Article

Comparison of Three Perturbation Molecular Dynamics Methods for Modeling Conformational Transitions

He Huang, Elif Ozkirimli, and Carol Beth Post

J. Chem. Theory Comput., **2009**, 5 (5), 1304-1314• DOI: 10.1021/ct9000153 • Publication Date (Web): 09 April 2009

JCTC Journal of Chemical Theory and Computation

Comparison of Three Perturbation Molecular Dynamics Methods for Modeling Conformational Transitions

He Huang,[†] Elif Ozkirimli,[†,‡] and Carol Beth Post*[,†]

Department of Medicinal Chemistry and Molecular Pharmacology, Markey Center for Structural Biology and Purdue Cancer Center, Purdue University, West Lafayette, Indiana 47907, and Department of Chemical Engineering, Bogazici University, Bebek, 34342, Turkey

Received January 9, 2009

Abstract: Targeted, steered, and biased molecular dynamics (MD) are widely used methods for studying transition processes of biomolecules. They share the common feature of adding external perturbations along a conformational progress variable to guide the transition in a predefined direction in conformational space, yet differ in how these perturbations are applied. In the present paper, we report a comparison of these three methods on generating transition paths for two different processes: the unfolding of the B domain of protein A and a conformational transition of the catalytic domain of a Src kinase Lyn. Transition pathways were calculated with different simulation parameters including the choice of progress variable and the simulation length or biasing force constant. A comparison of the generated paths based on structural similarity finds that the three perturbation MD methods generate similar transition paths for a given progress variable in most cases. On the other hand, the path depends more strongly on the choice of progress variable used to move the system between the initial and final states. Potentials of mean force (PMF) were calculated starting from unfolding trajectories to estimate the relative probabilities of the paths. A lower PMF was found for the lowest biasing force constant with BMD.

1. Introduction

Functionally important dynamic processes of proteins, such as folding/unfolding and allosteric conformational transitions, occur on the microsecond to millisecond time scale. Molecular dynamics simulation is a useful tool to elucidate the atomistic detail of protein dynamics; however, simulations of all-atom models are limited to the submicrosecond time range. To overcome this time scale problem, methods that utilize the principles of molecular dynamics, with some external perturbations to accelerate the reaction and guide the system toward a target state, have been developed.[1−6] Here, such methods are collected under the title of "perturbation molecular dynamics". They aim to identify possible transition pathways as well as energy barriers and metastable intermediates. Such pathways can then be further examined by thermodynamic simulation methods.

The most commonly used perturbation molecular dynamics methods are targeted molecular dynamics (TMD), steered molecular dynamics (SMD), and biased molecular dynamics (BMD). These methods share the common feature of guiding the transition between two end states through some progress variable (reaction coordinate), though they differ in the way that the progress variable is controlled. As originally introduced by Schlitter,[1,2] the TMD methodology imposes a time-dependent holonomic constraint on the rmsd to a target structure. SMD simulations were first used by Grubmueller[3] and Leech,[4] and were widely applied shortly after by Schulten et al.[7−10] SMD is akin to atomic force microscopy (AFM) in that a harmonic restraint based on a reference point moves the system toward the target when the reference point is updated. We note that in some later publications[11−16] the term TMD was associated with the rmsd progress variable

* Corresponding author e-mail: cbp@purdue.edu.
[†] Purdue University.
[‡] Bogazici University.

10.1021/ct9000153 CCC: $40.75
Published on Web 04/09/2009

with harmonic restraint rather than holonomic constraint. In the present paper, we denote the holonomic constraint as TMD and the harmonic restraint as SMD to distinguish the perturbation form regardless of the progress variable. BMD, also known as the adiabatic bias molecular dynamics, was originally proposed independently by Marchi and Ballone[5] and Paci and Karplus.[6] It provides the least perturbation among the three in that the system feels no force if it moves toward the target and the biasing potential is nonzero only if the system moves away from the target.

These three methods are commonly used to study transition processes of biomolecules[11−30] They are also used to complement AFM experiments that examine the mechanical properties of macromolecules.[10,31] Given the interest in these perturbation MD methods, an assessment of the effect of the choice of perturbation method, progress variable, and other simulation parameters on the resulting transition paths is of value.

In the present paper, we report a systematic comparison of the conformations and energetics of trajectories generated by the three perturbation molecular dynamics methods for two transition processes. The first example system is the unfolding of the B domain of staphylococcal protein A (BdpA). BdpA is a three-helix-bundle protein for which folding/unfolding has been studied extensively by experiments[32−34] and computer simulations.[35−38] Its unfolding is an example where the progress variable does not define a single target configuration. The second process is the conformational change between active and inactive structures of the kinase catalytic domain (CD) of Lyn, a member of the Src family of protein tyrosine kinases. This transition is an example that has defined target configurations at both end states. For TMD, SMD, and BMD paths generated for both systems, we examined the effect of different simulation conditions, including the choice of progress variable and perturbation strength. Our results suggest that, for the most part, the three perturbation MD methods generate similar transition paths for a given progress variable even though the time dependence of the progress variables differs substantially. On the other hand, the path depends strongly on the choice of the progress variable.

2. Methods

2.1. Three Perturbation MD Methods. Targeted molecular dynamics (TMD) introduces the most constrained perturbation among the three methods. While the other two methods add restraining potentials to guide the system, TMD imposes a holonomic constraint onto the dynamics of the system:[2]

$$\phi(\rho(x(t)), \rho_0(t)) = 0 \tag{1}$$

where $\rho(x)$ is the progress variable defined as a function of the coordinate, x, $\rho_0(t)$ is the reference value of ρ at time t, and ϕ is a function of ρ and ρ_0 which equals zero when $\rho = \rho_0$; for example

$$\phi(x) = \rho - \rho_0 \tag{2}$$

The constraint adds onto the system a constraining force

$$F^C = \lambda \nabla_x \phi \tag{3}$$

which keeps the progress variable ρ following the reference value $\rho_0(t)$ exactly. Here λ is a Lagrange parameter determined according to eq 1, and the reference value $\rho_0(t)$ moves at a constant rate v toward the target value:

$$\rho_0(t) = \rho_0(t_0) + v(t - t_0) \tag{4}$$

Steered molecular dynamics (SMD) corresponds closely to micromanipulation by AFM[39] when it uses a single interatomic distance as the progress variable. Computationally, it adds a full harmonic potential to restrain the progress variable around a reference value, which is moved to the target value at a constant rate v:

$$H(\rho) = \frac{\alpha}{2}(\rho - \rho_0)^2 \tag{5}$$

$$\rho_0(t) = \rho_0(t_0) + v(t - t_0) \tag{6}$$

where H is the biasing potential, α is the force constant, ρ is the progress variable, and ρ_0 is the reference point.

