JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

# Assessing the Accuracy of Two Enhanced Sampling Methods Using EGFR Kinase Transition Pathways: The Influence of Collective Variable Choice

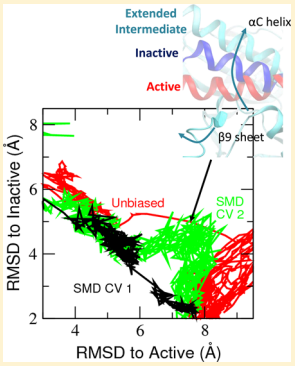
Albert C. Pan,*[,†] Thomas M. Weinreich,[†] Yibing Shan,[†] Daniele P. Scarpazza,[†] and David E. Shaw*[,†,‡]

[†]D. E. Shaw Research, New York, New York 10036, United States

[‡]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States

Supporting Information

**ABSTRACT:** Structurally elucidating transition pathways between protein conformations gives deep mechanistic insight into protein behavior but is typically difficult. Unbiased molecular dynamics (MD) simulations provide one solution, but their computational expense is often prohibitive, motivating the development of enhanced sampling methods that accelerate conformational changes in a given direction, embodied in a collective variable. The accuracy of such methods is unclear for complex protein transitions, because obtaining unbiased MD data for comparison is difficult. Here, we use long-time scale, unbiased MD simulations of epidermal growth factor receptor kinase deactivation as a complex biological test case for two widely used methods—steered molecular dynamics (SMD) and the string method. We found that common collective variable choices, based on the root-mean-square deviation (RMSD) of the entire protein, prevented the methods from producing accurate paths, even in SMD simulations on the time scale of the unbiased transition. Using collective variables based on the RMSD of the region of the protein known to be important for the conformational change, however, enabled both methods to provide a more accurate description of the pathway in a fraction of the simulation time required to observe the unbiased transition.




Observing pathways between stable protein conformations in atomic detail can provide crucial information for understanding biological mechanisms. The conformations visited along a pathway are transient, however, making their characterization difficult experimentally. One way to capture transition pathways at an atomic level of detail is using all-atom molecular dynamics (MD) simulation. In theory, one could run a long, unbiased simulation and simply watch such transitions occur. In practice, this can be a prohibitively intensive computational task, especially for thermally activated processes, in which a majority of simulation time is spent fluctuating within local free energy basins. As a result, in addition to the enhanced sampling MD methods that accelerate sampling of conformational space in general,[1−8] myriad methods—often, as here, also referred to as enhanced sampling methods—have been specifically developed to map and sample realistic transition pathways using a fraction of the simulation time needed to observe the unbiased events.[9−19]

Steered molecular dynamics (SMD) simulation[14,15] and related approaches[20] are among the most commonly used enhanced sampling methods for mapping out conformational change pathways. In SMD, a chosen collective variable (CV), such as the best-fit root-mean-square deviation (RMSD) of a subset of protein atoms to a model structure, is harmonically restrained to a target value that linearly increases or decreases with simulation time. A different class of methods takes an existing path and refines it using several conformational replicas along the pathway.[11,21,22] One such technique, the string method in collective variables,[22] evolves a "string"[11] of images defined in CV space such that it converges to a minimum free energy pathway (see Methods).

Although these and many more such path-finding methods are widely used, there have been very few tests of their accuracy in studies of complex biological systems. Previous studies have often successfully applied enhanced sampling methods to simple model systems. Sampling pathways in a simple model, however, can present qualitatively different challenges from doing so in a complex biological system such as a solvated protein with several hundred residues, represented in atomic detail.[23] Regardless of the sampling method used, comparing sampling results from simulations with experiments is not easy for such systems, and even when it can be done, the results are often hard to interpret. In particular, when comparing sampled paths with experimental data it may be difficult to distinguish errors due solely to the chosen sampling method, errors that arise from differing conditions between the simulation and the experiment, and errors in the physical model underlying the simulation (the "force field"). A direct comparison between transitions observed using an enhanced sampling technique and those seen in an unbiased MD simulation thus offers the

**Special Issue:** Free Energy Calculations: Three Decades of Adventure in Chemistry and Biophysics

**Received:** March 16, 2014
**Published:** June 3, 2014

potential for obtaining new insights into the characteristics of enhanced sampling techniques.

