Folding with Downhill Behavior and Low Cooperativity of Proteins

Guanghong Zuo,[1] Jun Wang,[1] and Wei Wang[1,2,*]
[1]*National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, China*
[2]*Interdisciplinary Center of Theoretical Studies, Chinese Academy of Sciences, Beijing, China*

ABSTRACT **The downhill folding observed experimentally for a small protein BBL is studied using off-lattice Gō-like model. Our simulations show that the downhill folding has low cooperativity and is barrierless, which is consistent with the experimental findings. As an example of comparison in detail, the two-state folding behavior of proteins, for example, protein CI2, is also simulated. By observing the formation of contacts between the residues for these two proteins, it is found that the physical origin of the downhill folding is due to the deficiency of nonlocal contacts which determine the folding cooperatively. From a statistics on contacts of the native structures of 17 well-studied proteins and the calculation of their cooperativity factors κ_2 based on folding simulations, a strong correlation between the number of nonlocal contacts per residue N_N and the factors κ_2 is obtained. Protein BBL with a value of $N_N = 0.73$ has the lowest cooperativity factor $\kappa_2 = 0.34$ among all 17 proteins. A crossover around $N_N^c \sim 0.9$ could be defined to separate the two-state folders and the downhill folder roughly. A protein would behave downhill folding when its $N_N = N_N^c$. For proteins with their N_N values are about (or slightly larger than) N_N^c, the folding behaves with low cooperativity and the barriers are small, showing a weak two-state behavior or a downhill-like behavior. Furthermore, simulations on mutants of a two-state folder show that a mutant becomes a downhill folder when its N_N is reduced to a value smaller than N_N^c. These could enable us to identify the downhill folding or the cooperative two-state folding behavior solely from the native structures of proteins. Proteins 2006;63:165–173.**

Key words: protein BBL; downhill folding; Gō-like model; cooperativity; the number of nonlocal contacts per residue

INTRODUCTION

Natural proteins fold to their unique folded structures or native states relying on the interactions encoded in their amino acid sequence.[1–5] A widely accepted folding mechanics is based on the funnel-like landscape theory.[5] The folding of protein molecules resembles a diffusive process on rugged landscape of free energy, and there is an ensemble of pathways to the native state.[6,7] Different proteins show different folding kinetics under different conditions. Many small or single-domain proteins behave with so-called two-state folding behavior, that is, hopping between the denatured states and the native state without any accumulation of intermediate states due to the existence of a main barrier between these two kinds of states.[8] The corresponding folding transition resembles to a first-order–like phase transition with high cooperativity.[9–11] For large or multidomain proteins, the folding shows a behavior with multiple states or intermediate states. However, the funnel theory also suggested another kind of folding, which proceeds without crossing free energy barrier along the reaction coordinate.[5,7,12] Recently, Garcia-Mira and colleagues observed experimentally such a folding behavior, namely the downhill folding, for a small protein BBL (PDB code: 1BBL).[13] They showed that the cooperativity of folding is low and there is no free energy barrier from their thermodynamic measurements. In contrast to the binary character of two-state folder, the order parameter Q of equilibrium states for this protein decreases continuously as the temperature increases, implying a conformational rheostat which is a new concept of functional states of proteins.[13–15] Such a protein is called a downhill folder. Very recently, Muñoz and coworkers further checked the robustness of the downhill folding and showed experimentally that singly and doubly labeled BBL do not aggregate, and unfold reversibly.[16]

Thus, some interesting issues have been raised. What features does the downhill folding show, and can we simulate the downhill folding using simplified or all-atomic models and characterize the folding mechanism? How do we interpret the difference between the experimental finding for protein BBL and other two-state proteins? More importantly, can we find out what are the physical reasons that make the downhill folder, such as protein BBL, show the downhill folding behavior? Can we identify a protein being a two-state folder, a downhill or downhill-like folder solely from the nature of its native structures? The answer to these questions will be helpful for under-

Grant sponsor: NSF; Grant numbers: 90403120, 10474041, 10204013, 10021001; Grant sponsor: NSM.

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

Received 7 June 2005; Revised 11 October 2005; Accepted 24 October 2005

Published online 13 January 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20857

standing the folding behaviors, and also the protein functions and protein design as well.

In this work, a study on the downhill folding from aspects of both molecular dynamics simulations and statistics on the contacts between residues for a number of proteins is reported. The Gō-like model[17–25] is used for our simulations. For protein BBL, a wide peak of specific heat and the absence of folding transition imply that the cooperativity is low. The free energy as a function of the similarity Q values to the native state clearly shows the absence of energetic barrier. As a comparison, the folding of two-state folders, for example, the chymotrypsin inhibitor II (CI2) and some other proteins, is also worked out. Our results show that the downhill folding is rather different from the two-state folding behavior, and the nonlocal contacts dominate essentially the cooperativity of folding. A statistics on two kinds of interactions, namely the local and nonlocal contact interactions, between the residues for the native structures of 17 proteins is made. In addition, the cooperativity factors κ_2 [26–28] of these 17 proteins are also calculated based on the folding simulations. A strong correlation between the numbers of nonlocal contacts per residue N_N and the cooperativity factor κ_2 is found, and a crossover around $N_N^c \sim 0.9$ could be defined to separate the two-state folder and the downhill folder roughly. This is further validated by folding simulations on mutants of a two-state folder which become a downhill folder when their $N_N < N_N^c$. Thus, a protein could be roughly classified into downhill folding or two-state folding only by checking its number of nonlocal contacts per residue, that is, the value of N_N.