Biased molecular dynamics (BMD) guides the change of the progress variable by penalizing a move in the undesired direction through a one-sided harmonic potential. At each time step, the reference point is updated to the previously sampled value that is closest to the target. The method is defined by the following equations assuming the system is moved in the direction in which the progress variable ρ increases:

$$H(\rho) = \begin{cases} \frac{\alpha}{2}(\rho - \rho_0)^2 & (\rho < \rho_0) \\ 0 & (\rho \geq \rho_0) \end{cases} \tag{7}$$

$$\rho_0(t) = \begin{cases} \rho_0(t - \Delta t) & (\rho < \rho_0) \\ \rho(t) & (\rho \geq \rho_0) \end{cases} \tag{8}$$

where Δt is the simulation time step. Among the three methods, BMD provides the least restrained perturbation to the molecular system in that progress in the direction toward the target occurs without external perturbation.

2.2. Molecular Systems. The molecular system of BdpA was derived from the NMR structure (PDB ID: 1BDD).[40] The C- and N-terminal loops were removed, and residues 10−55 were kept. A 2-ns equilibrium MD simulation was calculated with the CHARMM22 force field and the GBSW implicit solvent model[41] in CHARMM.[42] Four configurations from the equilibrium run separated by 500 ps were saved and used as initial structures of the perturbation MD runs. The CHARMM22 force field and GBSW solvation model were used in all perturbation MD runs for BdpA unfolding.

For the Lyn CD, the active and inactive coordinates were obtained by homology modeling based on crystallographic structures of the Lck kinase (PDB ID: 3LCK)[43] and Hck kinase (PDB ID: 1QCF),[44] respectively. Both structures were equilibrated with the CHARMM22 force field with approximately 9400 TIP3P waters in a periodic rhombic dodecahedral box at 298 K for 300−400 ps, as previously reported.[23] The CHARMM22 force field with explicit TIP3P water was also used in all perturbation MD runs for the Lyn CD conformational transition.

2.3. Progress Variables. For BdpA unfolding, the three perturbation MD methods were examined with two progress variables, either the end-to-end distance between the two

Table 1. Summary of BdpA Unfolding Trajectories

progress variable	method	length (ns)	force constant (kcal/(mol·Å²))	rate (Å/ps)	no.[a]	code
h		0.5	–	0.1	2	hT1
	TMD	2	–	0.025	2	hT2
		10	–	0.005	2	hT3
		0.5	50	0.1	2	hS1
	SMD	2	50	0.025	2	hS2
		10	50	0.005	2	hS3
		0.5	50	–	2	hB1
	BMD	2	20	–	2	hB2
		10	8	–	2	hB3
R_g		0.5	–	0.05	2	RT1
	TMD	2	–	0.0125	2	RT2
		10	–	0.0025	4	RT3
		0.5	5000	0.05	2	RS1
	SMD	2	5000	0.0125	2	RS2
		10	5000	0.0025	4	RS3
		0.5	5000	–	2	RB1
	BMD	2	400	–	2	RB2
		10	110	–	4	RB3

[a] Number of independent trajectories that were calculated.

terminal C_α atoms ($\rho = h$) or the radius of gyration based on all heavy atoms ($\rho = R_g$), which is defined as

$$R_g = \sqrt{\frac{1}{N}\sum_i^N \left| r_i - \frac{1}{N}\sum_j^N r_j \right|^2} \tag{9}$$

where r_i is the Cartesian coordinate of the ith heavy atom and N equals the total number of heavy atoms. Each combination of method and progress variable was used with three different perturbation strengths as reflected by the simulation lengths: 0.5, 2, and 10 ns. With TMD or SMD, the simulation length was controlled directly by the rate for updating the reference point, while in BMD, in which no such parameter is available, it was controlled by tuning the force constant so that the unfolding process happens in roughly the specified time. For each set of simulation conditions (perturbation method, progress variable, and simulation length), two to four (see Table 1 for details) trajectories were calculated with different initial coordinates and velocities.

The Lyn CD conformational transition for both activation (inactive to active CD) and deactivation (active to inactive) was simulated by all three perturbation MD methods with the progress variable mean square internal deviation (ρ = MSID) defined in terms of internal distances:

$$\mathrm{MSID} = \frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j>i}^{N}(d_{ij} - d_{ij}^0)^2 \tag{10}$$

where d_{ij} and d_{ij}^0 are distances between heavy atoms i and j in the current and target structures, respectively, and N is the number of atoms. This progress variable was originally used to investigate partially unfolded protein in combination with BMD by Paci et al.,[45] and later applied to Src kinase activation.[23] A similar progress variable defined upon internal distances was used by Markwick et al. together with a mass weighting.[46]

A second progress variable, root-mean-square deviation (ρ = rmsd), was examined with only the TMD method, and is defined as

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|r_i - r_i^0|^2} \tag{11}$$

where r_i and r_i^0 are Cartesian coordinates of the ith heavy atom in the superimposed current and target structures, respectively, and N is the total number of heavy atoms. The combination of rmsd as the progress variable and the holonomic constraint as the perturbation method coincides with the original TMD method used by Schlitter et al.[1,2] For each combination of method, progress variable, and direction, two to four (see Table 2 for details) trajectories were calculated with different updating rates, v, of the progress variable (TMD, SMD) or force constants, α (BMD).

2.4. Trajectory Averaged rmsd. The structural similarity of the BdpA unfolding trajectories was evaluated from a trajectory averaged rmsd calculated to measure the pairwise similarity between different trajectories. For each unfolding trajectory with either h or R_g as the progress variable, snapshots were binned according to R_g into 0.5-Å-wide bins. An average configuration $\bar{X}_k$ was calculated for each bin k to represent the snapshots within that bin. Snapshots within each bin have similar structures, as indicated by the average rms fluctuation of 2.0, 2.0, and 1.6 Å for TMD, SMD, and BMD 10 ns trajectories, respectively. The within-bin variances for shorter trajectories are expected to be even lower. Between trajectories i and j, the trajectory averaged rmsd is defined as

$$\overline{\mathrm{rmsd}}_{ij} = \frac{1}{M}\sum_{k=1}^{M}\mathrm{rmsd}(\bar{X}_k^i, \bar{X}_k^j) \tag{12}$$

where M is the number of bins and rmsd is calculated after superimposing all heavy atoms. This single value provides a metric for the overall similarity between two trajectories. Because the rmsd between two extended structures is always small, we ignored snapshots with R_g greater than 17 Å in this calculation.