In this paper, we use a conformational change in the epidermal growth factor receptor (EGFR) kinase to perform such a comparison. We found that common CV choices, based on the C$\alpha$ RMSD of the entire protein to a target structure, prevented these methods from reproducing an accurate pathway, even in SMD simulations on time scales comparable to those of the unbiased transition. When used with an RMSD selection that focused on the region of the protein known to be important for the conformational change, however, both methods produced a more accurate pathway in significantly less simulation time than the unbiased transitions.

Transitions from catalytically active to inactive conformations of EGFR kinase were recently observed in unbiased MD simulations.[24] Each transition, taking ~10 $\mu$s, occurred by way of highly extended intermediate conformations characterized by large fluctuations of the activation loop and the $\alpha$C helix. By many measurements, the active state differed more from several of these simulation-derived extended intermediates than from the inactive state. Such a pathway would thus be difficult to infer from an examination of the known crystal structures of these two states.

To gauge the success of SMD and the string method in accurately reproducing these transitions, we focused on how well the method's generated pathways captured two key features of the unbiased transitions: (i) the extent to which EGFR kinase undergoes the transition by way of an extended intermediate conformation and (ii) the fluctuations of the activation loop during the transition. In a typical unbiased transition, EGFR visits an extended intermediate conformation in which the $\alpha$C helix is displaced beyond the position it assumes in the inactive conformation and in which the $\beta$9 sheet, at the beginning of the activation loop, is between its location in the active state and the region where the two-turn helix will eventually form in the inactive state (Figure 1, inset).[24,25] Transitioning by way of this intermediate conformation explains the outward curve in the path of the unbiased transition when it is projected onto the 2D RMSD space, as shown in Figure 1. The entire activation loop also fluctuates significantly during the transition, with C$\alpha$ root-mean-square fluctuations (RMSF) as high as 13 Å.

We note that our unbiased simulations are unlikely to have exhaustively sampled the active to inactive pathway of EGFR kinase. In particular, the details of the transition (such as specific side-chain rotations) are likely unconverged. The two features mentioned above (transition by way of an extended intermediate and large RMSF fluctuations of the activation loop during the transition), however, are shared by all five unbiased transition pathways, and we believe that they form a good basis for comparison with enhanced sampling results (see Methods and Supporting Information Figure S1).

We first applied SMD by steering an active-like conformation of EGFR kinase toward an inactive-like conformation (see Methods), with the simulated kinase's similarity to the inactive-like conformation characterized by a single, straightforward CV: the whole-protein C$\alpha$ RMSD (residues 706–979), which excludes the flexible N and C termini. In these SMD simulations, the RMSD CV was harmonically restrained to a target value that decreased linearly over simulation time to 0 Å away from an inactive-like conformation (see Methods). We found that these SMD simulations could not accurately reproduce either of the two key features of the unbiased

