

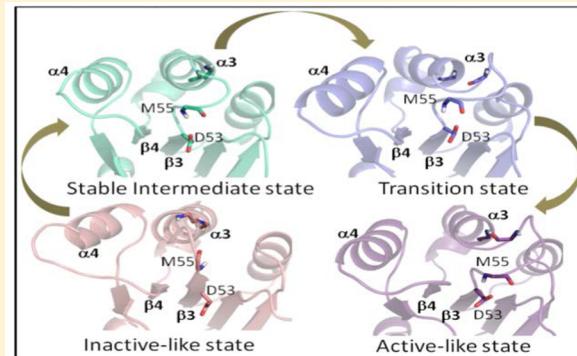
Conformational Transition in Signal Transduction: Metastable States and Transition Pathways in the Activation of a Signaling Protein

Rahul Banerjee,[†] Honggao Yan,^{*,‡} and Robert I. Cukier^{*,†}

[†]Department of Chemistry and [‡]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States

 Supporting Information

ABSTRACT: Signal transduction is of vital importance to the growth and adaptation of living organisms. The key to understand mechanisms of biological signal transduction is elucidation of the conformational dynamics of its signaling proteins, as the activation of a signaling protein is fundamentally a process of conformational transition from an inactive to an active state. A predominant form of signal transduction for bacterial sensing of environmental changes in the wild or inside their hosts is a variety of two-component systems, in which the conformational transition of a response regulator (RR) from an inactive to an active state initiates responses to the environmental changes. Here, RR activation has been investigated using RR468 as a model system by extensive unbiased all-atom molecular dynamics (MD) simulations in explicit solvent, starting from snapshots along a targeted MD trajectory that covers the conformational transition. Markov state modeling, transition path theory, and geometric analyses of the wealth of the MD data have provided a comprehensive description of the RR activation. It involves a network of metastable states, with one metastable state essentially the same as the inactive state and another very similar to the active state that are connected via a small set of intermediates. Five major pathways account for >75% of the fluxes of the conformational transition from the inactive to the active-like state. The thermodynamic stability of the states and the activation barriers between states are found, to identify rate-limiting steps. The conformal transition is initiated predominantly by movements of the $\beta_3\alpha_3$ loop, followed by movements of the $\beta_4\alpha_4$ -loop and neighboring α_4 helix region, and capped by additional movements of the $\beta_3\alpha_3$ loop. A number of transient hydrophobic and hydrogen bond interactions are revealed, and they may be important for the conformational transition.



1. INTRODUCTION

It is well established that conformational dynamics play a critical role in protein function,^{1,2} particularly in signal transduction and regulation,^{3–6} which typically utilizes an inactive to an active state conformational transition as a means of signaling. One of the most prevalent and important forms of signal transduction in living organisms is via a two-component system (TCS), which is frequently integrated into a variety of complex cellular signaling circuits.⁷ TCSs have been found in all three kingdoms of life and are particularly abundant in prokaryotes, including a variety of pathogenic bacteria, such as *Mycobacterium tuberculosis*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa*.⁸ TCS-mediated signal transduction is the predominant mechanism for bacterial sensing of extracellular conditions. For bacteria in the wild, TCS signal transduction is responsible for sensing a variety of environmental conditions, such as temperature, osmopressure, and nutrients. For bacteria in human and animals, TCS signal transduction play critical role in bacterial adaptation to the various compartments of their hosts and are important for their virulence and pathogenesis.

A typical TCS system consists of a protein histidine kinase (HK) and a cognate response regulator (RR). While the HK is responsible for sensing environmental changes, RR functions as a switch and initiates cellular responses upon its phosphorylation by HK. The essence of RR activation is its conformational transition from an inactive to an active state, which is stabilized by phosphorylation. Consequently, elucidation of the conformational dynamics of RR is fundamental for understanding the mechanism of TCS signal transduction. Because of the importance of TCS signal transduction, the mechanism of RR activation has been studied extensively by a variety of experimental and computational methods.⁹ In particular, atomic structures of the inactive and active states have been determined by crystallography and NMR for many RRs, with the active state usually represented by a complex with beryllium trifluoride, which mimics the phosphoryl group attached to a conserved aspartate. Molecular dynamics (MD) simulations have been used to investigate the process of RR conformational

Received: March 17, 2015

Revised: April 29, 2015

Published: May 6, 2015

transition with atomic details difficult to obtain by experimental methods. Two systems well studied by this computational approach are CheY, an RR involved in chemotaxis, and NtrC, an RR that regulates a set of genes in response to nitrogen-limiting conditions. Sophisticated MD simulations have been performed on CheY,^{4,10} but insights gained by these MD studies are limited to the motions of some key residues, as the overall conformational differences between the inactive and active states of the protein are small. All-atom MD simulations have been performed on NtrC by several groups.^{11–13} The salient mechanistic features derived from these studies are not all consistent, however, probably due to the limitations of the computational methodologies employed, such as targeted MD (TMD) simulations, and significant differences in the experimental structures used for these MD studies. The pathways and structural mechanisms of RR conformational transition are still largely unknown.

The TCS HK853 and RR468 from *Thermotoga maritima* has emerged as an excellent model system for studying conformational dynamics of TCS signal transduction.¹⁴ It is one of a few TCSs with their HK–RR complex structures determined.¹⁵ Furthermore, the structures of both the inactive and the active form of RR have been determined by X-ray crystallography at high resolution.¹⁵ Like other RRs, the conformational differences between the two forms of RR468 are mainly located in two loop regions, the $\beta 3\alpha 3$ and $\beta 4\alpha 4$ loops, while its core structure remains quite invariant (Figure 1A).¹⁵ The most

magnitude of conformational change from the inactive to the active state. For example, the RMSD between the inactive and active forms of CheY, an RR in bacterial chemotaxis, is only 0.88 Å for C α atoms, with the largest deviation at 3.66 Å, whereas the RMSD between the inactive and active forms of RR468 is 1.85 Å for C α atoms, with the largest deviation at 9.66 Å (8.58 Å for backbone atoms and 10.49 for all heavy atoms), which is located in the $\beta 3\alpha 3$ loop (Figure 1B).

In previous work,¹⁶ we have shown that, in MD simulations, the conformations of RR468 remain close to the starting conformations derived from the crystal structures of the inactive and active forms of the protein. While the multiple 100 ns simulations provided a great deal of understanding on the conformational plasticity of the protein, they were unable to provide insight into the process of conformational transition between the inactive and active states.¹⁶ TMD simulations were then used to investigate the critical interactions associated with the conformational transition of RR468 from its inactive to active state.¹⁶ Geometric analysis of the TMD data revealed that the conformational transition could be divided into three stages, and breaking and formation of both hydrogen bond and hydrophobic interactions are important for the conformational transition. However, the resolution of the conformational transition in these three stages is necessarily low and, more significantly, the important kinetic and thermodynamic information cannot be obtained via such an analysis due to the limitation of the TMD methodology.

To elucidate the mechanism and pathway of RR conformational transition in TCS signal transduction, in the present work, we performed 101 unbiased 15 ns production all-atom MD simulations of RR468 in explicit solvent, starting with snapshots from various points along the TMD trajectory,¹⁶ yielding an aggregate simulation time of more than 1.5 μ s. The MD data are analyzed using Markov state models (MSMs).^{17,18} Recently, MSMs have been successfully applied to the analysis of protein folding,^{19–26} conformational dynamics of native proteins,^{27–30} protein oligomerization,³¹ ligand binding,³² and enzymatic catalysis.^{33,34} The wealth of the unbiased MD data has allowed us to build a high-resolution Markov state model for the conformational transition of RR468 from the inactive to the active state. The MSM is composed of 2134 kinetically connected microstates, representing the local minima of the free energy landscape of the conformational transition. The microstates are then grouped into 24 metastable states based on their kinetic behaviors. Transition path theory (TPT) analysis of the metastable states has allowed us to identify the major pathways for the conformational transition via a set of intermediates, and calculate the fluxes from the inactive to the active state. Residue interaction network³⁵ and hydrogen bond analyses of the intermediates have allowed us to decipher the changes in structure and interaction during the activation of the RR.

