THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Configuration-Dependent Diffusion Dynamics of Downhill and Two-State Protein Folding

Weixin Xu,[†,‡,§] Zaizhi Lai,[†] Ronaldo J. Oliveira,[∥,⊥] Vitor B. P. Leite,[∥] and Jin Wang*[,†,¶]

[†]Department of Chemistry and Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794, United States

[‡]State Key Laboratory of Precision Spectroscopy, Department of Physics, East China Normal University, Shanghai 200062, China

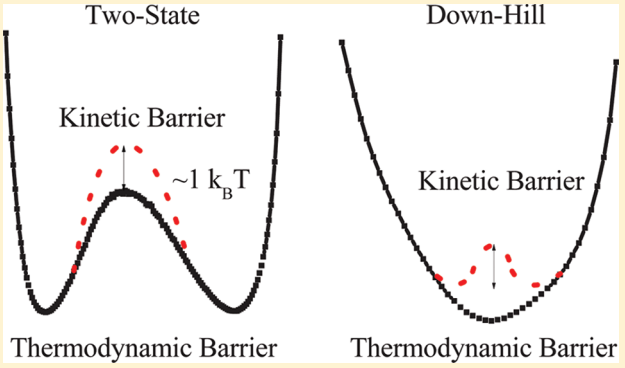
[§]Institute of Theoretical and Computational Science, Institutes for Advanced Interdisciplinary Research, East China Normal University, Shanghai 200062, China

[∥]Departamento de Física at Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, SP 15054-000, Brazil

[⊥]Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Campinas, SP 13083-970, Brazil

[¶]State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin 130021, China

ABSTRACT: Configuration-dependent diffusion (CDD) is important for protein folding kinetics with small thermodynamic barriers. CDD can be even more crucial in downhill folding without thermodynamic barriers. We explored the CDD of a downhill protein (BBL), and a two-state protein (CI2). The hidden kinetic barriers due to CDD were revealed. The increased ~1 k_BT kinetic barrier is in line with experimental value based on other fast folding proteins. Compared to that of CI2, the effective free-energy profile of BBL is found to be significantly influenced by CDD, and the kinetics are totally determined by diffusion. These findings are consistent with both earlier bulk and single-molecule fluorescence measurements. In addition, we found the temperature dependence of CDD. We also found that the ratio of folding transition temperature against optimal kinetic folding temperature can provide both a quantitative measure for the underlying landscape topography and an indicator for the possible appearance of downhill folding. Our study can help for a better understanding of the role of diffusion in protein folding dynamics.



1. INTRODUCTION

The energy landscape theory[1−3] has been an invaluable theoretical framework for understanding protein folding,[4−7] oligomerization,[8−11] and functional transitions.[12−14] According to this theory, the energy landscape associated with protein folding lacks large energetic traps and has an overall funnel shape where the native ensemble is the lowest energy state. Nevertheless, to reach the native state, proteins generally must overcome the underlying free-energy barriers, often giving rise to simple exponential kinetics for small proteins. In contrast to classical chemical kinetics, theory predicts that proteins may undergo downhill folding without an activation barrier under certain thermodynamic conditions.[15]

Recently, Muñoz and co-workers observed experimentally such a downhill folding behavior, for the protein BBL (PDB code: 1BBL).[16] They showed that the folding cooperativity of BBL is low and that there is no free-energy barrier by using differential scanning calorimetry (DSC), far-UV (UV) circular dichroism (CD), and fluorescence resonance energy transfer (FRET) experiments. Subsequently, more and more works were involved and contributed to the downhill protein folding regime,[17−23] although with different findings.[24,25]

The essential features of protein folding kinetics can often be captured approximately by diffusion on a low-dimensional free-energy surface. Such diffusion dynamics are usually dependent on configurations due to environmental changes along the energy landscape. The folding dynamics of small single-domain proteins with small thermodynamic barriers were successfully modeled by diffusion along an appropriate reaction coordinate, such as the fraction of native contacts.[26] This was verified by a recent experimental study that showed the kinetics of all the fast folding proteins discovered could be interpreted quantitatively as diffusion on a one-dimensional free-energy surface.[27] As for the above barrierless downhill folding protein BBL, the diffusion dynamics should deserve even more attention since the protein has no thermodynamic barrier to

Received: December 15, 2011
Revised: April 11, 2012
Published: April 12, 2012

   dx.doi.org/10.1021/jp212132v | J. Phys. Chem. B 2012, 116, 5152−5159

overcome along the free-energy surface; thus, diffusion determines the folding kinetics.

The diffusion dynamics can be well characterized by the diffusion coefficient (D) that is dependent on the underlying local energy barriers along the folding landscape topography. Diffusion coefficient measures the ability of local escape. It is quantified by the local escape rate or time from a particular location of the underlying energy landscape. So diffusion coefficient is determined by the local surrounding landscapes and reflects the local topography of the energy landscape. So, the diffusion calculated is the small length-scale average over local roughness. In practice, i.e., by molecular simulations, one measures the diffusion coefficient by the correlation time of the variables of interest. Since the correlation time is related to the escape time through a fluctuation dissipation theorem, the longer the correlation, the longer the local escape time. The local free-energy barrier along the landscape toward folding can be quite different, which leads to different diffusion coefficients. Many recent efforts have attempted to characterize D via experimental methods[28−35] and theoretical calculations.[26,36−41] As an example, the first experimental study showed that conformational dynamics slowed down with the degree of protein collapse using laser T-jump on a denatured protein.[42] These studies have found that diffusion is not constant as a protein folds to its native state, and that can give rise to kinetic barriers in addition to the thermodynamic barriers.[26,38−40]

In order to address how diffusion affects downhill protein folding, we calculated configurational diffusion coefficient by means of coarse-grained structure-based molecular dynamics simulations and compared the findings to a fast two-state folder CI2. The position dependence of the diffusion coefficient along different reaction coordinates (Q, R_g, and rmsd) was investigated. We found that diffusion coefficients for coordinates that directly probe fluctuations in Cartesian space, such as R_g and rmsd, have significant position dependence. Further, by considering the kinetic barrier induced by diffusion, the reconstructed effective free-energy profiles along both R_g and rmsd reveal kinetic intermediates that are hidden in thermodynamic analysis without diffusion for barrierless BBL. This provides evidence of the degree of roughness present in BBL. In addition, we found the temperature dependence of diffusion. It is also found that the ratio of folding transition temperature against optimal kinetic folding temperature can provide both a quantitative measure for the underlying landscape topography and an indicator for the possible appearance of downhill folding. It is hoped that the theoretical explorations presented here will contribute to a more complete understanding of the interplay between thermodynamics and diffusion on protein folding kinetics, especially for downhill folders.