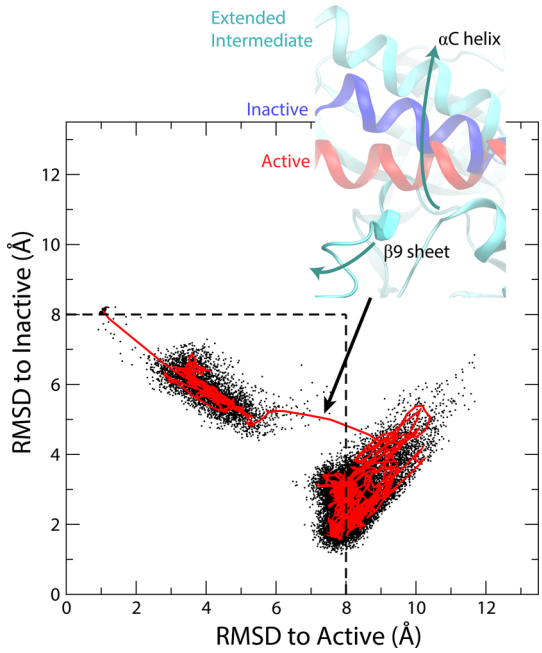



**Figure 1.** Transitions from active to catalytically inactive EGFR kinase conformations occur by way of an extended intermediate state. A representative transition from the active to a catalytically inactive conformation of EGFR kinase obtained from an unbiased MD simulation is shown as a projection (black dots) onto the RMSD to the active (PDB ID 2ITP)[32] and to the Src-like inactive (PDB ID 2GS7)[33] crystal structures. The red curve is a smoothed version of this transition, using a 50 ns running average. The molecular snapshot (inset) shows a typical example of an extended intermediate conformation (cyan), through which the unbiased transition occurs, where the $\alpha$C helix is displaced above its position in the inactive state and the $\beta$9 sheet transitions toward the region of the two-turn helix. The blue and red helices depict the position of the $\alpha$C helix in the inactive and active crystal structures, respectively. Here, the RMSD is calculated, as in Shan et al.,[24] using the C$\alpha$ atoms of residues in the $\alpha$C helix and $\beta$9 sheet region (residues 756–769 and 857–863) after aligning the protein to residues in the kinase domain (residues 703–861 and 896–979), excluding the flexible activation loop region. The residue numbering is based on the active crystal structure (PDB ID 2ITP). The transition by way of the extended intermediate conformation was similar in four other unbiased trajectories.

transitions mentioned above (Figures 2A and B, black line). In the 100 ns SMD simulations, for example, rather than transitioning by way of an extended intermediate conformation, the conformational change progressed along a more confined pathway, in which the $\beta$9 sheet and $\alpha$C helix transitioned in concert, and in which the $\alpha$C helix remained in the region delimited by the position of the inactive and active structures. The RMSF of the activation loop in the 100 ns SMD paths was also far below the typical values of the unbiased trajectories. Surprisingly, even SMD simulations as long as 10 $\mu$s—the typical transition time of an unbiased trajectory—were unable to accurately reproduce the correct transition path or the RMSF of the activation loop (Figure 2A and B, blue line). A typical example is shown in Figure 2A (blue line). Two additional 10-$\mu$s SMD simulation replicates that began in the same configuration, but had different initial velocities, showed similar results, as did three replicates with a weaker force constant (Supporting Information Figure S2).

The choice of CV can be important for the application of methods that sample transition paths,[20,22] so we investigated


