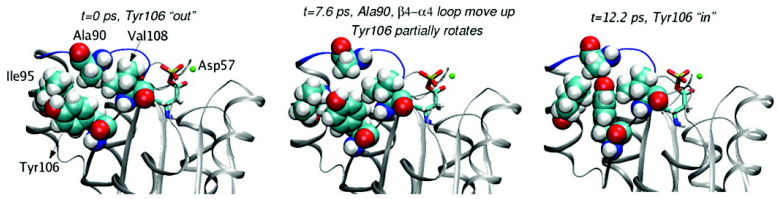
Article

Activation Mechanism of a Signaling Protein at Atomic Resolution from Advanced Computations

Liang Ma, and Qiang Cui

J. Am. Chem. Soc., **2007**, 129 (33), 10261-10268• DOI: 10.1021/ja073059f • Publication Date (Web): 26 July 2007



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 2 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 07/26/2007

Activation Mechanism of a Signaling Protein at Atomic Resolution from Advanced Computations

Liang Ma[†] and Qiang Cui*[,†,‡]

Contribution from the Graduate program in Biophysics and Department of Chemistry and Theoretical Chemical Institute, University of Wisconsin−Madison, 1101 University Avenue, Madison, Wisconsin 53706

Received May 1, 2007; E-mail: cui@chem.wisc.edu

Abstract: Advanced computational techniques including transition path sampling and free energy calculations are combined synergistically to reveal the activation mechanism at unprecedented resolution for a small signaling protein, chemotaxis protein Y. In the conventional "Y−T coupling" model for response regulators, phosphorylation induces the displacement of the conserved Thr87 residue through hydrogen-bond formation, which in turn makes it sterically possible for Tyr106 to isomerize from a solvent exposed configuration to a buried rotameric state. More than 160 *unbiased* activation trajectories show, however, that the rotation of Tyr106 does not rely on the displacement of Thr87 *per se*. Free energy calculations reveal that the Tyr106 rotation is a low-barrier process in the absence of the Thr87-phosphate hydrogen bond, although the rotation is stabilized by the formation of this interaction. The simulations also find that structural change in the $\beta4-\alpha4$ loop does not gate the Tyr106 rotation as suggested previously; rather, the rotation of Tyr106 stabilizes the activated configuration of this loop. The computational strategy used and mechanistic insights obtained have an impact on the study of signaling proteins and allosteric systems in general.

1. Introduction

Signaling proteins are often activated to perform their biological function through a localized event such as phosphorylation and ligand (ion) binding.[1] Understanding how such local modifications lead to striking transitions in the structure and therefore activity of signaling proteins is evidently of great value from both fundamental and biomedical perspectives. Due to the intrinsic dynamical nature of the process, a microscopic description of signaling protein activation with structural and energetic details has been difficult to obtain.[2−4] For example, recent NMR studies[5] of small signaling proteins in two-component systems suggested that the structural motifs to be activated have a small but non-negligible population in the active conformation prior to phosphorylation; the role of phosphorylation is to shift this population to become the dominant one rather than inducing new conformations. Such a "population shift" framework,[2] which has features of the Monod−Wyman−Changeux (MWC) model[6] for allostery,[6−9] emphasizes the dynamical nature of signaling proteins (and allosteric systems in general) and provides a rather different picture from the "push and pull" type of description as characterized by the stereochemical model for hemoglobin.[10] As discussed in detail in our recent work,[11] however, what is missing in the classical MWC,[6] KNF[7] models and the more recent "population shift" rendition is the actual *mechanism* for propagating the effect of ligand binding or phosphorylation over a long distance in structural and energetic terms. In the context of the "population shift" description of signaling protein activation, for example, it often remains unclear *how* phosphorylation shifts the population of the distant response site. For such a purpose, atomistic simulations[12] are needed to complement experimental studies. Indeed, as pointed out by Kern and Zuiderweg,[2] "*Regretfully, experimental evidence showing the intermediates through which an allosteric protein travels has not been forthcoming to our knowledge···longer-term unrestrained MD simulations will, in the near future, become feasible for the modeling of the pathways.*" As they commented, due to the long time scale ($>\mu$s−ms) of activation processes, most previous simulation studies had to use artificially accelerated molecular simulations[11,13,14] and very few studies[15,16] characterized the energetics of the processes involved.

[†] Graduate program in Biophysics.
[‡] Department of Chemistry and Theoretical Chemical Institute.

(1) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular biology of the cell*; Garland Publishing, Inc.: New York, 1994.
(2) Kern, D.; Zuiderweg, E. R. P. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748.
(3) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins* **2004**, *57*, 433.
(4) Swain, J. F.; Gierasch, L. M. *Curr. Opin. Struct. Biol.* **2006**, *16*, 102.
(5) Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D. *Science* **2001**, *291*, 2429.
(6) Monod, J.; Wyman, J.; Changenux, J.-P. *J. Mol. Biol.* **1965**, *12*, 88.
(7) Koshland, J. D. E.; Nemethy, G.; Filmer, D. *Biochemistry* **1966**, *1*, 365.
(8) Eaton, W. A.; Henry, E. R.; Hofrichter, J.; Mozzarelli, A. *Nat. Struct. Biol.* **1999**, *6*, 351.
(9) Szabo, A.; Karplus, M. *J. Mol. Biol.* **1972**, *72*, 163.
(10) Perutz, M. F. *Nature* **1970**, *228*, 726.
(11) Formaneck, M. S.; Ma, L.; Cui, Q. *Proteins* **2006**, *63*, 846.
(12) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679.
(13) Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11905.
(14) Ma, J.; Sigler, P. B.; Xu, Z.; Karplus, M. *J. Mol. Biol.* **2000**, *302*, 303.
(15) Banavali, N. K.; Roux, B. *Structure* **2005**, *13*, 1715.
(16) Radhakrishnan, R.; Schlick, T.; *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5970.