2. METHODS

2.1. Molecular Dynamics Simulations. The starting coordinates for the unbiased molecular dynamics simulations were taken from a 20 ns (TMD) simulation that spanned the inactive to active states.¹⁶ A total of 101 snapshots were selected along this TMD trajectory. For each of these snapshots, restraining potentials in the form of positional restraints on all protein heavy atoms were introduced first and then relaxed to zero in a stepwise manner over 2 ns. Then, unbiased simulations were performed for 15 ns, resulting in a

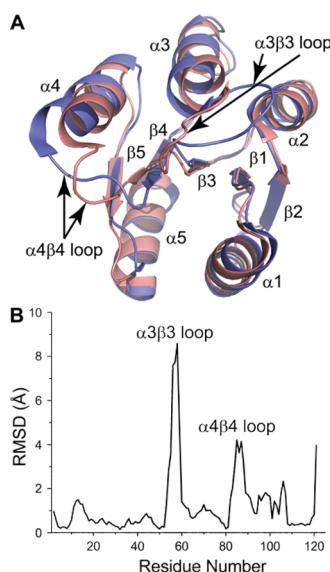


Figure 1. Structural comparison of the inactive and the active form of RR468. (A) Cartoon drawing of the superimposed structures with the inactive (PDB ID 3DGF) and the active (PDB ID 3GL9) form in light pink and blue, respectively. (B) Residue RMSDs between the inactive and the active form of RR468 calculated from the backbone atoms.

significant conformational difference is in the $\beta 3\alpha 3$ loop, which is adjacent to the site of phosphorylation, the conserved aspartate residue, D53. The conformational differences between the inactive and active forms at the $\beta 4\alpha 4$ loop region are mainly due to the $\alpha 4$ helix, which is slanted outward by $\sim 10^\circ$ at the N-terminus of the helix in the active conformation with respect to that in the inactive conformation. Structural comparison with other RRs reveals that, among RRs with the structures reported for both inactive and active states, RR468 has the largest

total of 1.515 μ s of simulation trajectory. Amber10³⁶ and the ff99SB force field were used for the MD simulations. In addition to protein, there were 7068 water molecules for solvation of the protein and 2 Na⁺ ions to neutralize the system (already present from the TMD step).¹⁶ The Particle-Mesh-Ewald³⁷ method was used to evaluate long-range electrostatic interactions. Langevin dynamics was used for temperature control during constant pressure simulation at 300 K. A time step of 2 fs was used with the SHAKE³⁸ option on all bonds containing H atoms. A conformational snapshot was saved every 10 ps so that 151 500 snapshots were accumulated for subsequent MSM analysis.

2.2. Markov State Modeling. Markov state models can be constructed based on kinetic transitions between discrete states, which represent a partitioning of phase space into metastable states.^{17,18} MSM building techniques include some clustering method in the first place to divide the conformational space with a very stringent criterion into clusters, referred to as microstates. Here, the high degree of structural similarity within microstates ensures relatively easy exchanges. The numbers of transitions among different microstates over an interval of length τ , the lag time, provides a transition count matrix $C(\tau)$ with elements $C_{ij}(\tau)$. This matrix can be related to a transition probability matrix $T(\tau)$ with elements $T_{ij}(\tau) = C_{ij}(\tau)/p_i^{\text{eq}}$, with the p_i^{eq} the equilibrium populations obtained from the expected values of the state populations over the trajectory.³⁹ Construction of the transition matrix establishes the kinetic relationship among the microstates. If a suitable range of lag times can be found where the dominant eigenvalues of $T(\tau)$ are independent of τ , the model should satisfy the Markov property. Relaxation time scales for transitions between different microstates are a function of the eigenvalues calculated after diagonalization of the transition matrix, $T(\tau)$. The Markov property can be established by plotting the relaxation time scales for the slowest processes, i.e., based on the dominant eigenvalues, against the lag time. Microstates are lumped into macrostates based on kinetics, whereby there are fast exchanges between microstate pairs within a macrostate, and slow exchange between microstates belonging to different macrostates. With the transition matrix and equilibrium probabilities available, transition path theory can be used to investigate the sequence of intermediates that the protein follows in transiting between given end point states.^{24,40}

2.3. Clustering and Microstates. The MSM analysis was performed using MSMBuilder2 (version 2.6.0).⁴¹ The 151 500 protein conformations from the 101 unbiased trajectories, as described in the Section 2.1, were clustered using a k -centers algorithm as implemented in MSMBuilder2. From the crystal structures of the active and the inactive form of RR468,¹⁵ it is evident that the conformations of the two states differ most in two regions, D53-D60 (the $\beta 3\alpha 3$ loop) and T83-E91 (the $\beta 4\alpha 4$ loop and part of $\alpha 4$). Consequently, the snapshots were aligned based on the backbone heavy atoms (N, Ca, C, and O) and C β atoms of the protein excluding those atoms of residues D53-D60 and T83-E91. They were clustered using RMSD as the distance metric. Backbone heavy atoms, N, Ca, C, O and C β atoms of residues only in the range D53-D60 were used for clustering, because the conformational differences between the two states are the largest in this $\beta 3\alpha 3$ loop region. Also, our previous work suggested that the charge–charge repulsions among a set of ionized residues spanning E88, E89, D90, and E91 at the N-terminus of the $\alpha 4$ helix produce a somewhat unstable region in the protein. The RMSD cutoff for the

clustering (maximum cophenetic distance) was 1.0 Å. Lag times 0.5, 1.0, 1.5, 2.0, and 3.0 ns were tried for MSM model building and estimation of the count ($C_{ij}(\tau)$) and transition ($T_{ij}(\tau)$) matrices, and equilibrium populations (p_i^{eq}) of the microstates. Reversible models that satisfy the detailed balance condition were calculated from the estimation of the most likely reversible matrices using Maximum Likelihood Estimation (MLE) as implemented in MSMBuilder2.⁴¹

2.4. Lumping of Microstates into Macrostates. The PCCA+ algorithm⁴² as implemented in MSMBuilder2⁴¹ was used to lump microstates into models with 8, 12, 16, 20, 24, 28, and 32 macrostates. The equilibrium populations of each macrostate were calculated by adding those of all the constituent microstates. Randomly chosen sets of 1000 conformations from each macrostate were used for subsequent structural characterization using AmberTools⁴³ and Wordom.³⁵

2.5. Transition Path Theory Analysis. The transition pathways, committors (splitting probabilities), and net fluxes for the conformational transitions were calculated using transition path theory (TPT)⁴⁰ as implemented in MSMBuilder2.⁴¹ The highest population microstates for the inactive-like and active-like macrostates were chosen, respectively, as initial and final states for the transition path calculation. Transition paths were initially obtained in terms of the microstates, and later mapped and replaced by the assigned macrostates for each microstate. Averages over the committor values for all constituent microstates were calculated and assigned to the corresponding macrostate.

2.6. Activation Free Energies from Transition State Theory. Activation barriers ΔG_{ij}^{\ddagger} for a transition from macrostate i to macrostate j can be estimated from transition state theory,⁴⁴ using a Kramers flux over population definition of rate constants, K_{ij} .^{44,45} In this approach, the state-to-state rate constants are given by $K_{ij} = (C_{ij}/\tau)/p_i^{\text{eq}}$, and this defines activation barriers via⁴⁶

$$\Delta G_{ij}^{\ddagger} = -k_B T \ln(C_{ij}/p_i^{\text{eq}}) \quad (1)$$

with C_{ij} being the number of transitions from macrostate i to macrostate j , and p_i^{eq} the number of snapshots in macrostate i . Transition state theory assumes that a system in state i that reaches the transition state crosses over irreversibly to state j .

2.7. Protein Structure Network Analysis. A protein structure network (PSN) analysis as implemented in the program Wordom³⁵ was done on each macrostates to assess changes in pairwise residue interactions. The interaction strength (I_{ij}) between residue pair i and j is defined as

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} \cdot 100 \quad (2)$$

Here, I_{ij} is the interaction percentage of residue i and j , n_{ij} is the number of side-chain atom pairs within a given distance cutoff (4.5 Å as a default), and N_i and N_j are, respectively, the normalization factors for residues i and j , that take into account the differences in size of the different residues as bigger residues are likely to make more contact pairs, n_{ij} .^{47,48} Pairwise interactions with interaction percentage 10.0% (I_{\min}) with an occurrence of $\geq 50\%$ in at least one macrostate were used for the analysis.

3. RESULTS

3.1. Construction of the Micro and Macrostates of the MSM. The first step of a discrete state-space MSM

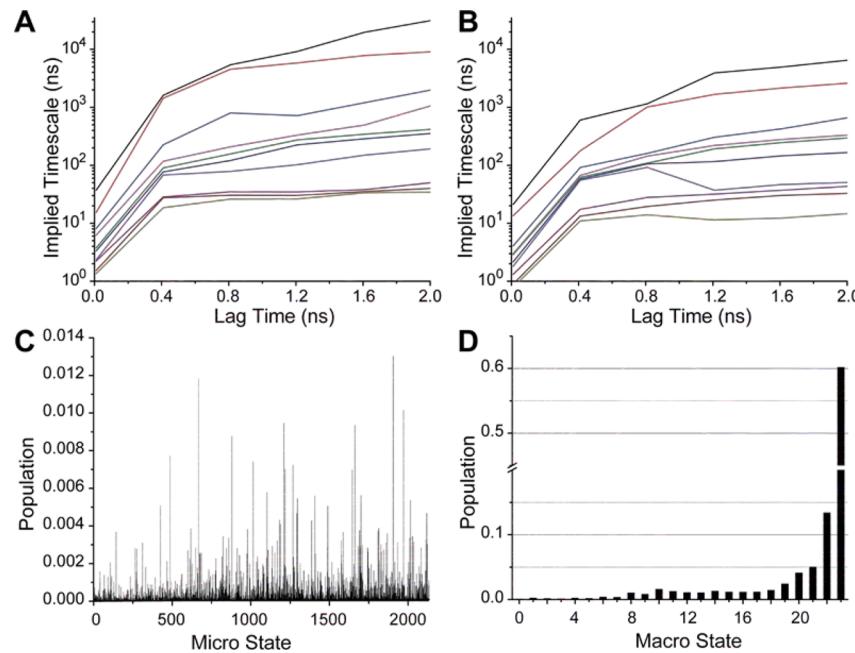


Figure 2. Relaxation time scale as a function of lag time derived from (A) microstate and (B) macrostate transition matrix. (C) Equilibrium population of microstates lumped into (D) 24 macrostates, indexed as 0–23. First 10 relaxation times as a function of lag time shows that there are 6 modes that are characteristically different in terms of the time scale in the range between 1 and 2 ns, as evident from the number of gaps in (A). A macrostate model with 24 metastable states appears to identify six slowest states in the same range as in (B).