MODEL AND METHODS

Gō-Like Model

The Gō-like energy[17] has been widely used to describe the interactions of residues both in lattice[29,30] and off-lattice model proteins.[18–25] Residues are represented by single beads centered in their Cα atoms, and all beads are connected into a polymer chain by virtual bonds. The interactions include terms related to the virtual bonds, angles, dihedral angles, and nonbonded pairs of the beads, that is, the residues. Such kind of Gō-like interactions has been successfully applied in the modeling of folding processes of small and single domain proteins.[18–20,31–34] The terms and parameters of the Gō-like interactions are similar to those used by Clementi and colleagues.[18] For a certain conformation of the protein, the total Gō-like potential energy is thus given by the expression:

$$\begin{aligned} V_{\text{total}} &= V_{\text{bonded}} + V_{\text{bond-angle}} + V_{\text{dihedral}} + V_{\text{nonbonded}} \\ &= \sum_{\text{bonds}}^{N-1} K_r(r - r_0)^2 + \sum_{\text{bond-angles}}^{N-2} K_\theta(\theta - \theta_0)^2 \\ &+ \sum_{\text{dihedrals}}^{N-3} \{K_\phi^{(1)}[1 - \cos(\phi - \phi_0)] + K_\phi^{(3)}[1 - \cos 3(\phi - \phi_0)]\} \\ &+ \sum_{i<j-3}^{\text{native}} \epsilon\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right] + \sum_{i<j-3}^{\text{non-native}} \epsilon\left(\frac{\sigma_0}{r_{ij}}\right)^{12}. \quad (1) \end{aligned}$$

Here, N is the total number of residues; r, θ, and ϕ are the bond length between two subsequential residues; bond-angle is formed by three values in the native structure of the protein; and r_{ij} is the spatial distance between two Cα atoms that have at least four residues between them along the chain. In the summation for all native contacts, a 10 to 12 Lennard–Jones (LJ) potential is used, and σ_{ij} is the Cα − Cα distance between the contacting residues i and j in the native structure. Here, the native contact is defined as the contact between two residues i and j in the native structure if any a heavy atom pair from each residue are less than 5 Å apart. In the summation for all non-native contacts, the parameter σ_0 introduces the excluded volume repulsion between residue pairs that do not belong to the given native contact set. In our simulations, $\sigma_0 = 4$ Å is used. The interaction parameters are taken as $K_\gamma = 100\epsilon$, $K_\theta = 20\epsilon$, $K_\phi^{(1)} = \epsilon$, $K_\phi^{(3)} = 0.5\epsilon$, respectively. During the folding, a native contact between two residues i and j is formed if the distance between their Cα atoms is shorter than κ times their native distance σ_{ij}. Here $\kappa = 1.2$ is used. Based on such a simplified model, the cooperative folding behavior of proteins can be reproduced in a semiquantitative manner, which suggests that the Gō-like potential is workable.[24,28]

Langevin Dynamics

We use Langevin dynamics[22,24,25,35] to simulate the folding process starting from a linear coil. The Langevin equation of motion is:

$$m\dot{\mathbf{v}}(t) = \mathbf{F}_{conf}(t) - \gamma\mathbf{v}(t) + \Gamma(t) \quad (2)$$

where $\dot{\mathbf{v}}$, $\mathbf{v}$, and m are velocity, acceleration, and mass of a bead, respectively; $\mathbf{F}_{conf} = \nabla E_p$ is the conformational force; γ is a friction (viscosity) constant. Here, we use $\gamma = 0.5\tau^{-1}$ with τ being the time scale. Γ is the random force, which is produced from the Gaussian distribution with a standard variance related to temperature by:

$$< \Gamma(t)\Gamma(t') > = 6\gamma k_B T\delta(t - t'). \quad (3)$$

Here, k_B is the Boltzmann constant, T is the temperature, t is the time, and $\delta(t - t')$ is the Dirac delta function. Every bead is subject to a random force at each integration time step. The components of random force are independently generated by setting $\Gamma_\alpha = \Phi_\alpha\sqrt{2m\gamma k_B T/\delta t}$ where Φ_α is a random value taken from a standard Gaussian distribution (zero mean and unit variance); α denotes the uncorrelated components of random force in the x, y, or z directions; and δt is the integration time step. At the beginning of every simulation, the initial velocities are assigned as zero.

We use the leap-frog algorithm to integrate Eq. (2). The time scale of the model protein here is always controlled by a quantity $\tau = \sqrt{m\lambda^2/\epsilon_0}$ with the length scale $\lambda = 3.8$ Å, the energy scale $\epsilon_0 = 1$. The time step of simulations is set as $\delta t = 0.005\tau$. Simulation times in this study are presented in units of δt. The energy parameter ϵ and temperature T are given, in units ϵ_0 and ϵ_0/k_B, respectively. The length is measured in units of λ. To simplify the notation, other units are chosen such as $m = 1$ and $k_B = 1$ in the present

simulation, as used by Thirumalai and coworkers.[36] An approximate correspondence between the model time and real protein kinetic time scales has been discussed by Thirumalai and coworkers.[36]

Weighted Histogram Analysis Method

The weighted histogram analysis method (WHAM) is used to calculate relevant thermodynamic quantities based on the statistical physics.[18,37,38] WHAM yields an optimal estimate of the density of state of the system. In the canonical ensemble at temperature T, the probability distribution, P_T, of potential energy E is given by

$$P_T(E) = (1/Z_c)w(E)e^{-E/k_BT} \quad (4)$$

and the probability distribution of other thermodynamic quantities, such as Q, follows

$$P_T(Q) = (1/Z_c)\sum_E w(Q,E)e^{-E/k_BT} \quad (5)$$

where w is the density of state of the system and is obtained by solving two self-consistent equations.[38] Z_c is the canonical partition function. The specific heat is defined by the energy fluctuation,[25] and the free energy at fix temperature $F_T(Q)$ is estimated by the logarithm of probability $P_T(Q)$.[18]

Quantification of the Cooperativity

In this work, a factor κ_2 is used to measure the cooperativity of the folding. Such a factor has been discussed by Chan and colleagues[26–28] and can be defined as

$$\kappa_2 = 2T_f\sqrt{k_BC_v(T_f)}/\Delta E_{\text{cal}} \quad (6)$$

where T_f is the folding transition temperature, $C_v(T_f)$ is the specific heat at T_f, and ΔE_{cal} is the calorimetric energy defined by the difference between the energy of denatured state and native state. A detailed discussion can be found in the related papers by Chan and coworkers.[26–28] In this article, the factor κ_2 of 17 proteins are calculated based on folding simulations.

