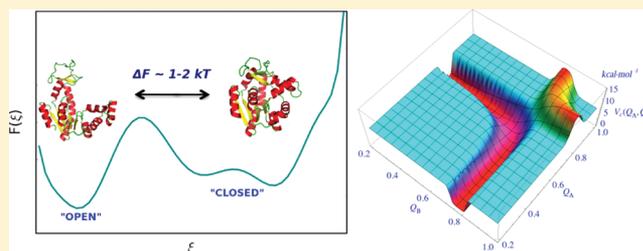


Computing Free Energy of a Large-Scale Allosteric Transition in Adenylate Kinase Using All Atom Explicit Solvent Simulations

Davit A. Potoyan,^{†,‡} Pavel I. Zhuravlev,[†] and Garegin A. Papoian^{*,†,‡,§}[†]Institute for Physical Science and Technology, [‡]Chemical Physics Program, and [§]Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, United States

ABSTRACT: During allosteric motions proteins navigate rugged energy landscapes. Hence, mapping of these multi-dimensional landscapes into lower dimensional manifolds is important for gaining deeper insights into allosteric dynamics. Using a recently developed computational technique, we calculated the free energy difference between the open and closed states of adenylate kinase, an allosteric protein which was extensively studied previously using both experimental and theoretical approaches. Two independent simulations indicate reasonable convergence of the computed free energy profiles.

The numerical value of the open/closed free energy difference is only $1-2 k_B T$, much smaller than some of the prior estimates. We also found that the conformations structurally close to the open form still retain many LID-NMP contacts, suggesting that the conformational basin of the closed form is larger than expected. The latter suggestion may explain the discrepancy in relative populations of open and closed forms of unligated adenylate kinase, observed in NMR and FRET experiments.



INTRODUCTION

It is well-known that energy landscapes underlying protein functional dynamics may be rugged.¹⁻⁸ The protein dynamics on these landscapes involve the interconversion among myriad of conformational states that take place on a spectrum of time scales from nanoseconds to seconds.⁴ The roughness of the landscape is a result of a many-body interaction between the protein chain, solvent molecules, and ions. In many cases, the ruggedness means glassy dynamical behavior and lack of self-averaging, preventing a generalized statistical description of the functional dynamics.^{9,10} Therefore, computing detailed maps of the energy landscapes is indispensable for understanding mechanisms of functional processes of proteins, such as allosteric regulation,^{9,11} native state dynamics,^{3-5,9,12} or biopolymer translocation.¹³ One approach to mapping of the energy landscapes is to start from calculating the free energy difference between two specific protein conformations, for example, two allosteric states. Then, in principle, one could extend this calculation to many more conformations of a protein, leading to a high resolution view into the energetic topography of the native basin.

Among a number of proteins used to gain insights into the interplay between protein dynamics and function, adenylate kinase (ADK) stands out as one of the most extensively studied, in the context of allosteric transitions. The structures of ADK's allosteric states, known as the open and closed forms, have been solved long ago.¹⁴ Also the dynamics of conformational change in ADK has been a topic of numerous experimental¹⁵⁻²² and theoretical²³⁻²⁸ studies. However, the free energy difference between the two main allosteric states of ADK has not yet been computed using rigorous approaches based on explicit solvent force fields. Meanwhile, this free

energy difference is one of the important aspects underpinning the nature of ADK catalytic action. We calculate it in this work.

Extracting free energy differences from all atom simulations of proteins presents a formidable challenge. The difficulty largely stems from a large number of degrees of freedom which encumbers full sampling of the thermodynamic states. In a prior work, structurally based umbrella sampling free energy calculation for ADK was carried out with the implicit solvent.²⁹ As elaborated in a recent publication,³⁰ the corresponding collective coordinate for umbrella sampling^{29,31} used to compute conformational free energy differences has significant shortcomings, and may potentially lead to artifacts. A key problem of the widespread structural coordinates like $\Delta \text{rmsd}_{AB}(X) = (\text{rmsd}_{XA} - \text{rmsd}_{XB})$ (the difference between RMSDs of an intermediate structure X from the reference structures A and B) is the fact that these coordinates have an unacceptably high degree of structural degeneracy. Namely, in these coordinates a single value maps into a substantial number of unrelated conformations, which defeats the whole purpose of measuring conformational free energy difference between two distinct states, since the very definition of states is inconsistent. To illustrate the structural degeneracy we have used $\xi(Q_A, Q_B) = \Delta Q$ as a reaction coordinate to map the conformational space of two structurally highly similar states of a small model protein (see Figure 1). As one can see from the Figure 1, even in this favorably picked case the structural degeneracy is quite large, with states that are structurally dissimilar to both conformations lumped into the reference basins. To address this problem,