construction involves dividing the trajectory data into a set of microstates, based on structural similarity, with a fine-grained cluster cutoff. As noted in Section 2.2, diagonalization of the transition probability matrix obtained from the microstates as a function of the lag time yields relaxation time scales for the various modes of $T(\tau)$. These relaxation time scales correspond to the time scales for transitions among the microstates. They show a strong dependence at short lag times, but should become invariant at longer lag times.⁴⁹ The relaxation time scales for multiple lag times were examined to confirm that the relaxation times become independent of τ , i.e., the states are Markovian on this time scale. As shown in Figure 2A, the relaxation time scale becomes relatively invariant to the choice of lag time in the range between 1 to 2 ns, and the model is Markovian for the choice of lag times within the range. For a choice of lag time more than 2 ns, count matrix elements ($C_{ij}(\tau)$) approach their limiting behavior and relaxation time estimation becomes unreliable.⁵⁰ A microstate model with lag time 1 ns was chosen for subsequent macrostate model generation for three reasons: (1) the microstate model is Markovian at this selection of lag time, (2) it contains the maximum possible data compared to models built from larger lag time microstate models, (3) each individual equilibrium simulation (15 ns) is sufficiently longer than the chosen lag time.

There are 2134 states in the microstate model with a lag time of 1 ns. These microstates were then lumped into varied numbers of macrostates using the PCCA+ algorithm,⁴² and their relaxation time scales versus lag time were plotted for various numbers of macrostates. Relaxation time scales for the macrostate model with 24 states follow that of the microstate model (Figure 2B). The first 10 relaxation times as a function of lag time calculated from 24 macrostates also identified roughly the same number of processes that are distinct in terms of the time scale. Equilibrium populations of the constituent

microstates were added to get the macrostate populations (Figure 2C and D). The equilibrium population of macrostate 23, accounting for ~60% of the total population, outnumbers those of all other states.

We extended the macrostate implied time scale plot to a lag time of 8 ns, as displayed in Figure S1 in the SI. The longest, most important eigenvalues show a reasonable consistency. It should be noted that the number of microstates found by the clustering algorithm, with the clustering criterion always fixed at 1 Å RMSD, for consistency, decreases as the lag time increases. Our 101 individual windows for the construction of the data pool only have 15 ns of data per window. Thus, while there is a great deal of data, it cannot be used for very long lag times.

One method of assessing the Markov state property is to use the Chapman–Kolmogorov (CK)⁵¹ condition that for a discrete time and state Markov process is $T(k\tau) = T^k(\tau)$ ($k = 1, 2, \dots$).^{39,52–54} As just noted, obtaining the transition matrix for very long lag times is not feasible with the data used here. A limited time check was carried out by direct multiplication of the transition matrix for lag time $\tau = 1$ ns and evaluation of diagonal elements of the product matrix. The diagonal elements correspond to time decays of state populations. Note that the matrix multiplication will involve all elements of the transition matrix including low-population macrostates that cannot be accurately obtained. Supporting Information Figure 2 presents time decays for some of the macrostates and show reasonable agreement of $T_{SS}(k\tau) = [T^k(\tau)]_{SS}$. Other diagonal elements satisfy the CK condition less well.

3.2. Two Distance-Based Metrics Characterize the Conformational Transition of RR468. In order to determine which of the macrostates are like the inactive and active states that span the conformational transition, we identified two sets of inter-residue distances that are characteristic of these two states (Table S1). One set of distances is for the $\beta 3\alpha 3$ loop and the other for the $\beta 4\alpha 4$ loop and $\alpha 4$ helix. The identification of

these characteristic distances was based on their average C α -atom distances in the active and inactive states, d_A^{ij} and d_I^{ij} , respectively, and their standard deviations, δ_A^{ij} and δ_I^{ij} , as measured from the 100 ns simulations starting from the crystal structures of the inactive and active forms of the protein. The characteristic distances are those average distances with their differences between the active and inactive states larger than 2 times the sum of their standard deviations:

$$|d_A^{ij} - d_I^{ij}| > 2(\delta_A^{ij} + \delta_I^{ij}) \quad (3)$$

This condition ensures that there is no overlap in the distance distribution as calculated from the active- and inactive-form MD simulations.

Then, the average distance, d_M^{ij} , between the i th and j th C α atoms was calculated for each macrostate, M. A new parameter was defined to assess the similarity of the macrostate to the active/inactive state based on these characteristic distances:

$$\Delta Q_M = \sum_{ij} f_M^{ij} \quad (4a)$$

where

$$\begin{aligned} f_M^{ij} &= +1 \text{ when } d_A^{ij} - 2\delta_A^{ij} < d_M^{ij} < d_A^{ij} + 2\delta_A^{ij} \\ &= -1 \text{ when } d_I^{ij} - 2\delta_I^{ij} < d_M^{ij} < d_I^{ij} + 2\delta_I^{ij} \\ &= 0 \text{ otherwise} \end{aligned} \quad (4b)$$

Two such parameters were calculated, one for the conformation of the $\beta 3\alpha 3$ loop, $\Delta Q_M^{\beta 3\alpha 3}$, based on 12 characteristic distances between the $\beta 3\alpha 3$ loop and core residues, and the other for the coupling of the $\beta 3\alpha 3$ loop with the $\beta 4\alpha 4$ loop and $\alpha 4$ helix, $\Delta Q_M^{\alpha 4}$, based on 12 characteristic distances between $\beta 3\alpha 3$ -loop and $\beta 4\alpha 4$ -loop/ $\alpha 4$ -helix residues (Table S1).

A connectivity map among the macrostates was derived from their transition matrix (Figure 3). Using these two parameters, the macrostates most similar to the active and inactive states can be identified and the degrees of similarity calculated for all macrostates (Figure 3 and Table S2). Macrostate 23 is identical to the inactive conformation in the sense that all the distances fall within the distributions of the distances characteristic of the inactive state, with both $\Delta Q_M^{\beta 3\alpha 3}$ and $\Delta Q_M^{\alpha 4}$ reaching -12. Macrostate 16 is most similar to the active form, where both $\Delta Q_M^{\beta 3\alpha 3}$ and $\Delta Q_M^{\alpha 4}$ adopt the value +6. The condition for the equilibrium simulations starting from the TMD generated seeds favor the inactive state, so it is not surprising that we were able to identify a state (macrostate 23) where all pairwise distances match that of the inactive form. In the absence of a phosphoryl group and Mg²⁺, the system is unlikely to be stable in the active conformation. Nevertheless, for all pairwise distances, f_{16}^{ij} adopts values of either +1 or 0; i.e., there is not a single pairwise distance in macrostate 16 that matches the inactive form distribution. It is worth noting that in the active form simulation with or without the phosphoryl group or Mg²⁺ the system remains close to the initial structure.¹⁶ So the active form conformation is a stringent condition, and any initial seed obtained from the TMD simulation is unlikely to match the active form in all aspects. Side-chain atom distances that are characteristic of the active and inactive forms further support the selection of macrostates 23 and 16 as inactive and active-like conformations, respectively.

3.3. The Top Five Pathways Account for over 75% of the Fluxes from the Inactive to the Active Conforma-

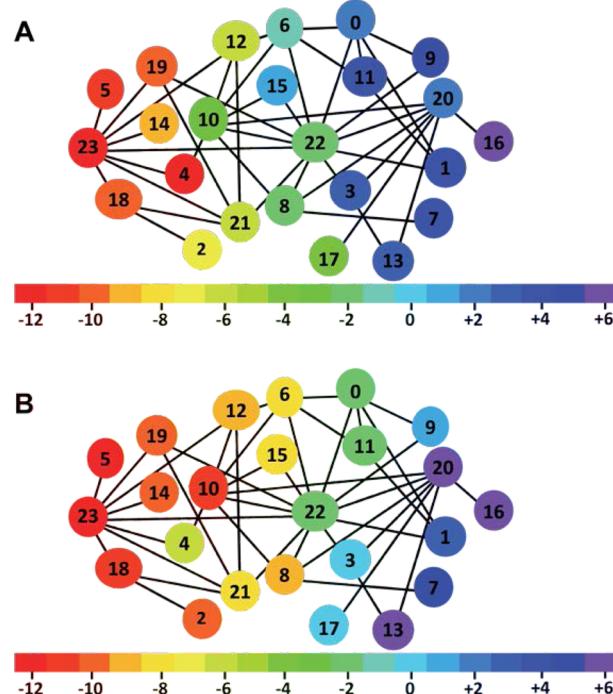


Figure 3. Connectivity map of the macrostates and their degrees of similarity to the inactive and active states. The degrees of similarity are color coded according to the $\beta 3\alpha 3$ -loop dependent parameter, $\Delta Q_M^{\beta 3\alpha 3}$, in panel A and the $\alpha 4$ helix dependent parameter, $\Delta Q_M^{\alpha 4}$, in panel B. The connectivity between the macrostates was derived from the transition matrix.