Computational Mutations

Mutation is a useful method for protein science. For some specific mutations of a protein, the stable structures of the related mutants may not be available from the PDB database although they can be solved experimentally using X-ray or NMR techniques. For our theoretical studies, because only some sketchy structures of mutants are needed, a computational approach to the structures of mutants would be a good alternative of experiments. In our work, the structures with global minimized energy are defined as the stable structures of the mutants using simulated annealing, which is realized by all-atomic simulation software Amber.[39] A similar realization using Insight-II for finding stable structures has also been reported in literature.[40]

In detail, the procedure of computational mutations for a certain protein is as follows: For our interested proteins, for example, protein monomeric λ-repressor, we regularly prefer to mutate the residues with large side-chains into residues with small ones. Thus, the number of the nonlocal contacts for residues could be reduced. Because only a few residues are mutated, the structure of the mutant may change slightly compared with the native structure of the wild type. For every mutated sequence, the unfolding–annealing and minimization start from the native structure of the wild type. [The mutated sequence is threaded onto the native structure of the wild type.] This initial structure begins its unfolding at 500 K until its value of cRMSD = 10 Å (here, cRMSD is the root mean square deviation of the coordinates of Cα atoms of protein[41]), and then a gradually annealing to room temperature within 3000 ps is made. A 4000-cycle minimization is further carried out after the annealing to get a structure with the minimal energy, and the resultant structure is taken as a candidate of the stable/native structure of the mutant. Here, the process of unfolding–annealing and minimization is repeated several times (4 times for mutants M2 to M6 and 50 times for mutant M1) for finding a structure with energy as low as possible. Among all the candidates, the one with the lowest energy is chosen as the final stable structure of the mutant. Note that the Amber 1999 force field,[39] periodic boundary condition, constant pressure, and all other default parameters of Amber are used in our simulations.

It is noted that taking the native structure of the wild type as the starting structure of the mutants for our computational mutations is quite reasonable. Threading of the mutated sequences onto different unfolded conformations of the native structure of the wild type shows that the energies of these conformations after minimization are larger than the lowest energy of the mutated sequence threaded onto the native structure of the wild type after minimization. Here, over 10 unfolded conformations with their related values of 10 Å < cRMSD < 20 Å are used. Therefore, a partial unfolding of the native structure of the wild type with cRMSD = 10Å for process of unfolding–annealing and minimization to find the stable structure of the mutant is workable. These actually indicate that all the related conformations are around the basin of the native structure of the wild type on the energy landscape. Thus, although the number of nonlocal contacts is reduced, the stable structure of the mutant does not change so significantly (see Fig. 5).

RESULTS AND DISCUSSIONS

Folding Behaviors

The folding kinetics of protein BBL is characterized by the fraction of native contacts Q versus the time at various temperatures. At temperature $T = 0.8T_f$, $1.0T_f$, $1.1T_f$, and $1.2T_f$, after an extremely short relaxation the system reaches different equilibrium states with Q = 0.9, 0.7, 0.45, and 0.3, respectively (see Fig. 1). It is found that the Q values fluctuate around these equilibrium states, and there is a maximal fluctuation at $T = T_f$ [Fig.1(b)]. {Here, the value of T_f could be defined at the peak of specific heat [see Fig. 2(a)]}. There is no cooperative hopping between native and denatured states, so no obvious folding transi-

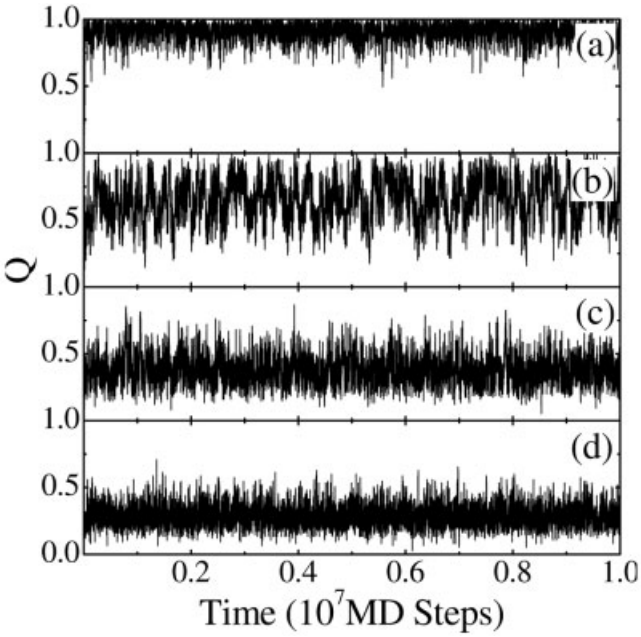

Fig. 1. Four folding trajectories for protein BBL at different temperatures: (**a**) $T = 0.8T_f$, (**b**) $T = 1.0T_f$, (**c**) $T = 1.1T_f$, and (**d**) $T = 1.2T_f$. Here Q is the fraction of native contacts.

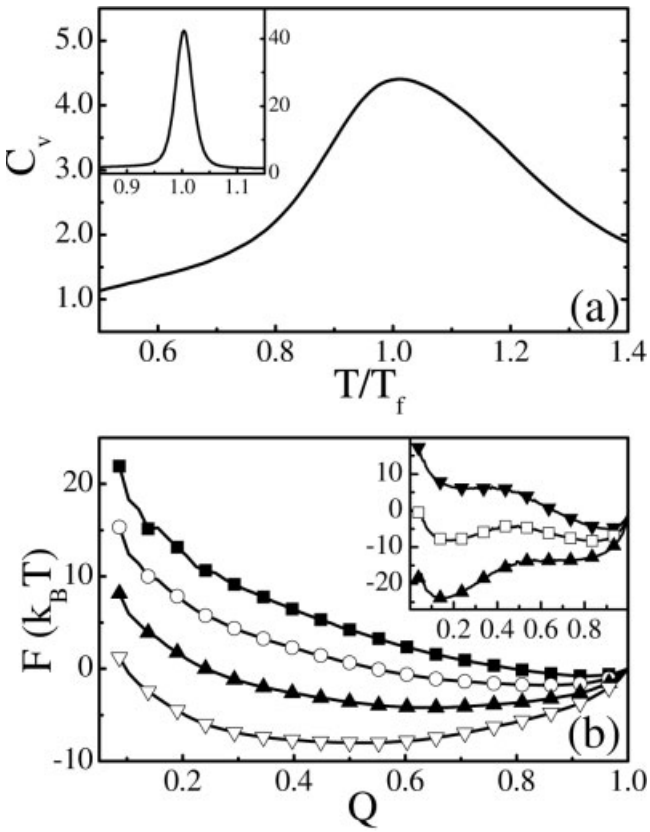

Fig. 2. The thermodynamics of folding. (**a**) The specific heat versus temperature T for protein BBL. The inset shows the specific heat for protein CI2. (**b**) Free energy profile for protein BBL at four temperatures $T = 0.8T_f$ (solid squares), $T = 0.90T_f$ (circles), $T = 1.00T_f$ (solid triangles), and $T = 1.10T_f$ (down triangles). The inset shows the free energy profile for protein CI2 at temperatures $T = 0.95T_f$ (solid down triangles), $T = 1.00T_f$ (squares), and $T = 1.05T_f$ (solid triangles). Here T_f for BBL and CI2 are 0.85 and 1.05, respectively.