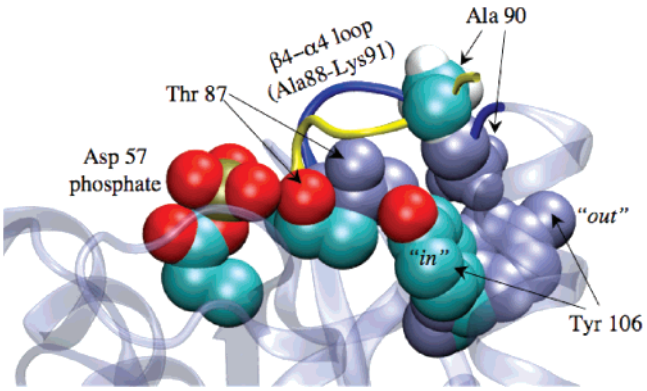


Figure 1. Comparison of the inactive[21] and active[26] structures of CheY. Overlay of key residues between the phosphorylation (Asp 57) and response sites (Tyr 106). Residues in the active structure are colored according to atom types, while those in the inactive structure are colored as ice-blue. The inactive and active configurations of the $\beta4-\alpha4$ loop are colored as dark-blue and yellow, respectively. The overall structure of the inactive state is shown in the transparent form as background. The steric interaction between Ala 90 and Tyr 106 in the inactive structure is clearly visible. All structural figures are made using VMD.[51]

To illustrate that modern computational techniques can indeed start to help reveal unbiased activation pathways of signaling proteins and provide new mechanistic insights, we study a small but representative signaling protein, the *E. coli* chemotaxis Y protein (CheY). CheY is a 129 residue prototypical response regulator in two-component signal transduction systems.[17] It is activated through phosphorylation, and the most important conformational change in CheY upon activation is the rotation of the Tyr106 side chain from a solvent exposed orientation to a fully buried state (Figure 1). Tyr106 is separated from the phosphorylation site of Asp57 by nearly 10 Å, and between them is Thr87. Since Thr87 is highly conserved among response regulators in two-component systems, the conventional description for CheY activation is the "Y−T coupling" model:[18] phosphorylation of Asp57 displaces Thr87 due to a hydrogen-bonding interaction, which in turn allows the rotation of Tyr106. Since partial activity has been observed for the wild type CheY[19] and the T87A mutant[20] in the absence and presence of phosphorylation, respectively, the "Y−T coupling" model has been questioned. In particular, since the $\beta4-\alpha4$ loop (Ala 88 to Lys 91) also undergoes a major displacement upon activation, it has been speculated[11,21] that this loop in fact gates the rotation of Tyr106 and the role of phosphorylation and Thr87 is to select a specific loop configuration, which is reminiscent of the "population shift" model. Due to the presence of a significant energy barrier for Tyr106 rotation, artificial bias had to be used in the simulations[11] and therefore the results were instructive but inconclusive.

Specifically, we use transition path sampling (TPS)[22−24] to harvest more than 160 *unbiased* activation trajectories to reveal the activation mechanism of CheY. Although TPS has been successfully applied to obtain reactive trajectories between stable basins for many systems of varying degrees of complexity,[16,22−24] it is difficult to reach conclusive mechanistic statements for biomolecular systems based on hundreds of reactive trajectories. Accordingly, we supplement the TPS analysis with free energy (potential of mean force, PMF) simulations for several dimensions identified as important by the TPS results. Combined together, these results help clarify the activation mechanism of CheY as a prototypical signaling protein with unprecedented resolution.

2. Results

2.1. Transition Path Sampling. The rotation of the Tyr106 side chain is the most significant conformational change upon activation and regarded to correlate with CheY activity.[25] In TPS, therefore, the side chain dihedral angle, χ, of Tyr106 is selected as the order parameter that characterizes the two basins associated with the activation process. The length of the transition trajectories is set to ~25 ps, which is determined as sufficiently long based on the analysis of τ_{mol} (see Supporting Information). As repeatedly emphasized in the literature,[22−24] the time scale of these transition trajectories should not be confused with the time span of the entire activation process, which includes the waiting time for the system to accumulate enough energy in the relevant degrees of freedom to overcome rate-limiting barriers (e.g., of Tyr106 rotation in CheY). The quality of path sampling is examined by calculating the autocorrelation function of χ at a particular time slice in the transition regions of the reactive trajectories in the path space. The results show that trajectories become uncorrelated every ~20th trajectory (see Supporting Information), which suggests that the sampling of 163 trajectories is statistically meaningful for the current purpose.

The two relevant basins are defined by the order parameter χ between ~30°−90° and ~150°−210°, respectively. In all activation trajectories, the Tyr106 rotation is observed and the rotameric state does not change (up to ~1 ns) once rotation has completed (data not included), which indicates that χ is indeed an appropriate order parameter. To illustrate the specific sequence of events associated with Tyr106 rotation, four representative configurations along a randomly chosen transition path are shown in Figure 2. At the beginning of the path (Figure 2a), CheY is in the inactive state with Tyr106 being solvent exposed and Ala 90 in the $\beta4-\alpha4$ loop pointing down to block the rotation of Tyr106. After ~8 ps (Figure 2b,c), Tyr106 partially rotates in, and correspondingly, Ala 90 and the $\beta4-\alpha4$ loop move upward to allow the rotation to proceed; note, however, that the $\beta4-\alpha4$ loop remains essentially in the inactive configuration. Finally by ~12 ps, Tyr106 completely rotates in and becomes fully buried in the cavity under the $\beta4-\alpha4$ loop (Figure 2d).

To collectively analyze the behavior of the system during the transition, we project snapshots from all harvested transition paths to the 2-D plane of the χ angle and several variables noted in our previous simulation study.[11] These variables include the position of Ala 90 (Figure 3a), which may sterically gate the rotation of Tyr106, two variables that characterize the config-

(17) Stock, A. M.; Robinson, V. L.; Goudreau, P. N. *Annu. Rev. Biochem.* **2000**, *69*, 183.
(18) Cho, H. S.; Lee, S. Y.; Yan, D. L.; Pan, X. Y.; Parkinson, J. S.; Kustu, S.; Wemmer, D. E.; Pelton, J. G. *J. Mol. Biol.* **2000**, *297*, 543.
(19) Barak, R.; Eisenbach, M. *Biochemistry* **1992**, *31*, 1821.
(20) Appleby, J. L.; Bourret, R. B. *J. Bacteriol.* **1998**, *180*, 3563.
(21) Simonovic, M.; Volz, K. *J. Biol. Chem.* **2001**, *276*, 28637.
(22) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. *Adv. Chem. Phys*. **2002**, *123*, 1.
(23) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.
(24) Hagan, M. F.; Dinner, A. R.; Chandler, D.; Chakraborty, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100*, 13922.
(25) Zhu, X. Y.; Rebello, J.; Matsumura, P.; Volz, K. *J. Biol. Chem.* **1997**, *272*, 5000.