tional Transition. With the MSM built and the transition matrix available, we analyzed the transition pathways for the conformational transition using TPT.^{17,40} The pathways are characterized by reactive fluxes and committor (splitting probability) values. Committor values, ranging from 0 to 1, are the probabilities of the intermediate states transiting to the final state without revisiting the initial state.⁴⁰ The inactive-like and active-like macrostates, 23 and 16, respectively, were chosen as the initial and final states for the conformational transitions. Twenty-two different pathways were found by the TPT analysis. The five top pathways are schematized in Figure 4. All the rankings and reactive fluxes are listed in Table S3, and the committor values for individual macrostates are in Table S2. Note that only 12 out of the 24 macrostates are involved in

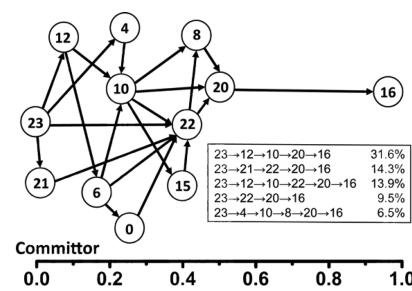


Figure 4. Pathways connecting macrostates 23 (inactive-like) and 16 (active-like). There are a total of 22 pathways that involve 10 intermediate states; the relative fluxes (%) of the top five pathways are listed in the inset. Committor probabilities for all the macrostates were calculated by transition path theory and indicated by their horizontal positions.

these pathways. All pathways go through macrostate 20, which has a committor value of 0.5, i.e., with equal probabilities of transiting to macrostate 16 with and without revisiting macrostate 23. All other intermediate states have a committor value less than 0.5, i.e., with a higher probability of revisiting macrostate 23 than that for reaching macrostate 16.

The top five pathways, each with a reactive flux of >5%, account for 75.8% of the total reactive fluxes from macrostate 23 to 16 (Figure 4), and the very top pathway, 23 → 12 → 10 → 20 → 16, alone accounts for nearly 32%. Each of the top five pathways contains only 2–4 intermediates.

Activation barriers ΔG_{ij} to crossing between pairs of states, i and j , can be estimated from eq 1. They provide information on potential rate limiting steps. Figure 5 presents activation

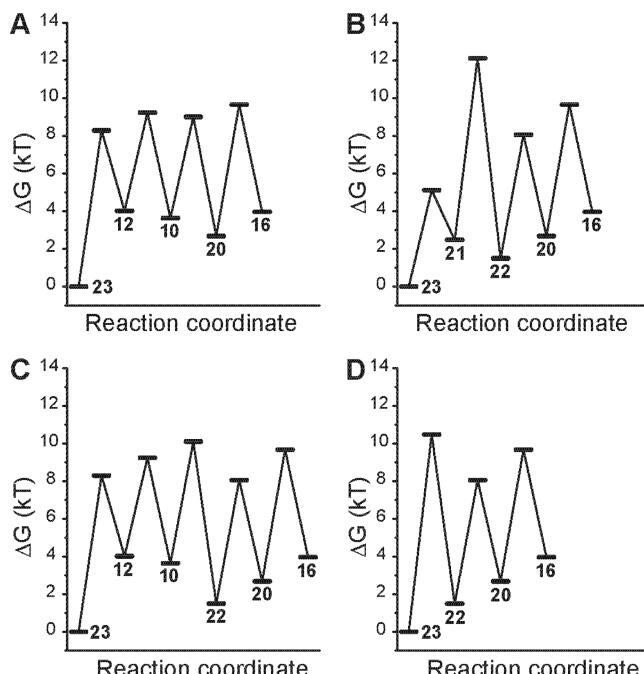


Figure 5. State-to-state activation barriers ΔG_{ij} estimated from transition state theory (eq 1) for the first four top pathways; respectively, panels A, B, C, and D. In the forward (inactive to active) direction, the largest activation barrier for the top pathway (panel A) is for the 23 to 12 transition, the first step. Other largest barriers are for the 23 to 22 and 21 to 22 transitions.

barriers for the top four pathways in Figure 4. In the forward direction (inactive to active), the rate-limiting step is the first step (23 to 12) for the top pathway. Other large activation barriers are for the 21 to 22 and 23 to 22 transitions. Large activation barriers are mainly but not always associated with large committor changes. For example, the ΔG_{ij} for the 23 to 22 and for 21 to 22 transitions are large, and they correspond to a committor change ~ 0.4 . For the 23 to 12 transition, however, ΔG_{ij} is also large but the committor change is small. Transition state theory rates (and corresponding activation energies) provide local information connecting two specific states. A committor is a global descriptor for an intermediate between initial and final states.

A mean first passage time (mfpt) analysis^{54–56} can be used to estimate the time scales of transitions for the major pathways that are displayed in Figure 4. A mfpt provides the first time a final state F is reached from some initial state I passing through, in principle, all the intermediate states. For the two top

pathways 23 → 12 → 10 → 20 → 16 and 23 → 21 → 22 → 20 → 16 listed in Figure 4, the mfpt values between these macrostate pairs range from ~ 750 ns to $\sim 5,500$ ns. The longest mfpt among this set of macrostates is for the complete inactive (23) to active (16) macrostate transition with mfpt ~ 7000 ns. These values, summarized in Tables SI 6 and SI 7, are long compared with the lag time used in the MSM construction, and suggest that the transitions of interest here occur on a microsecond time scale. For some of the other macrostates, there are mfpts on the order of a microsecond, but it should be noted that a mfpt between any pair of states depends on all the intervening states, and for the low probability states, numerical errors will be introduced.

3.4. The Inactive-to-Active Conformational Transition is Mainly Initiated by $\beta 3\alpha 3$ -Loop Movements. With the important transition pathways established, we can now examine the conformational transition in structural terms. The conformational transition involves mainly two regions of the protein, one containing the $\beta 3\alpha 3$ loop and the other containing the $\beta 4\alpha 4$ loop and the N-terminus of the $\alpha 4$ helix. As described in Section 3.2, the conformations of the macrostates and their degrees of similarity to the inactive and active conformations can be measured by two parameters, $\Delta Q_M^{\beta 3\alpha 3}$ for the $\beta 3\alpha 3$ loop and $\Delta Q_M^{\alpha 4}$ for the $\beta 4\alpha 4$ loop and $\alpha 4$ helix. The conformations of the intermediates in the top five pathways as measured by these two parameters are plotted in Figure 6. The results shows

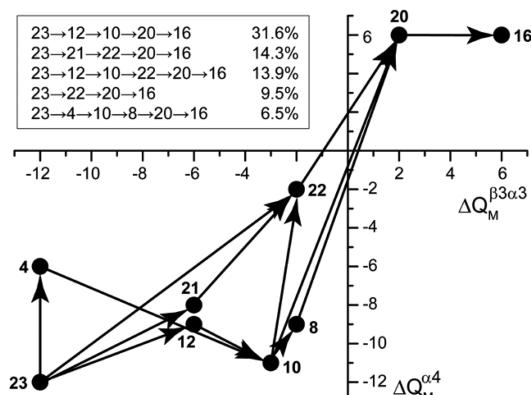


Figure 6. Conformations of the macrostates involved in the top five transition pathways from the inactive to the active state. The $\beta 3\alpha 3$ loop and its coupling with the $\beta 4\alpha 4$ loop and $\alpha 4$ helix are measured by the two parameters $\Delta Q_M^{\beta 3\alpha 3}$ and $\Delta Q_M^{\alpha 4}$, respectively, as defined in section 3.2. The relative fluxes (%) of the top five pathways are listed in the inset.

that the top three pathways, which accounts for 59.8% of the reactive fluxes, are initiated by mainly $\beta 3\alpha 3$ -loop movements, with only small movements in the $\beta 4\alpha 4$ loop and $\alpha 4$ helix. The large activation barriers shown in Figure 5 do mainly correspond to large, two-dimensional motions in this conformational space, as shown in Figure 6. In particular, the 23 to 12, 23 to 22, 23 to 21 and 21 to 22 transitions all involve concerted motions and have large activation barriers.

The fifth top pathway, 23 → 4 → 10 → 8 → 20 → 16, starts with a movement in the $\beta 4\alpha 4$ loop and $\alpha 4$ helix (macrostate 4), but in the next step (macrostate 4), the $\beta 4\alpha 4$ loop and $\alpha 4$ helix move back to nearly the original conformation, and the $\beta 3\alpha 3$ loop moves to about the halfway point to the active state. Then the $\beta 4\alpha 4$ loop and $\alpha 4$ helix move to the active-like conformations in two steps (macrostates 8 and 20). In the

fourth top pathway, $23 \rightarrow 22 \rightarrow 20 \rightarrow 16$, which accounts for 9.5% of the reactive fluxes, the two regions move extensively and concurrently in a single step to nearly the halfway point (macrostate 22) to the active-like conformation.

Of the six minor pathways, with reactive fluxes of 1–5% and ranked 6–11 (Table S3), pathways 6–8 and 10 are initiated with mainly $\beta\alpha_3$ -loop movements. Pathways 9 and 11 are similar to pathway 5, with the first step involving macrostate 4 and the second step macrostate 10, i.e., the $\beta\alpha_4$ loop and α_4 helix region moving first but moving back to nearly the original conformation in the next step and the $\beta\alpha_3$ loop moving to about the halfway point to the active state. With these major and minor pathways taken together, 81.9% of the total reactive fluxes are initiated with mainly $\beta\alpha_3$ -loop movements, and only 9.5% of the total fluxes with concurrent movements of the two regions.

