

The Atomistic Mechanism of Conformational Transition of Adenylate Kinase Investigated by Lorentzian Structure-Based Potential

Juyong Lee,^{†,‡} Keehyoung Joo,^{§,||} Bernard R. Brooks,[‡] and Jooyoung Lee^{*,†,§}

[†]School of Computational Sciences, Korea Institute for Advanced Study, Dongdaemun-gu, Seoul 130-722, Korea

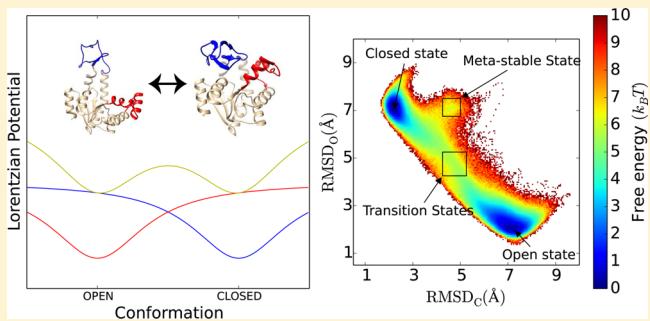
[‡]Laboratory of Computational Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20852, United States

[§]Center for In Silico Protein Science, Korea Institute for Advanced Study, Dongdaemun-gu, Seoul 130-722, Korea

^{||}Center for Advanced Computation, Korea Institute for Advanced Study, Dongdaemun-gu, Seoul 130-722, Korea

ABSTRACT: We present a new all-atom structure-based method to study protein conformational transitions using Lorentzian attractive interactions based on native structures. The variability of each native contact is estimated based on evolutionary information using a machine learning method. To test the validity of this approach, we have investigated the conformational transition of adenylate kinase (ADK). The intrinsic boundedness of the Lorentzian attractive interactions facilitated frequent conformational transitions, and consequently we were able to observe more than 1000 structural interconversions between the open and closed states of ADK out of a total of 6 μ s MD simulations. ADK has three domains:

the nucleoside monophosphate (NMP) binding domain, the LID-domain, and the CORE domain, which catalyze the interconversion between ATP and ADP. We identified two transition states: a more frequent LID-closed-NMP-open (TS1) state and a less frequent LID-open-NMP-closed (TS2) state. The transition was found to be symmetric in both directions via TS1. We also obtained an off-pathway metastable state that was previously observed with physics-based all-atom simulations but not with coarse-grained models. In the metastable state, the LID domain was slightly twisted and formed contacts with the NMP domain. Our model correctly identified a total of 14 out of the top 16 residues with highest fluctuation by NMR experiment, thus showing excellent agreement with experimental NMR relaxation data and overwhelmingly better results than existing models.



INTRODUCTION

To carry out their functions, most proteins undergo controlled conformational transitions; examples include movements of motor proteins,^{1,2} allosteric regulations,³ selective filtering of membrane channels,⁴ and enzymatic reactions.⁵ Identifying the atomistic mechanism of the conformational transition(s) of a protein is essential in understanding its functional mechanism, which can lead to a better or altered engineering of the protein. As a model system to study the molecular mechanism of a large-scale conformational change of a protein, adenylate kinase (ADK) has been widely studied experimentally as well as theoretically.^{5–7} ADK is a phosphotransferase enzyme found in various organisms from all three kingdoms of life and plays an essential role in cellular energy homeostasis by catalyzing the interconversion between ADP and ATP. The enzyme consists of three domains, an ATP-binding (LID) domain (residues 122–159), an AMP-binding (NMP) domain (residues 30–60), and a CORE domain (residues 1–29, 61–121, and 160–214). ADK is experimentally established to undergo a large conformational change between its open and closed states (see Figure 1). The LID and NMP domains are highly dynamic, and the structural rearrangement of these two domains was shown to be the rate-determining step of the

enzymatic reaction.⁸ It is also known that ADK can adopt both open and closed conformations even in the absence of ligands and that many local movements accompany the global conformational transition.^{5,9} However, critical unsettled issues remain for the molecular mechanism of ADK: What is the temporal order of domain movement? Are there intermediate states? Which residues are critical residues for the conformational change?

All-atom molecular dynamics (MD) simulations have played an important role in studying the dynamics of proteins. In general, however, the time scale of a typical molecular dynamics simulation of a protein is limited to the microsecond range, which is not long enough to simulate most enzymatic reactions, typically occurring in the range of milliseconds to seconds. To overcome this time scale mismatch, coarse-grained (CG) structure-based models based on the folding funnel hypothesis have been widely used.^{10,11} In these approaches, the protein structure is modeled as a chain of beads representing the trace of $C\alpha$ atoms. In addition, attractive interactions are imposed between $C\alpha$ atom pairs in spatial proximity in the native

Received: March 20, 2015

Published: June 10, 2015

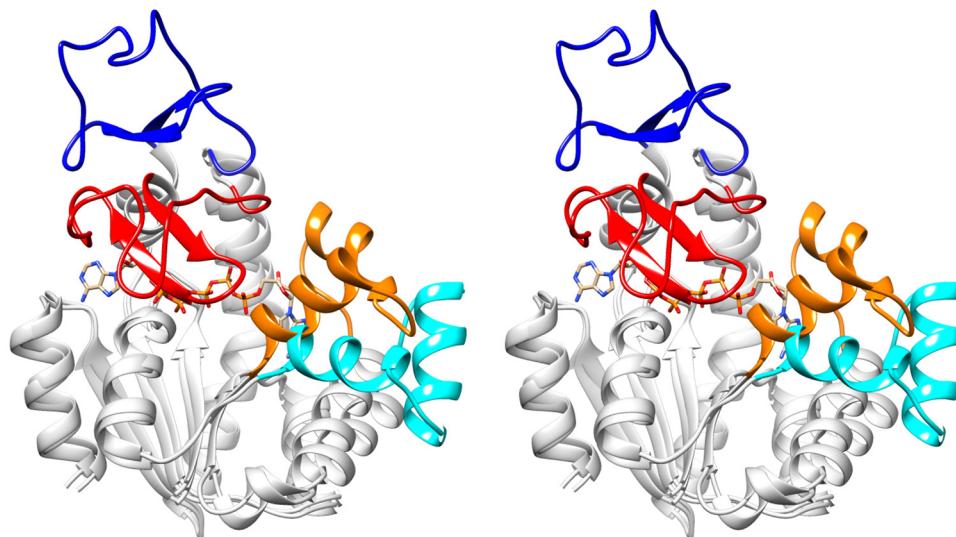


Figure 1. Open and closed structures of the adenylate kinase. Open (PDB ID: 4AKE) and closed (PDB ID: 1AKE) structures of the *E. coli* adenylate kinase are shown in stereoview. The stable and conserved CORE domain (residues M1–I29, D61–V121, and Q160–G214) of two structures are shown in light gray. The LID (residues G122–D159) and NMP domains (residues S30–T60) of the closed form are colored in red and orange, respectively. The LID and NMP domains of the open form are colored in blue and cyan, respectively. The inhibitor ApSA, which mimics ATP and AMP, is located in a pocket formed by the LID and NMP domains.

structure. The energy minimum distance of each attractive interaction is set according to the native structure. For ADK, many computational studies using the CG structure-based model have been performed to explain the origin and pathway of its large conformational change.^{12–20} A comprehensive review on the computational studies on ADK can be found elsewhere.²¹

With two X-ray crystal structures of ADK (open and closed), two single-well attractive interactions from these conformations should be integrated to simulate the transition between them. Maragakis and Karplus proposed a mixed elastic network model to find the minimum-energy path connecting the two conformations.¹² In the study by Best et al.,²² a Boltzmann weighted mixing scheme was proposed to study the conformational transition of an arc repressor mutant between helical and extended structures. These methods and their variants have been successfully applied to describe the conformational changes of large proteins.^{18,19,23–27} However, these CG structure-based attractive interaction mixing approaches have three potential technical problems. First, when energy minima of harmonic attractive interactions are separated from each other by a large distance, it may lead to an excessively large energy barrier that can deter the sampling of a large-scale conformational change. Second, the mixing of harmonic or Lennard-Jones (LJ) like attractive potentials excessively penalizes any structures that deviate from the linear interpolation between two input structures. This will excessively prevent the sampling of any structures far away from the input structures, including potential (unknown) intermediate states. Third, the CG model may not be suitable for research in drug discovery or protein design where atomic details as well as side chains may play important roles.

To address these issues, we developed an all-atom structure-based method for the study of the structural transition of proteins by utilizing Lorentzian attractive terms. The variation of each Lorentzian term is estimated by evolutionary information using a machine learning method.²⁸ The efficiency and adequacy of the proposed model is assessed by simulating

the conformational transition of ADK where there is good experimental data to test our predictions. With the current approach, we could observe over 1000 occurrences of spontaneous structural interconversion between the open and closed states of ADK in a simulation of 6 μ s. A comparison of B-factors from the simulation and crystal structures shows that our model is more consistent with experiments than previously reported. In addition, the backbone fluctuation estimated by the current study almost exactly matches with the NMR relaxation experiment.²⁹

We identified two transition states: a LID-closed and NMP-open (TS1) state and a LID-open and NMP-closed (TS2) state. Additionally, a metastable state was also observed, which is similar to the state observed in previous all-atom physics-based simulations.^{30,31} TS1 is observed to be more favorable by 1.5 kcal/mol than TS2 from the free energy calculation. Transition state analysis illustrates that the structural change of the NMP domain is accomplished by localized structural variation involved with only a small number of backbone dihedral angles at hinge regions. Contrarily, the structural change of the LID domain is observed to occur via weak unfolding of hinge regions. These results demonstrate that the proposed approach can be successfully applied to the study of various conformational transitions of macromolecules to provide new insights.