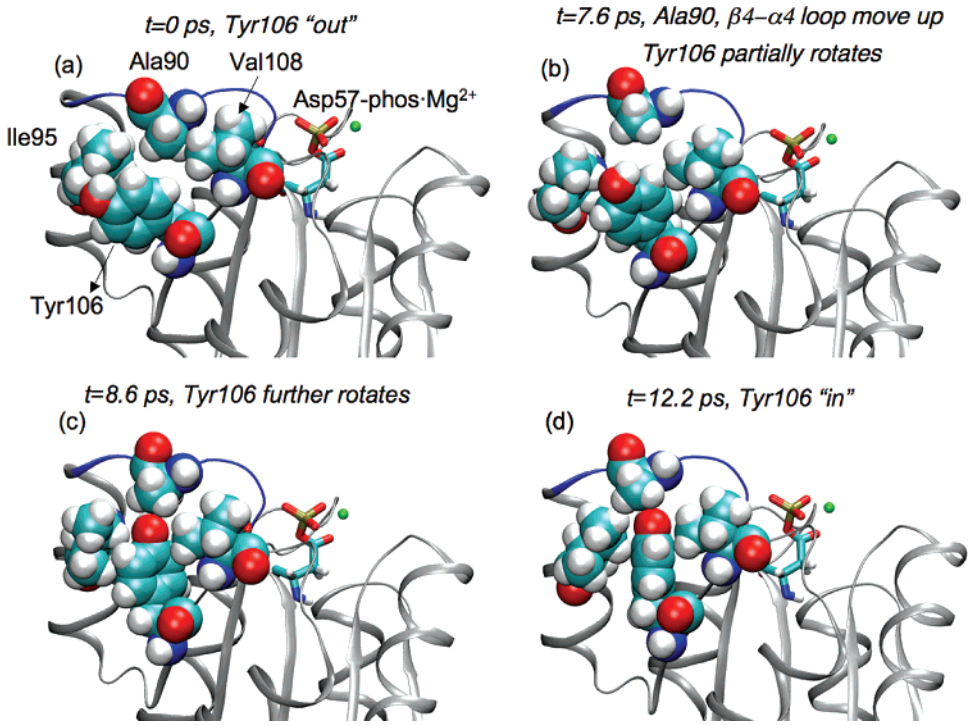


Figure 2. Four configurations along an exemplary activation trajectory. Several important residues including Tyr 106, Ala 90, Ile 95, and Val 108 are shown in the van der Waals scheme; the phosphorylated Asp 57 is shown in the licorice form; the $\beta4-\alpha4$ loop is shown as the blue ribbon. The rest of the protein is shown in cartoon, and only regions close to Tyr 106 are shown.

uration of the $\beta4-\alpha4$ loop (Figure 3b,c) and the width of the "doorway" toward the buried site under the $\beta4-\alpha4$ loop (Figure 3d), which is formed by Ile 95 and Val 108. All these variables show bimodal distribution along the χ angle, which again supports the use of χ as the order parameter. As shown in Figure 3a, Ala 90 systematically deviates further from the inactive position as the Tyr106 rotates toward the buried site, which is easily explained by the observation that the inactive position of Ala 90 sterically hinders the Tyr106 rotation. Systematic variations in the $\beta4-\alpha4$ loop configuration are also transparent. As shown in Figure 3b, the $\beta4-\alpha4$ loop moves toward the active configuration as Tyr106 rotates in, as reflected by the decreasing ΔRMSD of backbone atoms in the $\beta4-\alpha4$ loop with respect to χ; as defined in Materials and Methods, a smaller ΔRMSD indicates that the $\beta4-\alpha4$ loop is closer to the active configuration. Similarly, in Figure 3c, the pseudo-dihedral angle formed by the Cα atoms of the four residues in the $\beta4-\alpha4$ loop increases slightly (i.e., closer to the active configuration) as the Tyr106 rotates in. Regarding the doorway formed by Ile 95 and Val 108 as identified in the previous biased MD simulations,[11] it is interesting that the width of the "door" in fact tends to decrease as Tyr106 rotates in (Figure 3d), such that the van der Waals interactions between these residues and Tyr106 increases (Figure 3e). Therefore, unlike the proposal based on the biased MD simulations,[11] the function of these "doorway" residues is to stabilize the buried conformation of Tyr106 rather than gating its rotation; this is because the average width of the "doorway" is 7.5 Å, which does not limit the rotation of Tyr106. Finally, as shown in Figure 3f, the hydrogen bond between Asp57-phosphate and Thr87 is unformed in all transition paths with the shortest oxygen−oxygen distance reaching about 4 Å. Therefore, it is clear that Tyr106 rotation does not rely on the formation of this hydrogen bond.

Although the TPS results are instructive, it is difficult to reach a firmly conclusive statement about the mechanism of CheY activation because the number of harvested transition trajectories is still small. Indeed, even with a minimal model for CheY activation, at least three processes are implicated: rotation of the Tyr106 side chain, transition of the $\beta4-\alpha4$ loop, and hydrogen-bond formation between Asp57-phosphate and Thr87, which in principle can proceed in 3!=6 different sequences. Although TPS demonstrates that the side chain rotation of Tyr106 *may* occur independent of the Thr87-phosphate hydrogen-bond formation, the results do not guarantee that this is the only or even the dominant pathway. Alternative transition trajectories with very different orders of events may be separated by high barriers and thus have not been sampled. Therefore, it is useful to supplement the TPS results with free energy simulations, which we turn to next.

2.2. Free Energy Simulations. Variables associated with the key structural transitions during CheY activation are chosen as the reaction coordinates in the free energy simulations: ΔRMSD of all non-hydrogen atoms of the $\beta4-\alpha4$ loop, the side chain dihedral angle of Tyr106, χ, and the hydrogen-bond distance D_{hb} between the hydroxyl oxygen of Thr87 and the phosphate oxygen of Asp57. Umbrella sampling calculations are carried out on several 2-D projections composed of two of the three reaction coordinates (Figure 4). All of the simulations are performed with Asp57 phosphorylated (see Materials and Methods).