tion. The curve of specific heat is wide and the peak is not sharp. These are consistent with the absence of cooperative transition, indicating that the cooperativity of folding for protein BBL is rather low. For each temperature, the free energy profile shows only one minimum around the equilibrium Q value which is dependent of the temperature [see Fig.2(b)]. The equilibrium Q value increases (or the minimal position of the free energy) monotonically from its value $Q = 0.3$ at the denatured state to its value $Q > 0.9$ at folded state, behaving as a conformational rheostat. There is no metastable state in the whole range of temperatures, and the free energy profiles show no energetic barrier clearly. These relate to the characteristics of downhill folding observed in experiment.[13]

As a comparison, the folding behavior of protein CI2 is simulated [see the insets in Fig. 2(a,b)]. The folding of CI2 shows a typical two-state behavior, that is, below $T_f = 1.05$ the Q value reaches to $Q > 0.9$, and above T_f the Q value is about Q = 0.2, while around T_f the Q value changes between the folded and unfolded states. The specific heat has a much sharper peak, which is about 10 times higher than that of protein BBL [see Fig.2(a)]. The free energy profile shows a large barrier around $Q = 0.5$ [see Fig.2(b)]. These are also observed for SH3 domain and other proteins (data not shown). The folding cooperativity of the two-state folders is very high.

Local and Nonlocal Contacts

Cooperativity is a main signature of the folding mechanism. What results in the difference of the cooperativity between the downhill folders and the two-state folders? For the Gō-like model, the energetic effects are considerably optimized, while the topology of a protein is believed to be the controlling factor for the thermodynamics and kinetics of the folding.[42] The native topology is characterized by the distribution of contacts which can be divided into local contacts and nonlocal ones. A contact is classified as local (or nonlocal) contact[43,44] if the distance of the two contacted residues along the sequence, that is, L_c, is smaller (or larger) than a certain value, say five or six residues. Our criterion for this classification is defined as $L_c^c = 6$. Note that the distance along the sequence is relevant to the contact order.[43] Previously, it was argued that the nonlocal interactions dominate the cooperativity of the protein folding[44,45] because the nonlocal interactions introduce some global ordering into the system.

For proteins BBL and CI2, it is found that the nonlocal and local contacts have different distributions during the folding, and the evolutions of the degrees of these contacts formed at the transition temperature T_f are also different (see Fig. 3). For protein BBL, the local contacts with $L_c \leq 6$ and some nonlocal contacts with Lc > 8 –12 are formed quickly after a short initial stage [Fig.3(a)]. The nonlocal contacts with $L_c = 13$–22 are half formed, and those with $L_c > 23$ are basically not formed. The local and nonlocal contacts reach their equilibrium values at about 4×10^4

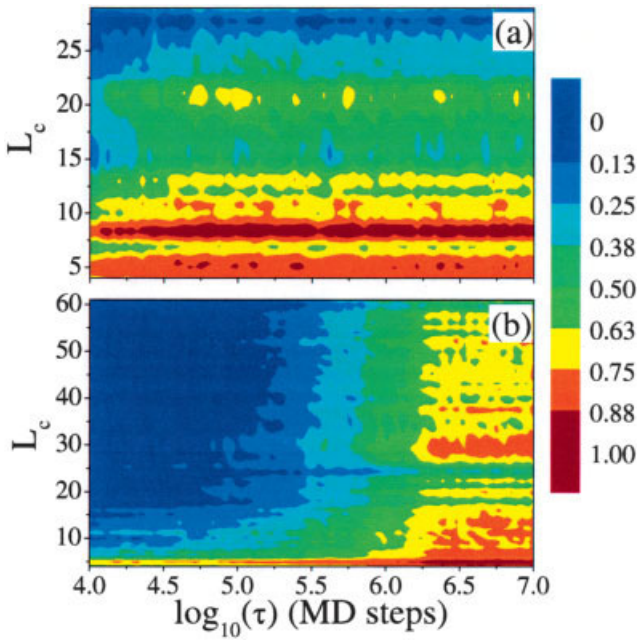

Fig. 3. The formation probabilities of contacts with different loop lengths L_c during the folding. The colors show the probabilities of contacts averaged over 100 trajectories for protein (**a**) BBL and (**b**) CI2 at their T_f.

MD steps almost simultaneously. There is basically no further formation of contacts as the time evolves. At temperature below or above T_f, the evolutions and distributions of the nonlocal contacts are similar. However, the situation is quite different for protein CI2. The local contacts are formed first, and all the nonlocal contacts begin to be formed cooperatively at about 6.2×10^6 MD steps [Fig. 3(b)]. This formation time is apparently about 150 times longer than that of the local contacts. The nonlocal contacts show an all-or-none behavior when the temperature is below or above the transition temperature T_f.

Clearly, these relate to the folding mechanisms of two kinds of proteins. It is noted that the formation of the local contacts for two proteins is similar, indicating that the formation of local contacts may be independent of the folding manners. Nevertheless, from our results, two proteins with well-defined folding mechanisms behave distinctly in their nonlocal contacts. Such differences in the features of the nonlocal contacts may act as an indicator for their folding behaviors. Furthermore, the topological structure of a protein may determine the formation of the nonlocal contacts which dominates the folding behavior.

The formation of nonlocal contacts can greatly limit the conformational searching of the protein molecule due to the geometrical restrictions. A few of nonlocal contacts may essentially outline the global topology of the chain. Therefore, the nonlocal contacts tend to make the convex shape of the landscape which induces a folding with a cooperative manner. These contacts generally relate to the so-called folding nucleus. The richness of nonlocal contacts could strengthen the cooperativity, and the deficiency of the nonlocal contacts will decrease the barrier, even waive off the barrier totally. Microscopically, the larger the number of nonlocal contacts per residue, the more stable the local packing structure is, although the packing of structural units of the protein is initiated by the formation of local contacts. Thus, the folding is not in a gradual mode, but is cooperatively progressed. However, it is found that for the downhill folder there is no nucleus, so no barriers. Some nonlocal contacts are formed early and are not broken, and some never be formed. Nevertheless, for protein CI2, the formation of nonlocal contacts is flexible with an all-or-none feature, thus a two-state behavior.