RESULTS

Two Lorentzian Attractive Interactions Provide a Double-Well Potential with a Bounded Energy Barrier. We devised a new structure-based potential to stabilize the native contacts of a protein by using the Lorentzian function, defined as $f(x) = cx^2/(x^2 + a^2)$. Near its minimum, the function behaves similarly to the harmonic potential, but it is bounded away from the minimum. A differentiable double-well potential can be generated by simply adding two Lorentzian functions located at two separate minimum values. A comparison of the two-state potential generated by the Lorentzian potential and

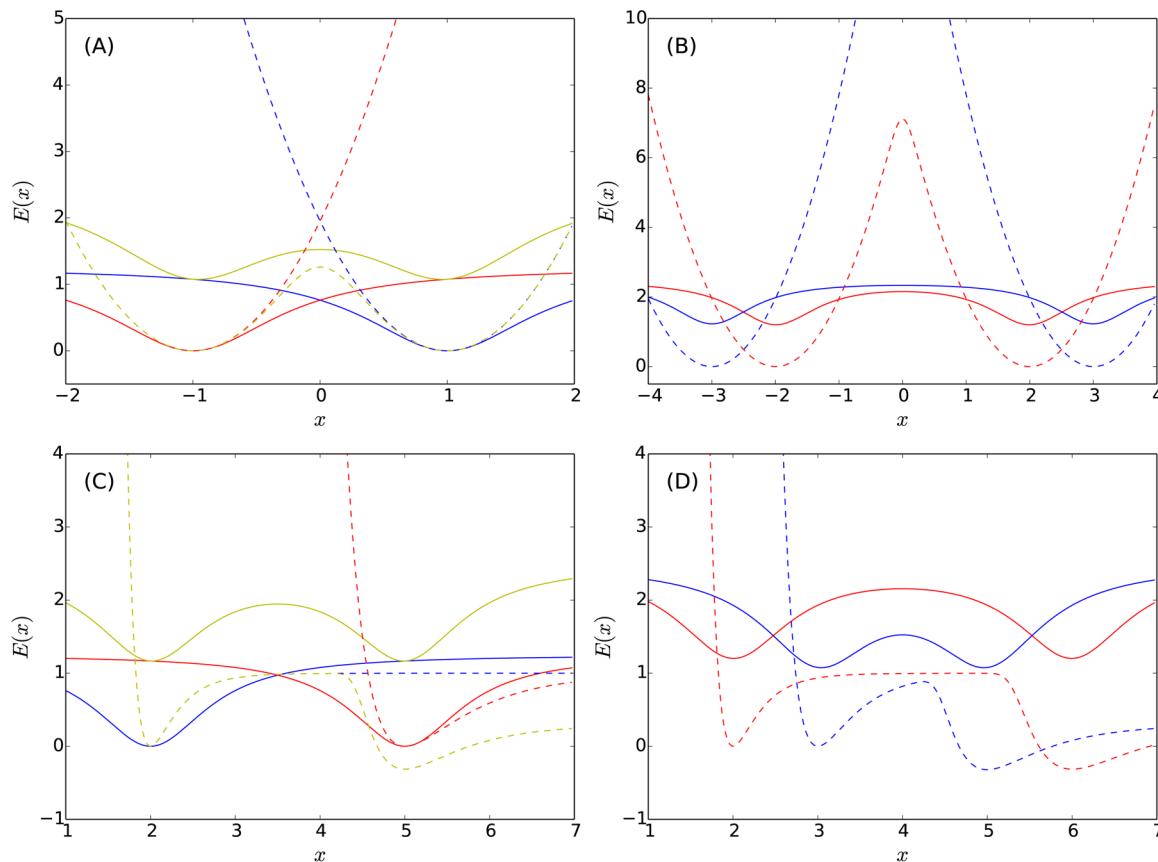


Figure 2. Two-state structure-based Hamiltonians with Lorentzian and harmonic attractive interactions. A comparison of a Lorentzian-based two-state attractive interaction (bold lines) to a mixed harmonic and a mixed Lennard-Jones-like 12–10 attractive interaction by the Boltzmann mixing scheme²² (dashed lines) is shown. The functional form of the Lorentzian attractive interaction has the minimum at x_0 , and the width of σ is $L(x; x_0, \sigma) = \sigma^{-1}(x - x_0)^2 / ((x - x_0)^2 + \sigma^2)$. The functional form of the 12–10 attractive interaction which has the minimum at x_0 is $E(x; x_0) = 5.0 (x_0/x)^{12} - 6.0 (x_0/x)^{10}$. When $|x - x_0| \ll 1$, L is reduced to $\sigma^{-3}(x - x_0)^2$. (A) Two Lorentzian functions, $L(x; \pm 1, 1.0)$ (red and blue solid lines), and two harmonic functions, $H(x; \pm 1, 1.0)$ (dotted lines), as well as their mixed forms (yellow lines) are shown. (B) Lorentzian and mixed harmonic potentials with two sets of minima, ± 2 and ± 3 , are shown. Note that the barrier height of the mixed harmonic function can increase dramatically as the separation between two minima increases while that of the mixed Lorentzian function is intrinsically bounded regardless of the separation. (C) Two Lorentzian functions, $L(x; x_0 = 2, 1.0)$ and $L(x; x_0 = 5, 1.0)$ (red and blue solid lines), and two 12–10 functions, $E(x; x_0 = 2)$ and $E(x; x_0 = 5)$ (dotted lines), as well as their mixed forms (yellow lines) are shown. (D) Lorentzian and mixed 12–10 attractive interactions with two sets of minima, $(2.0, 6.0)$ and $(3.0, 5.0)$, are shown. Note that the minima of the mixed Lorentzian functions are not adjusted to zero to emphasize that it is obtained by simple addition.

the Boltzmann weighted mixing of two harmonic/LJ-like potentials is shown in Figure 2.

One technical problem with the harmonic potential is that the barrier height of the mixed potential depends on the distance separation between minima, which can lead to a large energy barrier when the two distances are significantly separated. In contrast, the boundedness of the Lorentzian function guarantees a small energy barrier for the mixed potential, as shown in Figure 2B. The mixed Lorentzian functions (solid lines in Figure 2B) with two minima at ± 2 and ± 3 are of energy barriers of similar heights, while the energy barrier of a mixed harmonic function (dashed lines in Figure 2B) increases rapidly as the distance separation increases. When LJ-like potentials are mixed, the shape of the mixed potential is not symmetric, which can introduce an unwanted bias favoring of one of the original end states. In addition, for a given residue pair, an LJ-like potential would discourage the residue pair from searching a shorter distance than the minimum of native values (see Figure 2D). This limitation may exclude the sampling of important conformations that deviate from the linear interpolation between two input structures.

The intrinsic variability of the potential, the width and the depth of the Lorentzian potential (σ in eq 2), is determined by the machine learning method, Sigma-RF, which considers evolutionary information and was developed for accurate template-based protein structure modeling.²⁸ To test the effect of predicted intrinsic variability, we calculated the correlation coefficient between the B-factor from PDB and that from the MD simulation of ADK (open state) using either predicted σ or constant σ (see Table 1). We observe that the open state ensemble sampled with the predicted σ reproduces the

Table 1. Pearson Correlation Coefficient between the Native B-Factors and the Ones Measured from MD Simulation Trajectories

potential	variability	open (all-atom)	open (only $C\alpha$)
Lorentzian	predicted	0.783	0.847
	constant	0.591	0.576
Harmonic	predicted	0.760	0.810
	constant	0.692	0.699

experimental B-factor more accurately than that with the constant σ . The correlation coefficient of all heavy atoms using the predicted σ values is 0.783, while that using the constant σ is 0.591. When only $C\alpha$ atoms are considered, the corresponding values are 0.847 and 0.576. We also performed the same comparison using the harmonic potential model. Using the harmonic potential, the predicted σ also led to a better agreement with experimental results than the constant σ . Note that the value of the B-factor correlation obtained by the harmonic model with the constant σ (~ 0.7 in Table 1) is similar to previously reported values, 0.680¹³ and 0.787.³²

Lorentzian Structure-Based Model Facilitates Numerous Conformational Transitions. With our new model, multiple conformational changes of ADK were observed without using any biasing forces.^{33,34} RMSD plots of two 20 ns MD simulations are shown in Figure 3. RMSD values

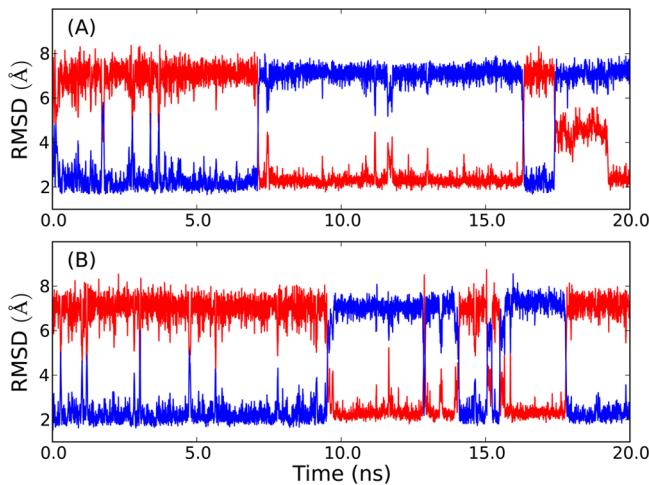


Figure 3. Spontaneous transitions between the open (OS) and the closed (CS) states of the adenylyl kinase are observed along with a metastable state (MS). Two typical patterns of RMSD variations along two 20 ns MD trajectories are shown, where RMSD values measured from the open (closed) structure of ADK is shown in red (blue). Multiple spontaneous transitions are observed in B. In A, in addition to the multiple transitions between OS and CS, MSs whose RMSD values are about 4.5 Å from the closed state and about 7.0 Å from the open state are observed from 17.5 to 19 ns.

calculated against the crystal structures of open (4AKE) and closed (1AKE) states are respectively shown in red and blue. When ADK is in the open state, RMSD values measured from the open structure ($RMSD_O$) fluctuate near 2 Å, while those from the closed one ($RMSD_C$) fluctuate near 7 Å. When ADK switches its conformation to the other reference state, corresponding RMSD values switch accordingly. In most of Figure 3 (except the 17.5–19 ns time window from Figure 3B), mirror conversion between red and blue data is observed, strongly indicating that the structural variation is mainly interconversion between two reference states along interpolated structures between them. When we define a complete structural transition by switching between the open state ($RMSD_O < 3$ Å and $RMSD_C > 6$ Å) and the closed state ($RMSD_O > 6$ Å and $RMSD_C < 3$ Å), a total of 1096 structural transitions are observed during the collection of 6 μ s MD simulations. It should be noted that, experimentally, it was observed that the transition occurs in the millisecond time scale.³⁵ In the MD simulation, the transition frequency is

accelerated in two ways: by considering only the native interaction as in the other structure-based models and by lowering the energy barrier between the two native states utilizing the Lorentzian mixing scheme.

In Figure 3, three and eight complete conformational transitions are respectively observed for A and B. In addition, many incomplete transitions and a slow transition are also observed. We observe numerous peaks reaching in the range of 4–5 Å, but failing to go through complete transitions, which correspond to the cases where the protein arrives near the transition state but returns to the prior state. We note that our MD models thus allow one to quantify the frequency of successful vs frustrated transition attempts. The example of a slow transition is observed in the time window of 17.5–19 ns from Figure 3A. Judging from the significant amount of residence time, this slow transition appears to go through a qualitatively different pathway from the other fast transitions, indicating a potential metastable intermediate state.