Figure 4a shows the 2D free energy surface for ΔRMSD and D_{hb} with the Tyr106 side chain "out" (solvent exposed). There are two local minima: basin I_O^U with ΔRMSD > 0 (inactive) and D_{hb} > 5.5 Å whereas basin I_O^F with ΔRMSD > 0 and D_{hb} < 3.0 Å. In other words, with Tyr106 "out", the $\beta4-\alpha4$ loop prefers to remain inactive, and the Thr87-phosphate hydrogen

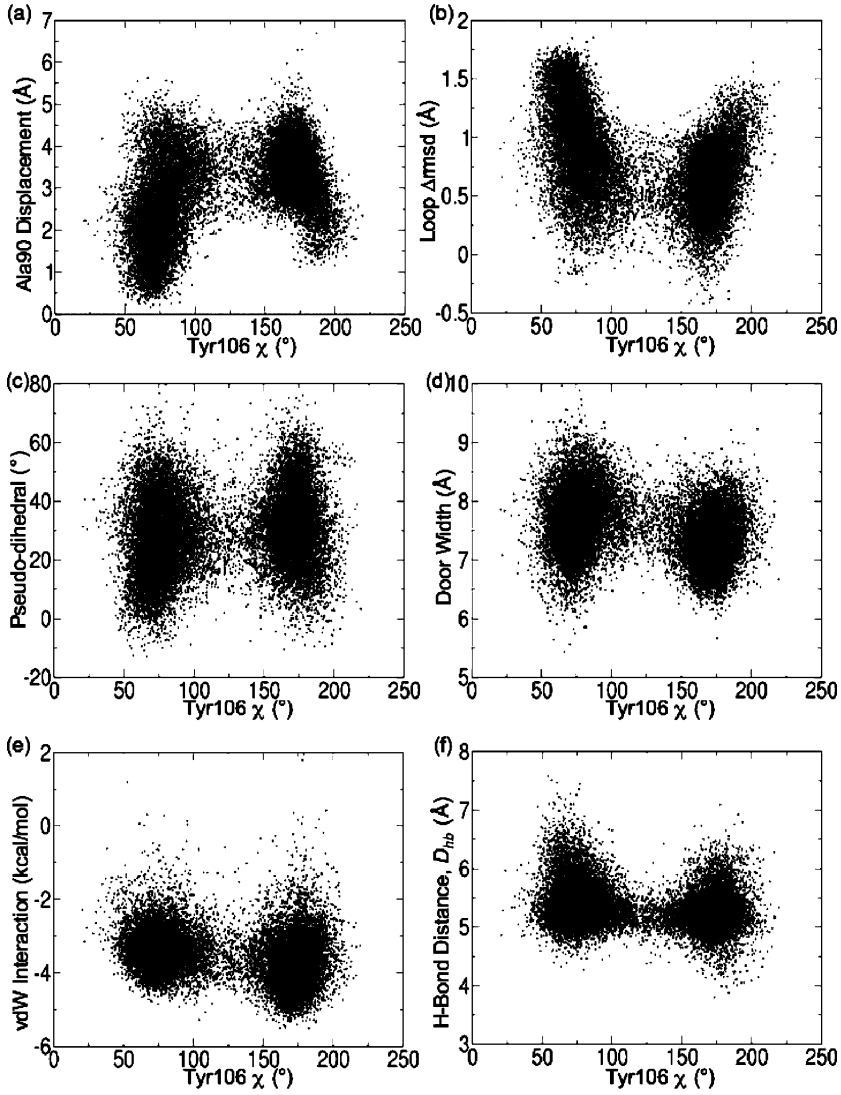


Figure 3. Projection of all 163 transition paths onto the two-dimensional plane of the Tyr 106 χ angle and several variables of interest. (a) The displacement of the Ala 90 side chain from the inactivated position; (b) ΔRMSD of the $\beta4-\alpha4$ loop relative to the inactive and active configurations; (c) the pseudo-dihedral angle of the $\beta4-\alpha4$ loop; (d) the width of the "doorway" formed by Ile 95 and Val 108; (e) van der Waals interaction between the "doorway" residues and Tyr 106; (f) the shortest O−O distance between Thr 87 and the phosphate.

bond is locally stable although formation of this hydrogen bond is thermodynamically unfavorable (∼5.3 kcal/mol) with a significant barrier of ∼7.0 kcal/mol.

Figure 4b shows the 2D free energy surface for ΔRMSD and D_{hb} with the Tyr106 sidechain "in" (buried under the $\beta4-\alpha4$ loop). There are also two local minima: basin I_I^U with ΔRMSD > 0 and D_{hb} > 4.0 Å and basin I_I^F with a broad distribution around ΔRMSD ≈ 0 and D_{hb} < 3.0 Å. Interestingly, with Tyr106 buried, the Thr87-phosphate hydrogen-bond formation becomes favorable by ∼2.0 kcal/mol with a much reduced barrier of ∼2.3 kcal/mol, which clearly demonstrates that Tyr106 rotation and the Thr87-phosphate hydrogen-bond formation are energetically coupled. Moreover, once this hydrogen bond is formed, the active configurations of the $\beta4-\alpha4$ loop become much more accessible; i.e., the energy landscape along ΔRMSD becomes much flatter with the hydrogen bond formed.

Although the comparison of Figure 4a,b clearly illustrates the energetic coupling between Tyr106 rotation and Thr87-phosphate hydrogen-bond formation, and the TPS results suggest that Tyr106 rotation does not have to rely on the displacement of Thr87, whether Tyr106 rotation is intrinsically a low-barrier process remains to be clarified. This is analyzed by the 2D free energy surface for ΔRMSD and Tyr106-χ with the Thr87-phosphate hydrogen bond unformed (D_{hb} > 4.0 Å). Figure 4c clearly shows two local minima: basin I_I^U with ΔRMSD > 0 and χ > 180° (Tyr106 "in") whereas basin I_O^U with ΔRMSD > 0 and χ < 90° (Tyr106 "out"). Consistent with the TPS results, even without major displacement of Thr87, Tyr106 isomerization has only a modest barrier of ∼4.8 kcal/mol although it is energetically unfavorable by ∼2.1 kcal/mol.

