

Kinetic transition in model proteins with a denatured native spinodal

Jun Wang, Ke Fan, and Wei Wang*

National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China

(Received 7 November 2001; published 11 April 2002)

The relaxation kinetics of three-dimensional lattice model proteins with $G\bar{o}$ potential is studied. A kinetic transition from an exponential behavior to a nonexponential one with a denatured native spinodal is characterized. The transition temperatures T_k , obtained from simulations and a semiquantitative estimation, are found to be the same. The change in free energy landscape during the transition is discussed microscopically by studying the detailed folding processes of various paths. The connection of T_k with the foldability is also studied by a Z -score-like quantity T_f/T_k .

DOI: 10.1103/PhysRevE.65.041925

PACS number(s): 87.10.+e, 87.15.He

Proteins move on a rough energy landscape resulting from their heteropolymeric nature and their huge number of conformations [1]. A thermodynamic folding transition from disordered denatured states to an ordered native-dominated state occurs at temperature T_f [2]. Below T_f , the relaxation of the system toward the native state on the landscape is complex and, thus, rich in physics. Recently, many studies have shown the existence of nonexponential behavior [3–12]. Such behavior is either suggested to relate to a glassy kinetics with freezing phenomenon [3–6], kinetic traps [7], or argued to be a downhill folding kinetics [8–10]. Anyhow, a kinetic transition of relaxation from an exponential form to a nonexponential one does exist. However, what kind of exact scenario happens during the kinetic transition is still not clear. The physical nature and microscopic mechanism of the kinetic transition need to be clarified. The underlying reasons of the change of folding kinetics may provide us some insights into the mechanism of protein folding.

In this paper, we report a simulation study on the kinetic transition and the relaxation of model proteins in a wide temperature range below T_f . Some evidence is presented to show the existence of a denatured native spinodal at temperature T_k . Above T_k , the system exhibits a coexistence of the flexible liquidlike denatured component and the rigid solidlike native state. Below T_k , the metastable denatured component of the free energy landscape changes to an unstable one, and the whole free energy landscape becomes a single funnel that makes the folding a downhill kinetics. The kinetic transition from an exponential behavior to a nonexponential one thus may be relevant to the spinodal rather than a glass transition due to the absence of freezing. This settles the nature of the kinetic transition, suggesting the important role of the diffusive feature of the model proteins. The correlation between a Z -score-like quantity T_f/T_k and the folding time is also studied, which shows the effect of the kinetic transition on the folding of proteins.

In the landscape, at high temperature with $T \leq T_f$, the main barrier, resulting from the conformational entropy, between the denatured states and the native state is high with respect to the temperature [see Figs. 1(a) and 1(b)]. The two-state barrier-crossing events induce an exponential kinetic

behavior. Conversely, at low temperature, the barrier mainly results from the energetics of the protein [see Figs. 1(c) and 1(d)]. Thus, the relaxation is complex and shows a nonexponential kinetic behavior due to nonuniform distribution of various energy barriers. Physically, there might be a transition between these two kinetic behaviors at a temperature T_k , and this transition is settled as the diminishing of the effective barrier from conformational entropy [see Fig. 1(c)].

Let us make a study of three-dimensional lattice model proteins with $G\bar{o}$ potential [13], i.e., the native contacts with the interactions ϵ_n are stronger than the non-native ones ϵ_{nn} . The model proteins with $G\bar{o}$ potential are argued to be a minimally frustrated one [14] since in this model the interactions between monomers that form native contacts are attractive. That is, after setting a native structure, the contacts defined as the native contacts are all assumed to have the same interaction potential with ϵ_n , and the $G\bar{o}$ potential is the main driving force for the folding of the model chain. It is noted that such a $G\bar{o}$ model has become popular in recent years (see Ref. [15] for a review, and references therein). The Metropolis Monte Carlo (MC) algorithm [16] is used. To obtain long-time relaxation, the time threshold is set at 8×10^9 Monte Carlo steps. About 5000 runs are processed at each temperature, and the units of energy and temperature are set as the value of ϵ_n .