Energetics and Kinetics of the Transition. We obtained the free energy profile of the ADK model using two sets of reaction coordinates: (1) two RMSD values measured from two reference crystal structures (4AKE and 1AKE), $RMSD_O$ and $RMSD_C$ (Figure 4A), and (2) the fractions of native contacts of two reference structures (Figure 4B). With the 2D RMSD reaction coordinate, two stable basins corresponding to the open (OS) and closed (CS) states are observed. A transition state (TS) connecting OS and CS ($4.25 \text{ \AA} < RMSD_C < 5.25$ and $4.25 \text{ \AA} < RMSD_O < 5.25 \text{ \AA}$) and an off-pathway off-diagonal metastable (MS) state near CS ($4.25 \text{ \AA} < RMSD_C < 5.0$ and $6.75 \text{ \AA} < RMSD_O < 7.5 \text{ \AA}$) are also identified. OS is the most stable state with the highest observed population, and the relative free energy values of the other three states (CS, TS, and MS) are respectively 0.3, 5.4, and 6.4 kcal/mol. MS is about 1.0 kcal/mol higher in its free energy than TS, showing that TS is approximately 5 times more frequently visited than MS.

We also generated a free energy profile using the fractions of native contacts from two reference structures, which is often used to describe the progress of protein folding (Figure 4B). Native contacts are defined as residue pairs whose $C\beta$ atoms are within 8.0 Å in the crystal structures of OS and CS. We observed two free energy minima corresponding to OS and CS, which are separated by a barrier. The barrier size (~ 3 kcal/mol) is smaller than that of the 2D-RMSD free energy profile because conformations on transition pathways are projected onto a smaller region on the free energy surface due to relatively small changes in the numbers of native contacts. In addition, MS is not identified when the fractions of native contacts are used. During the conformational transitions of ADK, the three subdomains do not undergo significant intradomain structural changes, and mostly interdomain contacts form and break.

To obtain a quantitative picture of transitions between the states defined above, we did a clustering analysis by using a network model³⁶ (see Materials and Methods), which revealed that TS can be further divided into two conformational states: the LID-closed and NMP-open (TS1) state and the LID-open and NMP-closed (TS2) state, while MS remained as a single conformational state (Figure 5). The population ratio of TS1 to TS2 indicates that TS1 is more favorable by 1.5 kcal/mol than TS2. In MS, the LID domain is slightly distorted and rotated toward the NMP domain along the rotation axis given by the seventh α -helix. Based on the clustering result, we counted all transitions among five states: OS and CS, TS1 and TS2, and

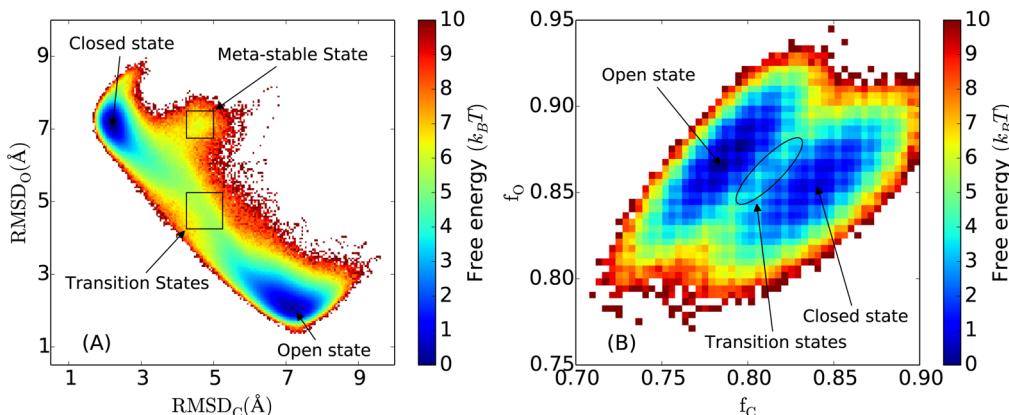


Figure 4. Free energy profiles of the conformational transition of ADK. Free energy profiles of conformational transition of ADK are illustrated using two sets of reaction coordinates: (A) RMSD values from OS and CS and (B) the fractions of native contacts of OS and CS. In A, the conformations in the ranges of $4.25 \text{ \AA} < \text{RMSD}_O < 5.25$ and $4.25 \text{ \AA} < \text{RMSD}_C < 5.25 \text{ \AA}$ are considered as the transition state ensemble. The conformations in the range of $6.75 \text{ \AA} < \text{RMSD}_O < 7.5$ and $4.25 \text{ \AA} < \text{RMSD}_C < 5.0 \text{ \AA}$ are considered as the metastable state ensemble. The barrier height is about 5.4 kcal/mol, and the free energy of the metastable state is 6.4 kcal/mol.

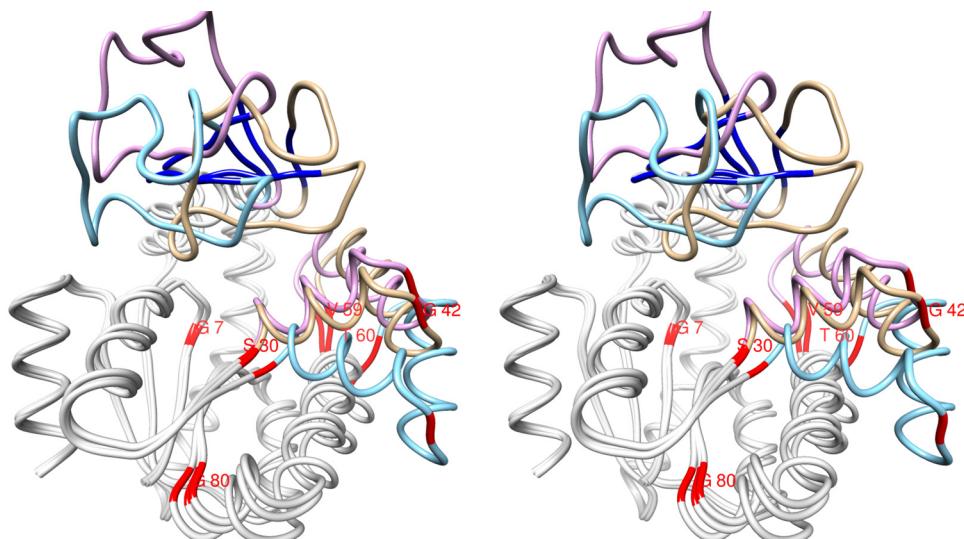


Figure 5. Structures of transition and metastable states. Average structures of two transition and metastable ensembles are shown in stereoview. The dominant transition state, TS1, is the LID closed and NMP opened state (cyan), and it occupies 92.4% population density of the transition state ensemble. The less populated transition state, TS2, corresponds to the LID opened and NMP closed state (purple). In the metastable state, the distances between LID and NMP domains are closer than the other states, and the LID domain is slightly distorted (gold). The residues that are important for the conformational changes of NMP (G7, S30, G42, V59, T60, and G80) and LID (R123-H126 and T154-R156) domains are highlighted in red and blue, respectively.

MS (Figure 6). We observe that the probability of TS1/TS2 to be followed by OS is roughly similar to that by CS, which is consistent with the transition state theory suggesting that TS1 and TS2 are properly identified as transition states. More evenly bifurcating TS1/TS2 is estimated to be located slightly toward CS/OS.

One may argue that the observation of similar forward and reverse transitions itself does not guarantee that TS1 and TS2 are true transition states because transitions to TS1 and TS2 may have come from OS or CS, but they may all return to their initial states, which does not form a true transition pathway. Similar cases are found in the early stage of Figure 3B. The conformation reached the transition state region and returned to its initial state a few times until a true transition occurred just before 10 ns. However, we observed that only a small fraction of 300 independent simulations remained at their initial states without a true transition. Thus, the TS1 and TS2 ensembles

must include a substantial number of conformations that are actually on or near true transition pathways.

To confirm that the conformations in the TS1 and TS2 ensembles contain true transition states, we performed shooting simulations.^{37,38} The shooting simulation results show that many conformations in the TS1 and TS2 ensembles are actually on the transition pathway (Supplementary Figure 1). About 35% and 64% of conformations in the TS1 and TS2 ensembles, respectively, have nonzero probabilities of being on a transition path. Although a small number of conformations are exactly at the true transition states where the probabilities of reaching OS and CS are equal, it demonstrates that significant numbers of conformations in the TS1 and TS2 ensembles are reactive (i.e., they form transition pathways between two states).³⁸ In summary, the identified TS1 and TS2 states are good approximations of the true transition states of the conformational transition of ADK.

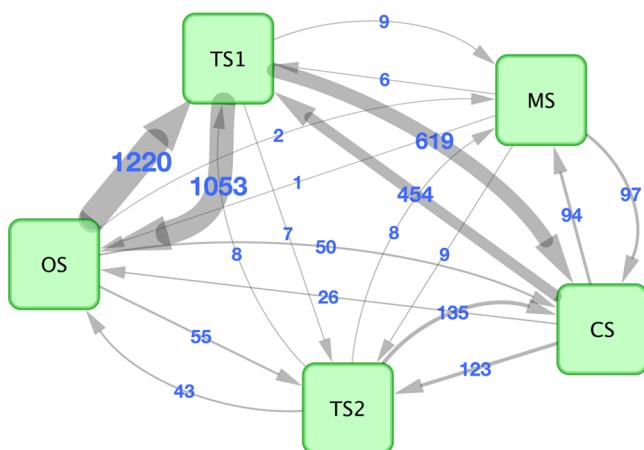


Figure 6. Quantitative map of conformational transition of adenylate kinase. The numbers of transitions between the native, transition, and metastable states observed during 6 μ s simulations are shown.

Two major pathways of ADK are identified for structural interconversion between OS and CS, the one through TS1 (LID-closed and NMP-open) being the major one and the other one through TS2 (LID-open and NMP-closed) the minor one. Starting from OS, 92.1% of the cases choose TS1 and the rest (7.9%) TS2. Starting from CS, about 13.5% of the cases transit to MS, most of which eventually return back to CS, indicating that MS is the off-pathway intermediate. Eventually, starting from CS, 79.3% of the cases move to OS via TS1 and the rest (20.7%) via TS2. Rare transitions among TS1, TS2, and MS are also observed.