2. MODELS AND METHODS

2.1. Proteins Studied. The proteins selected for this study are a two-state folder Chymotrypsin Inhibitor-2 with 64 amino acids (CI2, PDB code: 1COA) and a downhill folder BBL with 37 amino acids (PDB code: 1BBL), of which the cartoon structures are shown in Figure 1. Simulations for these two small proteins are maneuverable.[18,43−45]

2.2. Structure-Based C_α Model. The energy landscape theory can be idealized as being devoid of energetic roughness, which enables the use of structure-based models.[6,11,43,46] In this work, proteins are coarse-grained in a C_α level of simplification, and the Hamiltonian that gives the interaction energy is

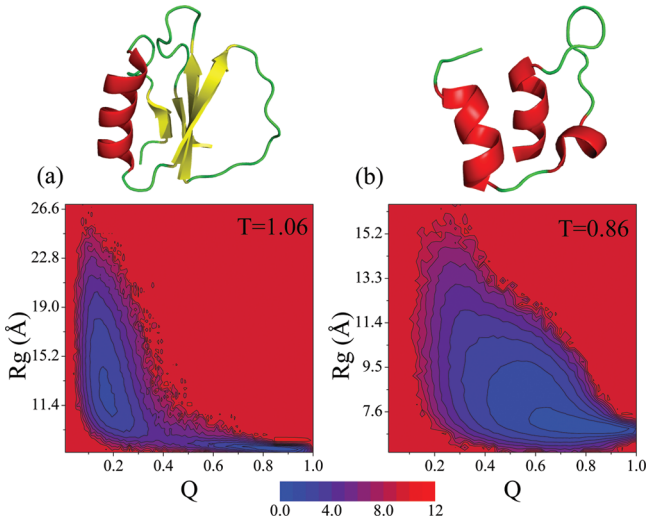


Figure 1. Contour plots of the free-energy landscape at folding transition temperatures projected onto various reaction coordinates, for (a) two-state protein CI2 and (b) downhill protein BBL, respectively.

structure-based. The following approaches are identical to the methods of ref 38. Proteins are modeled with residues represented as single beads located at the C_α-atom positions. In this model, the underlying interactions are biased according to native structure. The functional form of the potential for a given structure Γ with respect to its native structure Γ_o is

$$V(\Gamma,\Gamma_o) = \sum_{\text{bonds}} \epsilon_r(r-r_o)^2 + \sum_{\text{angles}} \epsilon_\theta(\theta-\theta_o)^2 + \sum_{\text{dihedrals}} \epsilon_\phi\left\{[1-\cos(\phi-\phi_o)] + \frac{1}{2}[1-\cos(3(\phi-\phi_o))]\right\} + \sum_{\text{contacts}} \epsilon_C\left[5\left(\frac{\sigma_{ij}}{r}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r}\right)^{10}\right] + \sum_{\text{non-contacts}} \epsilon_{NN}\left(\frac{\sigma_{NN}}{r}\right)^{12} \quad (1)$$

where $\epsilon_r = 100\epsilon_0$, $\epsilon_\theta = 20\epsilon_0$, $\epsilon_\varphi = \epsilon_0$, $\epsilon_C = \epsilon_0$, $\epsilon_{NN} = \epsilon_0$, $\sigma_{NN} = 4.0$ Å ,and ϵ_0 is the energy interaction defined per contact. The parameters r_o, θ_o, φ_o (dependents on i), and σ_{ij} are values extracted from the native coordinate structure for a given protein obtained by experimental techniques such as small-angle X-ray scattering and nuclear magnetic resonance (NMR). In this equation, adjacent beads interact via harmonic interactions. Angles formed by monomer i, $i + 1$, and $i + 2$ are also maintained by a harmonic potential energy function. Dihedral angles formed by 4 adjacent residues are given cosine potential energy functions. The native contact pairs of residues are assigned as an attractive interaction of a 10−12 Lennard-Jones (LJ) potential. All residue pairs that are not in contact in the native structure, interact via nonspecific repulsion to prevent chain crossing. Here, two residues are defined as a native contact if the distance between any pair of non-hydrogen atoms belonging to these two residues, respectively, is shorter than 5 Å in the native conformation.

dx.doi.org/10.1021/jp212132v | J. Phys. Chem. B 2012, 116, 5152−5159

The fraction of native contacts Q, of a given structure during a simulation at a time t is defined as the fraction of number of native contacts present on the structure Γ at t over the total number of native contacts present on the native structure Γ_o. A contact is formed when noncovalent beads i and j are at a distance $1.2\sigma_{ij}$ or less.

On the basis of the funneled nature of the energy landscape of protein folding, structure-based models provide effective means of capturing the functionally important long-time and large-length scale motions of macromolecules such as proteins and DNAs. In our simulations, the coarse-grained structure-based model was utilized, which provides an understanding of both long-time and large-length scale dynamics of BBL and CI2. Notably, as mentioned previously by Whitford et al., this model lacks high energetic and structural resolutions, and it is impossible to partition geometric effects from energetic ones due to the coarse-grained nature of the model.[47] As for fast-folding proteins BBL and CI2, the associated energy landscape lacks large energetic traps, and the folding dynamics can be modeled by diffusion along an appropriate reaction coordinate. In this case, there is no need to partition geometric effects from energetic ones. Our coarse-grained structure-based model used here is reasonable.

2.3. Diffusion Coefficient Simulations. In order to calculate the coordinate dependence of diffusion coefficient $D(Q)$, molecular dynamics simulations were performed restricting the simulations to enhance sampling around a specific point of the coordinate Q. In order to do that, a biasing harmonic potential was added to the original form of the potential (eq 1), in a manner similar to the well-known umbrella sampling technique.[48] The biasing energy potential is given as

$$V_{\text{bias}}(Q^*) = K_Q(Q - Q^*)^2 \tag{2}$$

with K_Q being the strength parameter of the bias included (in units of k_BT), and Q^* is the region of interest to probe. For each Q^* fixed, D can be obtained as a function of Q to investigate the reaction coordinate dependence of diffusion.