3.5. Macrostates 20 and 22 are Key Intermediates in the Inactive to Active Conformational Transition.

Macrostate 20 is the most important intermediate in the conformational transition, as all transition pathways, displayed in Figure 4, go through this intermediate, and it has the largest committor value. It can be viewed as the “gateway” to the final transition to macrostate 16, the active-like state. Macrostate 22 is also important in the conformational transition; it has the highest population (13.4%) among all intermediates (Figure 2), and the majority of the pathways (68.2%) and fluxes (57.4%) go through this intermediate before reaching macrostate 20, on the way to the final state.

3.5. A Structural Properties of Macrostate 22. Macrostate 22, which, in terms of committor value, is almost halfway to macrostate 16, adopts “−2” values on both the $\Delta Q_M^{\beta\alpha_3}$ and $\Delta Q_M^{\alpha_4}$ scales (Figure 6 and Table S2). The formation of macrostate 22 from macrostate 23 corresponds to a large motion along both scale directions. This large, “direct” transition only contributes to the fourth pathway in Figure 4, with its 9.5% relative flux. In terms of activation barriers, it is the rate-limiting step, as shown in Figure 5, panel D. The macrostate 23 to 22 transition requires breaking and forming many pairwise residue interactions (Figure 7 and Table S4), as shown by PSN analysis (see Section 2.7). Pairwise residue interactions that are broken or nearly so in this transition include E28/K117, A38/L68, I54/V58, M56/A84, F62/L65, E88/R104, and I113/K117. Pairwise residue interactions that are formed in this transition include D10/P57, A38/V64, L52/G61, I54/A94, M55/V58, and M55/K85.

Macrostate 22 is quite distinct in character relative to the inactive and active-like macrostates, 23 and 16. The PSN analysis shows that the pairwise residue interactions formed in macrostate 22—A38/V64, I54/A94, M55/V58, and M55/K85—are transient interactions, i.e., they are either absent or have very low frequencies in macrostates 16 and 23 (Figure 7 and Table S4). Furthermore, several hydrogen bonds are broken or nearly so in this transition, including those between V58 and F62, D62 and T63, and G87 and E91 (Figure 8 and Table S5).

The backbone conformation of the $\beta\alpha_3$ loop in macrostate 22 is of special interest, as this loop is closest to the site of phosphorylation. In this macrostate, the backbone carbonyl group of M55 from the $\beta\alpha_3$ loop approaches three highly conserved, active site carboxylate residues, D9, D10, and D53. The rotation of the I54 backbone Ψ torsion angle from its value in the initial inactive form (macrostate 23) accounts for the change in the $\beta\alpha_3$ loop backbone conformation (Figure 9A).

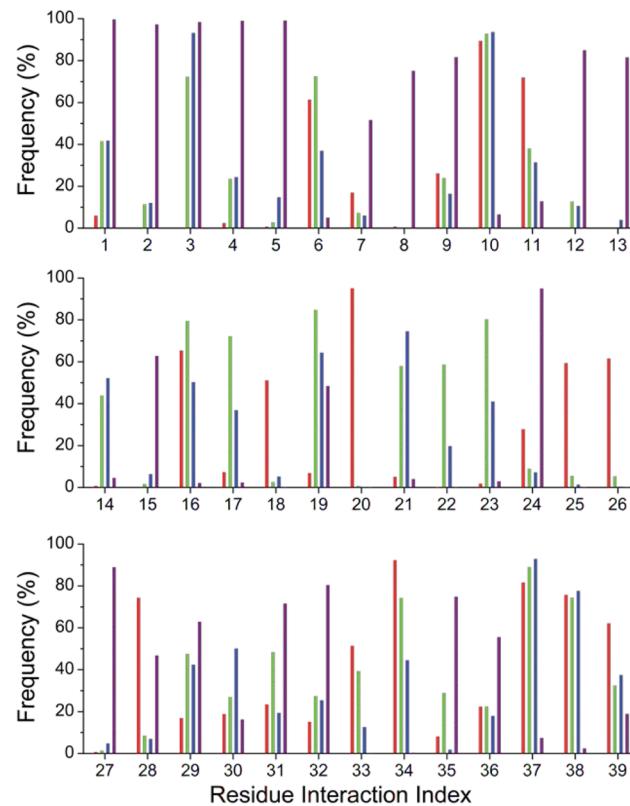


Figure 7. Pairwise residue interaction changes. Pairwise residue interactions and their strengths were calculated as described in the Methods section. The cutoff for interaction strength is 10%. The frequency of occurrence was calculated by the number of occurrences in a macrostate divided by its population. The frequencies of the pairwise residue interactions in macrostates 23, 22, 20, and 16 are represented by red, green, blue, and purple columns, respectively. Only the interactions with >50% occurrences in one or more but not all critical states are plotted. The residues involved in the 39 pairwise interactions are defined in Table S4.

In this regard, it is important to recognize that a divalent cation is necessary for stabilization of the active conformation. In RR468, Mg^{2+} is present in the active form crystal structure, along with the phosphoryl mimic BeF_3^{-15} . The six coordination positions of the Mg^{2+} ion are occupied in part by the three conserved aspartate residues, D9, D10, and D53. One of these aspartates, D9 interacts through a water molecule. The backbone carbonyl group of M55 from the $\beta\alpha_3$ loop and two water molecules are also within the coordination sphere of the Mg^{2+} . This active site architecture is very common among RRs.⁹ During the simulation of RR468 in the absence of a Mg^{2+} ion, all the conserved aspartates in the active site remain close to each other with the exception of the carboxylate group of D10, which points outward by 180° rotation of side-chain χ_1 , due to repulsion with that of residue D53. By contrast, in macrostate 22, the $\beta\alpha_3$ loop backbone comes close to the active site and gets even closer in macrostate 16, with the backbone amide oxygen of M55 ready to serve as a coordination ligand for the Mg^{2+} (Figure 9B,C).

3.5. B Structural Properties of Macrostate 20. Macrostate 20 has values of 2 and 6 on the $\Delta Q_M^{\beta\alpha_3}$ and $\Delta Q_M^{\alpha_4}$ scales, respectively (Figure 6 and Table S2). Its $\Delta Q_M^{\alpha_4}$ value is the same as that of macrostate 16, the final state, indicating that the conformations of the $\beta\alpha_4$ loop and α_4 helix are nearly the same between the two macrostates. Pairwise residue

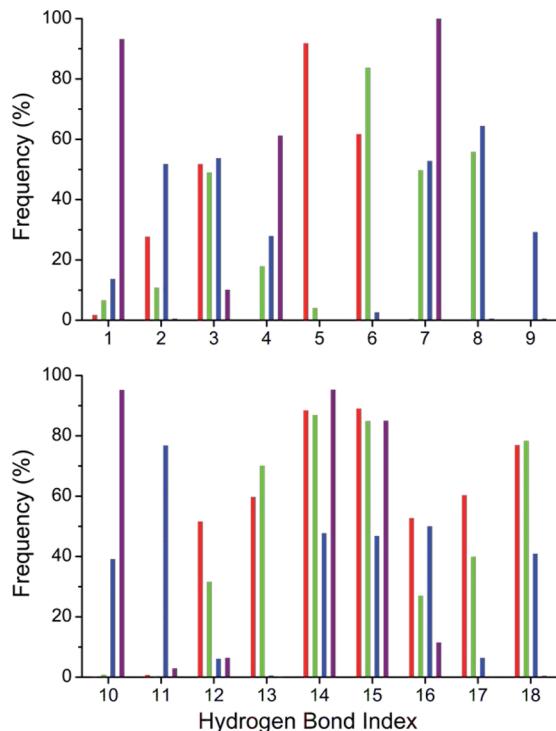


Figure 8. Hydrogen bonding changes. Hydrogen bonds are defined by D–A distances ($<3.5\text{ \AA}$) and D–H–A angles ($>120^\circ$). The frequency of occurrence was calculated by the number of occurrences in a macrostate divided by its population. The frequencies of the hydrogen bonds in macrostates 23, 22, 20, and 16 are represented by red, green, blue, and purple columns, respectively. Only hydrogen bonds with $>50\%$ occurrences in one or more but not all critical states are plotted.

interactions that are broken or formed, or nearly so, include I17/P109, A38/V64, M55/V58, M55/K85, L65/A99, and E91/R104 (Figure 7 and Table S4). The hydrogen bond between E91 and R104 is nearly broken in the transition from macrostate 22 to 20 (Figure 8 and Table S5). The hydrogen bonds between F62 and K66 and between F63 and K67 are significantly weakened. Two new hydrogen bonds are formed: one between D9 and S11 and the other between D60 and F62. Both hydrogen bonds are transient in the sense that they are absent in macrostates 23, 22, and 16 (Figure 8 and Table S5).