Table 2. Summary of Lyn CD Activation/Deactivation Trajectories

progress variable	method	direction[a]	length (ns)	force constant (kcal/(mol·Å⁴))	rate (Å/ps)	color[b]
MSID			0.16	–	0.1	black
	TMD	A → I	1.0	–	0.016	red
		I → A	0.16	–	0.1	green
			1.0	–	0.016	blue
		A → I	0.16	1000	0.1	black
	SMD		1.0	1000	0.016	red
		I → A	0.16	1000	0.1	green
			1.0	1000	0.016	blue
			1.9	1000	–	black
		A → I	1.5	2000	–	red
			1.5	3000	–	yellow
	BMD		1.3	5000	–	brown
			1.9	1000	–	green
		I → A	1.9	2000	–	blue
			1.7	3000	–	gray
			1.0	5000	–	purple
rmsd		A → I	0.2	–	0.03	black
	TMD		1.0	–	0.006	red
		I → A	0.2	–	0.03	green
			1.0	–	0.006	blue

[a] A, active state; I, inactive state. [b] Coloring for Figure 8a.

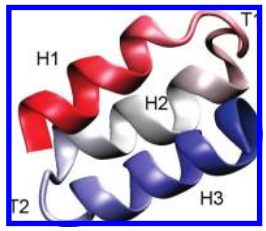

Figure 1. The native structure of BdpA[40] shown in cartoon representation. Helices H1, H2, and H3 are colored in red, white, and blue, respectively. This figure was generated with Visual Molecular Dynamics (VMD).[49]

2.5. Potential of Mean Force Calculation. The potential of mean force (PMF) was calculated with an initial path defined by a R_g-perturbed BdpA unfolding trajectory by using umbrella sampling[47] and the weighted histogram analysis method (WHAM).[48] For each trajectory, 41 snapshots were taken as the initial coordinates for umbrella sampling by choosing coordinates with R_g values closest to an equally spaced R_g series ranging from 9.75 to 15.75 Å in increments of 0.15 Å. Each of the 41 umbrella windows was simulated for 400 ps using a harmonic umbrella potential with force constant 10 kcal/(mol·Å^2) to restrain R_g around the initial value. The last 200 ps of the sampling was analyzed by WHAM to reconstruct the PMF profile with respect to R_g. The effect of initial-coordinate bias for a given trajectory was examined by choosing a different set of initial coordinates with a shifted R_g series ranging from 9.83 to 15.83 Å with 0.15 Å increments for the PMF calculations. Results for two trajectories showed that very similar PMF curves are obtained for each trajectory, indicating that the calculated PMF curve characterizes a trajectory rather than the specific selection of snapshots (data not shown).

3. Results

Tables 1 and 2 summarize the perturbation MD simulations carried out for the two transition systems. All simulations and analyses were carried out with the molecular dynamics program CHARMM.[42]

3.1. BdpA Unfolding. BdpA is a three-helix-bundle protein with a highly symmetrical topology. The three helices (H1, H2, and H3) of comparable length are joined by two turns (T1 and T2) in an antiparallel alignment to form two helix−turn−helix motifs (see Figure 1).

As listed in Table 1, a total of 42 unfolding trajectories of BdpA were generated with three perturbation MD methods (BMD/SMD/TMD), two progress variables (end-to-end distance h/radius of gyration R_g), and three different simulation lengths (0.5 ns/2 ns/10 ns). Figure 2 shows the time profiles of the progress variables for representative trajectories generated by the three methods. With TMD, the progress variables scale linearly with time as expected from the holonomic constraint. SMD also generated nearly linear progression curves due to the linear updating of the reference point, but the actual progress variables fluctuate about the linear line. With BMD, the dynamics of the progress variables appear more similar to natural fluctuation in the sense that their values change nonlinearly over time. For example, in the R_g-perturbed BMD simulation (Figure 2, right bottom panel), the progress variable R_g increases slowly for the first 8 ns until it reaches ∼11.5 Å, after which it rises with a much steeper slope. This nonlinear progression can be utilized to identify possible free energy barriers, because a barrier impedes spontaneous fluctuations along the progress variable and thus longer sampling time is needed to cross it.[20,21]

Effect of Perturbation Methods and Progress Variable on Determining Path. R_g and h were used to guide the unfolding transition of BdpA. Unfolding trajectories, from TMD, SMD and BMD, were compared by the trajectory averaged rmsd values ($\overline{\text{rmsd}}_{ij}$), for which the snapshots from each trajectory were binned according to their R_g values. The calculated unfolding trajectories were thus examined as paths connecting conformations in space rather than time evolution of the system. To compare two trajectories, the rmsd was calculated between average structures from each R_g bin, and the rmsd values were averaged over all bins (see Methods for details). The resulting $\overline{\text{rmsd}}_{ij}$ value measures the overall spatial similarity between two trajectories, and the all-against-all evaluation is plotted in the matrix in Figure 3 according to the code shown in Table 1. This pairwise similarity matrix shows that the 42 trajectories naturally fall into three clusters. Interestingly, all h-perturbed trajectories fall into cluster A, all R_g-perturbed trajectories except RB3 fall into cluster B, and all R_g-perturbed, RB3 trajectories fall into cluster C. The groups of trajectories are clearly distinct from each other: The $\overline{\text{rmsd}}_{ij}$ value within any cluster is around 4 Å while that between two different clusters averages near 11 Å.

It is apparent from the distinct clustering in Figure 3 that, for the case of BdpA unfolding, the global geometry of the transition path is largely determined by the progress variable being h or R_g, and not by the perturbation method. Comparison of trajectories within a cluster provides no evidence that trajectories generated by either TMD, SMD, or BMD differ. The $\overline{\text{rmsd}}_{ij}$ values calculated between unfolding trajectories generated with different perturbation methods are equivalent to values obtained by comparing multiple runs by the same method and simulation length.

Inspecting all trajectories with the visualization program VMD[49] identified the global features shared by all members from the same cluster. These common features are shown in Figure 4 by two representative trajectories for unfolding from each cluster. In cluster A trajectories, the protein unfolds by first extending and unwinding H1 and H3 from the ends. Breaking of the H1−H2 and H2−H3 interhelical contacts follows as a result of the stretching at the ends by perturbing h. The protein then extends into a linear chain. In cluster B trajectories, the H1−H2 hairpin always opens up with a flipping of H1 at an early stage of unfolding. The opening up of the H2−H3 hairpin occurs later. In contrast, these two events happen in the opposite order in cluster C, where H3 flips to lose contact with H2 earlier than the separation of H1 and H2.