The relaxation behavior is studied based on the residual probability of denatured states $P_d(t) = \int_t^\infty f(t') dt'$ [6,12] where $f(t)$ is the distribution of the first passage time (FPT) τ_f for reaching the native state. At high temperature, the system relaxes in an exponential manner, $P_d(t) \sim e^{-kt}$ with k the folding rate, while at low temperature, the residual probability $P_d(t)$ has a nonexponential form, but its exact form is still under study. Intuitively, the nonexponential relaxation

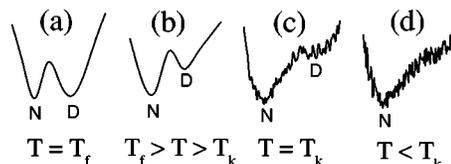


FIG. 1. The free energy landscapes at different temperatures. T_f marks the folding transition from disordered denatured state to ordered native state, and T_k marks the kinetic transition from metastable denatured state to unstable denatured state.

*Email address: wangwei@nju.edu.cn

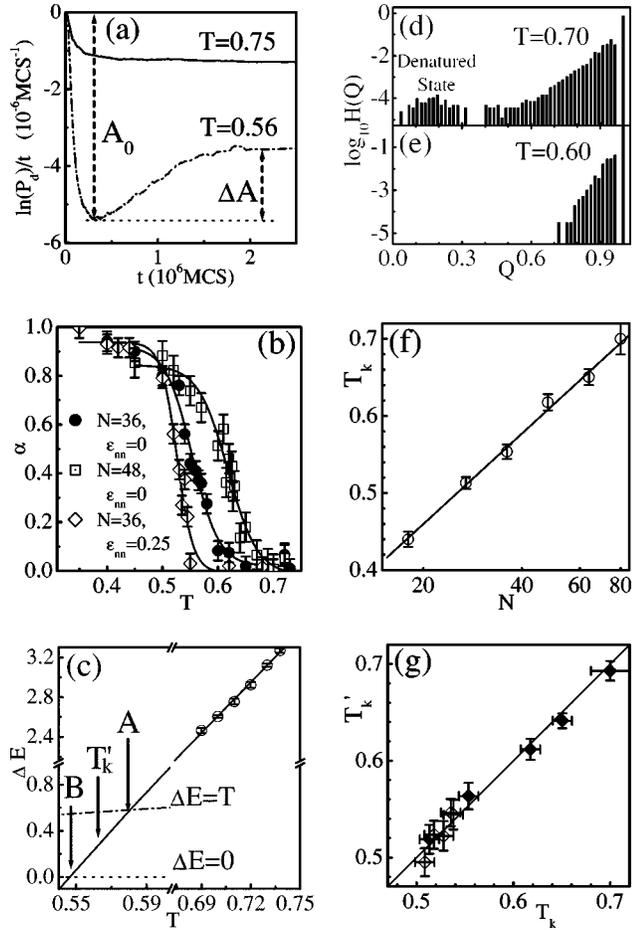


FIG. 2. (a) Time dependence of the slope $\ln[P_d(t)]/t$. A_0 defines the minimum after initial fast drop and ΔA defines the difference between long-time saturation and the minimum A_0 . (b) Bend factor α vs temperature T , for three different chains. (c) Effective barrier ΔE vs temperature T . The solid line is a linear fit. Point A and point B mark the beginning and the end of the transition region, and T'_k defines the transition. (d) and (e) Histogram statistics of states $H(Q)$ along Q for a chain with $N=48$. (f) The length dependence of transition temperature (N in logarithmic scale). (g) The correlation between T_k and T'_k . The solid diamonds represent data for different chains with $N=18, 27, 36, 48, 64, 80$ with $\epsilon_{nn}=0$. The open diamonds are for the chains $N=36$ with different $\epsilon_{nn}=0, 0.1, 0.15, 0.2, 0.25, 0.3$, respectively.