dx.doi.org/10.1021/ct500223p | J. Chem. Theory Comput. 2014, 10, 2860–2865

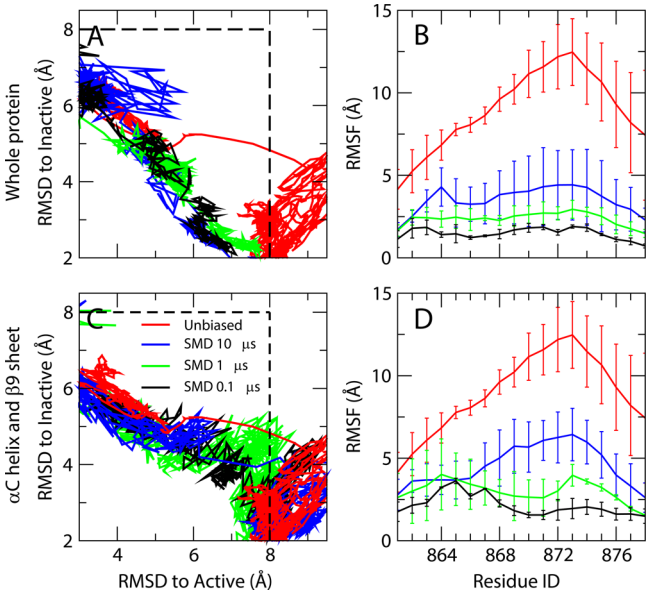



**Figure 2.** Accuracy of SMD in reproducing the features of an unbiased transition depends sensitively on the choice of CVs and the length of the simulation. (A, C) Projections of representative unbiased and SMD transition paths onto the 2D RMSD plot of Figure 1 are shown. The red curve is the path of the unbiased transition, and the black, blue, and green curves show SMD simulations of various lengths. (A) SMD simulations using a whole-protein C$\alpha$ RMSD selection as a CV (C$\alpha$ atoms of residues 706−979) all fall short of approaching the extended intermediate region delineated by the unbiased path (red curve), even in simulations as long as 10 $\mu$s. On the other hand, (C) SMD simulations using a more restricted selection from the $\alpha$C helix and $\beta$9 sheet (C$\alpha$ atoms of residues 756−769 and 857−869) approach that region much more closely, even in simulations as short as 100 ns. Results from independent SMD runs, as well as SMD runs with a weaker force constant, are consistent with the results shown here (Supporting Information Figures S2 and S3). (B, D) The corresponding RMSF curves of the activation loop, which undergoes fluctuations as high as 13 Å in the unbiased simulations (red curve), are shown. Each curve is averaged over three independent trajectories during the transition through the extended intermediate, and error bars represent standard errors (see Methods for more details of the calculation, including how the transition region was defined). The RMSF levels in the SMD runs never quite approach those seen in the unbiased runs; the magnitude of fluctuations increases with SMD simulation length, and the restricted selection produced overall higher RMSF levels than did the whole-protein selection. We note that the RMSD projection used in (A) and (C) is similar, but not identical to, the restricted RMSD selection used as a CV: In the plots (as described in the caption to Figure 1) the protein is first aligned to a larger selection of the kinase domain before the RMSD of the restricted selection—which involves residues in the $\alpha$C helix and $\beta$9 sheet—is determined; in the CV, the restricted selection is used both for alignment and RMSD determination.

how sensitively the success of SMD in determining a realistic pathway depended on this choice. A CV that closely approximates the true reaction coordinate of the process (i.e., the coordinate that drives the transition[12,26]) is usually an ideal choice. In the case of kinase activation, the $\alpha$C helix and activation loop are known to be regions of the protein important for the transition,[24,25] so we performed another series of SMD simulations in which we directly steered a single RMSD CV that included the C$\alpha$ atoms of just the $\alpha$C helix and $\beta$9 sheet (residue IDs 756−769 and 857−863), as opposed to those of the entire protein. Using this restricted RMSD selection, we found that a 100 ns SMD simulation reproduced well the gross features of the unbiased transition paths. In particular, these SMD simulations resulted in paths which transitioned by way of conformations much closer to the extended intermediate region of the unbiased path (Figure 2C, see also Supporting Information Figure S3). The RMSF of the activation loop also slightly improved—especially near the most mobile region, around residue 873—because less of the protein was being restrained, but the RMSF was still below that of the unbiased transitions (Figure 2D).

We proceeded to conduct a similar comparison study using the "on-the-fly" version of the string method in collective variables, an improved and simplified version of the method that allows string images and conformational replicas to be evolved concurrently.[21] The method was seeded with paths that were generated from fast, 10 ns SMD simulations. These initial transition paths deviated greatly from the unbiased path, especially in the portion of the transition in which the protein was in an extended intermediate conformation (Figure 3A and