Structural Characteristics of Transition States. To assess the validity of our model, we compared the extent of backbone dihedral angle fluctuations observed in our computations with available experimental data. Using the NMR relaxation technique, amplitudes of backbone amide bond fluctuations on the picosecond to nanosecond time scale

have been measured²⁹ for ADK. In the experiment, more dynamic residues were represented by smaller order parameter values of S^2 . In Figure 7, we show the NMR data (red circles) of $1 - S^2$ measured at the residue level (see Figure S13(c) in Henzler-Wildman et al.²⁹), where higher values correspond to more fluctuating residues. For direct comparison, we calculated S^2 values (S_{sim}^2) from the obtained trajectories using the Lipari-Szabo formula:³⁹

$$S_{\text{sim}}^2 = \langle [3(\hat{\mu}_{\text{NH}}(0)\hat{\mu}_{\text{NH}}(t))^2 - 1]/2 \rangle \quad (1)$$

where $\hat{\mu}_{\text{NH}}(t)$ is the backbone N–H vector at time t and t is set to 2 ns in this study. We also measured two additional quantities: the sum of ϕ/ψ fluctuations $\text{VAR} = (\langle \phi^2 \rangle - \langle \phi \rangle^2)^{1/2} + (\langle \psi^2 \rangle - \langle \psi \rangle^2)^{1/2}$ and the sum of ϕ/ψ deviations of the average TS1 state from two native states $\text{DEV} = |\langle \phi_{\text{TS1}} \rangle - \langle \phi_{\text{OS}} \rangle| + |\langle \phi_{\text{TS1}} \rangle - \langle \phi_{\text{CS}} \rangle| + |\langle \psi_{\text{TS1}} \rangle - \langle \psi_{\text{OS}} \rangle| + |\langle \psi_{\text{TS1}} \rangle - \langle \psi_{\text{CS}} \rangle|$.

It should be noted that the Lipari–Szabo formula provides rigorous grounds for direct comparison between our MD simulation results and the NMR fluctuation data while the VAR and DEV were additionally used to compare our simulation results to existing CG-model studies where changes of pseudobackbone angles were measured due to the lack of atomic details^{13,18,19} (Figure 7). Most of the experimentally identified dynamic residues are well represented in our calculations. The correlation coefficients between the NMR data²⁹ and $1 - S_{\text{sim}}^2$, VAR, and DEV are 0.426, 0.381, and 0.459, respectively, representing a reasonably good agreement between experimental and computational results. The fluctuating stability of the CORE domain (residues 1–29, 61–121, and 160–214) shows particularly good agreement.

In Table 2, the top 16 fluctuating residues measured by NMR are listed in descending order. For comparison, from our simulation, the same number of topmost fluctuating residues based on $1 - S_{\text{sim}}^2$, VAR, and DEV are also listed. We observe that, within three-residue separation, the NMR residues cover

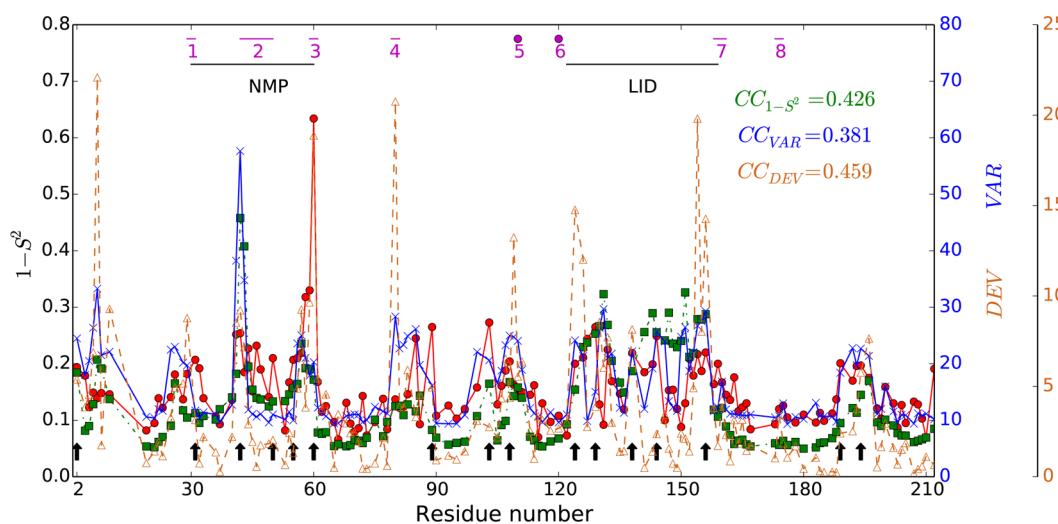


Figure 7. Comparison of backbone fluctuation measured by NMR and the simulation. $1 - S^2$ (shown in red circle) corresponds to the amplitude of backbone amide bond fluctuation on the picosecond to nanosecond time scale measured by an NMR relaxation experiment,²⁹ ranging from 0 (completely rigid) to 1 (free tumbling). For comparison, the $1 - S_{\text{sim}}^2$ values from the MD simulation (green squares; eq 1), the sum of ϕ/ψ angle fluctuations ($\text{VAR} = (\langle \phi^2 \rangle - \langle \phi \rangle^2)^{1/2} + (\langle \psi^2 \rangle - \langle \psi \rangle^2)^{1/2}$), and the sum of ϕ/ψ angle deviations of the average TS1 from two average native states ($\text{DEV} = |\langle \phi_{\text{TS1}} \rangle - \langle \phi_{\text{OS}} \rangle| + |\langle \phi_{\text{TS1}} \rangle - \langle \phi_{\text{CS}} \rangle| + |\langle \psi_{\text{TS1}} \rangle - \langle \psi_{\text{OS}} \rangle| + |\langle \psi_{\text{TS1}} \rangle - \langle \psi_{\text{CS}} \rangle|$) are shown. The residues without NMR data are not shown. Black arrows indicate the top 16 most fluctuating residues determined by NMR.²⁹ Eight hinge regions defined by simple structural comparison between the open and closed structures of ADK are indicated by numbered purple dots (separate residues) and lines.

Table 2. A Comparison of the Top 16 Highly Fluctuating Residues Identified from the Experiment and the Simulation^a

type	residues
NMR experiment	60, 103, 89, 129, 42, 144, 156, 138, 50, 31, 55, 108, 189, 124, 194, 2
$1 - S_{\text{sim}}^2$	42, 151, 131, 147, 143, 156, 127, 57, 102, 139, 7, 2, 196, 108, 135, 26
VAR	42, 7, 131, 156, 80, 102, 151, 85, 144, 30, 57, 108, 2, 195, 124, 26
DEV	30, 7, 80, 154, 60, 124, 109, 42, 84, 130, 138, 196, 12, 102, 2, 89
hinge regions ^b	29–30, 42–50, 59–61, 79–81, 110, 120, 158–161, 173–177

^aThe residues are listed in the descending order of fluctuation. For the NMR data, more dynamic residues represented by smaller S^2 values less than 0.81 are shown. The residues adjacent to the peaks less than four residues apart are considered to be redundant and not shown. For the simulation results, the residues, which agree with the experiment within the residue number difference of 3, are shown in bold. $S_{\text{sim}}^2 = \langle [3(\hat{\mu}_{\text{NH}}(0) \hat{\mu}_{\text{NH}}(t))^2 - 1]/2 \rangle$ with $t = 2$ ns. VAR = $\langle (\phi^2) \rangle - \langle \phi \rangle^2)^{1/2} + \langle (\psi^2) \rangle - \langle \psi \rangle^2)^{1/2}$. DEV = $|\langle \phi \rangle_{\text{TS1}} - \langle \phi \rangle_{\text{OSl}}| + |\langle \phi \rangle_{\text{TS1}} - \langle \phi \rangle_{\text{CSl}}| + |\langle \psi \rangle_{\text{TS1}} - \langle \psi \rangle_{\text{OSl}}| + |\langle \psi \rangle_{\text{TS1}} - \langle \psi \rangle_{\text{CSl}}|$.²⁹ ^bThe hinge regions were identified based on the variation of pseudo-angles and pseudo-dihedral angles between OS and CS.²⁹

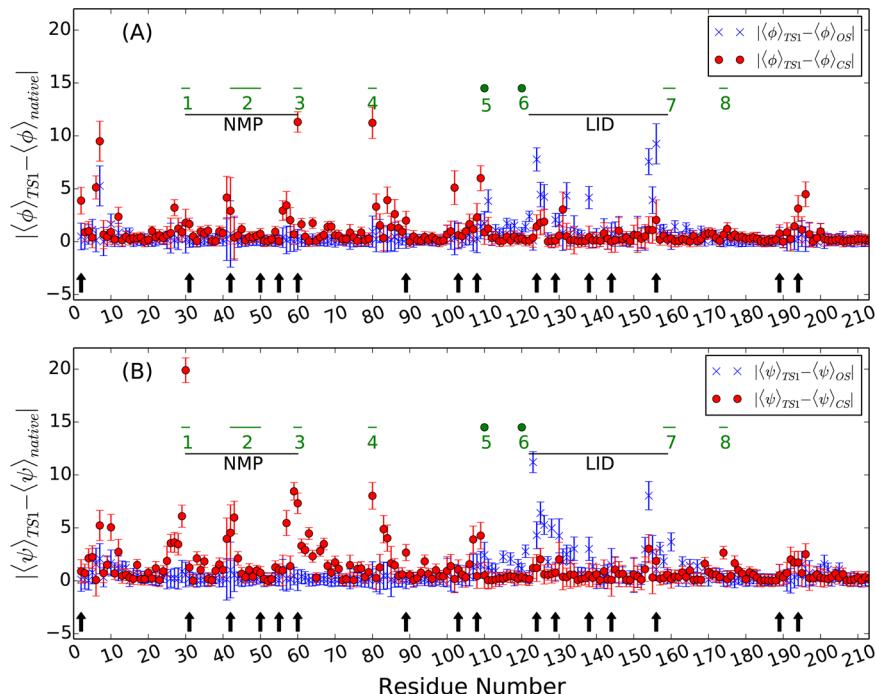


Figure 8. Deviations of ϕ and ψ angles of TS1 from OS and CS. Deviations of ϕ (A) and ψ (B) angle ensemble averages of TS1 from OS and CS are shown. Deviations from OS and CS are shown in blue and red, respectively. Eight hinge regions defined by the variation of pseudo angles and pseudo dihedral angles are indicated as green numbered dots and lines.²⁹ The LID domain covers the residue number from 122 to 159 and the NMP domain from 30 to 60. Black arrows indicate the top 16 highly fluctuating residues determined by the NMR experiment.²⁹

13 of $1 - S_{\text{sim}}^2$, 11 of VAR and 12 of DEV residues (shown as the bold-face font in Table 2). On the other hand, $1 - S_{\text{sim}}^2$, VAR, and DEV residues cover 12 NMR residues, respectively. A total of 14 out of the top 16 most fluctuating residues agree between experimental results and our model. Out of two disagreeing residues (50 and 189), 189 is two residue separations away from 191, which is ranked 19th according to VAR. This result represents an overwhelmingly improved explanation of NMR relaxation data than was possible with existing CG models. We note that all top eight of the most fluctuating residues from NMR (60, 103, 89, 129, 42, 144, 156, 138) are correctly identified by our computation within one residue separation (see Table 2).