In addition, Q only takes a discrete set of values due to the fact that contacts are either formed or not, thus any function of Q will be discontinuous. To avoid discontinuous functions during the molecular dynamic simulation, Q is slightly modified to produce the biasing potential from eq 2 with defined derivatives. The redefinition of Q^b takes the form of a tanh function as

$$Q^b = \frac{1}{N_Q}\sum_{\text{contacts}}\frac{1}{2}\{1 + \tanh[10(r - 1.2\sigma_{ij})]\} \tag{3}$$

with σ_{ij} previously defined as the native C_α–C_α distances, and N_Q is the total number of native contacts. Typically, a contact is defined as two residues at a distance about $1.2\sigma_{ij}$, so this tanh function acts effectively as a step function not changing the discrete notion of contacts formed during simulations.

The effects of the coordinate dependence of the diffusion coefficient were studied by varying the reaction coordinate through the folding process. For a fixed Q^*, the diffusion coefficient is estimated using the quasiharmonic diffusive approximation[49] as

$$D(Q^*) = \frac{\Delta Q^2(Q^*)}{2\tau_{\text{corr}}(Q^*)} \tag{4}$$

where ΔQ^2 is the mean square fluctuation in Q, and τ_{coor} is the relaxation time directly associated with the decay of the correlation function of Q

$$C(Q_0, \Delta) = \frac{\langle Q_0(t)Q_0(t+\Delta)\rangle - \langle Q_0{}^2(t)\rangle}{\langle Q_0{}^2(t)\rangle - \langle Q_0(t)\rangle^2} \tag{5}$$

Each simulation was performed at a fixed temperature T and Q^*. The reaction coordinate was collected as a function of time ($Q(t)$) for the calculations of averages on ΔQ^2 and $C(Q_0,\Delta)$.

Besides $Q(t)$, it is also possible to extract from the trajectories of other reaction coordinates that are more closely related to experimental measurements, such as root-mean-square deviation (rmsd(t)) and radius of gyration ($R_g(t)$), and determine their relationship with the computational $Q(t)$. If the diffusion coefficient is calculated for a given coordinate, in this case $D(Q)$, it is possible to obtain an approximation for the diffusion coefficient for another reaction coordinate, for example $D(R_g)$, using the transformation[40]

$$D(R_g) = D(Q)\left(\frac{dR_g}{dQ}\right)^2 \tag{6}$$

where $|dR_g/dQ|$ is the Jacobian of the variable transformation. The relationship $Q(R_g)$, calculated near the folding transition temperature T_f, is a monotonic function, and a quartic polynomial fit is employed to the data in order to obtain its inverse $R_g(Q)$ and the Jacobian.

The thermodynamic free-energy profile was calculated by combining simulations performed at multiple temperatures using the Weighted Histogram Analysis Method.[50]

The mean-first-passage time (MFPT) for folding (τ), i.e., the time to reach the folded state starting from a particular Q, can be obtained once the free-energy profile and the diffusion coefficient are given. The folding time is calculated, according to the energy landscape theory,[37,51] by

$$\tau(Q) = \int_Q^{Q_f}\frac{dQ'}{D(Q')}\exp\left[\frac{F(Q')}{k_BT}\right]\times\int_{Q_u}^{Q'}dQ''\exp\left[\frac{-F(Q'')}{k_BT}\right] \tag{7}$$

where $D(Q)$ is the diffusion coefficient as a function of the reaction coordinate Q and $F(Q)$ is the thermodynamic free energy at temperature T. Q_u and Q_f represents the unfolded and folded state, respectively.

The effective free-energy profile, taking into account the diffusion coefficient correction of kinetics from the above folding time formula, is obtained by[49]

$$F_{\text{eff}}(Q) = F(Q) - k_BT\ln\left(\frac{D(Q)}{D_0}\right) \tag{8}$$

where D_0 is the diffusion coefficient at the unfolded state. Equation 8 can also be applied to obtain $F_{\text{eff}}(R_g)$ and F_{eff}(rmsd).

3. RESULTS AND DISCUSSION

3.1. Free Energy Landscape and Barrier Height. To explore the kinetic folding mechanisms of two-state folder CI2 and downhill folder BBL, the two-dimensional (2D) projections of free-energy landscape at folding transition temperatures were constructed and shown in Figure 1. Here, two reaction coordinates are used: fraction of native contacts (Q) and the

dx.doi.org/10.1021/jp212132v | J. Phys. Chem. B 2012, 116, 5152–5159

radius of gyration (R_g). For CI2 (Figure 1a), there are two dominantly populated states: the unfolded state centered around Q = 0.2 and R_g = 12 Å, and the folded state centered around Q = 0.86 and R_g = 8.0 Å. In agreement with experimental findings, it shows no evidence of the existence of intermediate states. The free-energy barrier separating two basins is ~5 k_BT, indicating a highly cooperative folding. As for BBL (Figure 1b), it shows only one broad populated state, indicating that the folding cooperativity is rather low. This barrierless folding behavior is consistent with the first experimental assertion that BBL is a downhill folder,[16] which showed a gradual melting of native structure, and thus, only one single conformational ensemble at all denaturing conditions could be experimentally detected. Similarly, our simulations also reveal that only one unique structural distribution can be found with respect to certain reaction variable and the folded ensemble gradually shifts to unfolded ensemble as the simulation temperature changes from low to high values. In addition, different reaction coordinates may describe different conformational landscapes. Although it is not known whether the projected thermodynamic surface reflects faithfully the underlying kinetics and barrier height, it is clear that BBL has a lower folding cooperativity (indicating lower free-energy barrier) as compared with the classical two-state folding protein CI2. In this case, the folding kinetics of BBL are not expected to be obviously determined by the free-energy barrier or so-called transition state. The local escape capability through diffusion may play a more important role, which deserves further investigations.