shows a bend in $\ln(P_d)$ time t , which can be characterized by the variation of its slope $\ln(P_d)/t$. Quantitatively, a factor $\alpha = \Delta A/A_0$ can be used to define the bend of the relaxation $P_d(t)$ against t . Here the quantities ΔA and A_0 are defined in Fig. 2(a). For an exponential decay, there is a saturation of $\ln(P_d)/t$ in short time, which corresponds to $\alpha=0$ [the solid line in Fig. 2(a)]. Conversely, a nonexponential decay generally has an increase in the slope $\ln(P_d)/t$, which gives out a nonzero ΔA and, thus, α [the dotted line in Fig. 2(a)]. Clearly our description of the quantity α does not depend on the exact forms of the nonexponential decays. Figure 2(b) shows the bend factor α temperature for three cases. Obviously, there is a transition of α from about 0 at high temperature to about 1 at low temperature. That is, a transition

occurs for the kinetics of the system from an exponential behavior to a nonexponential one, as also implied in Refs. [4,6,7] with various implementation. The transition temperature T_k is defined as the median one in the curves. It is noted that there is still such a transition by fitting the data with the stretched exponential form. However, the stretched exponential form is not good for all the data with different temperatures, especially for those long-time relaxations at temperatures far below T_k . In addition, the mean first passage time, i.e., $\langle \tau_f \rangle$, follows the Arrhenius behavior of temperature dependence (data not shown) below T_k [17,18] though the relaxation becomes nonexponential [18].

Actually, as shown in Fig. 1, the height of barrier decreases as the temperature decreases. At high temperatures, the effective barrier ΔE can be estimated from the folding time $\langle \tau_f \rangle = \tau_0 \exp(\Delta E/T)$, where τ_0 is selected practically as the shortest folding time during all simulations. Above T_k , entropy-related property makes the barrier ΔE varies linearly with respect to the temperature [see the points in Fig. 2(c)]. The kinetics starts to become nonexponential when the height of the barrier equals to the temperature, i.e., $\Delta E \sim T$, and, when the barrier diminishes, the transition may saturate. Thus, a transition region can be estimated as the region between two intersections of $\Delta E = \Delta E(T)$ with the diagonal line $\Delta E = T$ [Point A in Fig. 2(c)] and the horizontal line $\Delta E = 0$ [Point B in Fig. 2(c)]. The median value of the temperature of this region gives out a good evaluation of transition temperature T'_k [Fig. 2(c)]. The diminishing of the apparent barriers implies the cease of stability of the denatured state. As a result, this kinetic transition can be understood as a denatured native spinodal in protein system due to a similar picture to the liquid-gas spinodal [19]. Below T_k , denatured state cannot exist as a metastable phase such as supercooled liquid. Moreover, the nonexponential kinetics directly relates to the downhill feature of free energy landscape at low temperature. To illustrate the spinodal behavior, two histograms for states with different native similarities Q [20] are shown [see Figs. 2(d)–2(e)]. Above T_k , there is a small peak at $Q \approx 0.2$ for the denatured state, suggesting the existence of the denatured state as a local minimum. Below T_k , the peak disappears, implying the instability of the denatured state.

The size dependence of T_k is also studied. All model chains with different sizes $N=18, 27, 36, 48, 64, 80$ have the similar transitions, and T_k increases with the chain size logarithmically, $T_k \propto \ln N$ [see Fig. 2(f)]. This relationship is similar to the increase of T_{min} in other works [21], and may be attributed to the increase of the energetic effect since larger proteins generally have larger energetic barriers. In addition, a good correlation between the evaluated values T'_k and the simulated values T_k of the transition temperatures clearly supports our understanding on the change of the landscapes [see Fig. 2(g)].

Now, let us interpret the folding processes microscopically. The model chain moves in the state space from a random coil state to the native state. A path can be depicted by a series of states and their links. The probability for the system to follow such a path is $P_{path} = \prod_{link} P_{link}$, with P_{link} defining the probability of choosing a specific link [22,23]