By way of comparison, Whitford et al.¹³ used a harmonic pseudoangle potential, a pseudodihedral-angle potential based on native values, and a native-contact-distance structure-based CG model using a 12–10 potential. They reported that a number of local regions (60–70 and 120–125 with high strain energy; 10–20, 30–35, 80–90, and 170–180 with some strain energy) should be unfolded for the structural transition of

ADK. However, the agreement between their results and the NMR data²⁹ is poor, covering only four residues (T60, T89, T31, and R124) out of 16 from Table 2.

Similarly, Daily et al.¹⁸ used a Boltzmann weighted mixing of two CG models^{22,40} where the native structure information is used for pseudo-bond-length, pseudo-bond-angle, and long-range residue contacts. They also used native-structure-independent statistical pseudo-dihedral-angle potentials.⁴⁰ They identified highly fluctuating residues by measuring p_{folded} , the fraction of a pseudo-dihedral angle in the same rotamer state with the native state (within 60°), and four out of eight dynamic hinge regions were correctly identified (hinges 2, 3, 4, and 7). In Henzler-Wildman et al.,²⁹ hinge regions were defined based on the large pseudoangle and pseudodihedral-angle differences between the open and closed structures of ADK, and they are listed in Table 2. Although it is plausible to assume that the above-defined hinge regions agree with highly backbone-fluctuating residues, they should not be considered to dictate the definite identity of such residues. When we identified the top 20 (four more than the 16 used in Table 2)

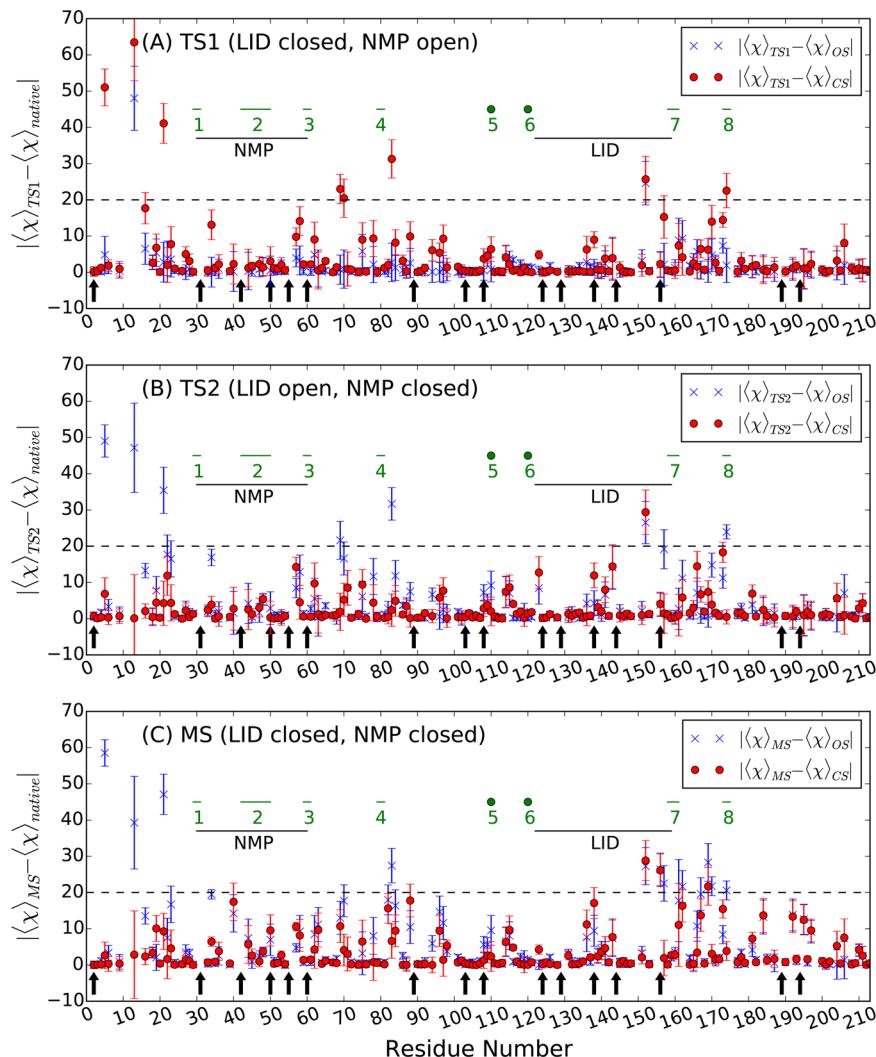


Figure 9. Deviations of χ_1 angles of TS1, TS2, and MS from OS and CS. Deviations of χ_1 angle ensemble averages of TS1 (A), TS2 (B), and MS (C) from OS and CS are shown. Deviations from the closed and open states are shown in red and blue, respectively.

fluctuating residues from Figure 3a of Daily et al.¹⁸ ($p_{\text{folded}} < 0.85$), they covered seven residues (42, 189, 156, 194, 144, 60, and 138) within a three-residue separation, which is an improvement over the result of Whitford et al.¹³ but still covers only 50% of the current study.

What can we learn about the nature of the transitions? From the structural comparison of the dominant transition state, TS1, with two native states, we observe that the movement of the NMP domain is mostly involved with the backbone dihedral angles of five separate residues (S30, G42, S43, T60, and G80), which are all in the previously defined hinge regions of the NMP domain (see Figure 8). These residues either highly fluctuate (large errorbars in Figure 8) or deviate from the native states (large y -axis values in Figure 8) in the TS1 ensemble. On the other hand, the movement of the LID domain is mostly involved with consecutive residues at the head (R123–H126) and the tail (T154–R156) of the LID domain. In order to reach MS, we find that L107–F109 should be perturbed in addition to the head and tail LID domain residues. L107–F109 were found to be flexible (see Figure 7) in the NMR experiment,²⁹ and the hinge 5 residue (D110) is located just on its right. It should be noted that most of the backbone dihedral angles are perturbed within the boundary of the open

and closed state values in our simulation, which indicates that the opening and closing of ADK occurs without significant unfolding of the molecule. The only exception is observed with the significant structural disruption of ADK (in both TS1 and TS2) in the β -strand region of R123–H126.

A comparison of the χ_1 angles of TS1 and TS2 with those of native states shows that most side-chain conformations of residues in the CORE domain depend on NMP's movement (Figure 9). When TS1 (LID closed and NMP open) is visited, the χ_1 angles of L5, M21, K69, E70, L83, and M174 deviate more than 20° from CS (Figure 9A) but more or less neutral to the OS direction. However, the χ_1 angles of K13 and E152 deviate from both CS and OS. We note that no side chains are observed to deviate significantly only from OS. When TS1 is formed—LID is closed and NMP is open—the χ_1 angles of L5, M21, K69, E70, L83, and M174 are perturbed more than 20° from those of the closed state (Figure 9A). When compared with TS2 (LID open and NMP closed), the trends are reversed, except that the χ_1 angle of K13 deviates only from OS (Figure 9B). These results imply that the movement of NMP is more correlated with the fine structure of the CORE domain than that of LID. To adopt the MS conformation, the side chains of the residues located around the tail of the LID domains (R156,

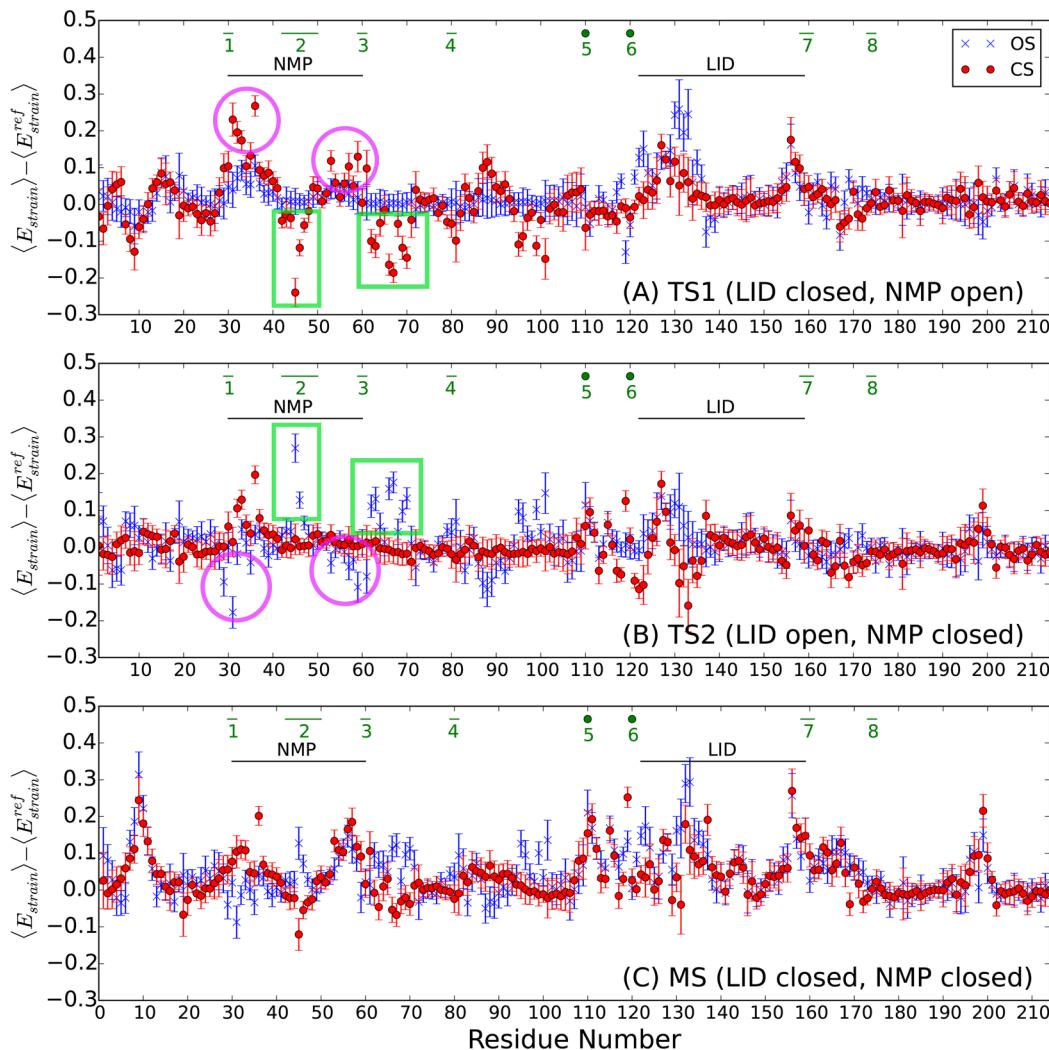


Figure 10. Comparison of strain energy of the transition/metastable and native state ensembles. Differences between the strain energy of the transition, metastable, and native state ensembles decomposed on a per residue basis are shown. Deviations from the closed and open states are shown in red and blue, respectively. The plots A and B correspond to TS1 and TS2, respectively. The plot C corresponds to the MS. In the top panel, two purple circles indicate residues from hinges 1 and 3, which are energetically excited when the NMP domain is open. Energetically stabilized regions (hinge 2 and residues 62–69) are shown inside two green boxes. In the second panel, the situation is reversed as indicated by circles and boxes.