The $\beta 3\alpha 3$ loop in macrostate 20 exhibits enhanced flexibility. The flexibility of the loop may be due to the unwinding of the adjacent structural elements. Previous equilibrium simulations found that D60 is part of the $\alpha 3$ helix in the inactive conformation, but it is part of the $\beta 3\alpha 3$ loop in the active conformation.¹⁶ Loss of backbone interaction between D60 and V64 at the early stage of the TMD simulation also points toward the weakening of the N-terminus of the $\alpha 3$ helix.⁵⁰ The number of occurrences of the characteristic backbone hydrogen bonds between D60 and V64, F62 and K66, and T63 and K67 at the N-terminus of the $\alpha 3$ helix were calculated for all the crucial macrostates (Figure 8 and Table S5). The N-terminus of the $\alpha 3$ helix is significantly weaker in macrostate 20 as compared to the other macrostates.

3.6. Macrostate 16 Is Stabilized by Many Hydrophobic Interactions. The transition from macrostate 20 to 16 mainly involves movements of the $\beta 3\alpha 3$ loop (Figure 6), which results in a net gain of many pairwise residue interactions as indicated by the PSN analysis (Figure 7 and Table S4). In comparison with macrostate 20, macrostate 16 loses five

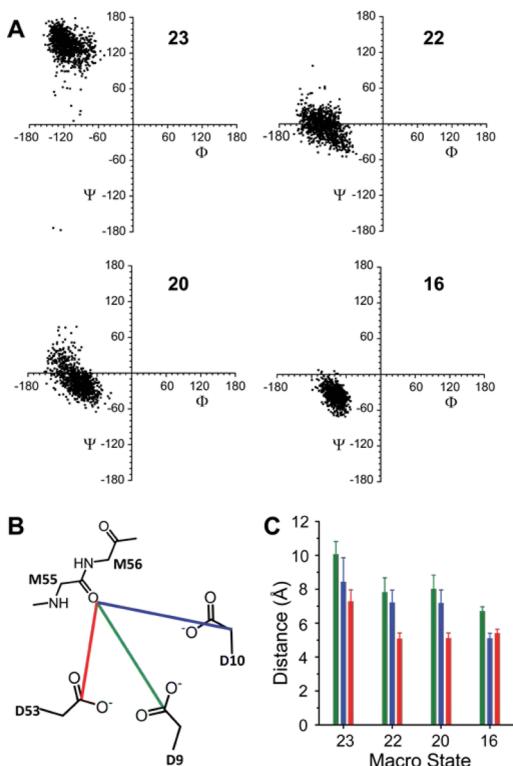


Figure 9. Conformational changes of the $\beta 3\alpha 3$ loop. (A) Changes in the backbone torsion angle of I54. In macrostate 23, the I54 Ψ angle approximates the inactive form value. As a result, the M55 carbonyl group is away from the active site. In states 22, 20, and 16, the changes in the I54 Ψ angle bring the M55 carbonyl group close to other active site residues. (B) The $\beta 3\alpha 3$ loop backbone comes close to the active site, comprised of the three conserved aspartate residues D9, D10, and D53, in macrostates 22, 20, and 16, unlike in macrostate 23. (C) The average distance between the M55 carbonyl group and side-chain atoms of D9, D10, and D53. The distances are defined in panel B.

pairwise residue interactions, L22/I113, N34/D60, I54/A94, M103/F107 and I113/K117, but it gains 16 pairwise residue interactions, D9/K105, D10/M56, S11/K105, L14/P106, V18/L82, L22/V51, N34/P57, N34/M59, Q36/M59, M55/K105, D60/V64, L65/A94, P78/R100, T83/D90, R100/L119, and K101/H118. Half of these pairwise interactions involve the $\beta 3\alpha 3$ loop and adjacent residues, which stabilize the conformation of this loop. By definition, these pairwise residue interactions are dominated by hydrophobic interactions.^{47,48} Consequently, the conformation of macrostate 16 is stabilized by many new hydrophobic interactions. In this regard, it is noted that 12 of the 16 newly formed pairwise residue interactions are characteristic of macrostate 16, i.e., they appear only infrequently, or not observed at all in other macrostates. There are also significant changes in hydrogen bonding in the transition from macrostate 20 to 16 (Figure 8 and Table S5). Three hydrogen bonds are lost, but five are gained. The lost hydrogen bonds are those between D9 and S11, N34 and I37, D60 and F62, and G87 and E91. The gained hydrogen bonds are those between D9 and K105, D53 and M55, D60 and G61, T62 and K66, and T63 and K67. The hydrogen bond between D9 and K105 may be particularly important for the stabilization of the conformation of macrostate 16, because this hydrogen bond is present only in macrostate 16, not in other macrostates, and these two residues are far apart in sequence.

4. DISCUSSION

The key to understanding the mechanisms of biological signal transduction is to elucidate the conformational dynamics of its signaling proteins, as the activation of a signaling protein is fundamentally a process of conformational transition from an inactive to an active state. Simulating the activation process by unbiased MD is challenging, however, as such a process is usually much longer than simulation times that can be achieved by currently affordable computational resources. To overcome the time scale issue faced by unbiased MD simulation methods, TMD simulations have been used to investigate RR activation in two-component signal transduction.^{11,13,16} Although these TMD simulations have revealed important structural features in RR activation, the kinetic and thermodynamic information is lost in these biased simulations. Consequently, the transition pathways obtained are necessarily simplistic and have a high uncertainty. To overcome these limitations, in this study, we performed extensive unbiased MD simulations using snapshots from our TMD simulation⁵⁰ as the starting points, and obtained in aggregate more than 1.5 μ s of unbiased MD data. The complexity of the wealth of the MD data was resolved by MSM and TPT analyses, which yielded decipherable metastable states and transition pathways.

While a direct comparison of these computational results with experiment is not possible at present, the population distribution of RR468 estimated by the MSM analysis is comparable to that measured for NtrC by NMR. According to an NMR relaxation dispersion measurement and a two-state analysis,⁵⁷ the active population of the unphosphorylated NtrC is 14%, and the inactive population is 86%. Using such a two-state model, with a cutoff of -2 on both the $\Delta Q_M^{\beta\alpha_3}$ and $\Delta Q_M^{\alpha_4}$ scales, the active population of RR468 includes macrostates 0, 1, 3, 7, 9, 11, 13, 16, 20, and 22, and accounts for 22% of the total population. If macrostate 22, a macrostate on the border of the cutoff on both scales, is excluded, the active population of RR468 is 9% of the total population. Both population estimations are close to that measured for NtrC by NMR.⁵⁷ This comparison indicates that the population distribution of RR468 estimated by the MSM analysis is reasonable.

Our results offer a comprehensive picture of RR activation. The results reveal that the active-like conformation has a significant population (Figure 1 and Table S2), in support of a "population shift" model for RR activation. In contrast to a single pathway and a few intermediates revealed by previous TMD simulations,^{11,13,16} our MSM and TPT analyses of the extensive unbiased MD data show that the activation of RR468 involves a network of metastable macrostates (Figures 3 and Table S2) and follows a series of interconnected transition pathways (Figure 4 and Table S3). Ten of these metastable macrostates serve as intermediates in the transition pathways, and five major pathways account for >75% of the fluxes of the conformational transition from the active to the active state. All pathways go through one intermediate, macrostate 20, to reach the final state, macrostate 16 (Figure 4). The transition from macrostate 20 to 16 has a high barrier, but not the highest, however, and consequently is not the rate-limiting step in the transition pathways. The highest activation barriers belong to the transitions to macrostate 22 (Figure 5). Unsurprisingly then, the major pathway, 23 \rightarrow 12 \rightarrow 10 \rightarrow 20 \rightarrow 16, which accounts for ~32% of the total reactive fluxes, does not go through macrostate 22. The step with the highest barrier in this pathway is the first step, from macrostate 23 to 12. However,

macrostate 22 has the highest population (Figure 1 and Table S2) and the lowest free energy among all intermediates, and the majority of transition pathways and fluxes go through this intermediate.

With the macrostates, their transition pathways, pathway strengths, state committer values and free energies, and state-to-state activation barriers determined, we then investigated how RR activation occurs in structural terms. A summary of some of the structures from inactive to active conformations is shown in Figure 10.

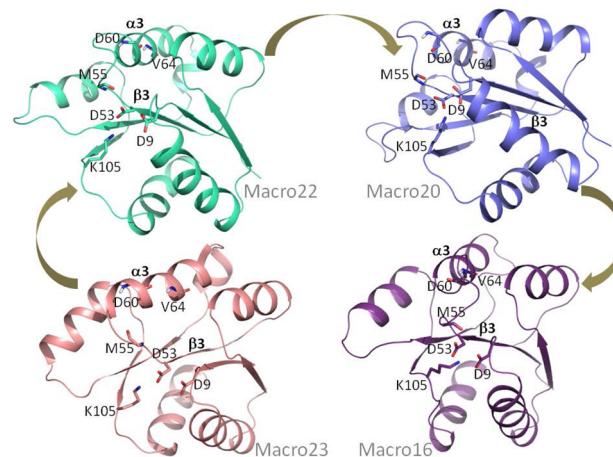


Figure 10. Salient features of major macrostates in terms of their active site reorganization, changes in α_3 helix conformation, and changes in interactions among the proximal loop residues. Formation of the salt bridge between K105 and D9 facilitates other interactions among proximal loop residues. The presence of this salt bridge indicates substantial interactions among proximal loop residues, resulting in reduced solvent accessibility of the active center D53. In macrostate 23 (bottom left) the backbone carbonyl group of M55 points away from the active site residue, D53. The α_3 helix is longer with D60 being part of the helix. The side chain of K105 ($\beta_3\alpha_3$ loop) is away from that of the D9 ($\beta_1\alpha_1$ loop). In macrostate 22 (top left), the backbone carbonyl group of M55 points toward the active site residue, D53 in a way conducive to bivalent metal ion binding. The other two features, however, remain invariant from the preceding macrostate 23. In macrostate 20 (top right), loss of characteristic helix interactions between D60 and V64 results in enhanced flexibility of the $\beta_3\alpha_3$ loop. The orientation of the backbone carbonyl group of M55 remains similar to that of the macrostate 22. In macrostate 16 (bottom right), the proximal loops have enhanced interactions, as indicated by the presence of the K105-D9 salt bridge. The salt bridge is a critical feature of active conformation of the protein and is not present in macrostates 23, 22, or 20.