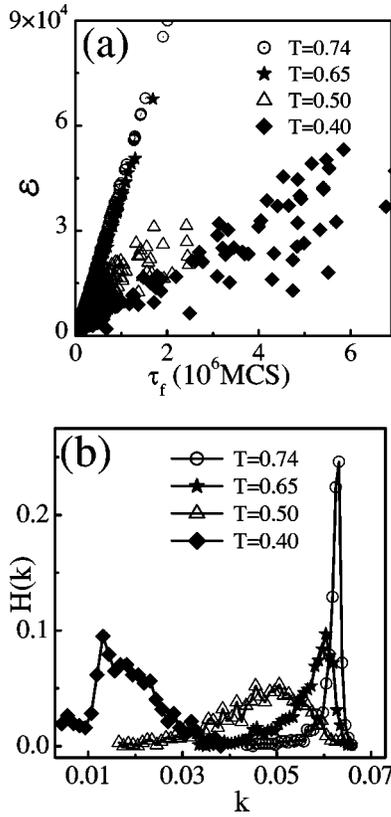


FIG. 3. (a) The distribution for \mathcal{E} on folding time τ_f for different paths of an $N=36$ chain. (b) Histogram statistics $H(k)$ of $k(=\mathcal{E}/(\tau_f T))$.

and a normalization $\sum_{path} P_{path} = 1$ for all paths. In the MC simulations, P_{link} can be expressed as P_{MC}/z , where P_{MC} is the probability for a corresponding step and z equals the number of all possible moves for all sites in the chain from present conformation to others in the conformational space. Here z , related to the chain length, is generally similar for different states. Note that the probabilities in P_{MC} for uphill steps with barriers $\epsilon_i > 0$ contribute largely to P_{link} , and those to stay in a state have an effect to decrease z , resulting in a small effective factor $z_{eff} < z$ [24]. Considering the summation of all local barriers $\mathcal{E} = \sum \epsilon_i$ and the folding time τ_f along a path as main features of the folding processes, we have $P_{link} \sim \exp(-\epsilon_i/T)/z_{eff}$, then $\sum_{path} P_{path} = \sum_{path} \prod_{link} (e^{-\epsilon_i/T}/z_{eff}) = \sum_{path} \exp(-\mathcal{E}/T - \tau_f \ln z_{eff}) = \sum_{\mathcal{E}, \tau_f} n(\mathcal{E}, \tau_f) \exp(-\mathcal{E}/T - \tau_f \ln z_{eff}) = 1$, where $n(\mathcal{E}, \tau_f)$ defines the density of paths (DOP) for certain \mathcal{E} and τ_f . It is assumed that the most probable path has a probability \mathcal{P}_m much larger than other paths, that is, $\mathcal{P}_m \sim n'(\mathcal{E}, \tau_f) \exp(-\mathcal{E}/T - \tau_f \ln z_{eff}) \approx 1$ with a DOP $n'(\mathcal{E}, \tau_f)$. At high temperatures, the paths with larger DOP take an important role in the folding. For this case, the DOP can be estimated as $n'(\mathcal{E}, \tau_f) \sim z^{\tau_f}$ resulted from the number of all possible paths within τ_f steps of random walks with z kinds of choices for each site in state space. Therefore, the most probable path satisfies $\mathcal{P}_m \sim \exp(-\mathcal{E}/T + \tau_f \ln z') \approx 1$ with $z' = z/z_{eff} > 1$, that is, one has $\mathcal{E}/\tau_f \approx T \ln z' \propto T$. Thus, there is a linear correlation between the total barriers \mathcal{E} and τ_f as veri-

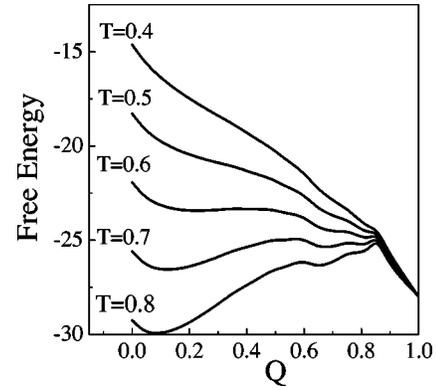


FIG. 4. The calculated free energy (F) landscapes at different temperatures for a 27-mer chain with Gō-type interaction. The entropy for different Q are obtained by entropic sampling. The entropy of the native state (with $Q=1$) is set as 0 due to its uniqueness. The similar behavior to Figs. 1(a)–1(d) can be clearly seen.