Received: October 17, 2011

Revised: December 31, 2011

Published: January 2, 2012

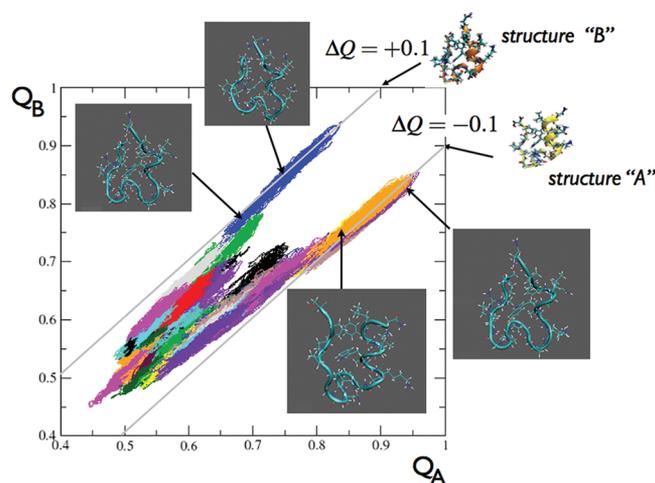


Figure 1. Illustration of a reaction coordinate degeneracy problem for defining conformational basins. On the plot are shown the conformational states of a model protein (trpcage, PDB ID 1L2Y) for which the $\Delta Q = Q_A - Q_B$ reaction coordinate was tested. Different colors correspond to the trajectories from different umbrella windows. The selected snapshots highlight the structural heterogeneity in the conformational ensembles of the basins A and B.

we have recently developed a rigorous technique that permits calculation of conformational free energies between two arbitrary polymer conformations from explicit solvent all atom simulations, where by two conformations we mean two well-defined regions of the polymer molecule phase space.³⁰ Our method largely consists of identifying a reaction coordinate which correlates with conformational transformation. Combining our coordinate with the enhance sampling simulations (such as umbrella sampling) provides a route for computing free energies or any other relevant equilibrium thermodynamic parameters regarding the transition. As demonstrated in the current work, this technique can be applied to large proteins, over two hundred residues, simulated in explicit solvent. Other approaches which attempt to accomplish the same goal use uncontrolled approximations^{32–34} or create unphysical intermediate states³⁵ which makes their application to complex explicit solvent bimolecular systems rather challenging (see ref 30 for a more detailed discussion of alternative techniques).

The proposed conformational reaction coordinate smoothly morphs one allosteric state into another, preserving structural locality near the allosteric basins.³⁰ The latter is a critical requirement missing in many commonly used alternative approaches. Locality means that if for a conformation X the value of the variable $\xi(X)$ is close to the value of the variable in the closed state $\xi(\text{closed})$, then X is structurally similar to the closed state. The same statement is also true for the open state.³⁰ Another important ingredient of the technique is the confinement potential which alleviates the problem of sampling the degrees of freedom transverse to the umbrella sampling variable. Its effect on the calculation amounts to enveloping the dynamic trajectories of the system into a phase space “tube” without affecting the neighborhood of the two allosteric states. The tube prevents sampling of the phase space regions of high conformational entropy which contain unfolded and other nonrelevant states, as it is not necessary in the case when only free energy difference between the two conformations is calculated. One of the goals of this paper is to demonstrate

that this new method, which was previously applied³⁰ to a 20-residue protein, Trp-Cage, straightforwardly scales up to 1 order of magnitude larger proteins.

Adenylate kinase (ADK) is an important member of kinase family of proteins which catalyzes a key metabolic step of phosphoryl transfer between ADP molecules: $\text{ADP} \cdot \text{Mg}^{+2} + \text{ADP} \leftrightarrow \text{ATP} \cdot \text{Mg}^{+2} + \text{AMP}$. Structurally, ADK consists of three domains: LID, NMP, and CORE.¹⁴ The CORE domain comprises the bulk of the protein and is relatively static during allosteric transition; meanwhile, NMP and LID form contacts with entering substrate by covering the bound substrate and preventing it from diffusing into the solvent environment. Many of the previous computational studies focused on the mechanistic aspects of the collective domain motions, which is correlated with the enzymatic activity.¹⁸ More specifically those studies identified the key rate limiting step of LID domain opening when bound to a ligand.^{25,36}

In this work we find the free energy difference between the states in the absence of ligand. In significant contrast to the previously reported high (tens of $k_B T$) values of the free energy difference from implicit solvent simulations,²⁹ we find that free energies of both states are comparable, with the difference on the order of $\sim 1-2 k_B T$. In addition, we find that, quite unexpectedly, interfacial contacts between LID and NMP domains, characteristic of the closed state, start to form even when the conformation is close to the open state. In other words, the transition from the closed to the open basin might be characterized by a late transition state.

FREE ENERGY CALCULATION TECHNIQUE

The central question addressed by our method³⁰ can be stated as follows: given two experimentally determined structures, A (open) and B (closed), how can we estimate the free energy difference ΔF_{AB} between them? Here we assume a time scale separation between fast sampling of the similar conformations around both open and closed states, and slow transitions between the states. Hence, the neighborhoods of the transition end points form corresponding basins of attraction (conformational basins), having a specific size depending on the molecular structure and intermolecular interactions. Thus, the conformational free energy difference that we are after is in fact the free energy difference between the conformational basins of A and B. The size of the basins should be chosen from the considerations external to our calculation technique.