Finally, to explicitly define the relationship between Tyr106 rotation and Thr87-phosphate hydrogen-bond formation, the 2D free energy is calculated for χ and D_{hb}, with the $\beta4-\alpha4$ loop in the inactive configuration (because the complete loop transition can be shown to occur last; see below). As shown in Figure 4d, there are four basins that correspond to the four expected metastable states: I_O^U, I_O^F, I_I^U, I_I^F; the relative stabilities of the four basins are not exactly identical to those sampled in the three 2D-PMFs discussed above because the overall structure of the protein has subtle differences (see Figure 4 caption). Nevertheless, the thermodynamic coupling between Tyr106

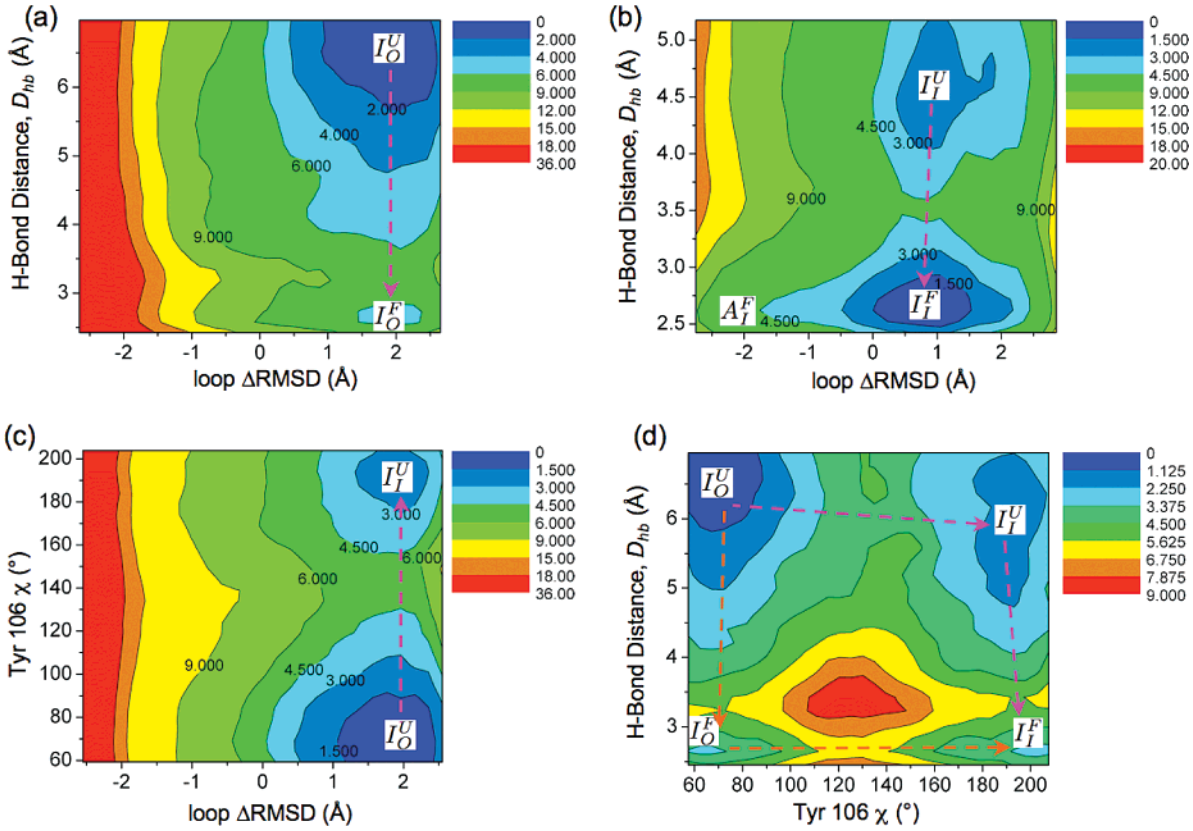


Figure 4. Two-dimensional potentials of mean force (in kcal/mol) for CheY for: (a) Δrmsd of the $\beta4-\alpha4$ loop and the Thr87-phosphate hydrogen-bond distance, D_{hb}, with the Tyr106 side chain solvent exposed; (b) ΔRMSD and D_{hb} with the Tyr106 side chain buried; (c) ΔRMSD and the Tyr106 side chain dihedral angle χ with the Thr87-phosphate hydrogen bond unformed; (d) χ and D_{hb}. The arrows indicate possible pathways of activation. For each plot, the lowest-energy basin is set to be the reference (0.0) of energy. The basins are labeled with notations such as I_O^U, which indicates inactive (I) $\beta4-\alpha4$ loop, unformed (U) Thr87-phosphate hydrogen bond, and out configuration (O) of the Tyr106 side chain. The PMFs in (b) are calculated using the 1F4V[26] X-ray structure while the other three are calculated with the 1JBE[21] X-ray structure.

rotation and Thr87-phosphate hydrogen-bond formation is qualitatively captured; e.g., the Tyr106 rotation is energetically unfavorable without the Thr87-phosphate hydrogen bond and becomes nearly thermoneutral with the hydrogen-bond formed. More importantly, the two extreme pathways (Tyr106 rotation occurs first vs Thr87-phosphate hydrogen bond forms first) are separated by a high barrier, which argues against a concerted mechanism, and both of them have a barrier lower than 7 kcal/mol.

The free energy results can be pieced together to explain the activation mechanism of CheY. As shown in Figure 5, the activation process can be described in terms of movement on a cube with three dimensions corresponding to the three reaction coordinates. The starting point is the inactive state I_O^U with the $\beta4-\alpha4$ loop inactive (I), Tyr106 "out" (O), and the Thr87-phosphate hydrogen bond unformed (U). The expected ending point, labeled as A_I^F with the current notation, is the active state with all key structural motifs adopting the active configuration. According to Figure 4a,b, the activation of the $\beta4-\alpha4$ loop is energetically very unfavorable (>9.0 kcal/mol uphill) and therefore has to occur after Tyr106 rotation and formation of the Thr87-phospate hydrogen bond. The latter two events are thermodynamically coupled (Figure 4a, b, and d), but kinetically, either can proceed first due to the similar barriers (4.8 vs 7.0 kcal/mol).