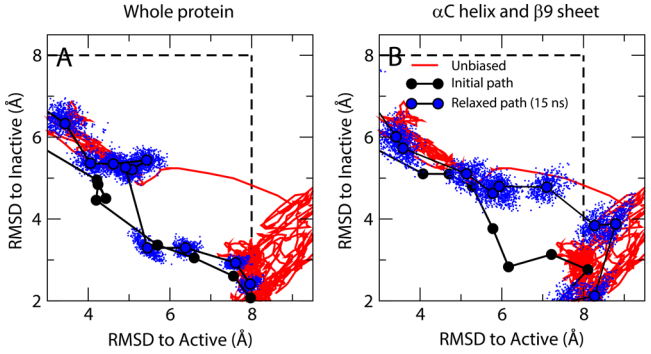



**Figure 3.** The string method can quickly refine a poor initial path when the CVs are well chosen. Shown here are projections of results from the on-the-fly string method onto the 2D RMSD plot of Figure 1 using (A) the whole protein and (B) the $\alpha$C helix and $\beta$9 sheet RMSDs as CVs. Black circles represent conformational replicas of an initial transition path obtained from 10 ns SMD runs. The blue circles represent the same path after refining with the on-the-fly string method for 15 ns per image and are averages of the final positions of the string in CV space from 14.5 to 15 ns; the blue dots represent the spread around the final points during those 500 ps. The initial paths deviate extensively from the unbiased transition (red curve) near the extended intermediate region. The string method−refined path using the RMSD of the $\alpha$C helix and $\beta$9 sheet as a CV, however, approaches this extended conformation very closely. In contrast, the string method−refined path using the whole-protein RMSD as a CV stays relatively close to the initial path.

B, black curves). Because the string method allows the flexibility of using multiple CVs, and because it relaxes to the minimum free energy path in the space of prescribed CVs, we refined these paths using *two* CVs. In addition to the C$\alpha$ RMSD to the inactive structure, which was used to generate the SMD paths, the C$\alpha$ RMSD to the active crystal structure was also used. As we did for the SMD runs, we performed two string calculations: one using the whole-protein RMSD selection and the other using the RMSD selection restricted to the $\alpha$C helix and $\beta$9 sheet.

Similarly to the SMD simulations, the string method refinement using the restricted selection (Figure 3B) was more accurate in reproducing the unbiased path than was the whole-protein selection (Figure 3A), particularly in the extended intermediate region. The string method sampled


dx.doi.org/10.1021/ct500223p | J. Chem. Theory Comput. 2014, 10, 2860−2865

the transition region more thoroughly than did SMD or unbiased MD (Supporting Information Figure S5). As in the SMD simulations, however, the activation loop RMSF in the string method–refined pathways did not approach the levels it reached in the unbiased simulation (Supporting Information Figure S4). This is not surprising, given the short simulation length of these runs; the sampling of slower orthogonal degrees of freedom was likely incomplete. The string method in collective variables, however, is relatively insensitive to the number of CVs used.[22,27] Assuming that additional CVs representing these degrees of freedom could be identified, incorporating them into the string method would thus be straightforward, and doing so might help improve sampling of the activation loop even on short time scales. Interestingly, the whole-protein RMSD selection did not facilitate the discovery of accurate transition paths even though two CVs were used (the RMSDs to the active and inactive states).

In summary, there have been few comparisons of enhanced sampling MD methods and unbiased MD simulations in complex biological systems, but such comparisons can be valuable tests of the accuracy of enhanced sampling techniques. We have assessed the accuracy of two popular enhanced sampling methods—SMD and the string method—in reproducing the unbiased transition of EGFR kinase deactivation by comparing the results of these methods to unbiased transitions observed in long-time scale simulations. We found that both methods are sensitive to the choice of CV used in mapping the pathway: With common CV choices, based on the RMSD of the entire protein, both methods produced pathways that differed significantly from the unbiased pathway, even when the simulation length of the SMD runs approached that of the unbiased pathway. With an RMSD selection that focused on the region of the protein known to be important for the conformational change (the $\alpha$C helix and $\beta$9 sheet), however, both SMD and the string method captured key features of the path of the unbiased transition, and they did so in significantly less simulation time than was needed in unbiased MD.