While the conformational differences between the inactive and the active forms of RR468 are largest in the $\beta_3\alpha_3$ loop, the $\beta_4\alpha_4$ -loop and α_4 helix region is most flexible in the inactive state, as this region has the highest fluctuations in the equilibrium MD simulation of the inactive form of the protein.¹⁶ However, geometric analysis of the major and significant minor pathways indicates that the conformational transition from the inactive to the active state is initiated predominantly (~82% of the total reactive fluxes) by movements of the $\beta_3\alpha_3$ loop. About 10% of the reactive fluxes are initiated by concurrent movements of the $\beta_3\alpha_3$ loop and the $\beta_4\alpha_4$ -loop and α_4 helix region. In either case, the next stage of the conformational transition, the transition to the final intermediate (macrostate 20 with a committer value of 0.5),

mainly involves movements of the $\beta 4\alpha 4$ -loop and $\alpha 4$ helix region (Figure 6) and the weakening of the N-terminus of the $\alpha 3$ helix. The conformation of the $\beta 4\alpha 4$ -loop and $\alpha 4$ helix region of macrostate 20 is almost identical to that of macrostate 16, and the final step of the conformational transition involves mainly movements of the $\beta 3\alpha 3$ loop (Figure 6). Our results show that the most flexible region of a protein does not necessarily initiate the major conformational transition pathways.

As expected, activation of RR468 involves the breaking of interactions existing only in the inactive state and the formation of interactions existing only in the active state. However, activation of the protein also involves transient interactions, interactions that are present only in the intermediates but not in the initial and final states. There are six residue interaction pairs present only in macrostates 22 and/or 20, but not in macrostates 23 and 16 (Figure 7 and Table S4). Three hydrogen bonds are found only in macrostates 22 and/or 20, but not in macrostates 23 and 16 (Figure 8 and Table S5).

RR activation is usually described in a "population shift" framework.^{4,57,58} This framework emphasizes the existence of an active state population in equilibrium with the inactive state, in the absence of ligand binding or covalent modification. It views activation as a shift of the equilibrium to favor the active state, by ligand binding or covalent modification. The presence of an active form without phosphorylation and a shift of the equilibrium between the inactive and the active form have been demonstrated for the RR NtrC.⁵⁹ Our results support the "population shift" framework for RR activation and offer a reasonable estimation of the active population in the absence of phosphorylation and Mg²⁺ binding. However, our results also indicate that while the conformation of the inactive form (macrostate 23) is the same as the crystal structure of the inactive form of RR468, as measured by the $\Delta Q_M^{\beta 3\alpha 3}$ and $\Delta Q_M^{\alpha 4}$ scales (Figure 3 and Table S2), the conformation of the "active" form (macrostate 16) of the unphosphorylated protein is not identical to the crystal structure of the activated protein as mimicked by its complex with beryllium trifluoride and Mg²⁺. It is evident that phosphorylation and Mg²⁺ binding result in conformational adjustments. RR activation is best described by a combination of the "population shift" and the "induced fit"^{60–64} framework. The former describes the larger conformational changes from the inactive to an active-like state, which occurs in the absence of phosphorylation and Mg²⁺ binding, and the latter describes the smaller conformational adjustments caused by phosphorylation and Mg²⁺ binding.

In summary, we have offered in this study a comprehensive description of the activation of a signaling protein. It includes thermodynamic information such as populations of the various states in the conformational transition, kinetic information such as transition pathways, transition probabilities, and reactive fluxes, and activation barriers, as well as structural information such as the formation of transient hydrophobic and hydrogen bond interactions. RR468 represents a large number of RRs in two-component systems, as it has a typical RR fold (a PDBeFold⁶⁵ search found ~150 RR structures with an RMSD within 1.8 Å for C α atoms). It will be of interest to simulate other RRs using the strategy we employed in this study to see whether these RRs follow a similar mechanism of activation. Our simulation results can also be used to formulate hypotheses and test them by experiments. For example, the existence of the active-like conformation in the absence of phosphorylation and Mg²⁺ binding and its population can be

determined by NMR. The roles of the transient interactions in the RR activation can be investigated by site-directed mutagenesis in combination with NMR and biochemical analysis as demonstrated for NtrC.⁵⁷

ASSOCIATED CONTENT

S Supporting Information

Figure S1 shows the implied time scale versus lag time, τ . Figure S2 has information on validation of the MSM. Table S1 provides distances between residues that were used to calculate $\Delta Q_M^{\beta 3\alpha 3}$ and $\Delta Q_M^{\alpha 4}$ parameters. Table S2 lists the population, number of microstates, committer (splitting) probabilities, and $\Delta Q_{\beta 3\alpha 3}^M$, $\Delta Q_{\alpha 4}^M$ parameter values for the 24 macrostates. Table S3 provides transition pathways and reactive fluxes between macrostate 23 (inactive-like) and 16 (active-like). Table S4 summarizes the pairwise residue interaction changes in the conformational transition. Table S5 summarizes the hydrogen bonding changes in the conformational transition. Tables S6 and S7 present data on mean first passage times between important macrostates. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b02582.

AUTHOR INFORMATION

Corresponding Authors

*(H.Y.) Address: Department of Biochemistry and Molecular Biology, Michigan State University, E. Lansing, MI 48224-1322. E-mail: yan@msu.edu. Phone: 517-353-5282.

*(R.I.C.) Address: Department of Chemistry, Michigan State University, E. Lansing, MI 48224-1322. E-mail: cukier@chemistry.msu.edu. Phone: 517-355-9715 × 263.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was in part supported by the National Institutes of Health Grant R01GM084402 and National Science Foundation Grant CHE-0957201. The authors would like to thank Dr. Kaillathe Padmanabhan at the Macromolecular Computer Facility and Dr. Dirk Colbry at the High Performance Computer Center at Michigan State University for technical support.

REFERENCES

- (1) Henzler-Wildman, K. A.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450*, 964–972.
- (2) Karplus, M.; Kuriyan, J. Molecular Dynamics and Protein Function. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6679–85.
- (3) Kern, D.; Zuiderweg, E. R. P. The Role of Dynamics in Allosteric Regulation. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748–757.
- (4) Ma, L.; Cui, Q. Activation Mechanism of a Signaling Protein at Atomic Resolution from Advanced Computations. *J. Am. Chem. Soc.* **2007**, *129*, 10261–10268.
- (5) Smock, R. G.; Giersch, L. M. Sending Signals Dynamically. *Science* **2009**, *324*, 198–203.
- (6) Grant, B. J.; Gorfe, A. A.; McCammon, J. A. Large Conformational Changes in Proteins: Signaling and Other Functions. *Curr. Opin. Struct. Biol.* **2010**, *20*, 142–147.
- (7) Jung, K.; Fried, L.; Behr, S.; Heermann, R. Histidine Kinases and Response Regulators in Networks. *Curr. Opin. Microbiol.* **2012**, *15*, 118–124.
- (8) Wuichet, K.; Cantwell, B. J.; Zhulin, I. B. Evolution and Phylogenetic Distribution of Two-Component Signal Transduction Systems. *Curr. Opin. Microbiol.* **2010**, *13*, 219–25.