In the current simulations, even following the Tyr106 rotation and Thr87-phosphate hydrogen-bond formation, a largely inac-

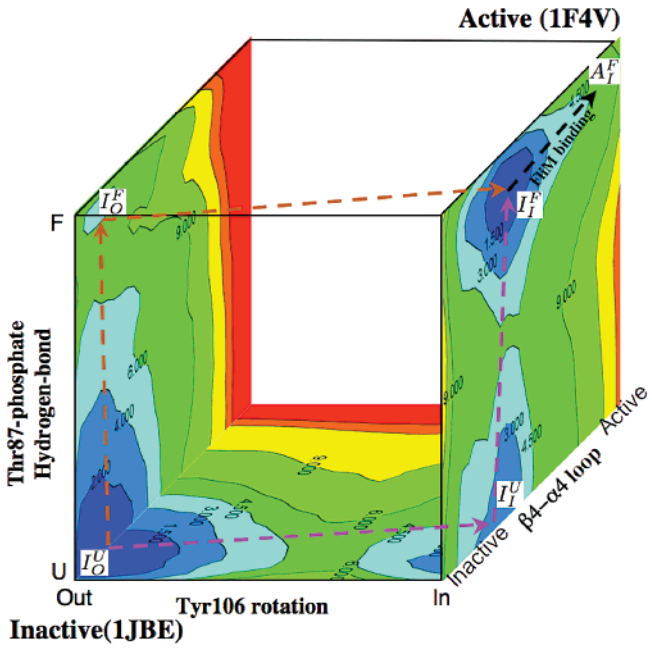


Figure 5. A three-dimensional scheme that illustrates the energetics and possible pathways for CheY activation. The three PMFs are based on Figure 4, with the energy reference of each plot adjusted such that the energetics in different PMFs are consistent; the PMF in Figure 4d is not shown for clarity. The expected fully active state, A_I^F, is not a local free-energy minima in the simulations, presumably due to the absence of the FliM peptide in the model.

tive configuration (ΔRMSD $\sim$0) of the $\beta4-\alpha4$ loop is still preferred energetically over the fully active configuration. In other words, the most active-like structure identified in the simulations (I_I^F in Figure 5) differs from the crystal structure of the fully active CheY(1F4V[26]) regarding the configuration of the $\beta4-\alpha4$ loop. However, we note that the crystal structure of the active CheY is bound to the FliM peptide; it is possible that the fully active $\beta4-\alpha4$ loop is better stabilized by the binding of the FliM peptide, which is not included in the simulation model. Nevertheless, the calculated free energy difference between states I_O^U and I_I^F is ~ -0.1 kcal/mol, which is very reasonable considering the absence of the FliM peptide in the simulations.

3. Discussion

In this study, TPS and free energy simulations are combined synergistically to probe the activation mechanism of a prototypical signaling protein at the atomic resolution, which has been difficult to achieve with experiments.[2] By themselves, a limited number of unbiased transition trajectories are not sufficient for reaching conclusive statements regarding the activation mechanism. However, even a modest number (hundreds) of unbiased activation trajectories are valuable for revealing genuinely relevant degrees of freedom during the transition and for generating configurations in the barrier region. This information can be effectively incorporated into the subsequent free energy simulations, which explicitly address the energetic coupling between different local structural changes and the likely sequences of events.

3.1. "Y−T Coupling" vs "Population Shift". As stated in the Introduction, there are two competing models for the activation of CheY and other response regulators in two-component systems upon phosphorylation: (i) the traditional "Y−T coupling" model, which assumes that the displacement of Thr87 due to hydrogen-bond formation with the phosphoryl group allows the Tyr106 rotation; (ii) the "population shift" model, which argues that the Tyr106 rotation may occur prior to the Thr87 displacement (or even phosphorylation) and the role of phosphorylation (and hydrogen-bonding with Thr87) is to shift the population of Tyr106 sidechain rotamers toward the active (buried) state. Our TPS results clearly show that Tyr 106 rotation *may* occur independent of the Thr87-phosphate hydrogen-bond formation; this is in agreement with the observation of the "meta-active" CheY X-ray structure in which the Tyr106 adopts the buried conformer without any major displacement of Thr87 or phosphorylation.[21] Free energy results further confirm that Tyr106 rotation in the absence of the Thr87-phosphate hydrogen bond is at least as kinetically competitive as a pathway in which the hydrogen bond forms first. Therefore, our results clearly argue against the traditional "Y−T" coupling scheme in its most strict sense. The results are consistent with the "population shift" description because the Tyr106 rotation is marginally unfavorable in energy ($\sim$2.1 kcal/mol in Figure 4c) without the Thr87-phosphate hydrogen bond and becomes more favorable ($\sim$5.2 kcal/mol in Figure 5 according to the thermodynamic cycle in the scheme) with that hydrogen-bond formed. As an additional connection with experimental studies of CheY activation, we note that the study of Schuster et al.[27] found that the coupling between phosphorylation and substrate−peptide binding in CheY is only modestly reduced in the T87A mutant, confirming that Thr 87 does not play a decisive role in CheY activation. In fact, significant coupling remains even in the Y106A mutant,[27] which is consistent with the idea that Tyr 106 isomerization is only modestly stabilized by phosphorylation through Thr87. In the NMR study of NtrC,[5] motion associated with the equivalent Tyr was observed to persist in both the unphosphorylated and phosphorylated forms, which led the interpretation that Tyr rotation is "uncoupled" from phosphorylation. The current computational analysis supports that the rotation of Tyr may occur in the absence of phosphoylation, but it is coupled thermodynamically with phosphorylation.

3.2. Role of the $\beta4-\alpha4$ Loop. Our previous biased MD simulations[11] and the X-ray study that revealed the "meta-active" CheY[21] suggested that the $\beta4-\alpha4$ loop may gate the Tyr 106 rotation. In this work, the free energy and TPS results show that the $\beta4-\alpha4$ loop transition has a much higher energy cost and only minimal adjustments (mainly associated with the side chain orientation of Ala90) are needed to leave Tyr106 rotation sufficiently low in barrier. A recent work[28] statistically analyzed all CheY structures in the PDB. Their results show that Tyr 106 rotation is required for the activation of the $\beta4-\alpha4$ loop because the hydrogen-bonding interaction between the hydroxyl group of Tyr 106 and the backbone oxygen of Glu89 can only occur when Tyr 106 is buried. In other words, rather than the $\beta4-\alpha4$ loop gating the Tyr 106 rotation, it is more likely that Tyr106 rotation facilitates the activation of the $\beta4-\alpha4$ loop; we note that a recent equilibrium MD simulation of CheY also found that the conformation of the $\beta4-\alpha4$ loop depends on the orientation of the Tyr106 side chain.[29] Consistent with this point of view, the active configurations of the $\beta4-\alpha4$ loop become accessible in the simulations only after both Tyr 106 rotation and the Thr87-phosphate hydrogen-bond formation are completed (Figure 4b), and it is possible that the fully active configuration of this loop is only highly populated in the presence of the FliM peptide.