## METHODS

**Simulation Setup.** We performed all-atom MD simulations of EGFR kinase on Anton,[28] a special-purpose machine, following a protocol established previously.[24] In that protocol, protein, water, and ions were represented explicitly, using the CHARMM22* force field[29] and the TIP3P water model.[30,31] The cubic system was approximately 80 Å on each side and was solvated with water containing 150 mM NaCl. Simulations were performed in the constant temperature (NVT) ensemble with $T$ = 310 K, unless otherwise noted. In addition to the two transitions of 23 and 12 $\mu$s in length reported in ref 24, we ran nine simulations from the equilibrated active structure with different, random initial velocities, and we observed another successful transition of 15 $\mu$s from the catalytically active to the inactive conformation. Two more transitions (one of 15 $\mu$s and the other of 25 $\mu$s) were also observed, in which EGFR kinase successfully transitioned out of the locally disordered state by way of an extended intermediate conformation but which did not undergo the final conformational change of the $\beta$9 strand into the two-turn helix. The remaining six simulations, of 20.7 $\mu$s, 20.2 $\mu$s, 30 $\mu$s, 10 $\mu$s, 10 $\mu$s, and 10 $\mu$s in length, remained trapped in metastable states in which some or all of the hydrogen bonds between the $\beta$9 and $\beta$6 sheets persisted.

**Steered Molecular Dynamics (SMD).** SMD runs were started from the equilibrated active structure and pulled to the equilibrated inactive structure. These simulations employed a time-dependent harmonic biasing potential, $U(t)$:

$$U(t) = \frac{1}{2}k(\theta[x(t)] - \theta_0(t))^2$$

where $x$ is the configurational state of the system at time $t$, $k$ is a force constant, $\theta$ is the best fit RMSD of protein C$\alpha$ atoms between a simulation conformation and the equilibrated inactive crystal structure, and $\theta_0$ is a target RMSD value that decreases linearly over simulation time from 5.5 to 0 Å. We used two different C$\alpha$ atom selections in the SMD runs: one encompassing the whole protein except for several residues near the termini (residues 706–979) and one focused on the $\alpha$C helix and the $\beta$9 sheet (residues 756–769 and 857–863). SMD simulations ranged from 10 ns to 10 $\mu$s. Several SMD simulations were also run at constant temperature and pressure (NPT) with $T$ = 310 K and $P$ = 1 bar. No substantive differences were observed between NVT and NPT simulations. Further parameters and trajectory details are provided in Supporting Information Table S1.

**On-the-Fly String Method.** In the on-the-fly string method, an initial conformational change pathway is represented as a discrete set of conformational replicas, $x(s,t)$, harmonically tethered by way of a set of CVs to a "string," or curve, of images in CV space, $z(s,t)$. Here, $s$ is a pathway index that ranges from 0 to 1, and $t$ is time. This path is refined according to the following equations of motion (cf., section 4 addendum in ref 21):

$$m_i\ddot{x}_i(s,t) = f_i[x(s,t)] - \kappa\sum_{\beta=1}^{n}(\theta_\beta[x(s,t)] - z_\beta(s,t)) \times \frac{\partial\theta_\beta[x(s,t)]}{\partial x_i} + \text{thermostat terms}$$

$$\gamma\dot{z}_a(s,t) = \sum_{\beta=1}^{n}(\tilde{M}_{a\beta}[x(s,t)]\kappa(\theta_\beta[x(s,t)] - z_\beta(s,t)) + \lambda(s,t)z'_a(s,t)$$

where $m$ is the mass of atom $i$, $f$ is the force on each atom due to the underlying biomolecular force field, $\kappa$ is a force constant that couples the images and the replicas, $\alpha$ and $\beta$ are indices over CVs and range from 1 to $n$, which is the total number of CVs, $\gamma$ is a friction coefficient, $\tilde{M}$ is a metric tensor, and $\lambda z'$ is a reparameterization force that keeps all images equidistant from each other in CV space. In practice, the reparameterization force is imposed by periodically interpolating a line through the $z$ values and repositioning them equidistantly along that interpolation. In addition, we do not explicitly calculate the $\tilde{M}$ tensor, but we rather assume that the tensor is the identity matrix at all times. This may subtly affect the outcome of the refinement, but we do not expect it to alter how different CVs affect the accuracy of the refinement. Specifically, the equation above converges to the path defined as points where $(M\nabla F - k_BT \text{ div } M)^\perp = 0$, where $\nabla F$ is the mean force in CV space along the path and $M$ is the Boltzmann average of $\tilde{M}$ with $\theta(x)$ constrained to $z$. With $\tilde{M}$ set to be the identity matrix, the above equation finds paths in which $(\nabla F)^\perp = 0$, which differs from the true mean free energy path defined as $(M\nabla F)^\perp = 0$. Additional details can be found in ref 22.