3.2. Diffusion Coefficient Is Robust to Changes in Restraining Potential. The primary objective of this study was to determine how the diffusion coefficient D, in reaction coordinate space, changes during the folding process of the two-state folder CI2 and the downhill folder BBL, as well as the kinetic consequences. To ensure that the diffusion coefficient is a result of the underlying energy landscape, and not the biasing potential, several sets of simulations were performed, each with a different strength of the restraint. To calculate D, we employed a C_α structure-based model (see eq 1) with a restraining potential to ensure that each simulation sampled the phase space local to a particular value of Q (see Models and Methods for a full description). The restraining potential was harmonic, centered at Q^*, and was given a strength of K_Q. The value of K_Q must be in a range for which a quasiharmonic approximation is warranted and D is not dependent on K_Q.

To determine values of K_Q for which the quasiharmonic approximation is valid, a similar strategy in our previous simulation study on diffusion[39] was applied. According to our simulation results, it showed that, for different protein systems, the appropriate K_Q values are different. Thus, a properly detailed description is necessary for choosing the appropriate K_Q values. As in the literature,[39] we compared the probability distributions in Q and diffusion coefficients for a variety of K_Q values. Figure 2a,b shows the probability distributions for several values of K_Q (where the harmonic restraint is centered at Q^* = 0.5), each at the folding temperature in the unrestrained case. Figure 2c,d shows D as a function of K_Q for a wide range of Q values. In summary, an overly weak restraint may result in insufficient samplings within the phase space of interest, and an overly strong restraint may lead to artifacts by disallowing some possible routes of escape. On the basis of the data in Figure 2, we concluded that K_Q equal to 5.0 and 2.0 will provide reliable values for the configuration-dependent diffusion of CI2 and BBL, respectively.

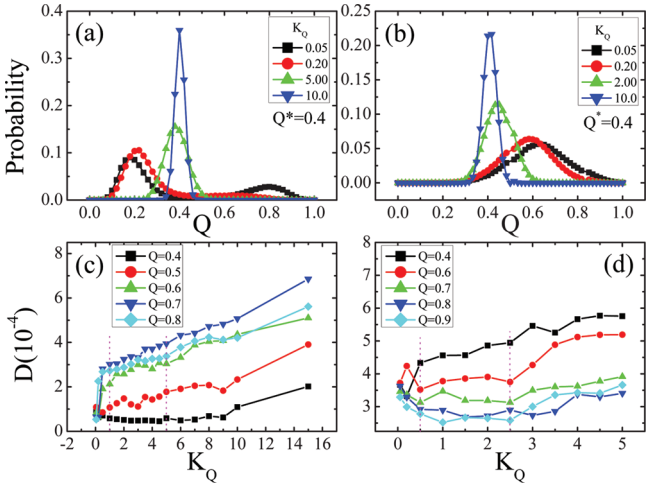


Figure 2. (a,b) Probability distributions in Q for biased ($K_Q > 0$, Q^* = 0.4) simulations with different strengths of the restraining potential K_Q. (c,d) The diffusion coefficient D is shown as a function of K_Q for five values of Q^*. Simulations were performed at the folding temperature of the pure structure-based model T_f.

3.3. Diffusion Coefficient Dependence on Q and Barrier Height. To understand the origins of the Q-dependence of the configuration-dependent diffusion coefficient, one must consider the fluctuations in Q and the decay time of these fluctuations $\tau(Q)$. All further results are reported for simulations performed at the folding transition temperature of the unrestrained simulations with K_Q = 5.0 and 2.0 for CI2 and BBL, respectively. In Figure 3a,b, the diffusion coefficients

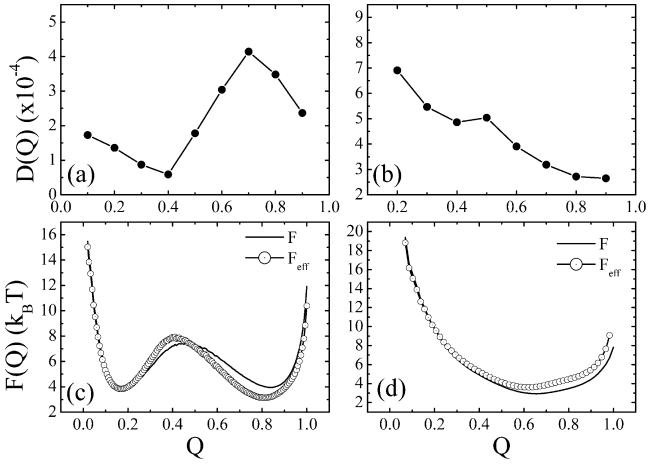


Figure 3. (a,b) Diffusion coefficients D calculated as a function of Q. All results were calculated with a restraining potential of strength K_Q = 5 for CI2 (a) and K_Q = 2 for BBL (b). (c,d) Thermodynamic free-energy profile $F(Q)$ and effective free-energy profile after the correction from $D(Q)$.

are shown as a function of the fraction of native contacts Q according to eq 4. For CI2, D, with a 7-fold variation, changes nonmonotonically as a function of Q. This is in agreement with earlier studies on the other two-state fast folding proteins[40] and analytic studies.[37,51] As for BBL, one can find that the diffusion coefficient in Q monotonically decreases as Q moves to native values, as shown clearly in Figure 3b. Here, $D(Q)$ varies by only

dx.doi.org/10.1021/jp212132v | J. Phys. Chem. B 2012, 116, 5152–5159

a factor of 3 over the range of the coordinate of Q. This relatively small variation in $D(Q)$ over the range of Q for both CI2 and BBL is in line with earlier results on the other two-state fast folding proteins.[38] The variations of the one-dimensional configuration-dependent diffusion coefficient indicate that the ruggedness of the energy landscape is not the same over the one-dimensional configuration space. $D(Q)$ describes the local moves over microscopic barriers that connect states with similar values of Q.