The stronger dependence of the transition path on the progress variable than the method is evident from a similarity shared by cluster B and C paths regardless of the big rmsd between them. In these two clusters, the opening up of one of the two hairpins (H1−H2 in B and H2−H3 in C) happens

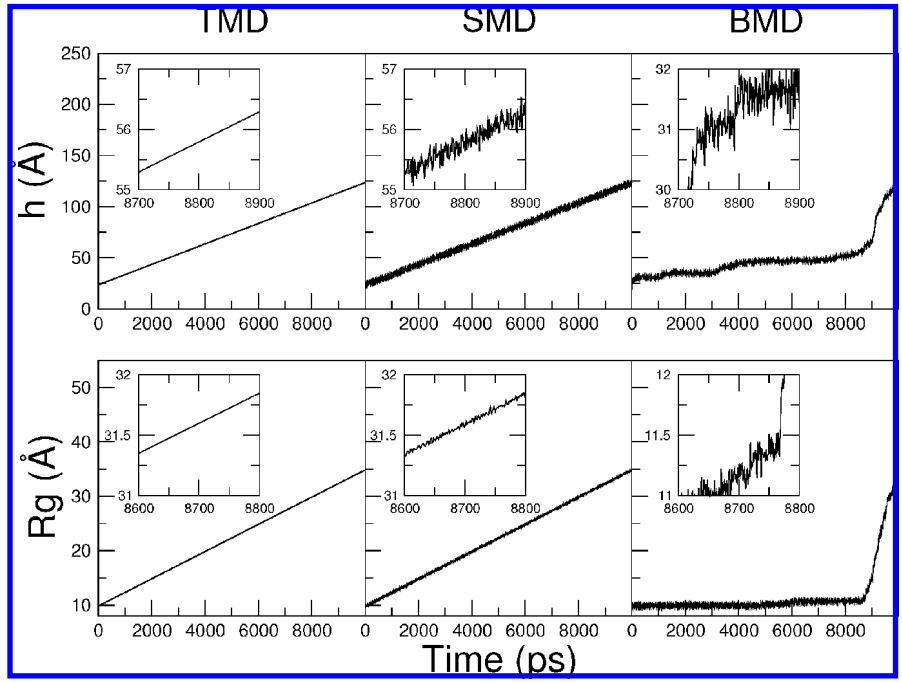

Figure 2. Progress variable as a function of time in representative BdpA unfolding trajectories generated by TMD, SMD, and BMD. The perturbed progress variable is the end-to-end distance (h) in the top panels and the radius of gyration (R_g) in the bottom panels. Insets show close-ups of a small portion of the progression curves.

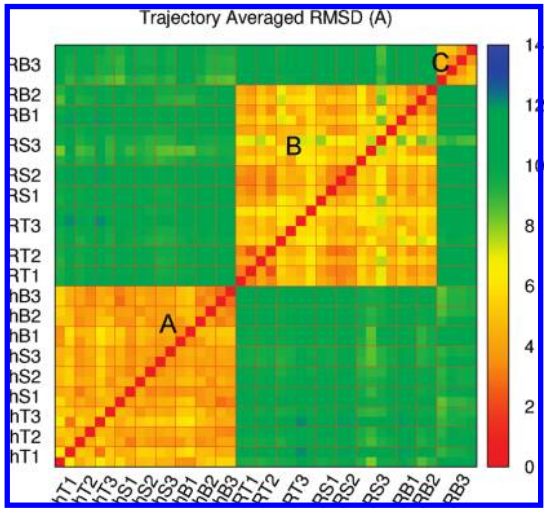

Figure 3. Pairwise trajectory averaged rmsd matrix for all BdpA unfolding trajectories. Each color square (pixel) shows the $\overline{\text{rmsd}}_{ij}$ value between a pair of unfolding trajectories. Trajectories are arranged in the same order on both axes so that the matrix is symmetric. Multiple trajectories with the same simulation condition are grouped together by thin lines in red, and the coding of simulation conditions is as in Table 1. Three clusters can be identified as follows: cluster A, all h-perturbed trajectories; cluster B, all R_g-perturbed trajectories except BMD, 10 ns ones; cluster C, R_g-perturbed, BMD, 10 ns trajectories.

as the first step of unfolding. This common feature is consistent with the external forces imposed by R_g perturbation in that the opening up of a hairpin will result in a big increase in R_g. In contrast, in all cluster A trajectories the unwinding of H1 and H3 from the ends precedes the breaking of the hairpin structure. This alternative pathway is also consistent with the pulling forces at the ends imposed by the end-to-end distance, h, perturbation.

Trajectories from the same cluster follow a common global trend but still display variability in structural details. One example of this variability is that H2 breaks in the middle (Figure 4, top right, $h = 60$ Å) and re-forms ($h = 76$ Å) before it is unwound in some but not all cluster A trajectories. Another example is that the unwinding of H1 accompanies H1−H2 separation in some but not all cluster B trajectories (Figure 4, middle left). Similar differences are also observed in cluster C trajectories (Figure 4, bottom). These local differences are not specific to a certain perturbation method. However, in general, the variability of R_g trajectories is greater than that of h trajectories in two aspects. First, R_g perturbation resulted in two distinct clusters (B and C) whereas h perturbation resulted in a single one (A). Second, R_g-perturbed trajectories have higher intracluster $\overline{\text{rmsd}}_{ij}$ values, as shown by more yellow-green elements in the similarity matrix in cluster B than in cluster A. This greater variability produced by the R_g bias is consistent with the more divergent nature of its corresponding forces.

BMD 10 ns R_g-Perturbed Trajectories Form a Distinct Cluster. Within R_g-perturbed trajectories, the RB3 trajectories generated by BMD in 10 ns (the longest time, with the smallest force constant) form a distinct group cluster C. As mentioned earlier, they differ from other R_g-perturbed trajectories in first opening up hairpin H2−H3 instead of H1−H2. In the three-helix-bundle structure of BdpA, both H1 and H3 form antiparallel interhelical tertiary interactions with H2, resulting in a nearly symmetrical topology. The interruption of one of these two tertiary interactions during unfolding can be considered a breaking of symmetry.[38] Assuming this symmetry to be perfect, its disruption should occur randomly at H1−H2 or H2−H3 as a result of the

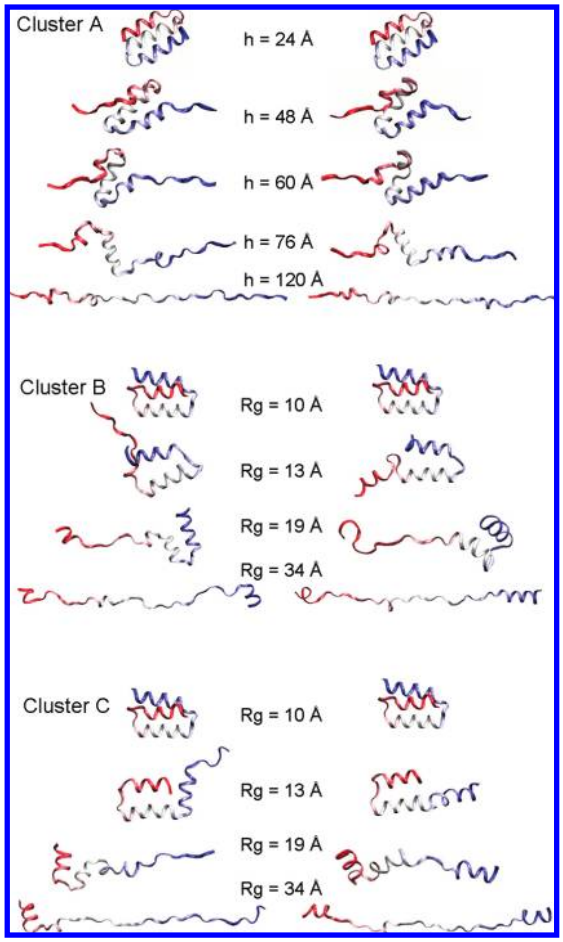

Figure 4. Snapshots unfolded to different extents for representative trajectories from each cluster. Helices H1, H2, and H3 are shown in red, white, and blue, respectively. These ribbon graphs show the sequence of key events involved in the unfolding, including the unwinding of individual helices and the interruption of different tertiary contacts. Two trajectories are shown for each cluster to exemplify variations within a cluster.