String method calculations were initiated from 13 frames extracted from 10 ns SMD runs, with $\kappa$ = 200 kcal mol$^{-1}$ Å$^{-2}$, for 15 ns per image, and with a reparameterization interval of 0.1 ps. The friction coefficient, $\gamma$, was 20 ps kcal mol$^{-1}$ Å$^{-1}$ for all images except the first and the last images, which were kept fixed with an infinitely large $\gamma$ (i.e., the initial $z$ values were kept constant and never updated). Two such calculations were done, each using two CVs: the C$\alpha$ RMSDs to the active (PDB ID 2ITP)[32] and Src-like inactive (PDB ID 2GS7)[33] crystal structures. One calculation used the whole-protein C$\alpha$ RMSD (residues 706−979), and the other calculation used the C$\alpha$ RMSD over the $\alpha$C helix and $\beta$9 sheet (residues 756−769 and 857−863). Both calculations converged within 5−10 ns (Supporting Information Figures S4A and C).

**RMSF calculation.** The RMSF curves of the activation loop (Figures 2B and 2D) were calculated by aligning trajectory frames to the C$\alpha$ atoms of residues 703−861 and 896−979 in the inactive crystal structure. Each curve was calculated over three independent simulations during the extended intermediate transition. This transition region was defined as being a trajectory segment beginning when a hydrogen bond between the $\beta$9 and $\beta$6 sheets (defined by the carbonyl oxygen of residue 832 and the backbone amide hydrogen of residue 862) surpassed 10 Å, and ending when the RMSD to the inactive crystal structure, as defined in Figure 1, fell below 2 Å.

## ■ ASSOCIATED CONTENT

### Supporting Information

Table of SMD simulations and five figures showing the results of additional unbiased MD transitions (Figure S1), additional SMD trajectory replicates (Figures S2 and S3) and the convergence, RMSF, and evolution of the string refinement calculations (Figures S4 and S5). This material is available free of charge via the Internet at http://pubs.acs.org/.

## ■ AUTHOR INFORMATION

### Corresponding Authors

*Email: Albert.Pan@DEShawResearch.com. Phone: (212) 403-8664. Fax: (646) 873-2664.
*Email: David.Shaw@DEShawResearch.com. Phone: (212) 478-0260. Fax: (212) 845-1286.

### Notes

The authors declare the following competing financial interest(s): This study was conducted and funded internally by D. E. Shaw Research, of which D.E.S. is the sole beneficial owner and Chief Scientist, and with which all authors are affiliated.

## ■ ACKNOWLEDGMENTS

The authors thank Michael Eastwood for helpful discussions and a critical reading of the manuscript, Kenneth Mackenzie for helpful discussions, and Berkman Frank for editorial assistance.