Statistics on the Native Contacts

To verify that the lack of the cooperativity for BBL is due to the deficiency of nonlocal interactions, let us calculate the statistics on the contact distribution and the folding behavior over some well-studied proteins.[46] The statistics on the number of nonlocal contacts per residue N_N, the number of local contacts per residue N_L, the ratio of them $R_{LN} = N_L/N_N$, and the related results of various features based on our simulations are listed in Table I. One can see clearly the differences between the downhill folder and other proteins. The downhill folder, such as the barrierless folder BBL, has a small number of nonlocal contacts per residues as $N_N = 0.730$, which is rather smaller than those of proteins of two-state folding or three-state folding, for example, $N_N = 1.875$ for CI2. As a result, the ratio $R_{LN} = 1.111$ of BBL is rather larger than those of the other proteins, for example, $R_{LN} = 0.250$ for CI2. It is obvious that protein CI2 has large propensity in nonlocal contacts, and protein BBL is deficient in nonlocal contacts. This statistics suggests that the number of nonlocal contact per residue is relevant to the folding mechanism of the downhill folder.

Note that the statistics listed in Table I is independent of the chain lengths. As an example, let us consider the β-hairpin, a short secondary structural element, which is rather short compared with BBL. The β-hairpin is argued to fold with two-state manner.[47] In our simulation, a barrier in the free energy profile is also found (data not shown). The statistics for the β-hairpin gives $N_N = 1$, which is slightly larger than that of the downhill folder BBL and is obviously smaller than those of other two-state folders. At the same time, the ratio R_{LN} is quite small and indicates a two-state folder. Actually, the free-energy barrier of β-hairpin is located at the region with rather small values of $Q = 0.1$.[47] Consequently, the folding trajectory is, in some sense, similar to the downhill folder after passing this barrier. In addition, some other proteins are also checked. It is found that some of these proteins would fold in a weak cooperative manner and have low barriers when their values of number of nonlocal contacts per residue N_N are around 1, such as the three-helix bundle,[48,49,50] the protein En-HD,[51–53] and so on.

Thus, whether a protein is a two-state folder or a downhill folder can easily be identified from the distribution of the contacts, namely the number of nonlocal contacts per residue, N_N, from the native conformation

TABLE I. The List of Various Statistical and Folding Features for 16 Cooperative Proteins and BBL

Index	Folding Behavior	Structure Class	Protein Name	PDB Code	Chain Length	R_{LN}	N_L	N_N	κ_2
1	Two-state	α	Monomeric λ-repressor[a]	1LMB	80	0.832	1.050	0.263	0.613
2		α	ACBP[a]	2ABD	86	0.856	1.105	1.291	0.603
3		α	Cytochrome c[a]	1HRC	104	0.457	0.769	1.683	0.590
4		α	En-HD[b]	1ENH	54	0.869	0.981	1.130	0.362
5		β	Tenascin[a]	2AIT	74	0.157	0.419	2.662	0.854
6		β	CspB[a]	1CSP	67	0.140	0.299	2.134	0.692
7		β	α-spectrin SH3[a]	1SHG	57	0.123	0.298	2.421	0.829
8		β	β-haipin[d]	–	16	0.267	0.250	0.938	0.512
9		α/β	CI2[a]	1COA	64	0.250	0.469	1.875	0.800
10		α/β	Ubiqution[a]	1UBQ	76	0.284	0.553	1.947	0.728
11		α/β	ADA2h[a]	1PBA	81	0.365	0.617	1.691	0.731
12	Three-state	α	Thre-helix Bundle[c]	1BDD	46	1.049	0.935	0.891	0.358
13		α	Barstar[a]	1BTA	89	0.526	0.899	1.708	0.762
14		β	CD2[a]	1HNG	98	0.135	0.306	2.265	0.743
15		α/β	Barnase[a]	1BNI	108	0.260	0.537	2.065	0.698
16		α/β	CheY[a]	3CHY	128	0.405	0.797	1.969	0.790
17	Downhill	α	BBL	1BBL	37	1.111	0.811	0.730	0.337

[a]Referred to Ref. 46
[b]Referred to Refs. 51–53
[c]Referred to Refs. 48–50
[d]β-haipin is the C-terminal fragment (41–56) of protein GB1.[47]
R_{LN} is the ratio of number of local contacts per residue; N_L to the number of nonlocal contacts per residue N_N; κ_2 is a value to quantify cooperativity of protein folding.[26–28] Note that protein barnase was described as a two-state folder in equilibrium denaturation,[57] while was argued as a three-state from kinetics.[58] However in our folding simulations, the results are more similar to the observation in Ref. 58 Thus, we classify protein barnase as a three-state folder, which is also consistent with the classification in Ref. 46

without any detailed folding simulations. That is, the statistics on the contacts can provide some judgment for possible folding mechanisms of proteins. Note other statistics with different definitions, for example, with different cutoff distances, for the contacts is also made (data not shown). It is found that they do not change our conclusion, only affect the value of N_N^c slightly (see following discussion).

Quantification of the Cooperativity

In order to get a quantitative relationship between N_N and the folding cooperativity, a factor κ_2 is used to measure the folding cooperativity. The results are listed in the last column of Table I. A strong correlation between N_N and κ_2 with a correlation coefficient $r = 0.878$ is found [see Fig.4(a)]. The values of κ_2 for various proteins are close to the fitting line. Clearly, protein BBL has the smallest value of N_N, that is, the lowest cooperativity among all the proteins studied. It is noted that the cooperativity of protein CI2 is high and the value of N_N is large. A plot of R_{LN} versus κ_2 is shown in Figure 4(b). It is found that the correlation coefficient is $r = -0.807$, and the points are dispersive relatively.

Thus, N_N can be used to characterize the cooperativity of the folding and is considered as an important topological parameter describing the proteins. From the cooperativity factors κ_2 and the statistics on the values of N_N shown in Table I, a crossover around $N_N = 0.9$ could be defined to separate the two-state folder and the downhill folder.

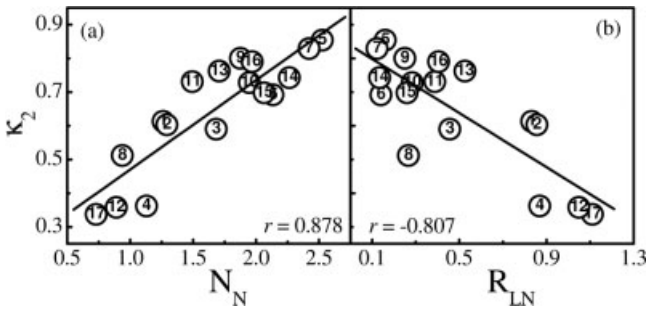

Fig. 4. The cooperativity factor κ_2 versus the number of nonlocal contact per residue N_N (**a**), and the ratio between the number of local contact and the number of nonlocal contact $R_{LN} = N_L/N_N$ (**b**) for 17 proteins listed in Table I. The number in the circle is the index of the related protein in Table I. *r* is the correlation coefficient.