chaotic nature of molecular dynamics. However, here we observe H2−H3 breaking only in RB3 trajectories. This observation suggests an underlying picture in which both the protein's internal interactions and the external perturbation forces are slightly asymmetric and favor different symmetry-breaking positions, and the balance between the two asymmetric factors determines which contact breaks first. With short simulation lengths, the perturbation strength was high and all perturbation methods generated H1−H2 breaking trajectories only. As the simulation time increased, the perturbation strengths for all three methods were decreased, which would potentially cause a change of the balance between the external perturbation and internal interaction. Interestingly, in the longest simulation time (10 ns), only BMD sampled the alternative pathway, suggesting that BMD gives a smaller effective perturbation in the same simulation time. This smaller effective perturbation can be explained by the nonlinear sampling of BMD. With BMD, a larger portion of the total sampling time is spent before barrier-crossing events. Because barrier-crossing events are the critical events in determining which path to follow, BMD renders a weaker perturbation during these important events and thus sampled the alternative pathway in the longest simulation time.

Cluster C Has Lower PMF Values Than Cluster B. The TMD, SMD, and BMD perturbation methods were further examined for their ability to generate physically relevant transition paths by an assessment of the energetics of the unfolding trajectories. Because instantaneous energies are sensitive to local structural changes and contain a high level of noise, they were not used to compare the energetics of the trajectories. Instead, a potential of mean force (PMF) curve was calculated for a trajectory by using umbrella sampling and WHAM with snapshots taken from the trajectory as initial coordinates. The PMF curves aim to characterize the path obtained from trajectories generated for set conditions of perturbation method and length of simulation. It is recognized that there is variation among the curves because of limited sampling in the energy averaging on a rugged conformational landscape. Nonetheless, the PMF curves were found to be useful for comparative purposes.

The PMF with respect to R_g were calculated for all R_g-perturbed trajectories (clusters B and C) over the R_g range of 9.5−16.0 Å (see Methods for details). Figure 5 shows the average PMF and its variance as a function of R_g for each simulation condition. The average and variance were calculated from multiple (two to four) independent trajectories for each simulation condition and are shown in Figure 5a−c. Only the averaged PMF is shown in Figure 5d for clarity. All PMF curves share the same overall shape. There is a minimum at $R_g = 9.8$ Å, which is the R_g of the native structure. There are no other minima along the curves, a behavior also observed in BMD unfolding of the Fn3 domain.[6]

Paths generated by RB3 trajectories (cluster C) have the lowest PMF for R_g values > 11 Å, indicating an energetically more favorable unfolding path. Nevertheless, a comparison of shorter (0.5 and 2 ns) trajectories reveals that those generated by BMD are not more energetically favorable than SMD or TMD generated trajectories. Interestingly, the structural difference between RB3 and other R_g-perturbed trajectories for R_g > 11 Å is the breaking of H2−H3 contact or that of H1−H2. It is likely that the difference in PMF between RB3 and other trajectories is a reflection of the two different unfolding pathways described earlier. The fact that RB3 paths have lower PMF curves is also consistent with the earlier discussed notion that BMD renders a smaller effective perturbation in the same simulation time, thus samples an alternative, lower free energy pathway.

3.2. Lyn Activation/Deactivation. The second example system on which the three perturbation methods were tested is the conformational transition of the Lyn kinase CD. This process is a transition between two end states with well-defined structures, in contrast to the unfolding of BdpA.

The catalytic domain of Src-family kinases, such as Lyn CD, adopts different conformations in the active and inactive states shown in Figure 6. Major conformational changes of the CD upon activation are the opening of the cleft between the N- and C-terminal lobes (N-lobe and C-lobe), the rotation and translation of the αC helix, and rearrangement of the

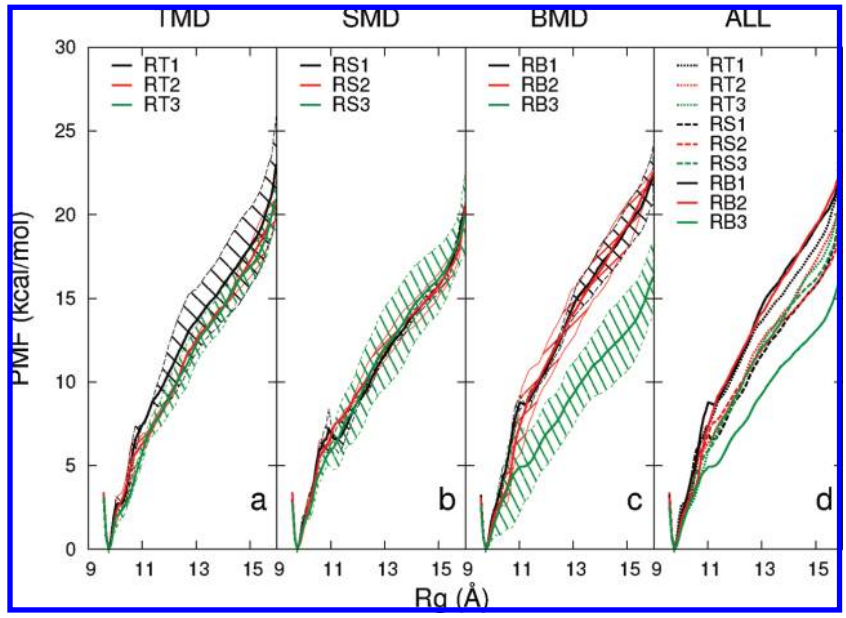

Figure 5. Potential of mean force (PMF) as a function of R_g for R_g-perturbed BdpA unfolding paths. The first three panels show the PMF curves, labeled by the codes shown in Table 1, for each time length for (a) TMD, (b) SMD, and (c) BMD. An average curve is calculated from PMF profiles of multiple trajectories (two to four) with the same simulation condition, and the shaded area shows the variance. (d) Average PMF curves.