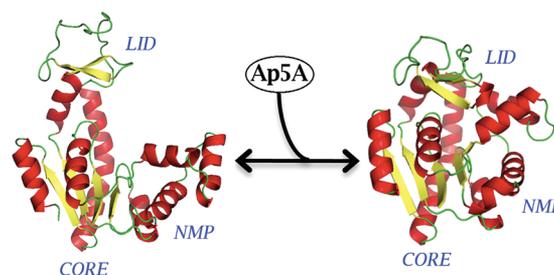


Figure 2. Crystallographic structures of open (4AKE) and closed (1AKE) forms of ADK are shown.¹⁴

The umbrella sampling variable (or reaction coordinate) $\xi(X)$ which is local near the allosteric states and continuously morphs one structure into another is described by the following

functional form:³⁰

$$\xi(X) = \exp \left[- \frac{(s(X,A) - s(A,B))^2 + (s(X,B) - 1)^2}{2\rho^2} \right] - \exp \left[- \frac{(s(X,B) - s(A,B))^2 + (s(X,A) - 1)^2}{2\rho^2} \right] \quad (1)$$

where $s(X,Y)$ is a measure of structural similarity between conformations X and Y , such as rmsd (root mean square deviation) or fraction of shared contacts (Q). When overlap parameter Q is used in defining $s(X,Y)$, then $s(X,Y) = 0$ means X and Y have nothing in common, and $s(X,Y) = 1$ means it is the same conformation. The conformational phase space, therefore, maps into a square with sides graded from 0 to 1. Regarding the variable ξ (see Figure 3) defined in eq 1 as elevation above this

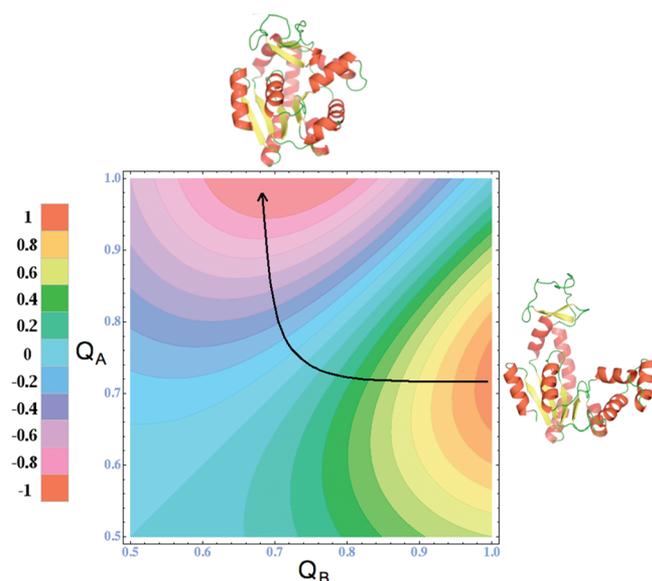


Figure 3. Contour plot of the reaction coordinate $\xi(X)$,³⁰ where X is an arbitrary point of the conformational space that maps into $Q_A(X)$ and $Q_B(X)$. The direction along which we sample conformations is indicated with a black arrow.

square, it corresponds to a positive circular Gaussian “peak” placed on the basin of conformation B and a negative circular Gaussian “pit” placed on conformational basin A.³⁰ The ρ in the equation is a parameter which controls the resolution of ξ and has to be decided based on the specific aims of the problem (see Table 1). If

Table 1. Parameter Values Used in the Auxiliary Potentials

parameter	numerical value
$Q(A,B)$	0.856
ρ	0.06
σ	1.5
k_{spring}	100 kcal mol ⁻¹
ε	10 kcal mol ⁻¹
κ	5000
a_1	0.76
a_2	0.88
R_1	0.0063
R_2	0.0044

its value is too large, the resolution coarsens and might not capture the subtle structural differences in the respective basins leading to

more or less random free energy estimates. On the other hand making resolution too high at best increases the computational demand, by requiring more windows and higher values for spring constants to adequately sample the entire pathway. For $s(X,Y)$, we chose the structural overlap parameter widely employed in studies of protein folding and spin glasses,³⁷ which is simply a generalized form of the fraction of shared contacts:

$$s(X,Y) = Q(X,Y) = \frac{1}{N} \sum_{i,j} \exp \left[- \frac{(r_{ij}^X - r_{ij}^Y)^2}{2\sigma^2} \right] \quad (2)$$

where σ sets the length-scale for the native contacts and is typically on the order of $C_\alpha - C_\alpha$ distance, e.g., ~ 1 . Hereafter, for the notational simplicity we will avoid the explicit indication of path variable X and instead will write the similarity as Q_A and Q_B . The immense volume of the phase space orthogonal to umbrella sampling variable ξ with varying spectrum of relaxation time scales at different points also represents a problem. Whenever the relaxation of transverse fluctuations exceeds the simulation times the trajectory may fall in an “entropic trap” and potentially not reach equilibration and convergence.