Exemplification by Mutants

Finally, to further illustrate the relationship between the parameter N_N and the cooperativity and to validate the crossover of N_N for different folding behaviors, a study based on mutation is made. If the downhill folder is solely determined by the value of N_N, one can change the two-state folder into a downhill one by reducing the related value of N_N via mutations, and vice versa. Taking an example, protein monomericλ-repressor (PDB code: 1LMB) has 80 residues and is an α-class protein with its $N_N = 1.263$. The mutation for this protein is easy since its value of N_N is close to that of N_N^c (see Table I).

Using the method of computational mutations, six mutants, that is, M1 to M6, have been obtained (see Fig. 5). The N_N values of M1 to M5 are distributed between those

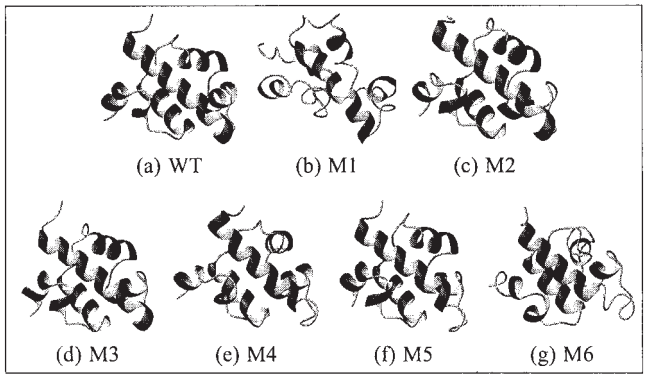

Fig. 5. The wild-type structure and the stable structures of protein monomeric λ-repressor and its six mutants. The mutations for mutant M1 are Leu50Gly, Phe51Ala, Leu65Gly, and Ile80Gly, and the mutations for mutant M2 to M6 include some of Gln33Ala, Leu50Ala, Phe51Ala, Leu57Ala, Leu65Ala, Phe76Ala, and Ile80Ala, respectively. Because only a few residues are mutated, the changes of the structures are small. The values of cRMSD between the stable structures of M1 to M6 and the structure of the wild type are 8.451 Å, 7.090 Å, 7.392 Å, 7.222 Å, 7.626 Å, and 6.858 Å, respectively.

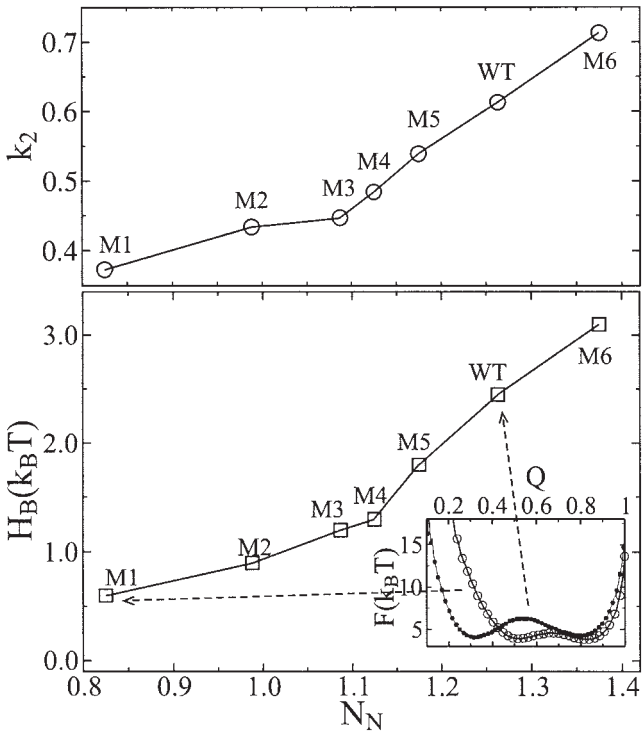

Fig. 6. The cooperativity factors κ_2 (circle) and the height of free energy barrier H_B (square) at T_f for the wild type (labeled as WT) and six mutant structures of protein Monomeric λ-repressor shown in Figure 5. The inset shows the free energy profiles for the wild type (solid circle) and the mutant M1 (circle).

of the native structure and M1 structure. Clearly, M1 has the smallest N_N value, that is, $N_N = 0.825$, and M6 has a N_N value larger than that of wild type, that is, $N_N = 1.375$. For each mutant, the related folding simulations are made using the Gō-like model, and the heights of free energy barriers at their related T_fs and cooperativity factors κ_2 are obtained (see Fig. 6). From Figure 6, it is seen that the heights of the free energy barriers and the cooperativity factors κ_2 for these six mutants decrease as the related values of N_N decrease. Note that the free energy barrier of the wild type obtained from our simulations is about 2.5 k_BT. For M1, the height of the barrier is only about 0.6 k_BT and the cooperativity is quite low with $\kappa_2 = 0.373$, indicating clearly a downhill or downhill-like folding behavior at temperature T_f. This is really consistent with our argument mentioned above. The inset in Figure 6 shows the free energy profile with a very small barrier after the mutation. Thus, we believe that when the value of N_N of a protein is smaller than N_N^c, the folding is of downhill and the height of the free energy barrier will be smaller than 1 k_BT, and when it is extremely smaller than N_N^c, such a barrier will disappear.[54] As a result, the protein can be a conformational rheostat.

Remarks on Recent Experiments

In order to make a further check, let us compare our results with recent related work. In a review article,[55] Eaton and coworkers made a summary on the fast and ultrafast folding behavior of proteins. They argued that generic N-residue single-domain proteins fold at or near the speed limit, predicted to be approximately N/100 μs by both experimental and theoretical approaches, and may have no free energy barrier. As listed in the review, indeed, the free energy barriers for some fast or ultrafast folding proteins are quite small. Thus, 12 fast folding proteins with $\tau_{\text{folding}} < 100$ μs were considered to be the candidates of downhill folder.[55] Among them, seven proteins fold very fast, with $\tau_{\text{folding}} < 10$ μs. Using our definition, their related values of N_N are calculated and plotted against the folding times τ_{folding} in Figure 7.