fied in our simulations [see Fig. 3(a)]. Generally, the probabilities of different paths P_{path} decay with the increase of folding time, $P_{path} \sim \exp(-k\tau_f)$. Here, the quantity k resembles the effective rate of folding, $k \sim \mathcal{E}/(\tau_f T)$. [Note that a constant $k_0 = \ln z_{eff}$ is omitted due to its independent of paths.] At high temperatures, since $k = \mathcal{E}/(\tau_f T) = \ln z'$ is a constant, there is an exponential distribution of the FPTs. Thus, the system behaves exponentially. A histogram statistics on the rate k shows a sharp peak at high temperatures [see Fig. 3(b)], which reveals the existence of the homogeneity for various diffusive processes. All these show the intrinsic connection between the barrier and the entropy of system, supporting the view that the entropic contribution to folding barrier is important in high temperature region. While at low temperatures, the paths with small \mathcal{E} are more preferable. These paths depend on the details of the energy landscape, and there may be no explicit relation between \mathcal{E} and τ_f [see Fig. 3(a)]. The correlation between \mathcal{E} and τ_f is poor and different ratios of \mathcal{E}/τ_f may appear for various paths, implying that k is not a constant. As a result, different decay times appear, which leads to the nonexponential behavior. As shown in Fig. 3(b), the distribution of the rate k becomes broad, and its peak shifts to low value as T goes below the transition. Thus, the broad distribution of the folding rate suggests the inhomogeneity of the landscape since different paths have distinct diffusion rates. Indeed, at high temperature, proteins experience more and larger barriers than they do at low temperature. Yet, as the temperature lowers down, the decrease of diffusion rate makes the folding times longer. This relates to the argument for Fig. 1. Therefore, a homogeneous diffusion on the landscape at high temperatures can be implied from the linear relation between \mathcal{E} and τ_f , whereas the diverse values of \mathcal{E}/τ_f indicate an inhomogeneous relaxation of proteins on the landscape at low temperatures. Such features, resulted from the homogeneous and inhomogeneous free energy landscape, are the underlying reasons for different kinetic behaviors.

At low temperatures, the native state of proteins occupying a large part of the population is also found, which shows the rigidity of the folded proteins. Thus, the protein behaves

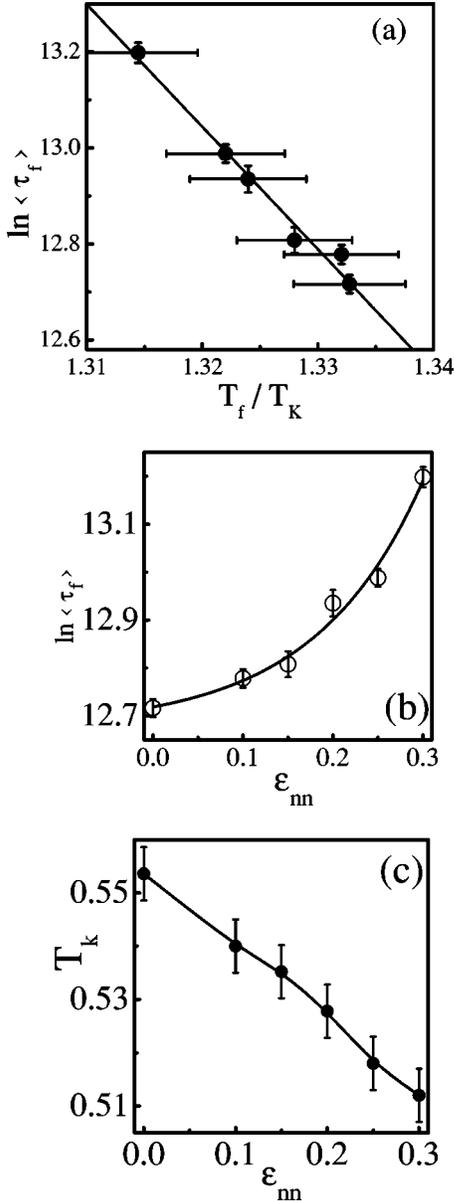


FIG. 5. (a) The folding times $\langle \tau_f \rangle$ (at $\langle Q \rangle = 0.95$) vs a Z-score-like quantity T_f/T_k for different values of $\epsilon_{nn} = 0$ to 0.3 (from bottom to top) for chains with $N = 36$. (b) The folding time $\langle \tau_f \rangle$ vs frustration ϵ_{nn} . (c) The transition temperature T_k vs frustration ϵ_{nn} .