Further, we reconstructed the effective free-energy profile after the correction from $D(Q)$ using eq 7, as shown in Figure 3c,d. Here, the correction part of $D(Q)$ is treated through a fitting process so that one can get a smooth curve of the effective free-energy profile. For CI2, the thermodynamic transition state is at $Q = 0.44$, and the barrier height is 3.5 k_BT. The equation $\exp[F(Q)/k_BT - \ln D(Q)]$, which determines the kinetic time, has an effective barrier height of 4.2 k_BT. The effective thermodynamic transition state is at $Q = 0.41$ (see Figure 3c). Thus, the kinetic transition state as a function of Q is shifted, and the barrier is slightly (20%) increased. While for BBL, in the reaction coordinate of Q, no free-energy barrier is detected along both the original and effective free-energy profiles. However, the effective free-energy profile is shifted up 1 k_BT around the folded state. Therefore, the folded state becomes less stable and the folding kinetics are expected to be slowed down. This indicates that diffusion makes the folding landscape less downhill in the Q-coordinate space, thereby increasing the effective kinetic barrier in an alternative way, other than adding the conventional thermodynamic barrier.

3.4. Alternative Reaction Coordinates and Barrier Height. The diffusion coefficient will in general depend on the positions along the folding coordinate, and this dependence may vary for different coordinates. Above, we explored the Q-dependence of the configuration-dependent diffusion coefficient. In the following, we use R_g and rmsd as alternative reaction coordinates (RC) to study the RC-dependence of diffusion coefficient (Figures 4 and 5).

According to the eq 6 and based on the relationships between Q and R_g (Figure 4a,b) and rmsd (Figure 5a,b), we obtained the diffusion coefficient as a function of R_g (Figure 4c,d) and rmsd (Figure 5c,d). For both CI2 (Figures 4c and 5c) and BBL (Figures 4d and 5d), there is a monotonic decrease in the diffusion coefficient $D(R_g)$ and D(rmsd) toward the folded state. Such a trend in the configuration-dependent diffusion coefficient is in line with experimental expectations.[22,30,52] This is likely in that the coordinates of R_g and rmsd better preserve the magnitude of fluctuations in Cartesian space. It is also noticed that the change in $D(R_g)$ and D(rmsd) is of around 1 order of magnitude, in contrast to the small variation of $D(Q)$. There is a 20-fold change in $D(R_g)$ and D(rmsd) for CI2 (Figures 4c and 5c), and a 30-fold difference for BBL (Figures 4d and 5d). The apparent discrepancies in position dependence between Q and R_g/rmsd can be understood from the relationship between the two coordinates (see Figures 4a/5a and 4b/5b). The unfolded state spans a wide range of R_g/rmsd, yet this is compressed into only small fluctuations in Q, whereas, for the folded state, the reverse is true. For the unfolded protein, large changes in conformation can occur with a change of only a few contacts, but even a small conformational change near the folded state will cause a large shift in Q.[40]

We are interested in the effects of diffusion in R_g/rmsd on kinetics by altering the free-energy profile projected onto the coordinates of R_g/rmsd. Figures 4e/5e and 4f/5f show the original and effective free-energy curves of CI2 and BBL, respectively. For CI2, the effective barrier height is ~3.0 k_BT, which is a higher kinetic barrier compared with the thermodynamic one of ~2.0 k_BT (see Figure 5e). Therefore, the contribution to the kinetic barrier purely from diffusion is 3.0 k_BT − 2.0 k_BT = 1.0 k_BT, which is significant relative to the thermodynamic barrier height. This indicates that the configuration-dependent diffusion can play a role in kinetics, especially when the thermodynamic barrier is relatively small (e.g., fast folding proteins). Interestingly, for BBL, diffusion results in a ~1.0 k_BT kinetic barrier height that is in the scale of thermodynamic perturbation, whereas the thermodynamic free-energy profile is barrier-free. In addition, the kinetic barrier difference for R_g and rmsd is 0.1 k_BT, which indicates a consistent folding kinetics with respect to different reaction coordinates.

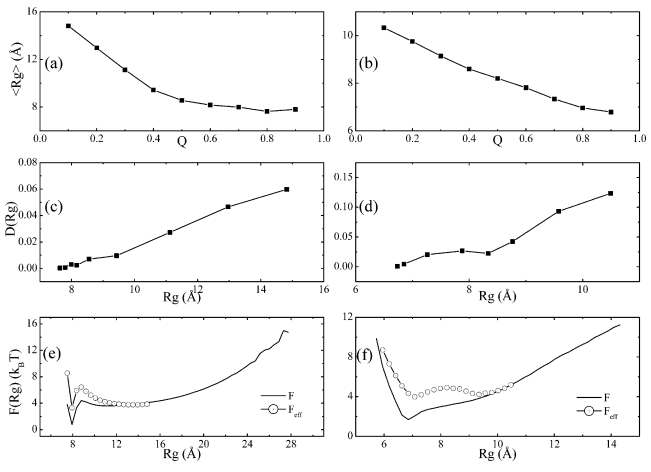


Figure 4. (a,b) Average values of R_g (radius of gyration) as a function of Q. (c,d) Diffusion coefficient in R_g. (e,f) Thermodynamic free-energy profile $F(R_g)$ and effective free-energy profile after the correction from $D(R_g)$. Time is shown in reduced units, and the temperature is the folding transition temperature T_f.

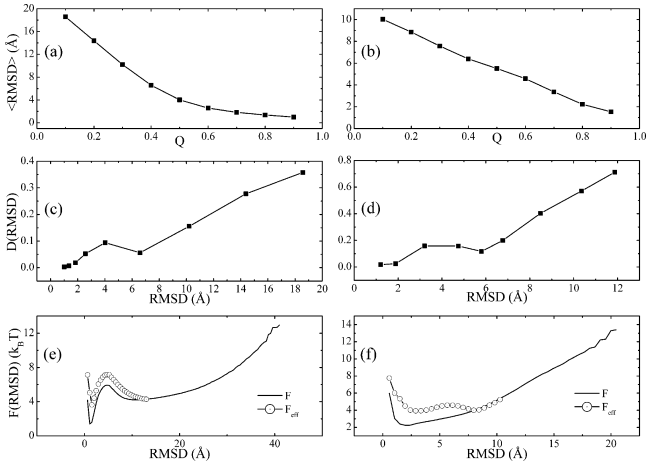


Figure 5. (a,b) Average values of rmsd (root-mean-square of distance) as a function of Q. (c,d) Diffusion coefficient in rmsd. (e,f) Thermodynamic free-energy profile F(rmsd) and effective free-energy profile after the correction from D(rmsd). Time is shown in reduced units, and the temperature is the folding transition temperature T_f.

dx.doi.org/10.1021/jp212132v | J. Phys. Chem. B 2012, 116, 5152−5159

Interestingly, there are several experimental works that have found the same observations in other proteins. For instance, Naganathan et al. experimentally compared the thermodynamic barrier heights for 15 proteins with the folding kinetic rates.[53] The analysis showed a good correlation but with a slope lower than 1. That means the effective barrier is higher from the kinetics, i.e., there are hidden barriers in addition to the thermodynamic one due to kinetics. Moreover, based on one-dimensional free-energy surface models, another two thorough experimental studies on two fast folding proteins revealed also higher kinetic barriers than the thermodynamic ones.[54,55] Notably, the kinetic barriers in these two cases are also about 1 k_BT higher, as found in the results of our theoretical model. In all, these experimental data support our simulation results.