Among the seven ultrafast folding proteins, three [tryptophan cage (1L2Y), BBA5 (1T8J), and Villin headpiece subdomain (1VII)] have their values of $N_N = N_N^c$, and other three [albumin binding domain (1PRB), α_3 D (2A3D), and B domain of protein A (1BDC)] have their values $N_N = N_N^c$. Therefore, the former three proteins could be potential candidates of downhill or downhill-like folders because they fold very fast with $\tau_{\text{folding}} < 10$ μs and the barriers of folding are quite small as listed by Eaton and coworkers.[55] Our simulations for proteins BBA5, HP-36, and Trp-cage are also consistent with such a prediction. Differently, the later three proteins show weak two-state folding behavior and have low barriers because they are in the crossover between the two-state and the downhill folding (see Ref. 55 and the references therein). Note that as argued by Sosnick and coworkers, proteins or peptides with folding speed $\tau_{\text{folding}} < 6$ μs are not enough to be barrierless. To fold in a downhill manner, the proteins or peptides should fold extremely rapidly.[56] Here, in this work, our criterion based on the structural features of proteins may roughly outline the crossover or the boundary of the potential candidates for downhill folding if the proteins fold fast enough and their $N_N = N_N^c$. It is worthy to note that although protein Cytochrome b_{652} (PDB ID: 1QQ3) has $N_N \leq N_N^c$ and $\tau_{\text{folding}} < 10$μs, its folding rate was measured in

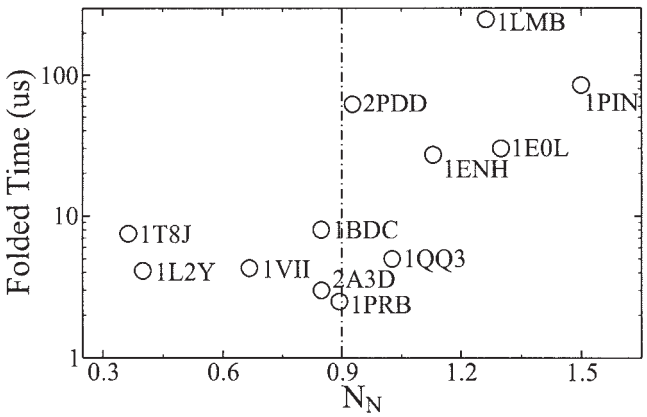

Fig. 7. The number of nonlocal contacts per residue N_N and the folding time $\tau_{folding}$ of 12 ultrafast folding proteins collected by Kubelka and colleagues.[12] The labels are the PDB identities of the protein structures. They are tryptophan cage (1L2Y), BBA5 (1T8J), villin headpiece subdomain (1VII), WW domain Pin (1PIN), WW domain FBP28 (1E0L), peripheral subunit binding (2PDD), albumin binding domain (1PRB), engrailed homeodomain (1ENH), α_3D (2A3D), λ-repressor (1LMB), B domain of protein A (1BDC), and cytochrome b_{562} (1QQ3). The dashed line indicates the rough crossover $N_N^c = 0.9$. Note that 1BDC is used here because it was discussed in Ref. 54 and its related N_N value is calculated based on the first structure among 10 structures in the file of 1BDC. Differently, in our simulations for Table I, the averaged NMR structure (corresponding to the code 1BDD of the same protein) is used.

2.2*M* GndHCl instead of the natural physiologic condition. This implies that the case may be different and complicated.

CONCLUSION

In this work, the downhill folding of protein BBL is simulated with the Gō-like model, and is compared with the two-state folding. There are obvious differences both in kinetics and in thermodynamics. The basic feature of the folding for protein BBL is of weak cooperativity, and there exists no obvious free energy barrier. The thermodynamic stable state of downhill folder, corresponded to the single minimum in the free energy landscape, changes continuously from high Q to small Q values as the temperature increases. That is to say, unlike some typical two-state proteins, downhill folder is more like a molecular rheostat. In addition, in our study the downhill folder candidates can be identified from the statistics on the features of contacts. If the number of nonlocal contacts per residue is smaller far than the crossover around 0.9, the protein can be identified as a candidate of downhill folder. On the contrast, for the two-state proteins, the nonlocal contacts are always rich, which leads to a high cooperativity of folding. Therefore, the content of the nonlocal contacts per residue could act as an important quantity to predict the folding behaviors of proteins.

ACKNOWLEDGMENT

We are thankful for suggestions by H. S. Chan and V. Muñoz. J. Wang also thanks the support of FANEDD.

REFERENCES

1. Anfinsen CB. Principles that govern the folding of protein chains. Science 1973;181:223–230.
2. Fersht AR. Characterizing transition states in protein folding: an essential step in the puzzle. Curr Opin Struct Biol 1995;5:79–84.
3. Brooks CL III. Simulations of protein folding and unfolding. Curr Opin Struct Biol 1998;8:222–226.
4. Mirny L, Shakhnovich EI. Protein folding theory: from lattice to all-atom models. Annu Rev Biophys Biomol Struct 2001;30:361–396.
5. Onuchic JN, Luthey-Schulten Z, and Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
6. Socci ND, Onuchic JN, and Wolynes PG. Diffusive dynamics of the reaction coordinate for protein folding funnels. J Chem Phys 1996;104:5860–5868.
7. Bryngelson JD, Socci ND, Onuchic JN, and Wolynes PG. Funnels, pathways and the energy landscape of protein folding. Proteins 1995;21:167–195.
8. Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 1991;30:10428–10435.
9. Shakhnovich EI, Finkelstein AV. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. Biopolymers 1989;28:1667–1680.
10. Alves NA, Hansmann UHE. Partition function zeros and finite size scaling of helix-coil transitions in a polypeptide. Phys Rev Lett 2000;84:1836–1839.
11. Wang J, Wang W. Folding transition of model protein chains characterized by partition function zeros. J Chem Phys 2003;118:2952–2963.
12. Eaton WA, Muñoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J. Fast kinetics and mechanisms in protein folding. Annu Rev Biophys Biomol Struct 2000;29:327–359.
13. Garcia-Mira MM, Sadqi M, Fischer N, Sanchez-Ruiz JM, Muñoz V. Experimental identification of downhill protein folding. Science 2002;298:2191–2195.
14. Fabiana Y, Muñoz V. A simple thermodynamic test to discriminate between two-state and downhill folding. J Am Chem Soc 2004;126:8596–9697.
15. Muñoz V, Sanchez-Ruiz JM. Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. Proc Natl Acad Sci U S A. 2004;101:17646–17651.
16. Naganathan AN, Perez-Jimenez R, Sanchez-Ruiz JM, Muñoz V. Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. Biochemisty 2005;44:7435–7449.
17. Taketomi H, Ueda Y, Gō N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effects of specific amino acid sequence represented by specific interunit interaction. Int J Pept Res 1975;7:445–459.
18. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: what determines the structural details of the transition state ensemble and en-route intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 2000;298:937–953.
19. Clementi C, Jennings PA, Onuchic JN. Prediction of folding mechanism for circular-permuted proteins. J Mol Biol 2001;311:879–890.
20. Clementi C, Jennings PA, Onuchic JN. How native-state topology affects the folding of dihydrofolate reductase and interleukin-1 β. Proc Natl Acad Sci U S A 2000;97:5871–5876.
21. Karanicolas J, Brooks CL III. Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards nonnative interactions. J Mol Biol 2003;334:309–325.
22. Koga N, Takada SJ. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. J Mol Biol 2001;313:171–180.
23. Head-Gordon T, and Brown S. Minimalist models for protein folding and design. Curr Opin Struct Biol 2003;13:160–167.
24. Kaya H, Chan HS. Salvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? J Mol Biol 2003;326:911–931.
25. Hoang TX, Cleplak M. Molecular dynamics of folding of secondary structures in Gō-type models of proteins. J Chem Phys 2000;112:6851–6862.
26. Chan HS. Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. Proteins 2000;40:543–571.