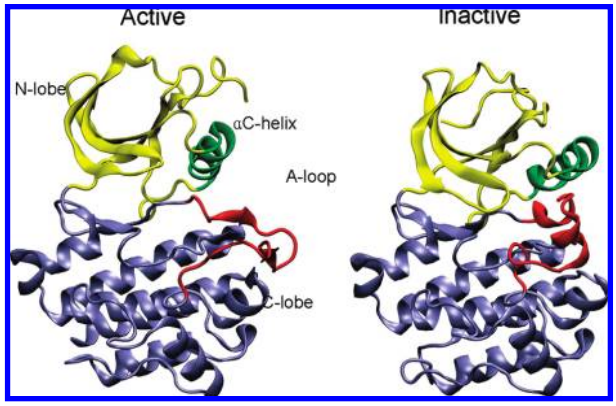

Figure 6. Comparison of the catalytic domain conformations of a Src kinase Lyn in the active (left) and inactive (right) states. The structures are from equilibrium molecular dynamics simulations[23] initiated with coordinates obtained by homology modeling with Lck[43] and Hck[44] crystal structures. Upon activation, the N-lobe (yellow) and C-lobe (blue) move apart, the helix C (green) moves and rotates inward, and the A-loop (red) deforms from two short helices to extended structure.

activation loop (A-loop). The 20-residue-long A-loop undergoes the most complex rearrangement. It adopts an extended conformation in the active state, whereas in the inactive state it folds back into the catalytic cleft, forming two short α helices.

Both activation and deactivation transition paths were calculated for the Lyn CD using the three perturbation MD methods. The two target structures were obtained by equilibrium MD with explicit water molecules initiated from homology modeled coordinates (see Methods). All three perturbation methods were applied with the MSID progress variable (see Methods for definition) to guide the transition to the target structure. Trajectories perturbing the root-mean-square deviation (rmsd) were also calculated with TMD. Details about the simulations are listed in Table 2.

Path Completeness. Unlike unfolding, the conformational transition process modeled in this example has a defined target configuration. By definition, TMD guarantees that the target configuration will be reached, while SMD only restrains the system to the target value with a defined force constant. With BMD, even the final reference point ρ_0 is not guaranteed to be the target value in a certain amount of time because it depends on spontaneous fluctuations of the system.

The final states of Lyn generated by the three methods are in accord with the above expectations. Figure 7 shows a superposition based on all heavy atoms of the target structure and the structure at the end of the perturbation MD run for representative deactivation simulations. TMD guided the transition to the exact target. With SMD, the final structure and the target configuration differ slightly (overall rmsd = 1.2 Å and A-loop rmsd = 2.8 Å), with differences in several backbone torsion angles in the C-terminal region of the A-loop. The final structure generated by BMD has the largest deviation (overall rmsd = 1.9 Å, A-loop rmsd = 3.4 Å) from the target of the three perturbation methods. The major deviation is again in the A-loop region but also involves differences at the secondary structure level. Specifically, the two short A-loop helices do not form.

Effect of Perturbation Methods and Progress Variable on Determining Path. Figure 8a shows the projection of all 20 trajectories onto two coordinates: Δrmsd of the A-loop and Δrmsd of the N-lobe. $\Delta\text{rmsd}_{\text{A-loop}}$ and $\Delta\text{rmsd}_{\text{N-lobe}}$ of a configuration are defined as the difference between the rmsd of this configuration to the active structure and the rmsd to the inactive structure for the heavy atoms of the A-loop (residues 406−526, c-Src numbering) and the N-lobe (resi-

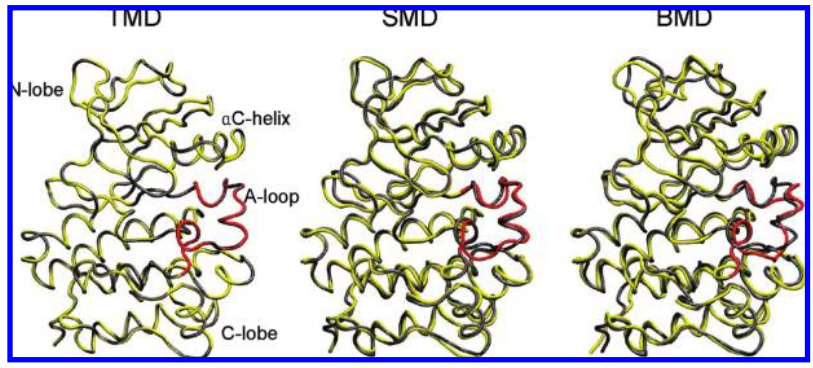

Figure 7. Final structures of Lyn CD deactivation trajectories generated by TMD, SMD, and BMD. These calculated final structures are shown in yellow except for the A-loop colored in red, whereas the native inactive structure, i.e., the target configuration, is shown in gray.

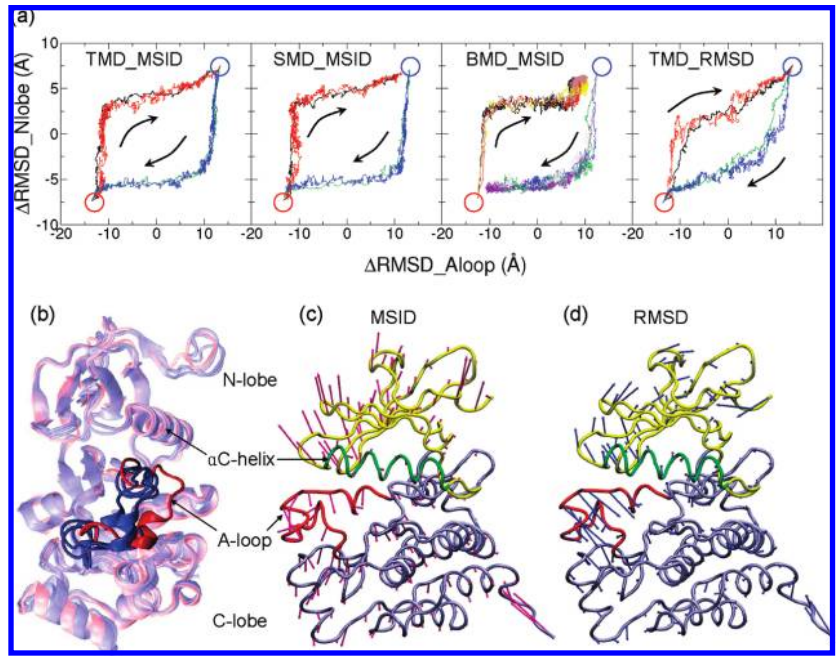

Figure 8. (a) Lyn CD activation/deactivation trajectories projected on the plane of Δrmsd of the N-lobe and Δrmsd of the A-loop. For each panel, multiple trajectories are shown in different colors. See Table 2 for color coding. Δrmsd = rmsd(to active) − rmsd(to inactive), superimposing the C-lobe of the protein. The active and inactive configurations are labeled by red and blue circles, respectively, and the transition directions are indicated by arrows. (b) Four snapshots with the N-lobe opened to the same extent from four Lyn CD activation trajectories generated by TMD−MSID, SMD−MSID, BMD−MSID and TMD−rmsd. For each trajectory, the first snapshot that reaches $\Delta rmsd_{N\text{-}lobe} = -5$ Å is shown. The three snapshots with MSID as the progress variable are colored in blue, whereas the one with MSID as the progress variable is colored in red. The A-loop region is shown in solid cartoon representation, while the rest of the protein is shown as transparent. (c, d) External forces acting on Lyn CD at the beginning of activation by perturbing MSID (c) or rmsd (d). The forces are averaged for each residue and scaled so that their average magnitudes are the same in both cases. rmsd induces relatively greater forces on the A-loop region compared to what MSID does.

dues 261−301 and 317−341), respectively, with the C-lobe superimposed.