K157, E162, V169, and E170) are perturbed in addition to the residues identified above (Figure 9C).

Energy Decomposition. We calculated the strain energy of ADK at the residue level and compared the results between TS1, TS2, and MS and OS and CS. We observe that the conformational-change mechanisms of LID and NMP are qualitatively different from each other (Figure 10). For TS1, we observe that the boundary parts (residues 120–135 and 150–160) of the LID domain are of higher strain energy than CS as well as OS (see Figure 10A). A similar pattern is also observed in TS2, although to a lesser degree.

In contrast, the NMP domain shows different strain energy characteristics of hinge regions. When the NMP domain is open/closed, hinge 1 (residues 30–36) and hinge 3 (residues 59–60) are energetically excited/stabilized (shown by red circles). However, the hinge 2 (residues 45–47) region and residues 62–69 are energetically stabilized/excited (shown by green boxes). Overall, these opposite strain energies cancel out each other, which makes the net energy contribution of the NMP domain in forming the transition states almost zero.

These results suggest that the major energy barrier of the conformational change of ADK is the destabilization of the LID domain. For MS, additional strain energies of residues 10, 196, and 197 are observed compared to TS, indicating that these residues play important roles in forming MS.

DISCUSSION

Using static protein structures to guide dynamic models of protein function is a challenge for the field. Here, we propose a new method by using Lorentzian structure-based functions and predicted intrinsic variability of each native contact. We first tested the validity of the proposed method by measuring the B-factor of the open state ADK and comparing it with the measured B-factor from PDB. In this study, the correlation coefficient between the simulated values and the measured ones from PDB is respectively 0.783 and 0.847 for all heavy atoms and $\text{C}\alpha$ atoms, which is significantly improved when compared to existing works.^{13,32} The predicted variability of the native atom-pair distance was shown to improve the results.

The good agreement between the backbone perturbation of our model and the NMR relaxation data²⁹ shows that the current Lorentzian structure-based potential model reproduces the dynamics of ADK at the residue level far more accurately than existing CG models. This improved agreement can be attributed to the fact that, in our model, (a) no native-structure-dependent dihedral angle potentials are used, (b) all-atom details are kept, and (c) the variability of the Lorentzian-type attractive interaction is properly estimated from a protein structure database.

Previously, Whitford et al.^{13,14} used a simple CG model where all interactions were specific to ADK's crystal structures by utilizing pseudo-bonds, pseudo-angles, pseudo-dihedral angles, and residue pair distance attractive interactions based on native values. They proposed that the cracking of high strain regions was necessary for the structural interconversion between the open and closed states of ADK. However, the high strain regions identified in their work did not agree with the NMR experiment.²⁹

Using a CG model with a native-structure-independent statistical dihedral potential,⁴⁰ Daily et al.¹⁸ correctly identified four out of eight dynamic hinge regions without observing cracking, which is to some extent consistent with our results. However, we are actually able to do much better: in the current study, almost all the experimentally determined dynamic regions of ADK are properly reproduced, and only a small amount of ϕ/ψ angle perturbation was observed in the TS ensemble. Keeping all atom details enabled us to compare the MD simulation results with the NMR relaxation experiment data. In addition, the absence of cracking (supported by us and Daily et al.¹⁸) suggests that the cracking represented by large deviation of pseudodihedral angles from both native states may arise from the highly native-specific potential, which imposes an unnecessarily high barrier for the conformational transition.¹⁸

The variability (the width and the depth) of the Lorentzian interaction is inferred from the corresponding structural information on homologous proteins using a machine learning approach.²⁸ Proper variability prediction may have played an important role in the close agreement of the current work with the experiment. The rationale of our approach is that if a residue pair in ADK and its corresponding pairs in other homologous proteins show a large variation in the inter-residue distance, it may indicate that the pair has evolved to be flexible, and a relatively large value is assigned to σ in eq 2. In contrast, if inter-residue distances show little variation, the pair may have evolved to be rigid for an evolutionary reason, and in this case, we assign a relatively small value to σ . Previously, this approach was successfully applied to the protein homology modeling and was shown to improve the model quality.^{28,41,42} Our results suggest that inferring this variability from homologous protein information may also provide insights into the dynamics of a protein.

The TS ensemble obtained by our simulation shows that only a small amount of ϕ/ψ perturbation (less than 10° for most residues) is necessary for the large-scale conformational change of ADK. For the movement of the NMP domain, perturbations of backbone dihedral angles are highly localized around four hinge regions, S30, G42–S43, T60, and G80. On the other hand, for the movement of the LID domain, ϕ/ψ angles from the head (R124–G130) and the tail (T155–Q160) residues are perturbed slightly by only about 5° (see Figure 8). These small perturbations of ϕ/ψ angles can not be adequately represented by a CG model, which could have led to the

incomprehensive conclusion that only a fraction of hinge regions are involved in the structural change.¹⁸ All these results suggest that the current approach can serve as a powerful tool to describe more realistic conformational transitions of proteins with atomic details.

The intrinsic nature of the boundedness of the Lorentzian function enabled us to observe more than 1000 times of conformational interconversions between the open and closed states of ADK in 6 μs without introducing any biasing forces, despite the highly rugged energy landscape involved in the all-atom representation of the molecule. It is essential to sample many independent transition events to obtain statistically meaningful results on the dynamics of a large molecule such as ADK. Most existing all-atom studies of ADK^{7,31,43,44} could observe only a few structural transitions (far less than 1000), from which it is difficult to conclude the existence of an alternative pathway or the relative dominance between competing pathways. For this reason, existing studies focused on utilizing coarse-grained models to sample numerous structural transitions of ADK.^{12–20} However, in these studies, due to the lack of atomic details, it was difficult to carry out analysis on the role of individual residues, which, we believe, led to a qualitatively different picture of the transition. For example, the current ϕ/ψ angle analysis shows that the opening/closing of the NMP domain can be achieved by a small number of localized backbone perturbations at hinge regions, which differs from the existing study¹³ where significant NMP domain unfolding was reported.

In this study, we found that the LID-closed–NMP-open state (TS1) is the dominant transition state of the conformational transition of apo-ADK in both directions. The LID-first mechanism of the open-to-closed transition has been supported by many other studies of both CG and all-atom models. However, there is a dispute on the mechanism of the close-to-open transition.^{12,13,16–18,43,45} Interestingly, the symmetric transition pathway via TS1 (supported by the current study) has been reported in the studies that used double-well potentials that have bounded energy barriers and are weakly dependent on native values,^{17,20,32} which can be considered to be philosophically similar to the current Lorentzian-based double-well potential. On the other hand, other multistate structure-based models whose energy functions are heavily dependent on the native values^{13,14,17} seem to prefer an asymmetric pathway. These results indicate that the dependence of an energy function to native values may play an important role in determining the dominant transition pathway. Therefore, when multiple transition states exist, special attention should be paid to investigate the dominance of pathways from a structure-based study. A more detailed discussion on pathway symmetry can be found elsewhere.⁴⁶

We found that the conformational transitions of the LID and NMP domains occur in a qualitatively different manner. The comparison of backbone dihedral angles shows that the flexibility of the LID domain is determined by the perturbation of several consecutive residues at the boundary parts of the domain, which consist of mainly arginine residues. Although most backbone dihedral angles deviate less than 15° from either native state, the loss of β-strand structure of residues 123–126 can be considered as “weak” partial unfolding of the region. On the other hand, the perturbation of the NMP domain is more localized and controlled by a few small-sized flexible residues of S30, G42, S43, T60, and G80, which adopt the coil secondary structure. This characteristic difference may arise from the

manner each domain interacts with adenine nucleotides. The LID domain can be easily refolded with the help of strong electrostatic interactions between two arginine residues (R123 and R156) and the phosphate group of ATP or ADP. This mechanism is similar to the “population shift followed by induced fit” mechanism discussed in earlier studies.^{47–50} On the contrary, the NMP domain interacts with the adenosine of AMP or ADP, which is weaker than the charge–charge interaction of the LID domain mentioned above. Therefore, the unfolding of the NMP domain may lead to the higher chance of local unfolding, which can make ADK more error prone.

One of the major differences between our results and previous CG model-based studies^{13,51} is that significant local unfolding was not observed during the conformational transition in our model. This difference is probably caused by the global mixing of two energy functions performed in previous studies while a local residue-wise mixing was performed in our study. Our results show that the NMP domain movement occurs via highly localized perturbations of a couple of ϕ/ψ angles in the hinge regions while the CG model predicted a large-scale unfolding of the region. The validity of either picture for the conformational transition of ADK can be tested by the hydrogen–deuterium (H/D) exchange measurements⁵² and/or unbiased all-atom MD simulations.^{53,54} Recently, the conformational transition of the kinase domain of the epidermal growth factor receptor was investigated with long-time MD simulations and H/D exchange measurements.⁵⁵

We identified a metastable state, where the LID domain is slightly distorted and rotated toward the NMP domain. The distortion of the LID domain is supported by the existence of other crystal structures of homologous ADK (1AK2 and 2AK2),⁵⁶ where the orientation of the LID domain is distorted in a similar fashion. Physics-based all-atom MD studies^{7,17,30,34,57} also observed a similar distortion of the LID domain, while it has not yet been reported in any CG studies. Particularly, the metastable state identified in this study is in close agreement with the conformation observed in a principal component analysis of a high temperature all-atom MD simulation.³⁰ This result suggests that MD simulations based on a conventional structure-based double-well potential might have failed to capture potential intermediate states due to an excessively high intrinsic penalty for conformations away from both reference structures. The current study demonstrates that the proposed potential based on the simple addition of Lorentzian structure-based attractive interaction terms and all-atom representation of the protein can bridge the gap between the all-atom physics-based model and the CG model.