To alleviate the difficulty of adequately sampling transverse fluctuations, we designed a confining potential V_c , which is added to the Hamiltonian and is meant to block the trajectory from escaping into high conformational entropy regions. Confinement potential V_c in $\mathcal{H}' = \mathcal{H} + V_c$ can be chosen such that it is equal to zero near the allosteric states, in which case the sought for free energy will not be affected by the virtue of the following equation:

$$\begin{aligned} F'_B - F'_A &= - \frac{1}{\beta} \ln \frac{\int_{\Gamma_B} e^{-\beta \mathcal{H}'} d\Gamma}{\int_{\Gamma_A} e^{-\beta \mathcal{H}'} d\Gamma} \\ &= - \frac{1}{\beta} \ln \frac{\int_{\Gamma_B} e^{-\beta \mathcal{H}} d\Gamma}{\int_{\Gamma_A} e^{-\beta \mathcal{H}} d\Gamma} = F_B - F_A \end{aligned} \quad (3)$$

where $\beta = 1/k_B T$, Γ represents the full conformational space and the Γ_A and Γ_B are the portions of full space that correspond to conformational states of A and B (where $V_c = 0$).³⁰ The particular form of confinement potential employed in the present work is chosen to have the following functional form (see Figure 4) which also illustrates the idea of phase space “tube”:

$$\begin{aligned} V_c(Q_A, Q_B) &= \varepsilon [2 - \tanh(\kappa((Q_A - a_1)(Q_B - a_1) - R_1)) \\ &\quad + \tanh(\kappa((Q_A - a_2)(Q_B - a_2) - R_2))] \end{aligned} \quad (4)$$

The potential on the (Q_A, Q_B) plane can be envisioned as two walls of height ε of hyperbolic shape (see Figure 4) with radii R_1 and R_2 . The κ defines steepness of these walls (or hardness of the phase space tube enveloping the dynamic trajectory). a_1 and a_2 are constants that determine positions of the hyperbolae. The optimal tube should allow many pathways connecting the open and closed states, but not excessively many. The particular form for the confining potential is not critical, since it is eventually canceled when computing the free energy difference between the conformations A and B. However, it should have a smoothly rising “walls”, which would push the trajectory within the confines of the tube without causing abrupt jumps in the trajectory. Ideally, the confining potential should enclose the actual transition pathway along with the reference basins. However, if only the free energy difference between the basins

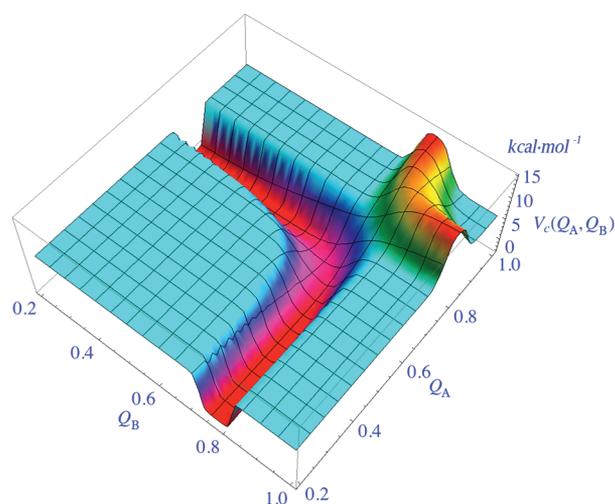


Figure 4. Potential V_c from eq 4 forms a “creek” confining the sampling trajectories inside. The shape of the “creek” is hyperbolic and the rise of its shores is hypertangential.

is calculated, it is not necessary. To find the shape of the tube we ran short (20–50 ps) simulations in each umbrella window, $U_i^{\text{umb}}(\xi) = k_{\text{spring}}(\xi(Q_A, Q_B) - \xi_i)^2$ without confinement, obtaining an approximation to a dominant transition pathway, or, at least, a preferential dynamics in a given ξ umbrella window. The general shape is indicated by the regions sampled by the preliminary short trajectories, which overall are localized in the upper right corner of the (Q_A, Q_B) plot (see Figure 3). The hyperbolic shape of the “tube” is a reasonable choice for the purpose of confining trajectories (irrespective of a protein) in the upper right corner, being one of the simplest smooth analytic curves, a quadratic polynomial. It is also natural to choose hyperbolic tangent for the smooth rise of the wall, the height of which is controlled by ε . One typically assigns some arbitrary high value for ε (e.g., 10.0 kcal/mol). The steepness κ on the other hand should be high enough (see Table 1) to prevent the trajectories from overcoming the wall, but low enough to not cause numerical instabilities when computing the forces near the boundaries.