27. Kaya H, Chan HS. Polymer principles of protein calorimetric two-state cooperativity. Proteins 2000;40:637–661.
28. Chan HS, Shimizu S, Kaya H. Cooperativity principles in protein folding. Methods Enzymol 2004;380:350–379.
29. Pande VS, Rokhsar DS. Folding pathway of a lattice model for proteins. Proc Natl Acad Sci U S A 1999;96:1273–1278.
30. Fan K, Wang J, Wang W. Kinetic transition in model protein with a denatured native spinodal. Phys Rev E 2001;64:041907.
31. Zhang J, Qin M, Wang W. Multiple folding mechanisms of protein ubiquitin. Proteins 2004;59:565–579.
32. Chen J, Wang J, Wang W. Transition states for folding of circular-permuted proteins. Proteins 2004;57:153–171.
33. Levy Y, Wolynes PG, Onuchic JN. Protein topology determines binding mechanism. Proc Natl Acad Sci U S A 2004;101:511–516.
34. Stoycheva AD, Brooks CL III, Onuchic JN. Gatekeepers in the ribosomal protein s6: thermodynamics, kinetics, and folding pathways revealed by a minimalist protein model. J Mol Biol 2004;340: 571–585.
35. Guo Z, Thirumalai D. Kinetics and thermodynamics of folding of a de novo designed four-helix bundle protein. J Mol Biol 1996;263: 323–343.
36. Veitshans T, Klimov D, Thirumalai D. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. Fold Des 1997;2:1–22.
37. Ferrenberg AM, Swendsen RH. Optimized Monte Carlo data analysis. Phys Rev Lett 1989;63:1195–1198.
38. Shea JE, Brooks CL III. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 2001;52:499–535.
39. Case DA, et al. Amber. San Francisco: University of California; 2002.
40. Nomura A, Sugiura Y. Contribution of individual zinc ligands to metal binding and peptide folding of zinc finger peptides. Inorg Chem 2002;41:3493–3698.
41. Kearsley SK. On the orthogonal transformation used for structural comparisons. Acta Cryst 1989;A45:208–210.
42. D. Baker. A surprising simplicity to protein folding. Nature 2000;405:39–42.
43. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain protein. J Mol Biol 1998;277:985–994.
44. Chan HS. Matching speed and locality. Nature 1998;392:761–763.
45. Abkevich VI, Gutin AM, Shakhnovich EI. Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. J Mol Biol 1995;252:460–471.
46. Jackson SE. How do small single-domain proteins fold? Fold Des 1998;3:R81–R91.
47. Muñoz V, Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanism of β-hairpin formation. Nature 1997;390:196–199.
48. Berriz GF, Shakhonovich EI. Characterization of the folding kinetics of a three-helix bundle protein via a minimalist langevin model. J Mol Biol 2001;310:673–685.
49. Bottomley SP, Popplewell AG, Scawen M, Wan T, Sutton BJ, Gore MG. The stability and unfolding of an IgG binding protein based on the bdomain of protein a from staphylocaoccusaureus probed by tryptophan substitution and fluorescence spectroscopy. Protein Eng 1994;7:1463–1470.
50. Zhou YQ, Karplus M. Interpreting the folding kinetics of helical proteins. Nature 1999;401:400–403.
51. Mayor U, Guydosh NR, Johnson CM, Grossmann JG, Sato S, Jas GS, Freund SMV, Alonsok DOV, Daggett V, Fersht AR. The complete folding pathway of a protein from nanoseconds to microseconds. Nature 2003;421:864–867.
52. Mayor U, Johson CM, Daggett V, Fersht AR. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. Proc Natl Acad Sci U S A 2000;97:13518–13522.
53. Mayor U, Grossmann JG, Foster NW, Freund SMV, Fersht AR. The denatured state of engrailed homodomain under denaturing and native conditions. J Mol Biol 2003;333:977–991.
54. It is found that there is a conformation with rather small value of N_N = 0.725 among all candidates for mutant M1 during the computational mutation. However, this conformation has the energy a little higher than that of the conformation shown in Figure 5 for M1, and is not chosen as the stable structure of M1. If we still define it as the stable structure, its folding is found to be of barrierless based on the folding simulations. This clearly indicates that the folding is of barrierless when the N_N value of a protein is smaller far than N_N^c.
55. Kubelka J, Hofrichter J, Eaton WA. The protein folding speed limit. Curr Opin Struct Biol 2004;14:76–88.
56. Meisner WK, Sosnick TR. Barrier-limited, microsecond folding of a stable protein measured with hydrogen exchange: implications for downhill folding. Proc Natl Acad Sci U S A 2004;100:15639–15644.
57. Taybovsky YI, Kedrov AA, Martsev SP. Independent folding and conformational changes of the barnase module in the VL-barnase immunofusion: calorimetric evidence FEBS Lett 2004;557:248–252.
58. Matouschek A, Kellis JT Jr, Serrano L, Bycroft M, Fersht AF. Transient folding intermediates characterized by protein engineering. Nature 1990;346:440–445.