CONCLUSIONS

In this work, we present a new all-atom structure-based method to study protein conformational transitions using Lorentzian attractive interactions based on native structures. The variability of each native contact is estimated based on evolutionary information using a machine learning method.²⁸ We applied the method to simulate the conformational transition of ADK. When compared with previous studies, the B-factor of the open state ADK was better reproduced by the current study with a correlation coefficient of 0.783 and 0.847 for all heavy atoms and C α atoms, respectively. The usage of predicted variability of the native atom-pair distance was shown to improve the results.

The intrinsic boundedness of the Lorentzian attractive interaction facilitated frequent conformational transitions, and

consequently we observed more than 1000 structural interconversions between the open and closed states of ADK out of a total of 6 μ s MD simulations. We identified two transitions states, the more dominant LID-closed–NMP-open (TS1) state and the less dominant LID-open–NMP-closed (TS2) state. The transition is shown to be symmetric in both directions via TS1. Our all-atom model shows that a rather small amount of ϕ/ψ angle perturbation can lead to the conformational interconversion of ADK. Transition mechanisms of LID and NMP are shown to be qualitatively different from each other. LID changes via perturbation of several consecutive residues located near its head and tail regions, while NMP changes via highly localized perturbations of separate hinge residues. Our results are consistent with the “population shift followed by induced fit” mechanism. We also observed the off-pathway metastable state that was previously observed with physics-based all-atom simulations but not with coarse-grained models, where LID is slightly twisted and forming contacts with NMP. Most importantly, a total of 14 out of the top 16 most fluctuating residues identified by the NMR relaxation experiment were correctly identified by our all-atom model, which is overwhelmingly better than existing results.

MATERIALS AND METHODS

Structure-Based Modeling with Lorentzian Attractive Potential. Here, we introduce a Lorentzian attractive potential to replace the conventional harmonic potential for structure-based modeling. For a given reference structure K , the structure-based term from the atom-pair distance r_{ij} is defined as

$$l^K(r_{ij}) = \frac{1}{\sigma_{ij}^K} \frac{(r_{ij} - r_{ij}^K)^2}{(r_{ij} - r_{ij}^K)^2 + (\sigma_{ij}^K)^2} \quad (2)$$

where r_{ij}^K is the distance between i and j in K and σ_{ij}^K is the predicted distance variation of r_{ij} near r_{ij}^K . The value of σ_{ij}^K determines the width and depth of the Lorentzian term in eq 2, and it is inferred from the statistics of known protein structures aligned using sequence information. To generate a double-well potential for the conformational transition of ADK, simple addition of two Lorentzian terms is carried out: $L(r_{ij}) = l^{K=1}(r_{ij}) + l^{K=2}(r_{ij})$, where reference structures 1 and 2 are taken from two PDB structures of ADK, 4AKE and 1AKE (Figure 1). The resulting Lorentzian potential $L(r_{ij})$ is bounded regardless of the value of r_{ij} in contrast to the potential generated by mixing two harmonic terms.^{12,15,22}

In our model, the stereochemistry and short-range Lennard-Jones (LJ) potentials are adopted from MODELLER^{58,59} and CHARMM,⁶⁰ and attractive pairwise interactions generated following the standard MODELLER^{58,59} procedure are represented by the Lorentzian term in eq 2. The complete Hamiltonian of our model is defined as follows:

$$\begin{aligned}
H = & E_{\text{stereochemistry}} + w_{\text{LJ}} E_{\text{LJ}} + w_{\text{contact}} E_{\text{contact}} \\
= & \sum_{\text{bond lengths}} K_b(b - b^0)^2 + \sum_{\text{bond angles}} K_\theta(\theta - \theta^0)^2 \\
& + \sum_{\text{dihedral angles}} K_\phi(1 + \cos(n\phi - \delta)) \\
& + \sum_{\text{improper angles}} K_\omega(\omega - \omega^0)^2 - \sum_{\text{residues}} \ln \\
& \sum_{i=1}^m \omega_i \frac{1}{2\pi\sigma_{\phi,i}\sigma_{\psi,i}\sqrt{(1-\rho_i^2)}} \\
& \times \exp \left\{ \frac{1}{(1-\rho_i^2)} \left[\frac{1 - \cos(\phi - \bar{\phi}_i)}{\sigma_{\phi,i}^2} \right. \right. \\
& \left. \left. - \rho_i \frac{\sin(\phi - \bar{\phi}_i)}{\sigma_{\phi,i}} \frac{\sin(\psi - \bar{\psi}_i)}{\sigma_{\psi,i}} + \frac{1 - \cos(\psi - \bar{\psi}_i)}{\sigma_{\psi,i}^2} \right] \right\} \\
& + w_{\text{LJ}} \sum_{\text{nonbonded pairs}} \epsilon_{ij} \left(\left(\frac{r_{ij}^{\min}}{r_{ij}} \right)^{12} - \left(\frac{r_{ij}^{\min}}{r_{ij}} \right)^6 \right) \quad (3) \\
& + w_{\text{contact}} \sum_{K=1,2} \sum_{\text{contacts} \in K} l^K(r_{ij}) \quad (4)
\end{aligned}$$

where the parameters of bond length, bond angle, dihedral/improper angles, and LJ terms are adopted from the CHARMM22 force field.^{61,62} The fifth term of eq 4 is a multiple Gaussian restraint term of the main-chain ϕ/ψ angles adopted from MODELLER that is derived from the statistics of known protein structures,⁵⁸ where m is the number of main-chain conformation classes in the Ramachandran plot, ω is the weight of each conformation class, σ and ρ are the standard deviation and the correlation coefficient between ϕ and ψ , and the bar indicates the average of ϕ and ψ angles. Following MODELLER, native contacts are grouped into four categories based on the pairwise atom types, C α –C α , backbone N–O, main chain–side chain, and side chain–side chain, using the cutoff distances of 14.5, 9.0, 8.0, and 5.0 Å, respectively. LJ interactions are calculated with the cutoff distance of 9 Å. The values of w_{LJ} and w_{contact} terms are tuned to be 0.66 and 0.024, respectively, so that the equilibrium constant between the open and closed states is about 1 as observed experimentally.³⁵

Intrinsic Variability Prediction. In this study, the width and depth of each Lorentzian potential term is estimated by using evolutionary information, which is inferred from the structural variations of known homologous proteins. Recently, we have developed a machine to predict the intrinsic variability of the distance between two atoms in a protein, which led to more accurate prediction than previously estimated by MODELLER.²⁸ Here, we briefly outline the prediction procedure. A precompiled set of 1100 nonredundant crystal structures of proteins, with a sequence identity cutoff of 20%, a crystal resolution cutoff of 1.6 Å, and an R-factor cutoff of 0.25, was adopted from the PISCES server.⁶³ For each protein in the set, its best template along with its alignment is identified by using our in-house fold recognition server, FoldFinder, which has been successfully utilized for protein homology modeling.^{42,64,65} Based on the alignment result, we measured the relative distance variation of all pairs of equivalent atoms within

the cutoff distances defined above. From the collection of these data, we have constructed a machine using the random forest machine-learning method⁶⁶ to estimate the distance variation. The machine takes, as inputs, the sequence alignment information between a target protein and its template protein and the structural information on the template. The list of nine input features for the prediction of intrinsic variability is shown in Table 3. The output is the estimated distance variation between two atoms in the target and in the template.

Table 3. Input Features for Intrinsic Variability Prediction

index	feature ^a
1	$ I - J $
2	$ I - J /N_{\text{res}}$
3	distance between atom i and j
4–6	predicted secondary structure propensity by PSIPRED ⁶⁷
7–9	predicted solvent accessibility by SANN ⁶⁸

^aAtoms i and j are included in residues I and J of the reference protein, respectively.

Clustering Analysis of Transition State Ensemble and Metastable Ensemble. Based on the obtained free energy profile, the transition state (TS) is defined as the conformations in the regions of $4.25 \text{ \AA} < \text{RMSD}_O < 5.25 \text{ \AA}$ and $4.25 \text{ \AA} < \text{RMSD}_C < 5.25 \text{ \AA}$. Similarly, the metastable state (MS) is defined as the conformations in the region of $6.75 \text{ \AA} < \text{RMSD}_O < 7.5 \text{ \AA}$ and $4.5 \text{ \AA} < \text{RMSD}_C < 5.25 \text{ \AA}$. For each state, we carried out a clustering analysis by constructing a network³⁶ among generated conformations as follows. First, 10% of each population is used to generate the network. Each selected conformation constitutes a node, and two nodes are connected if the RMSD value between them is less than 2 Å. Thus, a generated TS network consisted of two disconnected giant-component subnetworks, while only one giant-component subnetwork was identified for MS. From this analysis, two substates of TS, i.e., TS1 and TS2, were identified (see the section Energetics and Kinetics of the Transition).

Molecular Dynamics Simulation. In this study, we used the Tinker molecular modeling software⁶⁹ supplemented with Lorentzian structure-based potentials and a modified CHARMM22 force field⁶¹ without hydrogen atoms to implement the MODELLER^{58,59} system. All simulations were performed at 300 K. A Berendsen thermostat⁷⁰ and Beeman integrator⁷¹ were used to perform constant temperature simulations. The integration time step of 2 fs was used, and snapshots were recorded for each 1 ps interval. For the open structure of ADK, the PDB structure of 4AKE⁷² was taken as the reference structure. For the closed structure of ADK, the PDB structure of 1AKE⁷³ was taken. The energy of the native open structure was minimized by the L-BFGS algorithm⁷⁴ for 500 steps. Subsequently, a 2-ns-long MD simulation was performed to equilibrate the system. Starting from the last snapshot of the equilibration MD, 300 independent MD simulations were performed for 20 ns with separately assigned velocities based on the Gaussian distribution, which amounts to a total of 6 μs MD simulations. From the simulation, we observed a total of 1096 spontaneous transitions between the open and closed structures.