The tube parameters, a_1 , a_2 , R_1 , and R_2 (see Table 1), which determine the position and thickness of the tube, have to be chosen in such way as to leave the reference basins intact. Therefore, there is a minimally possible width of the tube. On the other hand, the width should not be too large, so the trajectories equilibrate on the time scales comparable to the length of the simulation and also allow for overlap of trajectories in the neighboring windows. Thus, to determine the parameters we first run preliminary short simulations by gradually raising the thickness of the tube until the two of the mentioned conditions are satisfied.

■ COMPUTATIONAL DETAILS

As starting structures we used atomic coordinates of the closed (1AKE) and open (4AKE) forms of ADK taken from the Protein Data Bank¹⁴ (see Figure 2). After stripping off the ligand coordinates from the raw crystallographic structure of the closed state, we immersed both conformations in the TIP3P water boxes and added Na^+ and Cl^- ions to mimic the physiological concentration of a cell ($\sim 140 \text{ mM L}^{-1}$). Total number of ions and water was chosen the same for both systems. All subsequent simulations were performed with the program NAMD,³⁸ utilizing CHARMM27 forcefield.³⁹ After

initial minimization steps with the constrained and unconstrained protein coordinates, we heated the systems to 300 K by performing 200 ps NVT simulations. NVT simulations were carried out in the contact to heat bath, emulated by Langevin dynamics with the friction coefficient of 4 ps^{-1} . After heating steps we equilibrated the density of the system by performing 2 ns NPT simulation.

Fully equilibrated systems were then replicated among 121 windows. Approximately half of the umbrella sampling simulations were initiated from the closed form and the other half from the open form. We accumulated a series of ~ 1.0 ns trajectories, carrying out two independent umbrella sampling simulations for each window. The simulations with added umbrella potentials were performed with the LAMMPS software.⁴⁰ Before productive simulations, we ran multiple short trajectories in all windows in order to probe the local landscape and find optimal spring constants that guarantee sufficient overlap of reaction coordinate distributions. We have found that uniform spring constant values (see Table 1) across the equally spaced windows ($\delta\xi = 0.02$) sufficed to generate excellent overlap between umbrella sampling variable distributions of neighboring windows. The shape of the confining wall has been chosen such that it leaves the basins of the reference conformations unchanged meanwhile enveloping all the intermediate ones with a high walled tube. Since there is a time scale separation of backbone motions and side-chain rotations we have included only C_α atoms in the definition of Q . We limited the summation in Q to only those pairs of C_α atoms that are closer than 12 Å to each other in either of the conformations. Such definition of a contact has been used in the coarse grained protein folding studies.⁴¹ By introducing distance cutoff in the definition of Q we significantly reduce the large number of conformational states which on average contribute equally to both basins and therefore act as a “noise”.

■ RESULTS AND DISCUSSION

To ascertain whether our simulations are capable of reproducing the sought free energy difference to a sufficient degree of accuracy, we have performed two independent simulations using different sets of initial conditions, where the initial atomic velocities were randomized in all windows, hence, producing trajectories that are mutually unrelated. As one can see in Figure 5, the resulting two free energy profiles are in semiquantitative agreement, showing similar basic features. Another interesting feature of the obtained free energy profiles is the well pronounced minima near the end points of the reaction coordinate. These correspond to the basins of open and closed conformations of ADK. In the case of ADK, there are no external considerations which dictate the choice of the physically meaningful size of the conformational basins. Therefore, one could consider ADK conformations falling within the corresponding minima of the end points as belonging to one basin, leading to approximately $|\Delta\xi| = 0.2$ defining the basin size.

After defining the reference basins, one can obtain the conformational free energy by integrating the potential of mean force along the reaction coordinate in the respective basins (defined with the boxed regions in Figure 5)

$$F_{\text{closed}} - F_{\text{open}} = -\frac{1}{\beta} \ln \frac{\int_{\Gamma_{\text{closed}}} e^{-\beta F(\xi)} d\xi}{\int_{\Gamma_{\text{open}}} e^{-\beta F(\xi)} d\xi} \quad (5)$$

where $F(\xi)$ is the profile in Figure 5. The free energy difference between the two allosteric states estimated from two

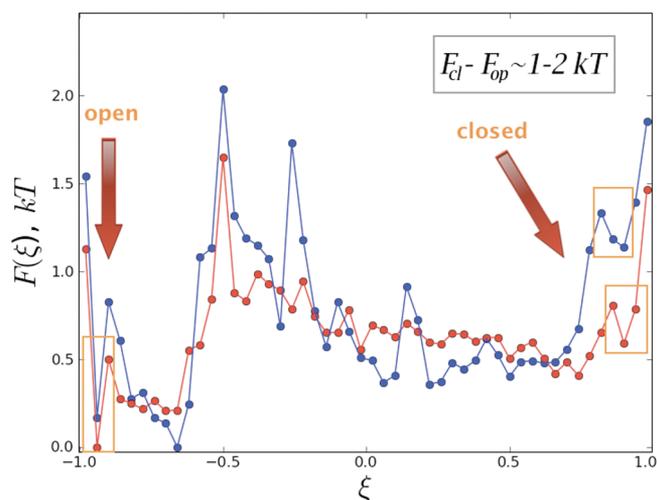


Figure 5. Free energy of ADK as a function of reaction coordinate ξ is plotted. Two different plots correspond to two completely independent simulations. The orange boxes indicate the definitions of the basins used for calculating the free energy differences according to eq 5.

uncorrelated trajectories was found to be $1.5 \pm 0.5 k_B T$. This small value of free energy difference is in marked contrast with the work of Arora et al.²⁹ where the potential of mean force as a function of rmsd was estimated to be tens of $k_B T$ higher in free energy upon approaching the basin containing the ensemble of closed conformations.