## ■ REFERENCES

(1) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles. *J. Chem. Phys.* **1992**, *96* (3), 1776−1783.
(2) Marinari, E.; Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhys. Lett.* **1992**, *19* (6), 451.
(3) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1), 141−151.
(4) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (20), 12562−12566.
(5) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919−11929.
(6) Maragliano, L.; Vanden-Eijnden, E. A Temperature Accelerated Method for Sampling Free Energy and Determining Reaction Pathways in Rare Events Simulations. *Chem. Phys. Lett.* **2006**, *426*, 168−175.
(7) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu. Rev. Biophys.* **2011**, *40*, 41−62.
(8) Wu, X.; Damjanovic, A.; Brooks, B. R. Efficient and Unbiased Sampling of Biomolecular Systems in the Canonical Ensemble: A Review of Self-Guided Langevin Dynamics. *Adv. Chem. Phys.* **2012**, *150*, 255−326.
(9) Elber, R.; Karplus, M. A Method for Determining Reaction Paths in Large Molecules: Applications to Myoglobin. *Chem. Phys. Lett.* **1987**, *139* (5), 375−380.
(10) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105* (5), 1902−1921.
(11) E, W.; Ren, W.; Vanden-Eijnden, E. String Method for the Study of Rare Events. *Phys. Rev. B* **2002**, *66* (5), 052301.
(12) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291−318.
(13) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in Free Energy Space. *J. Chem. Phys.* **2007**, *126* (5), 054103.
(14) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin−Biotin Rupture Force. *Science* **1996**, *271* (5251), 997−999.
(15) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular Dynamics Study of Unbinding of the Avidin−Biotin Complex. *Biophys. J.* **1997**, *72* (4), 1568−1581.
(16) Perilla, J. R.; Beckstein, O.; Denning, E. J.; Woolf, T. B. Computing Ensembles of Transitions from Stable States: Dynamic Importance Sampling. *J. Comput. Chem.* **2010**, *32* (2), 196−209.
(17) Paci, E.; Vendruscolo, M.; Dobson, C. M.; Karplus, M. Determination of a Transition State at Atomic Resolution from Protein Engeineering Data. *J. Mol. Biol.* **2002**, *324* (1), 151−163.
(18) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J., Ciccotti, G., Coker, D. F., Eds.; World Scientific Publishing Company: Singapore, 1998; pp 385−404.
(19) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. Efficient and Verified Simulation of a Path Ensemble for Conformational Change in a United-Residue Model of Calmodulin. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (46), 18043−18048.
(20) Huang, H.; Ozkirimli, E.; Post, C. B. A Comparison of Three Perturbation Molecular Dynamics Methods for Modeling Conformational Transitions. *J. Chem. Theory Comput.* **2009**, *5* (5), 1301−1314.
(21) Maragliano, L.; Vanden-Eijnden, E. On-the-Fly String Method for Minimum Free Energy Paths Calculations. *Chem. Phys. Lett.* **2007**, *446*, 182−190.
(22) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.
(23) Ovchinnikov, V.; Karplus, M.; Vanden-Eijnden, E. Free Energy of Conformational Transition Paths in Biomolecules: The String Method and its Application to Myosin VI. *J. Chem. Phys.* **2011**, *134* (8), 085103.
(24) Shan, Y.; Arkhipov, A.; Kim, E. T.; Pan, A. C.; Shaw, D. E. Transitions to Catalytically Inactive Conformations in EGFR Kinase. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (18), 7270−7275.
(25) Huse, M.; Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **2002**, *109* (3), 275−282.

(26) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. On the Transition Coordinate for Protein Folding. *J. Chem. Phys.* **1998**, *108* (1), 334−350.
(27) Miller, T. F., III; Vanden-Eijnden, E.; Chandler, D. Solvent Coarse-Graining and the String Method Applied to the Hydrophobic Collapse of a Hydrated Chain. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (37), 14559−14564.
(28) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y.; Towles, B. Millisecond-Scale Molecular Dynamics Simulations on Anton. *Proceedings of the Conference on High Performance Computing, Networking, Storage and Analysis (SC09)* ACM: New York, 2009.
(29) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100* (9), L47−L49.
(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(31) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, S.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586−3616.
(32) Yun, C. H.; Boggon, T. J.; Li, Y.; Woo, M. S.; Greulich, H.; Meyerson, M.; Eck, M. J. Structures of Lung Cancer-Derived EGFR Mutants and Inhibitor Complexes: Mechanism of Activation and Insights into Differential Inhibitor Sensitivity. *Cancer Cell* **2007**, *11* (3), 217−227.
(33) Zhang, X.; Gureasko, J.; Shen, K.; Cole, P. A.; Kuriyan, J. An Allosteric Mechanism for Activation of the Kinase Domain of Epidermal Growth Factor Receptor. *Cell* **2006**, *125* (6), 1137−1149.