as an ordered crystal rather than a polymorphic glassy liquid. Moreover, in our simulations, the chain folds along different paths to the native state even for $T \ll T_k$, and few paths are frozen in traps. Such a diffusive kinetics is also found in some other off-lattice simulations [12]. These suggest that the nonexponential behavior probably results not from the freezing in landscape, but from the downhill folding. This diffusive behavior may be relevant to the weak average attraction in $G\bar{o}$ model that is similar to real proteins, especially to the small single-domain proteins. The weak attraction prefers the spinodal to the glass transitions [19].

It is well known that the energy landscape of a protein is of multidimensions, and also is abundant of local energy minima [25,26]. For such a kind of complicated object, our

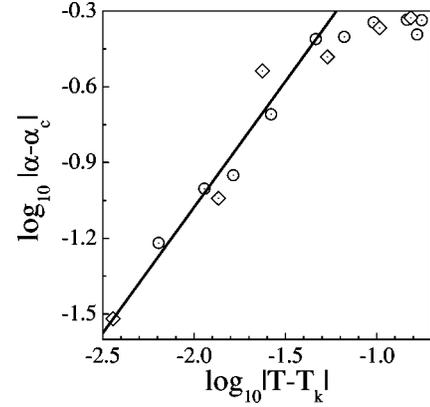


FIG. 6. The scaling relation of α vs the temperature T near the transition T_k for the 36-mer chain with $\epsilon_{nn} = 0$.

present knowledge only allows us to have a coarse-grained view on its topological structure. As a kind of approach, in Fig. 4 we plot the free energy profiles with respect to the Q coordinate from our simulations. Here, the quantity Q defines the similarity of a contact map C with native contact map $C^{(nat)}$ as we used in Fig. 2(d) and Fig. 2(e) (for the definition, for example, see Ref. [1]). From Fig. 4 we see that the free energy profiles really relate to our argument in Fig. 1. As the temperature T decreases, the main barrier disappears and the free energy profiles become downhill. Although the simplified reaction coordinate Q is used, the profiles shown in Fig. 4 characterize approximately the direction of the folding processes. Previously, changes of landscapes for different temperatures are speculated to interpret the experiments [8,9]. Here our studies on the features of landscapes show the existence of these landscapes and may provide a theoretical explanation for these experiments.

Now, let us discuss the foldability. For temperatures T with $T_k < T < T_f$, the apparent barriers make proteins fold cooperatively, i.e., all molecules fold in short times. Thus, a large interval between T_f and T_k may benefit the folding rates. To show this, the correlation between a Z-score-like quantity T_f/T_k and the folding time $\langle \tau_f \rangle$ is studied for 36-monomer chains with different interaction strengths of non-native bonds, ϵ_{nn} , which introduces some kind of energetic frustrations. T_f is identified as the temperature with largest variation of Q . From Fig. 5 one can see that large value of T_f/T_k has fast folding, and large frustration induces slow folding and low T_k . Generally, the non-native contacts enlarge the denatured state ensemble, and it needs lower temperature to destabilize these denatured states by decreasing the entropic contribution. Our results agree to such a general physical comprehension. It is also worthy to note that a scaling near the transition T_k , $|\alpha - \alpha_c| \sim |T - T_k|^\sigma$, with $\sigma \approx 0.5 \pm 0.2$ is approximately obtained, which indicates a phase-transition-like feature. This scaling behavior is shown in Fig. 6, but not all points fall in the scaling line.

Finally, as a remark, our studies propose an idea for exploring the change of the landscape based on the kinetic data in the high temperature region [as in Fig. 2(c)]. That is to say, for protein systems, the knowledge obtained in the high temperature region may be helpful to find some information

of the kinetics at low temperatures. The kinetic transition would be one such application. Of course, our prediction on the kinetics transition could be tested by further experiments.