We suggest that the significant discrepancy between the current and previous results may be explained by a combination of the following two factors: (1) the implicit solvent that has been used in the earlier work is known for introducing a strong conformational bias⁴² and (2) the reaction coordinate used in the previous study is not local near the allosteric states, which may lead to artifacts.³⁰ The free energy profile computed in the current work is in semiquantitative agreement with the single molecule FRET and NMR spin relaxation experiments which reported a relatively fast collective domain motions that take place on a nanosecond time-scale^{18,19} and a relatively rarer event of attaining catalytically competent closed conformation (e.g., the one that is consistent with X-ray structure for the closed state) on a microsecond time scale.¹⁷ From the kinetics perspective based on Kramer's theory,⁴³ if the closed state is reached within a microsecond or faster from the open state, then closed state's free energy (as well as the barrier for the transition) should be within several $k_B T$ from the open state (assuming a pre-exponential factor which is characteristic of polypeptide chains⁴⁴). NMR experiments with ligated ADK (achieved by excessive ATP concentration) also report²² the difference between the conformational basins to be on the order of $1 k_B T$. Finally, when coarse-grained computer simulations of ADK were parametrized by experimental data, the resulting difference in the free energies of open and closed states was also estimated²⁴ to be $1-2 k_B T$.

In addition, the obtained free energy profile offers a simple way of resolving the discrepancy concerning the dominant conformations that has been reported by NMR and single molecule FRET experiments. A few years ago, Hanson et al.²⁰ showed that in FRET experiments, closed-like conformations are sampled more often. In contrast, the solution NMR suggests a picture^{19,45} where open-like conformations are heavily populated and transitions to the closed form occur

occasionally. We explain the discrepancy in the following way: the conformational basin of the open state is narrower and well-defined, whereas the basin for the closed state extends further along the reaction coordinate. Therefore, based on this line of reasoning, there is not a unique way to define the closed form of ADK, and other external criteria should be invoked to make the definition consistent with the problem. To clarify the latter statement, if we take a more generous definition of the closed basin, by extending the limit of integration in eq 3 for the closed state beyond the boundaries of the yellow box in Figure 5, the free energy gap between closed and open states narrows considerably. We thus attribute the discrepancy between two different experimental techniques to the distance-sensitivity difference intrinsic to NMR and FRET that have skewed the conformational distribution in favor of one or the other form.

The transition path taken by the trajectories in our simulations may not correspond to the actual one (although the tube path was chosen based on short simulations which likely indicate some conformational preferences of the protein), but one can still infer useful information about the intermediate steps. For instance, the barrier in our profile provides the upper limit to the actual barrier, since cutting off parts of the phase space can only make some pathways unavailable to the trajectory, which, in turn, will remove the contribution of these cutoff pathways to lowering the free energy of the transition state.

Apart from the thermodynamics, one can gain further insight into the nature of conformational transition by analyzing the structural evolution across the umbrella sampling windows. Comparing pairwise interatomic distances in reference states, we have identified the evolution of interfacial contacts, the contacts that are being broken when going from closed to the open form on the interface of LID and NMP domains. Using the interfacial contacts we have computed Q_{int} for all conformations in each window and plotted the distribution of Q_{int} values against the order parameter, ξ , that has been used in the free energy simulations (see Figure 6). From the plot one can

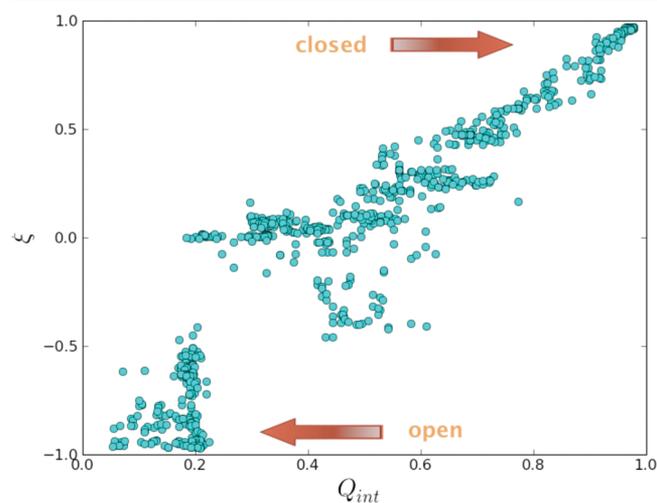


Figure 6. Each point on the plot corresponds to a snapshot from the simulation, for which a reaction coordinate ξ and a fraction of remaining interfacial contacts Q_{int} are calculated. Even close to the open state (in terms of ξ , and therefore structurally) many interfacial contacts are still present.

see the nonlinear dependence of the order parameter on the fraction of native contacts, with an abrupt transition that coincides with the location of the barrier in Figure 5.