3.5. Folding Kinetics Determined by Effective Free Energy Landscape. Figure 6 shows the mean-first-passage time (MFPT) τ as a function of the simulation temperature compared with the results of analytical models with Q-dependent diffusion. For both CI2 (Figure 6a) and BBL (Figure 6b), one can see that the kinetics in terms of MFPT from simulations has a U shape dependence on the simulation temperature. This is due to the fact that, at low temperature, the kinetic traps are dominant, and the folding kinetics are thus slowed down; as temperature increases, the kinetic traps are overcome by the enhanced thermal motions of the protein. As a result, the folding kinetics are sped up. While at high temperature, the folded state is unstable, and the folding kinetics also slow down significantly. We further calculated the MFPT according to the analytical eq 8 at several temperatures. We can see the kinetic time with Q-dependent diffusion is in reasonable agreement with simulations that are also shown in Figure 6a,b as diamonds. Such a good agreement in kinetic behavior has also been observed for the CspTm protein[38] and a lattice model protein.[26] Also, as in Figure 6a,b, position-dependent diffusion models reproducing the folding time (or folding rates) upon solvent friction constant with good agreement have been shown by other works as well.[56,57] Although configuration-dependent diffusion does not disturb the equilibrium distribution, it modifies the kinetic rate or flux through increasing the kinetic barrier height or destabilizing the folded ensemble, and the kinetic route or path through the shift of the kinetic barrier position.

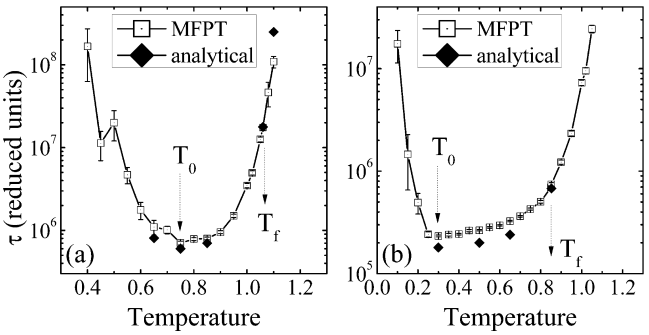


Figure 6. Mean-first-passage time (MFPT) in logarithmic scale versus temperature for (a) CI2 and (b) BBL, respectively. The open squares are from kinetic simulation runs, and the solid diamonds are using the analytical expression from eq 8.

As we can see, T_0 (the optimum folding temperature) is a quantitative signature below which the trapping effects start to set in. This can be clearly seen as the kinetics start to turn over and significantly slow down below T_0 in the MFPT versus temperature figure. Therefore, we can use T_0 as a scale for trapping. However, T_f is the folding transition temperature below which folding is stable and preferred and above which the folding is unstable and unfolding is preferred. The ratio of T_f/T_0 then provides a dimensionless measure of the stability of folding against trapping for a particular protein. This is also related to the underlying folding landscape topography, that is the ratio of biasing toward the native state or slope of the funnel against the roughness modularized by the size of the landscape.[58,59] So, the higher ratio of folding against trapping characterized by the ratio T_f/T_0 implies the steeper funnel or less trapping and faster kinetics.[51,60,61]

Interestingly, we noticed that the ratio of the folding transition temperature (T_f) over the optimal folding temperature (T_0), i.e., T_f/T_0, is 1.41 for CI2 and 2.84 for BBL, respectively. So, BBL has a significantly higher ratio of T_f/T_0 than CI2. We expect that the higher ratio of T_f/T_0 corresponds to a steeper funnelled folding energy landscape or less trapping, which, in an extreme case, leads to the downhill folding behavior without the thermodynamic barrier in the free energy landscape as shown for the case of BBL. Therefore, the ratio might provide us a quantitative indicator for the possible appearance of downhill folding. Interestingly, the roughness of a designed monomeric miniprotein (FSD-1ss) was experimentally estimated, which was characterized by $T_m/T_g \approx 1.3$.[62] The ratio of 1.3 is lower than the estimates for well evolved natural proteins.[63] As a matter of fact, the experimentally measured quantity T_m/T_g is closely linked to the theoretical ratio of T_f/T_0. T_f/T_0 is bigger than 1 for both BBL and CI2 from the theoretical calculations of our models, and the experimental results support our simulation results and theoretical prediction that T_m/T_g for folding is bigger than 1. This shows the steepness of the landscape toward the funnel dominates against traps. It indicates that T_f/T_0 is closely related to the underlying folding landscape topography.[58,59]

3.6. Influence of Temperature on Diffusion Coefficient. Finally, we studied the temperature dependence of the diffusion coefficient. Figure 7 shows the diffusion coefficient $D(T)$ versus temperature, which are integrated over R_g for (a) CI2 and (b) BBL. The $D(T)$ as a function of Q and rmsd are similar, which are not shown here. In most of the temperature regions, for both cases as shown in Figure 7, $D(T)$ changes monotonically as a function of the simulation temperature. That is, $D(T)$ increases with T because the local escape kinetics are faster as temperature is enhanced. This observation is completely in line with a previous experimental study of fast-folding kinetics interpreted as diffusion on a one-dimensional free-energy surface.[27] However, a mild local peak is noticed around $T = 0.5$ for CI2. The origin of the mild peak may come from the integration over R_g with discrete data points or insufficient simulation sampling.