Shooting Simulations. To provide more rigorous grounds that the TS1 and TS2 ensembles actually contain reactive transition state conformations, the probability that a microstate x is on a transition path $P(TP|x)$ was calculated by shooting

simulations.^{37,38} For each conformation \mathbf{x} in the TS1 and TS2 ensembles, 10 independent 500 ps MD simulations were performed with randomly assigned velocities using an Andersen thermostat.⁷⁵ $P(\text{TP}|\mathbf{x})$ is defined by sum of products of splitting probabilities

$$P(\text{TP}|\mathbf{x}) = \phi_{\text{OS}}(\mathbf{x}, \mathbf{p})\phi_{\text{CS}}(\mathbf{x}, -\mathbf{p}) + \phi_{\text{OS}}(\mathbf{x}, -\mathbf{p})\phi_{\text{CS}}(\mathbf{x}, \mathbf{p}) \quad (5)$$

where $\phi_{\text{OS}}(\mathbf{x}, \mathbf{p})$ and $\phi_{\text{CS}}(\mathbf{x}, \mathbf{p})$ are the fractions of trajectories starting from (\mathbf{x}, \mathbf{p}) that reach OS and CS first. Because stochastic MD simulations were performed with randomly assigned velocities, we assume that $\phi_{\text{OS}}(\mathbf{x}, \mathbf{p}) \sim \phi_{\text{OS}}(\mathbf{x}, -\mathbf{p})$ and $P(\text{TP}|\mathbf{x}) \sim 2\phi_{\text{OS}}(\mathbf{x})\phi_{\text{CS}}(\mathbf{x})$. Here, $\phi_{\text{OS}}(\mathbf{x})$ and $\phi_{\text{CS}}(\mathbf{x})$ are the probabilities to reach OS and CS first starting from \mathbf{x} with randomly assigned velocities. If, out of 10 simulations starting from \mathbf{x} , five reached OS first and the other five CS first, $P(\text{TP}|\mathbf{x})$ becomes 0.5, which corresponds to the true transition state. If all 10 trajectories reach OS first or CS first, $P(\text{TP}|\mathbf{x})$ becomes 0.0. Thus, \mathbf{x} with nonzero $P(\text{TP}|\mathbf{x})$ is considered to be reactive (i.e., form transition paths connecting two states).

AUTHOR INFORMATION

Corresponding Author

*E-mail: jlee@kias.re.kr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Masaki Sasai, Kunihiro Kuwajima, Yanwen Tan, and Steven Gross for helpful discussions. We thank Katherine Henzler-Wildman and Dorothee Kern for kindly providing the NMR relaxation data. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2008-0061987). We thank Korea Institute for Advanced Study for providing computing resources (KIAS Center for Advanced Computation Linux Cluster) for this work. We also would like to acknowledge the support from the KISTI Supercomputing Center (KSC-2012-C3-01).

REFERENCES

- (1) Vale, R.; Milligan, R. *Science* **2000**, *288*, 88.
- (2) Carter, A.; Cho, C.; Jin, L.; Vale, R. *Science* **2011**, *331*, 1159.
- (3) Goodey, N.; Benkovic, S. *Nat. Chem. Biol.* **2008**, *4*, 474.
- (4) Perozo, E.; Marien, D.; Cortes; Cuello, L. G. *Science* **1999**, *285*, 73.
- (5) Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hübner, C. G.; Kern, D. *Nature* **2007**, *450*, 838.
- (6) Arora, K.; Brooks, C. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18496.
- (7) Kubitzki, M. B.; de Groot, B. L. *Structure* **2008**, *16*, 1175.
- (8) Wolf-Watz, M.; Thai, V.; Henzler-Wildman, K.; Hadjipavlou, G.; Eisenmesser, E.; Kern, D. *Nat. Struct. Mol. Biol.* **2004**, *11*, 945.
- (9) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964.
- (10) Hyeon, C.; Thirumalai, D. *Nat. Commun.* **2011**, *2*, 487.
- (11) Takada, S. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130.
- (12) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807.
- (13) Whitford, P. C.; Miyashita, O.; Levy, Y.; Onuchic, J. N. *J. Mol. Biol.* **2007**, *366*, 1661.
- (14) Whitford, P. C.; Gosavi, S.; Onuchic, J. N. *J. Biol. Chem.* **2008**, *283*, 2042.
- (15) Chu, J.-W.; Voth, G. A. *Biophys. J.* **2007**, *93*, 3860.
- (16) Peng, C.; Zhang, L.; Head-Gordon, T. *Biophys. J.* **2010**, *98*, 2356.
- (17) Bhatt, D.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2010**, *6*, 3527.
- (18) Daily, M. D.; Phillips, G. N.; Cui, Q. *J. Mol. Biol.* **2010**, *400*, 618.
- (19) Daily, M. D.; Phillips, G. N.; Cui, Q. *PLoS Comput. Biol.* **2011**, *7*, e1002103.
- (20) Wang, Y.; Gan, L.; Wang, E.; Wang, J. *J. Chem. Theory Comput.* **2012**, *9*, 84.
- (21) Seyler, S. L.; Beckstein, O. *Mol. Simul.* **2014**, *40*, 855.
- (22) Best, R.; Chen, Y.; Hummer, G. *Structure* **2005**, *13*, 1755.
- (23) Zhu, F.; Hummer, G. *Biophys. J.* **2009**, *97*, 2456.
- (24) Zheng, W.; Brooks, B. R.; Hummer, G. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 43.
- (25) Zheng, W.; Thirumalai, D. *Biophys. J.* **2009**, *96*, 2128.
- (26) Tekpinar, M.; Zheng, W. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2469.
- (27) Zheng, W.; Auerbach, A. *PLoS Comput. Biol.* **2011**, *7*, e1001046.
- (28) Lee, J.; Lee, K.; Joung, I.; Joo, K.; Brooks, B. R.; Lee, J. *BMC Bioinf.* **2015**, *16*, 94.
- (29) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913.
- (30) Lou, H.; Cukier, R. I. *J. Phys. Chem. B* **2006**, *110*, 12796.
- (31) Adkar, B. V.; Jana, B.; Bagchi, B. *J. Phys. Chem. A* **2010**, *115*, 3691.
- (32) Lu, Q.; Wang, J. *J. Am. Chem. Soc.* **2008**, *130*, 4772.
- (33) Snow, C.; Qi, G.; Hayward, S. *Proteins: Struct., Funct., Bioinf.* **2007**, *67*, 325.
- (34) Beckstein, O.; Denning, E. J.; Perilla, J. R.; Woolf, T. B. *J. Mol. Biol.* **2009**, *394*, 160.
- (35) Hanson, J. A.; Duderstadt, K.; Watkins, L. P.; Bhattacharyya, S.; Brokaw, J.; Chu, J.-W.; Yang, H. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18055.
- (36) Lee, J.; Gross, S. P.; Lee, J. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2012**, *85*, 056702.
- (37) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.
- (38) Hummer, G. *J. Chem. Phys.* **2004**, *120*, 516.
- (39) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546.
- (40) Karanicolas, J.; Brooks, C. L. *Protein Sci.* **2002**, *11*, 2351.
- (41) Thompson, J.; Baker, D. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 2380.
- (42) Joo, K.; Lee, J.; Sim, S.; Lee, S. Y.; Lee, K.; Heo, S.; Lee, I.-H.; Lee, S. J.; Lee, J. *Proteins: Struct., Funct., Bioinf.* **2014**, *82* (Suppl 2), 188.
- (43) Brokaw, J. B.; Chu, J.-W. *Biophys. J.* **2010**, *99*, 3420.
- (44) Song, H. D.; Zhu, F. *PLoS One* **2013**, *8*, e68023.
- (45) Li, W.; Wolynes, P. G.; Takada, S. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3504.
- (46) Bhatt, D.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2011**, *7*, 2520.
- (47) Okazaki, K.-I.; Takada, S. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11182.
- (48) Itoh, K.; Sasai, M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 7775.
- (49) Silva, D.-A.; Bowman, G. R.; Sosa-Peinado, A.; Huang, X. *PLoS Comput. Biol.* **2011**, *7*, e1002054.
- (50) Matsunaga, Y.; Fujisaki, H.; Terada, T.; Furuta, T.; Moritsugu, K.; Kidera, A. *PLoS Comput. Biol.* **2012**, *8*, e1002555.
- (51) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12570.
- (52) Englander, S. W.; Kallenbach, N. R. *Q. Rev. Biophys.* **1983**, *16*, 521.
- (53) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. *Science* **2010**, *330*, 341.
- (54) Hynninen, A.-P.; Crowley, M. F. *J. Comput. Chem.* **2014**, *35*, 406.
- (55) Shan, Y.; Arkhipov, A.; Kim, E. T.; Pan, A. C.; Shaw, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 7270.

- (56) Schlauderer, G.; Schulz, G. *Protein Sci.* **1996**, *5*, 434.
(57) Formoso, E.; Limongelli, V.; Parrinello, M. *Sci. Rep.* **2015**, *5*, 8425.
(58) Fiser, A.; Do, R. K. G.; Šali, A. *Protein Sci.* **2000**, *9*, 1753.
(59) Fiser, A.; Šali, A. *Methods Enzymol.* **2003**, *374*, 461.
(60) Brooks, B. R.; et al. *J. Comput. Chem.* **2009**, *30*, 1545.
(61) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
(62) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86.
(63) Wang, G.; Dunbrack, R. *Bioinformatics* **2003**, *19*, 1589.
(64) Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 114.
(65) Joo, K.; Lee, J.; Lee, S.; Seo, J.; Lee, S.; Lee, J. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 83.
(66) Breiman, L. *Mach. Learn.* **2001**, *45*, 5.
(67) McGuffin, L.; Bryson, K.; Jones, D. *Bioinformatics* **2000**, *16*, 404.
(68) Joo, K.; Lee, S. J.; Lee, J. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 1791.
(69) Ren, P.; Ponder, J. *J. Phys. Chem. B* **2003**, *107*, 5933.
(70) Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684.
(71) Beeman, D. *J. Comput. Phys.* **1976**, *20*, 130.
(72) Müller, C.; Schlauderer, G.; Reinstein, J.; Schulz, G. *Structure* **1996**, *4*, 147.
(73) Müller, C.; Schulz, G. *J. Mol. Biol.* **1992**, *224*, 159.
(74) Zhu, C.; Byrd, R.; Lu, P.; Nocedal, J. *Acm. T. Math. Software* **1997**, *23*, 550.
(75) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.

■ NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on June 24, 2015. Modifications have been made to Table 2 and reference 36. The correct version was published on June 25, 2015.