered nowadays.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their critical, yet helpful comments. They acknowledge Shanghai Supercomputer Center for computing resources.

REFERENCES

1. Hagen SJ, Hofrichter J, Szabo A, Eaton WA. Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding. Proc Natl Acad Sci USA 1996;93:11615–11617.
2. Gruebele M. The fast protein folding problem. Annu Rev Phys Chem 1999;50:485–516.
3. Eaton WA, Muñoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J. Fast kinetics and mechanism in protein folding. Annu Rev Biophy Biomol Struct 2000;29:327–359.
4. Kubelka J, Hofrichter J, Eaton WA. The protein folding speed limit. Curr Opin Struct Biol 2004;14:76–88.
5. Sabelko J, Ervin J, Gruebele M. Observation of strange kinetics in protein folding. Proc Natl Acad Sci USA 1999;96:6031–6036.
6. Dyer RB. Ultrafast and downhill protein folding. Curr Opin Struct Biol 2007;17:1–10.
7. Yang WY, Gruebele M. Folding at the speed limit. Nature 2003;423:193–197.
8. Muñoz V. Conformational dynamics and ensembles in protein folding. Annu Rev Biophys Biomol Struct 2007;36:395–412.
9. Kubelka J, Chiu TK, Davies DR, Eaton WA, Hofrichter J. Sub-microsecond protein folding. J Mol Biol 2006;359:546–553.
10. Zhu Y, Alonso DOV, Maki K, Huang CY, Lahr SJ, Daggett V, Roder H, DeGrado WF, Gai F. Ultrafast folding of α_3D: A de novo designed three-helix bundle protein. Proc Natl Acad Sci USA 2003;100:15486–15491.
11. Bunagan MR, Yang X, Saven JG, Gai F. Ultrafast folding of a computationally designed Trp-cage mutant: Trp^2-cage. J Phys Chem 2006;110:3759–3763.
12. Bryngelson JD, Wolynes PG. Intermediates and barrier crossing in a random energy model (with applications to protein folding). J Phys Chem 1989;93:6902–6915.
13. Oliveberg M, Wolynes PG. The experimental survey of protein-folding energy landscapes. Quart Rev Biophy 2005;38:245–288.
14. Chan HS, Shimizu S, Kaya H. Cooperativity principles in protein folding. Methods Enzymol 2004;380:350–379.
15. Knott M, Chan HS. Criteria for downhill protein folding: Calorimetry. Chevron plot kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity. Proteins 2006;65:373–391.
16. Gruebele M. Downhill protein folding: evolution meets physics. C R Biol 2005;328:701–712.
17. Garcia-Mira MM, Sadqi M, Fischer N, Sanchez-Ruiz JM, Muñoz V. Experimental identification of downhill protein folding. Science 2002;298:2191–2195.
18. Ferguson N, Schartau PJ, Sharpe TD, Sato S, Fersht AR. One-state downhill versus conventional protein folding. J Mol Biol 2004;344:295–301.
19. Naganathan AN, Perez-Jimenez RP, Sanchez-Ruiz JM, Muñoz V. Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. Biochemistry 2005;44:7435–7449.
20. Ferguson N, Sharpe TD, Schartau PJ, Sato S, Allen MD, Johnson CM, Rutherford TJ, Fersht AR. Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. J Mol Biol 2005;353:427–446.
21. Naganathan AN, Doshi U, Fung A, Sadqi M, Muñoz V. Dynamics, energetics, and structure in protein folding. Biochemistry 2006;45:8466–8475.
22. Sadqi M, Fushman D, Muñoz V. Atom-by-atom analysis of global downhill protein folding. Nature 2006;442:317–321.
23. Ferguson N, Sharpe TD, Johnson CM, Schartau PJ, Fersht AR. Analysis of "downhill" protein folding. Nature 2007;445:E14–E15.
24. Zhou Z, Bai YW. Analysis of protein-folding cooperativity. Nature 2007;445;E16–E17.
25. Sadqi M, Fushman D, Muñoz V. Sadqi Reply. Nature 2007;445:E17–E18.
26. Dill KA, Shortle D. Denatured states of proteins. Annu Rev Biochem 1991;60:795–825.
27. Oliva FY, Muñoz V. A simple thermodynamic test to discriminate between two-state and downhill folding. J Am Chem Soc 2004;126:8596–8597.
28. Muñoz V, Sanchez-Ruiz JM. Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. Proc Natl Acad Sci USA 2004;101:17646–17651.
29. Naganathan AN, Sanchez-Ruiz JM, Muñoz V. Direct measurement of barrier heights in protein folding. J Am Chem Soc 2005;127:17970–17971.
30. Naganathan AN, Doshi U, Muñoz V. Protein folding kinetics: barrier effects in chemical and thermal denaturation experiments. J Am Chem Soc 2007;129:5673–5682.
31. Huang F, Sato S, Sharpe TD, Ying L, Fersht AR. Distinguishing between cooperative and unimodal downhill folding. Proc Natl Acad Sci USA 2007;104:123–127.
32. Zuo GH, Wang J, Wang W. Folding with downhill behavior and low cooperativity of proteins. Proteins 2006;63:165–173.
33. Garcia AE, Onuchic JN. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. Proc Natl Acad Sci USA 2003;100:13898–13903.

34. Zhou RH. Trp-cage: folding free energy landscape in explicit water. Proc Natl Acad Sci USA 2003;100:13280–13285.
35. Snow CD, Nguyen H, Pande VS, Gruebele M. Absolute comparison of simulated and experimental protein-folding dynamics. Nature 2003;420:102–106.
36. Mayor U, Guydosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, Freund SMV, Alonso DOV, Daggett V, Fersht AR. The complete folding pathway of a protein from nanoseconds to microseconds. Nature 2003;421:863–867.
37. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 1999;314:141–151.
38. Case DA, Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA. AMBER 8; San Francisco: University of California; 2004.
39. Zhang J, Qin M, Wang W. Folding mechanism of β-hairpins studied by replica exchange molecular simulations. Proteins 2006;62:672–685.
40. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. Phys Rev Lett 1989;63:1195–1198.
41. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH. Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 2004;13:1011.
42. Swope WC, Pitera JW. Describing protein folding kinetics by molecular dynamics simulations, part 1: Theory J Phys Chem 2004; 108:6571–6581.
43. Swope WC, Pitera JW. Describing protein folding kinetics by molecular dynamics simulations, part 2: Example applications to alanine dipeptide and a β-hairpin peptide. J Phys Chem 2004;108:6582–6594.
44. Snow CD, Qiu L, Du D, Gai F, Hagen SJ, Pande VS. Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. Proc Natl Acad Sci USA 2004;101:4077–4082.