4. CONCLUSIONS

Using a coarse-grained structure-based model, we calculated the configuration-dependent diffusion coefficient of a two-state folding protein CI2 and a downhill folding protein BBL as a function of various reaction coordinates: Q, R_g, and rmsd. We found that diffusion coefficients for coordinates that directly probe configurational fluctuations, such as R_g and rmsd, are markedly position dependent. There is a large decrease in configuration-dependent diffusion coefficient from unfolded to folded state for R_g and rmsd. Furthermore, by considering the

dx.doi.org/10.1021/jp212132v | J. Phys. Chem. B 2012, 116, 5152−5159

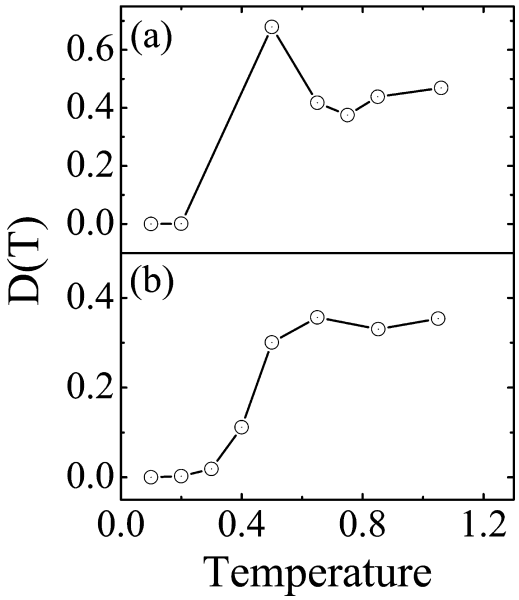


Figure 7. Diffusion coefficients D as a function of simulation temperatures, which are integrated over R_g for (a) CI2 and (b) BBL.

kinetic barrier induced by diffusion, the reconstructed effective free-energy profiles along both R_g and rmsd reveal kinetic barriers that are hidden in thermodynamic downhill free-energy profile of BBL. This provides evidence of the degree of roughness present in the free-energy landscape of BBL. As compared with two-state CI2, the kinetics of downhill BBL are found to be more affected by configuration-dependent diffusion. These findings on the fast two-state protein CI2 are consistent with both earlier simulations[38−40,64,65] and single-molecule fluorescence measurements.

The study on BBL by Wolynes and co-workers[21] found that, with the purely additive structure-based model, Muñoz's construct was barrierless,[16,19] while the full-length and truncated Fersht's constructs[24,25] had a free-energy barrier of ~2 k_BT. By addition of many-body interactions into the purely additive structure-based model, Muñoz's construct had a nonzero but negligible (0.5 k_BT) free-energy barrier, while Fersht's constructs yielded free-energy barriers of 4 k_BT. Recently, a full-fledged atomic-level MD study of several fast folding proteins has shown that the Fersht's constructs also lack a free-energy barrier along an optimized reaction coordinate.[66] The discrepancy between the two simulation studies in quantitative details may be due to the different simulation methods (one coarse grained and one atomic level) and the sampling. The latter full atomistic simulation with explicit solvents may give more accurate quantitative details. Using the same purely additive structure-based model, our simulations also lead to a barrier-free free-energy profile for Muñoz's construct (see Figures 3−5), but the folding cooperativity for Fersht's construct is much higher than that of Muñoz's construct (data not shown). Interestingly, the hidden nonzero but small kinetic barrier of Muñoz's construct was detected by taking into account the configurational dependent diffusion. Clearly, it is expected that, by considering diffusion, a relatively higher kinetic barrier would emerge for Fersht's construct.

This is the first practical investigation into the role of diffusion in downhill folding. For the two-state folder CI2, the thermodynamic barrier is still dominant for folding, while the diffusion component tunes the folding kinetics. The diffusion modifies the kinetic rate or flux through the emergence of the kinetic barrier height and changes the kinetic route or path through the shift of the effective free-energy barrier position. For downhill folder BBL, there is no free-energy barrier, and the folding kinetics is determined by diffusion. In addition, the diffusion coefficient shows a temperature-dependent behavior. Interestingly, in comparison with the two-state folder CI2, the downhill folder BBL shows a significantly higher ratio of the folding transition temperature (T_f) over the optimal kinetic folding temperature (T_0). This implies that the higher ratio of T_f/T_0 leads to a steeper funnel and less roughness of the underlying energy landscape and faster kinetics, which, in an extreme case, results in the downhill folding behavior without free-energy barrier. Here, to our structure-based simulation model, the roughness is topological roughness with entropy origin. This nature of protein folding confirms the predictions of our theoretical study on the nature of biomolecular folding and binding. We may use the ratio of T_f/T_0 as a quantitative indicator for the possible emergence of downhill folding. Our observations will hopefully facilitate a complete understanding of the energy landscape theory of protein folding.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jin.wang.1@stonybrook.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

J.W., Z.L., and W.X. thank NSF and NIH for support. R.J.O. and V.B.P.L. thank FAPESP, CAPES, and CNPq (Brazil) for support.

REFERENCES

(1) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *18*, 8721−8725.
(2) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524−7528.
(3) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70−75.
(4) Shoemaker, B. A.; Wang, J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 777−782.
(5) Nymeyer, H.; Garcia, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5921−5928.
(6) Clementi, C.; Nymeyer, H.; Onuchic, J. N. *J. Mol. Biol.* **2000**, *298* (5), 937−953.
(7) Xu, W. X.; Lai, T. F.; Yang, Y.; Mu, Y. G. *J. Chem. Phys.* **2008**, *128*, 175105.
(8) Levy, Y.; Cho, S. S.; Shen, T.; Onuchic, J. N.; Wolynes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2373−2378.
(9) Wang, J.; Lu, Q.; Lu, H. P. *PLoS Comput. Biol.* **2006**, *2*, e78.
(10) Lu, Q.; Lu, H. P.; Wang, J. *Phys. Rev. Lett.* **2007**, *98* (12), 128105.
(11) Wang, W.; Xu, W. X.; Levy, Y.; Trizacc, E.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5517−5522.
(12) Whitford, P. C.; Miyashita, O.; Levy, Y.; Onuchic, J. N. *J. Mol. Biol.* **2007**, *366*, 1661−1671.
(13) Schug, A.; Whitford, P. C.; Levy, Y.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (45), 17674−17679.
(14) Lu, Q.; Wang, J. *J. Am. Chem. Soc.* **2008**, *130* (14), 4772−4783.
(15) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167−195.
(16) Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Munoz, V. *Science* **2002**, *298*, 2191−2195.
(17) Yang, W. Y.; Gruebele, M. *Nature* **2003**, *423*, 193−197.