On the other hand, a naïve expectation for the open to close transition would be that of an abrupt disruption of contacts when the LID-NMP domains smoothly and linearly separate from each other by a few ngströms. However, contrary to what one would expect, a significant fraction of the interfacial contacts are retained even when the system has moved further away from the basin of the closed state (see Figures 6 and 7).

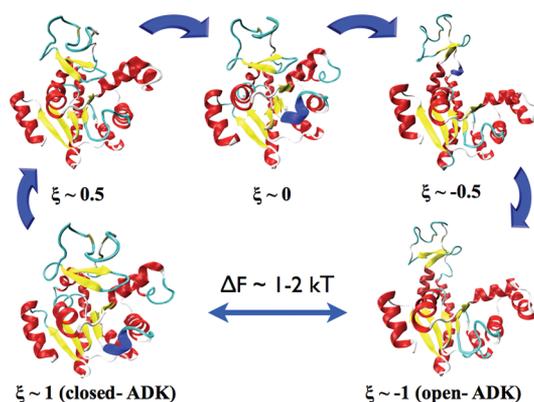


Figure 7. Intermediate structures along the pathway, selected from the appropriate histogramming windows. The transition from the closed to the open state goes through many closed like states followed by an abrupt opening of the NMP domain in the later stages.

This observation might be an artifact from confinement potential, but about 20% of interfacial contacts are retained close to the open state, where the influence of the confinement potential is low. This relation between the structure and interfacial contacts means that conformational change is structurally more intricate, involving partial breakage of contacts taking place gradually or in several stages.

The shallowness of the landscape near the closed state (corresponding to ξ being roughly in between 0 and 1) translates into a higher conformational entropy for the ensemble of closed-like conformations, which are qualitatively defined as collection of states with higher structural similarity to the reference closed state. This somewhat unexpected result is supported by the results of our local “conformational heterogeneity” calculations in the α -helices and the B factor distributions¹⁶ measured for the crystal structures of the two forms of the ADK. We quantify the so-called local “conformational heterogeneity” by the distribution of a variable Q_{helix} , which is calculated as in eq 2 but only the residues that are part of an α -helix are included in the sum. A broader distributions of Q_{helix} implies a higher “conformational heterogeneity” and hence a higher conformational entropy for the particular α -helical segment and vice versa. By computing the distribution of Q_{helix} in the basins of an open ($\xi \approx -1$) and the closed states ($\xi \approx 1$) we noticed that overall the helices in the closed form are characterized by a lower conformational heterogeneity compared to the open form (see Figure 8). In particular, the α -helix composed of residues: 43–53 in the NMP domain showed striking contrast in terms of the “conformational heterogeneity” of the two forms. Approximately a half of the indicated residues are in disordered state in the closed form of the ADK (according to the DSSP measure⁴⁶) and hence show a much higher conformational heterogeneity compared to the ordered form (see Figure 8), where the same residues are all part of an α -helix. From the protein physics viewpoint this observation can be explained by the local frustration in the

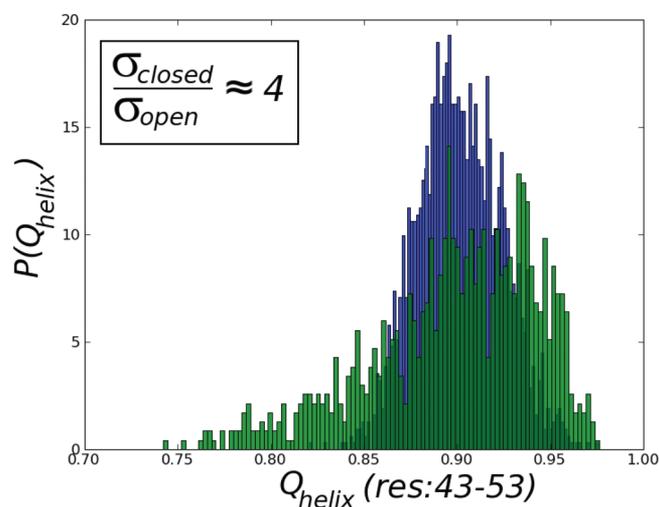


Figure 8. Evidence for the local frustration in the closed form of the ADK. The blue and green histograms are the distribution of Q_{helix} for the closed and the open forms respectively. σ_{open} and σ_{closed} denote the standard deviation in the respective distributions.

closed form of the ADK, where the multiple contacts between the domains are made at the expense of some of the contacts which stabilize the helices within the respective domains. From the thermodynamic point of view, one may entertain the idea that after the domain closure the free energy gained from forming the contacts is partially channeled into the helices by increasing their conformational heterogeneity. This essentially amounts to an “entropic transfer” in the ADK from the interdomain flexibility of the open form to the intrahelical disorder in the closed form. The net effect of this entropic transfer mechanism would be to diminish the free energy gap between the two corresponding allosteric states, which, in turn, would allow for easier regulation by small ligands of allosteric switching and the enzyme recycling.