- (9) Bourret, R. B. Receiver Domain Structure and Function in Response Regulator Proteins. *Curr. Opin. Microbiol.* **2010**, *13*, 142–149.
- (10) Formaneck, M. S.; Ma, L.; Cui, Q. Reconciling the “Old” and “New” Views of Protein Allostery: A Molecular Simulation Study of Chemotaxis Y Protein (CheY). *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 846–867.
- (11) Hu, X. H.; Wang, Y. M. Molecular Dynamic Simulations of the N-Terminal Receiver Domain of NtrC Reveal Intrinsic Conformational Flexibility in the Inactive State. *J. Biomol. Struct. Dyn.* **2006**, *23*, 509–517.
- (12) Khalili, M.; Wales, D. J. Pathways for Conformational Change in Nitrogen Regulatory Protein C from Discrete Path Sampling. *J. Phys. Chem. B* **2008**, *112*, 2456–2465.
- (13) Lei, M.; Velos, J.; Gardino, A.; Kivenson, A.; Karplus, M.; Kern, D. Segmented Transition Pathway of the Signaling Protein Nitrogen Regulatory Protein C. *J. Mol. Biol.* **2009**, *392*, 823–836.
- (14) Casino, P.; Rubio, V.; Marina, A. The Mechanism of Signal Transduction by Two-Component Systems. *Curr. Opin. Struct. Biol.* **2010**, *20*, 763–771.
- (15) Casino, P.; Rubio, V.; Marina, A. Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell* **2009**, *139*, 325–336.
- (16) Banerjee, R.; Yan, H. G.; Cukier, R. I. Conformational Transition of Response Regulator RR468 in a Two-Component System Signal Transduction Process. *J. Phys. Chem. B* **2014**, *118*, 4727–4742.
- (17) Noe, F.; Fischer, S. Transition Networks for Modeling the Kinetics of Conformational Change in Macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–62.
- (18) Bowman, G. R.; Huang, X.; Pande, V. S. Network Models for Molecular Kinetics and Their Initial Applications to Human Health. *Cell Res.* **2010**, *20*, 622–630.
- (19) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. Molecular Simulation of Ab Initio Protein Folding for a Millisecond Folder NTL9(1–39). *J. Am. Chem. Soc.* **2010**, *132*, 1526–1528.
- (20) Bowman, G. R.; Voelz, V. A.; Pande, V. S. Atomistic Folding Simulations of the Five-Helix Bundle Protein Lambda(6–85). *J. Am. Chem. Soc.* **2011**, *133*, 664–667.
- (21) Lane, T. J.; Bowman, G. R.; Beauchamp, K.; Voelz, V. A.; Pande, V. S. Markov State Model Reveals Folding and Functional Dynamics in Ultra-Long MD Trajectories. *J. Am. Chem. Soc.* **2011**, *133*, 18413–18419.
- (22) Voelz, V. A.; Jager, M.; Yao, S. H.; Chen, Y. J.; Zhu, L.; Waldauer, S. A.; Bowman, G. R.; Friedrichs, M.; Bakajin, O.; Lapidus, L. J.; Weiss, S.; Pande, V. S. Slow Unfolded-State Structuring in Acyl-CoA Binding Protein Folding Revealed by Simulation and Experiment. *J. Am. Chem. Soc.* **2012**, *134*, 12565–12577.
- (23) Dickson, A.; Brooks, C. L. Native States of Fast-Folding Proteins Are Kinetic Traps. *J. Am. Chem. Soc.* **2013**, *135*, 4729–4734.
- (24) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011–19016.
- (25) Bowman, G. R.; Pande, V. S. Protein Folded States Are Kinetic Hubs. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10890–10895.
- (26) Beauchamp, K. A.; Ensign, D. L.; Das, R.; Pande, V. S. Quantitative Comparison of Villin Headpiece Subdomain Simulations and Triplet-Triplet Energy Transfer Experiments. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 12734–12739.
- (27) Deng, N. J.; Zheng, W. H.; Gallicchio, E.; Levy, R. M. Insights into the Dynamics of HIV-1 Protease: A Kinetic Network Model Constructed from Atomistic Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 9387–9394.
- (28) Yang, S.; Banavali, N. K.; Roux, B. Mapping the Conformational Transition in Src Activation by Cumulating the Information from Multiple Molecular Dynamics Trajectories. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3776–81.
- (29) Bowman, G. R.; Geissler, P. L. Equilibrium Fluctuations of a Single Folded Protein Reveal a Multitude of Potential Cryptic Allosteric Sites. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11681–11686.
- (30) Beauchamp, K. A.; McGibbon, R.; Lin, Y. S.; Pande, V. S. Simple Few-State Models Reveal Hidden Complexity in Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17807–17813.
- (31) Kelley, N. W.; Vishal, V.; Krafft, G. A.; Pande, V. S. Simulating Oligomerization at Experimental Concentrations and Long Time-scales: A Markov State Model Approach. *J. Chem. Phys.* **2008**, *129*, 214707.
- (32) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10184–10189.
- (33) Da, L. T.; Wang, D.; Huang, X. H. Dynamics of Pyrophosphate Ion Release and Its Coupled Trigger Loop Motion from Closed to Open State in RNA Polymerase II. *J. Am. Chem. Soc.* **2012**, *134*, 2399–2406.
- (34) Silva, D.-A.; Weiss, D. R.; Avila, F. P.; Da, L.-T.; Levitt, M.; Wang, D.; Huang, X. Millisecond Dynamics of RNA Polymerase II Translocation at Atomic Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7665–7670.
- (35) Seeber, M.; Felline, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Cafisch, A.; Fanelli, F. Wordom: A User-Friendly Program for the Analysis of Molecular Structures, Trajectories, and Free Energy Surfaces. *J. Comput. Chem.* **2011**, *32*, 1183–1194.
- (36) Case, D. A.; Darden, T. A.; III, T. E. C.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *Amber 10*; University of California: San Francisco, CA, 2008.
- (37) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (38) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (39) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134*, 174105.
- (40) Vanden-Eijnden, E. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annu. Rev. Phys. Chem.* **2010**, *61*, 391–420.
- (41) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. MSMBuild2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* **2011**, *7*, 3412–3419.
- (42) Deufhard, P.; Weber, M. Robust Perron Cluster Analysis in Conformation Dynamics. *Linear Algebra Appl.* **2005**, *398*, 161–184.
- (43) Roe, D. R.; Cheatham, T. E. PTraj and CPPtraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
- (44) Hägggi, P.; Talkner, P.; Borkovec, M. Reaction-Rate Theory: Fifty Years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251–341.
- (45) Berezhkovskii, A.; Hummer, G.; Szabo, A. Reactive Flux and Folding Pathways in Network Models of Coarse-Grained Protein Dynamics. *J. Chem. Phys.* **2009**, *130*, 205102.
- (46) Jiang, X. W.; Zhong, A. H.; Chen, C. J.; Huang, Y. Z.; Xiao, Y. Network Approach to Identify the Folding Transition States of Peptides and Proteins. *Phys. Rev. E* **2012**, *051901*.
- (47) Brinda, K. V.; Vishveshwara, S. A Network Representation of Protein Structures: Implications for Protein Stability. *Biophys. J.* **2005**, *89*, 4159–4170.
- (48) Kanna, N.; Vishveshwara, S. Identification of Side-Chain Clusters in Protein Structures by a Graph Spectral Method. *J. Mol. Biol.* **1999**, *292*, 441–464.

- (49) Bowman, G. R.; Huang, X. H.; Pande, V. S. Using Generalized Ensemble Simulations and Markov State Models to Identify Conformational States. *Methods* **2009**, *49*, 197–201.
- (50) Banerjee, R.; Cukier, R. I. Transition Paths of Met-Enkephalin from Markov State Modeling of a Molecular Dynamics Trajectory. *J. Phys. Chem. B* **2014**, *118*, 2883–95.
- (51) van Kampen, N. *Stochastic Processes in Physics and Chemistry*; Elsevier: Amsterdam, 1981.
- (52) Li, M.; Duan, M.; Fan, J.; Han, L.; Huo, S. Graph Representation of Protein Free Energy Landscape. *J. Chem. Phys.* **2013**, *139*, 185101.
- (53) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 19011–19016.
- (54) Qiao, Q.; Bowman, G. R.; Huang, X. Dynamics of an Intrinsically Disordered Protein Reveal Metastable Conformations That Potentially Seed Aggregation. *J. Am. Chem. Soc.* **2013**, *135*, 16092–16101.
- (55) Bharucha-Reid, A. *Elements of the Theory of Markov Processes and Their Applications*; McGraw-Hill: New York, 1960.
- (56) Singhal, N.; Pande, V. S. Error Analysis and Efficient Sampling in Markovian State Models for Molecular Dynamics. *J. Chem. Phys.* **2005**, *123*, 204909.
- (57) Gardino, A. K.; Villali, J.; Kivenson, A.; Lei, M.; Liu, C.; Steinidel, P.; Eisenmesser, E. Z.; Labeikovsky, W.; Wolf-Watz, M.; Clarkson, M. W.; Kern, D. Transient Non-Native Hydrogen Bonds Promote Activation of a Signaling Protein. *Cell* **2009**, *139*, 1109–1118.
- (58) Stock, A. M.; Guhaniyogi, J. A New Perspective on Response Regulator Activation. *J. Bacteriol.* **2006**, *188*, 7328–7330.
- (59) Volkman, B. F.; Lipson, D.; Wemmer, D. E.; Kern, D. Two-State Allosteric Behavior in a Single-Domain Signaling Protein. *Science* **2001**, *291*, 2429–2433.
- (60) Csermely, P.; Palotai, R.; Nussinov, R. Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. *Trends Biochem. Sci.* **2010**, *35*, 539–546.
- (61) Hatzakis, N. S. Single Molecule Insights on Conformational Selection and Induced Fit Mechanism. *Biophys. Chem.* **2014**, *186*, 46–54.
- (62) Koshland, D. E., Jr. The Key-Lock Theory and the Induced Fit Theory. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 2375–2378.
- (63) Okazaki, K.-I.; Takada, S. Dynamic Energy Landscape View of Coupled Binding and Protein Conformational Change: Induced-Fit Versus Population-Shift Mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11182–11187.
- (64) Spyros, F.; BidonChanal, A.; Barril, X.; Javier Luque, F. Protein Flexibility and Ligand Recognition: Challenges for Molecular Modeling. *Curr. Top. Med. Chem.* **2011**, *11*, 192–210.
- (65) Krissinel, E.; Henrick, K. Secondary-Structure Matching (SSM), a New Tool for Fast Protein Structure Alignment in Three Dimensions. *Acta Crystallogr. D: Biol. Crystallogr.* **2004**, *60*, 2256–2268.