(18) Zuo, G. H.; Wang, J.; Wang, W. *Proteins* **2006**, *63*, 165−173.
(19) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2006**, *442*, 317−321.
(20) Liu, F.; Du, D.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2369−2374.
(21) Cho, S. S.; Weinkam, P.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 118−123.
(22) Lia, P.; Oliva, F. Y.; Naganathan, A. N.; Munoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 103−108.
(23) Liu, J. W.; Campos, L. A.; Cerminara, M.; Wang, X.; Ramanathan, R.; English, D. S.; Munoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 179−184.
(24) Ferguson, N.; Sharpe, T. D.; Johnson, C. M.; Schartau, P. J.; Fersht, A. R. *Nature* **2007**, *445*, E14−E15.
(25) Huang, F.; Sato, S.; Sharpe, T. D.; Ying, L. M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 123−127.
(26) Chahine, J.; Oliveira, R. J.; Leite, V. B. P.; Wang, J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (37), 14646−14651.
(27) Naganathan, A. N.; Doshi, U.; Munoz, V. *J. Am. Chem. Soc.* **2007**, *129*, 5673−5682.
(28) Kremer, W.; Schuler, B.; Harrieder, S.; Geyer, M.; Gronwald, W.; Welker, C.; Jaenicke, R.; Kalbitzer, H. R. *Eur. J. Biochem.* **2001**, *268* (9), 2527−2539.
(29) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76−88.
(30) Nettels, D.; Gopich, I. V.; Hoffmann, A.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2655−2660.
(31) Hoffmann, A.; Kane, A.; Nettels, D.; Hertzog, D.; Baumgartel, P.; Lengefeld, J.; Reichardt, G.; Horsley, D.; Seckler, R.; Bakajin, O.; Schuler, B. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 105−110.
(32) Karplus, M.; Weaver, D. L. *Protein Sci.* **1994**, *3* (4), 650−668.
(33) Kubelka, J.; Chiu, T. K.; Davies, D. R.; Eaton, W. A.; Hofrichter, J. *J. Mol. Biol.* **2006**, *359* (3), 546−553.
(34) Gruebele, M. *C. R. Biol.* **2005**, *328*, 701−712.
(35) Dhar, A.; Ebbinghaus, S.; Shen, Z.; Mishra, T.; Gruebele, M. *Biophys. J.* **2010**, *99*, L69−L71.
(36) Baumketner, A.; Hiwatari, Y. *Phys. Rev. E* **2002**, *66* (1), 011905.
(37) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93* (19), 6902−6915.
(38) Oliveira, R. J.; Whitford, P. C.; Chahine, J.; Leite, V. B. P.; Wang, J. *Methods* **2010**, *52*, 91−98.
(39) Oliveira, R. J.; Whitford, P. C.; Chahine, J.; Wang, J.; Onuchic, J. N.; Leite, V. B. P. *Biophys. J.* **2010**, *99*, 600−608.
(40) Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 1088−1093.
(41) Best, R. B.; Hummer, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16902−16911.
(42) Sadqi, M.; Lapidus, L. J.; Munoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12117C12122.
(43) Xu, W. X.; Wang, J.; Wang, W. *Proteins* **2005**, *61*, 777−794.
(44) Elcock, A. H. *PLoS Comput. Biol.* **2006**, *2*, e98.
(45) Zhang, J.; Li, W. F.; Wang, J.; Qin, M.; Wang, W. *Proteins* **2008**, *72*, 1038−1047.
(46) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K.; Onuchic, J. N. *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 430−441.
(47) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. *Proteins* **2009**, *75*, 430−441.
(48) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(49) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860−5868.
(50) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.
(51) Lee, C. L.; Stell, G.; Wang, J. *J. Chem. Phys.* **2003**, *118* (2), 959−968.
(52) Waldauer, S. A.; Bakajin, O.; Lapidus, L. J. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13713−13717.
(53) Naganathan, A. N.; M., S.-R. J.; Munoz, V. *J. Am. Chem. Soc.* **2005**, *127*, 17970−17971.
(54) Godoy-Ruiz, R.; Henry, E. R.; Kubelka, J.; Hofrichter, J.; Munoz, V.; Sanchez-Ruiz, J. M.; Eaton, W. A. *J. Phys. Chem. B* **2008**, *112* (19), 5938−5949.
(55) Fung, A.; Li, P.; Godoy-Ruiz, R.; Sanchez-Ruiz, J. M.; Munoz, V. *J. Am. Chem. Soc.* **2008**, *130*, 7489−7495.
(56) Hummer, G. *New J. Phys.* **2005**, *7*, 34−48.
(57) Sangha, A. K.; Keyes, T. *J. Phys. Chem. B* **2009**, *113*, 15886−15894.
(58) Plotkin, S. S.; Wang, J.; Wolynes, P. G. *J. Chem. Phys.* **1997**, *106*, 2932.
(59) Wang, J.; Verkhivker, G. M. *Phys. Rev. Lett.* **2003**, *90*, 188101.
(60) Lee, C. L.; Lin, C. T.; Stell, G.; Wang, J. *Phys. Rev. E* **2003**, *67*, 041905.
(61) Wang, J. *Chem. Phys. Lett.* **2006**, *540*, 418.
(62) Sadqi, M.; de Alba, E.; Perez-Jimenez, R.; Sanchez-Ruiz, J. M.; Munoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4127−4132.
(63) Chan, H. S.; Shimizu, S.; Kaya, H. *Methods Enzymol.* **2004**, *380*, 350−379.
(64) Yang, S.; Onuchic, J.; Levine, H. *J. Chem. Phys.* **2006**, *125* (5), 054910−054918.
(65) Best, R. B.; Hummer, G. *Phys. Rev. Lett.* **2006**, *96* (22), 228104−228108.
(66) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.