Trajectories generated with the same progress variable, MSID, using three perturbation methods are similar on the two-dimensional Δrmsd surface, as shown in the first three panels of Figure 8a. These trajectories follow two different major paths depending on their direction being activation or deactivation. The two motions take place in two distinct steps that follow each other, as indicated by a sudden change in slope once one change is completed. For either direction, the lobe−lobe motion occurs before the rearrangement of the A-loop and the forward path is not the opposite of the reverse path. This behavior indicates that the energy barrier for the hinge-like lobe−lobe motion is lower than that for A-loop rearrangement which involves a change in secondary structure, and suggests that the global geometry of a transition path is dependent not only on the perturbation force, but also on the ease of the motions induced by the perturbation force. The observation of globally similar paths for all MSID-perturbed trajectories suggests that the dependence of the path on the perturbation method is minor, even though BMD generated deactivation paths that slightly differ from those generated by TMD and SMD at the end (Figure 8a, panel 3) due to a delayed motion of the αC helix.

The fourth combination of perturbation method and progress variable, TMD with the rmsd progress variable, generated paths globally different from all MSID paths. On the two-dimensional surface, these paths progress more linearly rather than stepwise, which suggests that the N-lobe and the A-loop move toward the target structure with a higher degree of cooperativity. This result is again in accord with the notion that the global path is largely determined by the choice of the progress variable.

The difference in the linearity between paths generated with MSID and rmsd as the progress variable can be seen structurally in snapshots from representative activation trajectories. Figure 8b shows an overlay of four snapshots from activation trajectories generated by the four combinations of perturbation method and progress variable. Upon activation, both the N-lobe and the A-loop open up from a closed conformation, and the four snapshots were chosen so that their N-lobes are opened to the same extent: $\Delta\text{rmsd}_{\text{N-lobe}} = -5$ Å. Figure 8b shows that, in the snapshot generated by perturbing rmsd (red), the A-loop motion has progressed further toward the active conformation while A-loops of the other snapshots generated by perturbing MSID (blue) remain close to the inactive state. It is worth noting that both MSID and rmsd are global measurements of the distance between two structures, and therefore as progress variables they are less distinct than h and R_g. Nonetheless, the subtle distinction does result in visible differences between generated paths, again showing the sensitivity of the perturbation MD methods to the choice of progress variable.

To examine the reason for the generation of different paths with the two progress variables, the forces imposed by perturbation on MSID or rmsd on each residue at the beginning of the activation are shown in Figure 8c,d. The magnitudes of the forces are scaled so that their average magnitude for the molecule is the same. It can be seen in the visualization that rmsd induces relatively greater initial forces on the A-loop region compared to the forces imposed by MSID. The variation and difference in direction in the force vectors explains the observed differences between the paths. Simulations perturbing the rmsd progress variable were not carried out with BMD and SMD, but results similar to that observed by TMD are expected.

4. Conclusions and Discussion

In the present study, we compared TMD, SMD, and BMD methods for guiding conformational transitions with external perturbation. Two distinct transition systems were explored: the unfolding of BdpA with implicit solvation and a heterogeneous end state, and the activation of Lyn CD in explicit water with two well-defined end states. Results from the three perturbation methods with two choices of progress variable and varying simulation length or biasing force constant showed that the perturbation methods generate similar transition paths for a given progress variable, as defined by global structural features. On the other hand, the path was strongly dependent on the choice of progress variable. Specifically, choosing to perturb R_g or h for unfolding BdpA determines whether the protein unfolds by stretching from the ends or by opening up one of the two interhelical contacts. Also, choosing to perturb MSID or rmsd to guide the conformational change of Lyn CD affects the cooperativity of the movements of its N-lobe and A-loop.

The path dependence on the progress variable was observed to be related to the direction and relative magnitude of the perturbation force, and these were the dominant influence regardless of the perturbation method. The alternative forces induce different motions of the molecule in certain directions and, together with the intrinsic ease of these motions, determine the transition path. In most cases, it is not clear that a priori knowledge can be used to decide which progress variable is a good reaction coordinate.

Although the three methods generated similar paths in terms of the global structural features for a given progress variable in most cases, the RB3 BdpA unfolding simulations generated by BMD with the lowest force constant identified an alternative class of paths, whereas longer SMD and TMD simulations generated similar paths to those generated by shorter simulations. The fact that this alternative class of paths is more energetically favorable is likely due to the softer and nonlinear perturbation rendered by BMD. BMD realizes a nonlinear updating scheme for moving along the progress variable, and thus the perturbation strength varies along a trajectory. In simulations of the same total length, BMD spends a greater amount of time going up a barrier than going down a slope. Thus the perturbation strength is comparatively low with BMD when the system searches for an easy route to overcome a free energy barrier. It should be noted that this nonlinear updating scheme of BMD would not guarantee the sampling of globally lower free energy paths, because an effectively lower perturbation strength only helps find an easy route to cross a local free energy barrier, which may or may not be a part of the global optimal path. In fact, the blindness to global features of the free energy surface is a common problem shared by all three perturbation methods.

In the conformational transition of Lyn CD, the backward paths are not the reverse copies of the forward paths with all three perturbation methods. This apparent violation of the microscopic reversibility suggests that the nonequilibrium results obtained with the perturbation MD methods do not approach the equilibrium results at the time lengths examined in our simulations. With the perturbation methods in the nonequilibrium scenario, the sequence of events along a path is determined by a combined effect of the perturbation and the internal ease of certain motions. Under the perturbation, the parts that feel the strongest perturbation and the weakest hindrance move toward the target first. Due to this lack of reversibility, caution should be taken when interpreting perturbation MD results to infer features such as sequence of events. Nevertheless, some insight into the ease of a certain motion can be obtained with these perturbation methods.