3.3. General Implications to Signaling Protein Activation and Allostery. Although the current work focuses on a small signaling protein with relatively localized structural transitions, the results and observations have implications to the analysis of signaling protein activation and allosteric transition in general.

First, from the technical perspective, this study clearly demonstrates the advantages of combining TPS and free energy simulations over the biased MD approach or structural studies alone. Without concrete energetic information, even a collection of biased MD trajectories or a set of X-ray structures may lead to an incorrect conclusion about the importance of specific structural motifs and the causal relationship between processes. For example, our previous biased MD simulations[11] and a structural study[21] suggested that the $\beta4-\alpha4$ loop and the "doorway residues" (Ile 95, Val 108) play the role of gating Tyr106 rotation. This study, by contrast, finds that the "doorway residues" are unlikely crucial for the Tyr106 rotation and the $\beta4-\alpha4$ loop is, in fact, activated by the Tyr106 rotation.

(26) Lee, S. Y.; Cho, H. S.; Pelton, J. G.; Yan, D. L.; Henderson, R. K.; King, D. S.; Huang, L. S.; Kustu, S.; Berry, E. A.; Wemmer, D. E. *Nat. Struct. Biol.* **2001**, *8*, 52.
(27) Schuster, M.; Silversmith, R. E.; Bourret, R. B. *Proc. Acad. Natl. Sci. U.S.A.* **2001**, *98*, 6003.
(28) Dyer, C. M.; Dahlquist, F. W. *J. Bacteriol.* **2006**, *188*, 7354.
(29) Knaggs, M. H.; Salsbury, F. R., Jr.; Edgell, M. H.; Fetrow, J. S. *Biophys. J.* **2007**, *92*, 2062.

Although such a problem with the interpretation of biased MD results is often cautioned in the literature,[11,13,14] this work has provided a clear illustration on how artificially accelerated molecular dynamics simulations may lead to misleading results on subtle features of structural transition processes. The combination of TPS and free energy simulations provides a powerful alternative that should find value in the study of many allosteric systems including signaling proteins.

The detailed mechanistic discussion of CheY activation also reinforces the point emphasized by the "population shift" model of allostery[2−4] that the relevant structural transitions do not have to occur in a highly ordered fashion, propagating from the activation (e.g., phosphorylation) site toward the response site, as assumed in many historical models of allostery.[10] In other words, energetic coupling between different transitions does not necessarily imply any specific sequence of events, as exemplified in CheY by the relationship between the Tyr106 rotation and the Thr87-phosphate hydrogen-bond formation. Since biomolecules are highly dynamical systems,[30] local structural transitions *may* occur in different orders and the completion of one transition can facilitate other transitions through energetic couplings. For some transitions, such as the $\beta4-\alpha4$ loop displacement and Tyr106 rotation in CheY, a definite causal relationship indeed exists but is difficult to unravel based on experiments alone or biased simulations. In this regard, we emphasize that the "population shift" description, which is essentially a modern version of the MWC model,[6] does not, by itself, provide any mechanistic information regarding how different transitions are energetically coupled and through what causal relations. Therefore, revealing *how* population of the active configuration at the response site is shifted by events at the activation site is the major challenge and a key step toward developing new strategies that manipulate the function of signaling proteins and other allosteric systems. In this regard, although unbiased simulations for very large scale structural transitions remain challenging, advanced computational analysis can already complement experimental studies by bringing physical insights[12] regarding the energetic and dynamic characteristics[31] of these systems at multiple scales.[11,15,16,32,33] With further improvement of computational hardware and techniques,[34−36] molecular simulations will be an indispensable tool for the analysis of diverse signaling proteins and other "tunable" biomolecules.

4. Materials and Methods

4.1. Simulation Setup. Two crystal structures (PDB code 1JBE[21] and 1F4V[26]) are taken as the structures of inactive and active CheY, respectively. The hydrogen atoms are added with the HBUILD module[37] in the CHARMM program,[38] and the phosphate group and Mg^{2+} ion are added to Asp57 of the inactive structure to phosphorylate the inactive state *in silico*. The protein atoms and Mg^{2+} ion are described with the all-atom CHARMM22[39] force field for proteins, and the water molecules are described with the modified TIP3P model.[40] In all simulations including TPS and free energy simulations, we start with the phosphorylated structure and set up the system with the stochastic boundary condition. The CheY molecule is partitioned into a 20 Å inner region centered at the geometric center of heavy atoms of residue 87 to 106 and an outer region with the remaining portion of the system. The inner region is solvated with explicit water molecules and can be further divided into a 16 Å molecular dynamics region following the Newtonian equation-of-motion and a 16 to 20 Å buffer region that obeys Langevin dynamics with a thermal bath of 300 K.[41] All the atoms in the inner region are subject to a weak GEO type of restraining potential to keep them inside the inner sphere with the MMFP module of CHARMM; the effect of restraint on most inner region atoms is negligible. All protein atoms in the buffer region are harmonically restrained with force constraints derived directly from the B-factors in the PDB file.[41] Langevin atoms are updated heuristically during the simulation to consistently treat protein groups and water molecules that may switch regions during the simulation. All bonds involving hydrogen are constrained using the SHAKE algorithm[42] to allow a time step of 2 fs. Nonbonded interactions within the inner sphere are treated with the extended electrostatics model, in which groups beyond 12 Å interact as multipoles.[43] To account for the electrostatics between the inner and outer regions and the effect of solvation, the generalized solvent boundary potential (GSBP) approach developed by Roux and coworkers is used.[44,45] The static field due to the outer region atoms is evaluated with the linear Poisson−Boltzmann equation using a focusing scheme that places a 70 Å cube of fine grid (0.4 Å) into a larger 120 Å cube of coarse grid (1.2 Å). The reaction field matrix in GSBP is evaluated using 256 spherical harmonics. In the Poisson−Boltzmann calculations, the protein dielectric constant of 1, the water dielectric constant of 80, and salt concentration of 0.0 M are used. The optimized radii of Roux and Nina[46,47] are adopted to define the solvent−solute dielectric boundary. We note that previous studies have shown that the use of a GSBP protocol has little perturbation on processes localized at the center of the spherical region.[48] Indeed, if one compares the flexibility of the $\beta4-\alpha4$ loop, which is of major interest in this study, the behavior found in the GSBP simulations is very similar to previous unconstrained simulations with explicit solvent[11] (see Supporting Information).