In conclusion, we characterize the kinetic transition from exponential relaxations to nonexponential ones. This kind of transition occurs at the disappearance of the entropic barrier as temperature decreases, which resembles a denatured native spinodal, rather than a kinetic glass transition. The spinodal picture may be more suitable for understanding the kinetic transition for optimal proteins. We note that a similar

behavior has also been found in the off-lattice model of proteins; we will report on this elsewhere. Additionally, we find the stretched exponential fitting only for short chains, e.g., $N=36$, with the stretched exponent $\beta \approx 0.5$.

This work was supported by the National Natural Science Foundation of China (Grant No. 90103031, No. 10074030, and No. 10021001), and the Nonlinear Project of the NSM, and partly the Ke-Li Research Foundation. We thank H.S. Chan for helpful comments on the manuscript.

-
- [1] J.N. Onuchic, Z. Luthey-Schulten, and P.G. Wolynes, *Annu. Rev. Phys. Chem.* **48**, 545 (1997).
- [2] C.B. Anfinsen, *Science* **181**, 223 (1973).
- [3] H. Frauenfelder, S.G. Sligar, and P.G. Wolynes, *Science* **254**, 1598 (1991).
- [4] H. Nymeyer, A.E. Garcia, and J.N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5921 (1998).
- [5] M. Skorobogatiy, H. Guo, and M. Zuckermann, *J. Chem. Phys.* **109**, 2528 (1998).
- [6] N.D. Socci, J.N. Onuchic, and P.G. Wolynes, *Proteins: Struct., Funct., Genet.* **32**, 136 (1998).
- [7] H.S. Chan and K.A. Dill, *Proteins: Struct., Funct., Genet.* **30**, 2 (1998).
- [8] J. Sabelko, J. Ervin, and M. Gruebele, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6031 (1999).
- [9] M. Gruebele, *Annu. Rev. Phys. Chem.* **50**, 495 (1999).
- [10] D.J. Bicout and A. Szabo, *Protein Sci.* **9**, 452 (2000).
- [11] R. Metzler *et al.*, *Chem. Phys. Lett.* **293**, 477 (1998).
- [12] G. Hummer, A.E. García, and S. Garde, *Phys. Rev. Lett.* **85**, 2637 (2000).
- [13] N. Go, *Annu. Rev. Biophys. Bioeng.* **12**, 183 (1983).
- [14] J.D. Bryngelson and P.G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7524 (1987).
- [15] R. Guerois and L. Serrano, *Curr. Opin. Struct. Biol.* **11**, 101 (2001).
- [16] N.D. Socci and J.N. Onuchic, *J. Chem. Phys.* **101**, 1519 (1994).
- [17] M. Cieplak *et al.*, *Phys. Rev. Lett.* **80**, 3654 (1998).
- [18] A. Gutin *et al.*, *J. Chem. Phys.* **108**, 6466 (1998).
- [19] S. Sastry, *Phys. Rev. Lett.* **85**, 590 (2000).
- [20] N.D. Socci, J.N. Onuchic, and P.G. Wolynes, *J. Chem. Phys.* **104**, 5860 (1996).
- [21] M. Cieplak *et al.*, *Phys. Rev. Lett.* **83**, 1684 (1999).
- [22] J. Wang, J. Onuchic, and P.G. Wolynes, *Phys. Rev. Lett.* **76**, 4861 (1996).
- [23] H.S. Chan and K.A. Dill, *J. Chem. Phys.* **100**, 9238 (1994).
- [24] In our case, $P_{path} = (P_u)^{\tau_u} (P_s)^{\tau_s} / z^{\tau_f}$. Here, P_u means the MC probability for a step with energy increase and P_s is the one staying in the same state with $P_s = z - \sum_{move} P_{MC}$. As a result, we have $P_{path} = (P_u)^{\tau_u} / z_{eff}^{\tau_f}$ with $z_{eff} = z / P_s^{\tau_s / \tau_f}$. Generally, since $P_s > 1$ and $\tau_s / \tau_f \sim O(1)$ (only depending on T) for different paths, we always have $z_{eff} < z$ as verified in our simulations.
- [25] K.A. Dill and H.S. Chan, *Nat. Struct. Biol.* **4**, 10 (1997).
- [26] H. Frauenfelder and D.T. Leeson, *Nat. Struct. Biol.* **5**, 757 (1998).