Global and/or local structural properties of the proteins should be considered when choosing a perturbation MD method. As our simulations on the conformational transition of Lyn CD show, BMD trajectories can be trapped in local minima in the presence of significant energy barriers (such as the A-loop rearrangement), preventing the system from

arriving at the target conformation. TMD and, for the most part, SMD ensure the completeness of the transition, although the likelihood of the path followed by TMD has not been assessed here. SMD closely resembles the AFM experiment, and it is preferred when direct analogy is desired. Further use of SMD allows the application of the Jarzynski equality to calculate equilibrium free energy differences.[50] Jarzynski's equality requires the external perturbation to be a component of the Hamiltonian as an explicit function of time, a condition satisfied by the perturbation form of SMD but by neither TMD nor BMD.[51]

Abbreviations and Symbols. BMD, biased molecular dynamics; TMD, targeted molecular dynamics; SMD, steered molecular dynamics; ρ, progress variable; h, end-to-end distance; R_g, radius of gyration; rmsd, root-mean-square deviation; MSID, mean square internal deviation.

Acknowledgment. This work was supported by National Institutes of Health (NIH) Grants GM039478 and GM083605 (C.B.P.), and a Purdue University reinvestment grant. E.O. and H.H. were supported by Purdue Research Foundation Fellowships. The authors acknowledge contributions from Bonnie Co in the initial stages of this work.

References

(1) Schlitter, J.; Engels, M.; Kruger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291–308.

(2) Schlitter, J.; Engels, M.; Kruger, P. *J. Mol. Graphics* **1994**, *12*, 84–89.

(3) Grubmueller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997–999.

(4) Leech, J.; Prins, J.; Hermans, J. *IEEE Comput. Sci. Eng.* **1996**, *3*, 38–45.

(5) Marchi, M.; Ballone, P. *J. Chem. Phys.* **1999**, *110*, 3697–3702.

(6) Paci, E.; Karplus, M. *J. Mol. Biol.* **1999**, *288*, 441–459.

(7) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568–1581.

(8) Lu, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. *Biophys. J.* **1998**, *75*, 662–671.

(9) Krammer, A.; Lu, H.; Isralewitz, B.; Schulten, K.; Vogel, V. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 1351–1356.

(10) Marszalek, P. E.; Lu, H.; Li, H. B.; Carrion-Vazquez, M.; Oberhauser, A. F.; Schulten, K.; Fernandez, J. M. *Nature (London)* **1999**, *402*, 100–103.

(11) Law, R. J.; Munson, K.; Sachs, G.; Lightstone, F. C. *Biophys. J.* **2008**, *95*, 2739–2749.

(12) Swift, R. V.; McCammon, J. A. *Biochemistry* **2008**, *47*, 4102–4111.

(13) Zou, J.; Wang, Y.-D.; Ma, F.-X.; Xiang, M.-L.; Shi, B.; Wei, Y.-Q.; Yang, S.-Y. *Proteins: Struct., Funct., Bioinf.* **2008**, *72*, 323–332.

(14) Apostolakis, J.; Ferrara, P.; Caflisch, A. *J. Chem. Phys.* **1999**, *110*, 2099–2108.

(15) Bui, J. M.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15451–15456.

(16) Ferrara, P.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **2000**, *104*, 4511–4518.

(17) Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11905–11910.

(18) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 252–260.

(19) Paci, E.; Caflisch, A.; Pluckthun, A.; Karplus, M. *J. Mol. Biol.* **2001**, *314*, 589–605.

(20) Morra, G.; Hodoscek, M.; Knapp, E. W. *Proteins: Struct., Funct., Genet.* **2003**, *53*, 597–606.

(21) Li, Y.; Zhou, Z.; Post, C. B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7529–7534.

(22) Stultz, C. M. *Protein Sci.* **2006**, *15*, 2166–2177.

(23) Ozkirimli, E.; Post, C. B. *Protein Sci.* **2006**, *15*, 1051–1062.

(24) Levinson, N. M.; Kuchment, O.; Shen, K.; Young, M. A.; Koldobskiy, M.; Karplus, M.; Cole, P. A.; Kuriyan, J. *PLoS Biol.* **2006**, *4*, e144.

(25) Kastenholz, M. A.; Schwartz, T. U.; Hünenberger, P. H. *Biophys. J.* **2006**, *91*, 2976–2990.

(26) Perdih, A.; Kotnik, M.; Hodoscek, M.; Solmajer, T. *Proteins: Struct., Funct., Bioinf.* **2007**, *68*, 243–254.

(27) Ozkirimli, E.; Yadav, S. S.; Miller, W. T.; Post, C. B. *Protein Sci.* **2008**, *17*, 1871–1880.

(28) Tikhonova, I. G.; Best, R. B.; Engel, S.; Gershengorn, M. C.; Hummer, G.; Costanzi, S. *J. Am. Chem. Soc.* **2008**, *130*, 10141–10149.

(29) Zhong, W.; Guo, W.; Ma, S. *FEBS Lett.* **2008**, *582*, 3320–3324.

(30) Matrai, J.; Jonckheer, A.; Joris, E.; Kruger, P.; Carpenter, E.; Tuszynski, J.; Maeyer, M. D.; Engelborghs, Y. *Eur. Biophys. J.* **2008**, *38*, 13–23.

(31) Carrion-Vazquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. *Nat. Struct. Biol.* **2003**, *10*, 738–743.

(32) Sato, S.; Religa, T. L.; Fersht, A. R. *J. Mol. Biol.* **2006**, *360*, 850–864.

(33) Vu, D. M.; Myers, J. K.; Oas, T. G.; Dyer, R. B. *Biochemistry* **2004**, *43*, 3582–3589.

(34) Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001**, *8*, 552–558.

(35) Guo, Z.; Brooks, C. L.; Boczko, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10161–10166.

(36) Garcia, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898–13903.

(37) Cheng, S.; Yang, Y.; Wang, W.; Liu, H. *J. Phys. Chem. B* **2005**, *109*, 23645–23654.

(38) Itoh, K.; Sasai, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7298–7303.

(39) Binnig, G.; Quate, C. F.; Gerber, C. *Phys. Rev. Lett.* **1986**, *56*, 930–933.

(40) Gouda, H.; Torigoe, H.; Saito, A.; Sato, M.; Arata, Y.; Shimada, I. *Biochemistry* **1992**, *31*, 9665–9672.

(41) Im, W.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691–1702.

(42) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(43) Yamaguchi, H.; Hendrickson, W. A. *Nature (London)* **1996**, *384*, 484–489.

(44) Schindler, T.; Sicheri, F.; Pico, A.; Gazit, A.; Levitzki, A.; Kuriyan, J. *Mol. Cell* **1999**, *3*, 639–648.

(45) Paci, E.; Smith, L. J.; Dobson, C. M.; Karplus, M. *J. Mol. Biol.* **2001**, *306*, 329–347.

(46) Markwick, P. R. L.; Doltsinis, N. L.; Schlitter, J. *J. Chem. Phys.* **2007**, *126*, 045104.

(47) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187–199.

(48) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.

(49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14* (33−8), 27−8.

(50) West, D. K.; Olmsted, P. D.; Paci, E. *J. Chem. Phys.* **2006**, *125*, 204909.

(51) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.

CT9000153