4.2. Transition Path Sampling Protocols. Transition path sampling employs a combination of Molecular Dynamics (MD) and Monte Carlo (MC) simulations. It samples the path phase space using MD and accepts or rejects the paths using MC procedures to obtain the transition path ensemble.[22,23]

4.2.1. Generation of an Initial Trajectory. To generate the initial trajectory as the input for TPS, biased MD is carried out by imposing a harmonic constraint on the backbone atoms of residue 85 to 109 and the heavy atoms of residue 106 with a force constant of 1000.0 kcal/(mol·Å^2). As a result, the CheY molecule is pulled from the inactive/phosphorylated state to the active/phosphorylated state, resulting in a 20 ps non-natural activation trajectory. During this trajectory, the χ angle shows a dramatic transition from the inactive basin to the active basin at around 8−9 ps. Several intermediate structures in this region are selected, and 10 trial trajectories with each lasting 50 ps are carried out to roughly estimate their committor P_B (the probability to reach

(30) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A theoretical perspective of dynamics, structure, and thermodynamics*; Wiley and Sons: New York, 1988.
(31) Jardetzky, O. *Prog. Biophys. Mol. Biol.* **1996**, *65*, 171.
(32) Daily, M. D.; Gray, J. J. *Proteins* **2007,** *67*, 385.
(33) Yu, H.; Ma, L.; Yang, Y.; Cui, Q. *PLoS. Comput. Biol.* **2007**, *3*, 214.
(34) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919.
(35) Faradjian, A. K.; Elber, R. *J. Chem. Phys.* **2004**, *120*, 10880.
(36) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15859.
(37) Brunger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148.
(38) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(39) MacKerell, A. D., Jr. et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(41) Brooks, C. L.; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159.
(42) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(43) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667.
(44) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994,** *100*, 9050.
(45) Im, W.; Berneche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924.
(46) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239.
(47) Nina, M.; Im, W.; Roux, B. *Biophys. Chem.* **1999**, *78*, 89.
(48) Riccardi, D.; Cui, Q. *J. Phys. Chem. A* **2007**, *111*, 5703.

the second basin). One intermediate structure at 8.8 ps of the biasing trajectory shows a $P_B = 0.4$ and simulations are initiated from this structure both forward and backward in time with a velocity distribution at 300 K. This generates a 24.8 ps natural reactive trajectory that connects two basins, and this natural trajectory is used as the initial input for the subsequent TPS shooting algorithm.

4.2.2. Harvest of Unbiased Trajectories The shooting algorithm for stochastic dynamics as implemented[24] in CHARMM is employed to harvest unbiased activation trajectories starting with the initial trajectory described above. The trajectory is regarded as reactive if it starts from one basin and ends in the other with the χ angle making a transition between ~30°−90° and ~150°−210°. In total 410 trajectories are sampled and 163 reactive trajectories constituting the transition path ensemble are collected with an acceptance ratio of 39.7%. Each reactive trajectory lasts ~24.8 ps. Based on the analysis of τ_{mol} and correlation between different reactive trajectories (see Supporting Information), these trajectories are sufficiently long for describing the Tyr106 transition, which is the process of interest here. For more diffusive types of motions, such as the $\beta4-\alpha4$ loop transition, it is likely that substantially longer activation trajectories are needed; this once again highlights the importance of supplementing TPS studies with relevant free energy simulations.

4.3. Free Energy Simulations. Umbrella sampling is employed to explore the 2D free energy profiles composed of two of the three variables: ΔRMSD of all non-hydrogen atoms of the $\beta4-\alpha4$ loop, the Tyr106 χ angle, and the Thr87-phosphate hydrogen-bond distance, D_{hb}. Procedures used to generate the starting structures for the PMFs shown in Figure 4 are described in detail in the Supporting Information. Altogether, there are 416, 256, 280, and 182 windows for Figure 4a−d, respectively.

For each window, the restraining potential w_j on a reaction coordinate D_j (ΔRMSD, χ, or D_{hb}) is given as

$$w_j = K_j (D_j - D_{j,min})^2 \quad (1)$$

where $D_{j,min}$ specifies the position of the minimum of the umbrella potential in a specific window. The reaction coordinate ΔRMSD is defined as the difference between the heavy-atom-RMSD values for the $\beta4-\alpha4$ loop in an instantaneous structure ($\mathbf{X}_t$) relative to the two reference structures ($\mathbf{X}_{active}$ and $\mathbf{X}_{inactive}$),

$$\Delta\text{RMSD} = \text{RMSD}(\mathbf{X}_t - \mathbf{X}_{active}) - \text{RMSD}(\mathbf{X}_t - \mathbf{X}_{inactive}) \quad (2)$$

The force constants K_{RMSD}, $K_{D_{hb}}$, K_χ are gradually reduced from 1000 kcal/(mol·Å^2), 500 kcal/(mol·Å^2), 1000 kcal/(mol·rad^2), to 50 kcal/(mol·Å^2), 25 kcal/(mol·Å^2), 50 kcal/(mol·rad^2), respectively, in the 100 ps equilibration simulation for each window and then kept fixed for another 500 ps (Figure 4a−c) or 1 ns (Figure 4d) of production simulation. The reaction coordinate values from the last 300 ps (Figure 4a−c) or 200 ps (Figure 4d) of each window are used for analysis, and the 2-D Weighted histogram Analysis Method (WHAM)[49,50] is employed to obtain the 2-D free energy surfaces.

Acknowledgment. This work is partially supported by a grant from the National Institutes of Health (R01-GM071428-02) and a Research Innovation Award from the Research Corporation. Q.C. is an Alfred P. Sloan Research Fellow. Computational resources from the National Center for Supercomputing Applications at the University of Illinois are greatly appreciated.

Supporting Information Available: The details of the transition path sampling (TPS) and convergence of free energy simulations as well as the full reference for ref 39 are included. This material is available free of charge via the Internet at http://pubs.acs.org.

JA073059F

(49) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011.
(50) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339.
(51) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.