Other native contacts might be more significant to the transition than the interfacial ones. The retention of the interfacial contacts near the open state prompts one to speculate that the transition state ensemble is more likely to be located closer to the open state. From the kinetic point of view this would be an indication of an anti-Hammond type behavior: more structural reorganization is needed to reach the transition state, if one starts from the closed, than from the open state, while the free energy increase is, on the contrary, larger if one starts from the open state.^{47,48}

CONCLUSION

The free energy difference between the open and closed forms of ADK in the absence of ligands, computed in this work, turned out to be rather small ($\sim 1-2 k_B T$). Two fully independent calculations indicate that the results are reasonably converged and reproducible. From the biological perspective, the relatively small free energy difference between the allosteric states may facilitate fine control of allosteric transition by environmental perturbations and signaling. In addition, we found that even when some of the NMP-LID interfacial contacts are formed, the typical conformations are still structurally more similar to the open form, suggesting a late transition state for domain opening. Our conformational heterogeneity calculations further clarify the mechanistic and thermodynamic signature of the transition. We have found that

in closed state domains have higher conformational heterogeneity compared to the open form which acts as a counterweight balancing the entropic loss associated with domain closure. We rationalize these observation by providing a mechanism of entropic transfer, which is a way for the allosteric transition to lower the free energy gap between the end point states.

We conclude that the recently developed method for calculating conformational free energy differences³⁰ can be applied to systems of real biological importance, for instance, large proteins like adenylate kinase. In order to calculate a free energy profile corresponding to the actual kinetics of the allosteric transition, the confinement tube trajectory should be dynamically updated, instead of being statically defined, allowing the system to find the dominant path of minimum free energy. The latter goal will be pursued in our subsequent studies.

AUTHOR INFORMATION

Corresponding Author

*E-mail: gpapoian@umd.edu.

ACKNOWLEDGMENTS

All of the simulations in the present contribution have been carried out on the supercomputers Deepthought at the University of Maryland and Topsail at the UNC-CH. We acknowledge financial support of the Camille and Henry Dreyfus Foundation, National Science Foundation under the Grant CHE-0846701.

REFERENCES

- (1) Frauenfelder, H.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 5129–34.
- (2) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598–1603.
- (3) Zhuravlev, P. I.; Papoian, G. A. *Curr. Opt. Struct. Biol.* **2010**, *20*, 16–20.
- (4) Henzler-Wildman, K. A. *Nature* **2007**, *450*, 913–916.
- (5) Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14408–14413.
- (6) Frauenfelder, H.; Parak, F. G.; Young, R. D. *Annu. Rev. Biophys. Biochem.* **1988**, *17*, 451–79.
- (7) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. *J. Mol. Biol.* **2005**, *346*, 1121–1145.
- (8) Zhuravlev, P. I.; Materese, C. K.; Papoian, G. A. *J. Phys. Chem. B* **2009**, *113*, 8800–12.
- (9) Zhuravlev, P. I.; Papoian, G. A. *Q. Rev. Biophys.* **2010**, *43*, 295–332.
- (10) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
- (11) Kern, D.; Zuiderweg, E. R. P. *Curr. Opt. Struct. Biol.* **2003**, *13*, 748–57.
- (12) Materese, C. K.; Goldmon, C. C.; Papoian, G. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 10659–10664.
- (13) Makarov, D. E. *Acc. Chem. Res.* **2009**, *42*, 281–9.
- (14) Müller, C. W.; Schulz, G. E. *J. Mol. Biol.* **1992**, *224*, 159–77.
- (15) Kern, D.; Eisenmesser, E. Z.; Wolf-Watz, M. *Methods Enzymol.* **2005**, *394*, 507–24.
- (16) Müller, C. W.; Schlauderer, G. J.; Reinstein, J.; Schulz, G. E. *Structure* **1996**, *4*, 147–56.
- (17) Wolf-Watz, M.; Thai, V.; Henzler-Wildman, K.; Hadjipavlou, G.; Eisenmesser, E. Z.; Kern, D. *Nat. Struct. Mol. Biol.* **2004**, *11*, 945–9.
- (18) Shapiro, Y. E.; Kahana, E.; Tugarinov, V.; Liang, Z.; Freed, J. H.; Meirovitch, E. *Biochemistry* **2002**, *41*, 6271–81.
- (19) Shapiro, Y. E.; Meirovitch, E. *J. Phys. Chem. B* **2006**, *110*, 11519–24.

- (20) 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*, 18055–60.
- (21) Adén, J.; Wolf-Watz, M. *J. Am. Chem. Soc.* **2007**, *129*, 14003–12.
- (22) Olsson, U.; Wolf-Watz, M. *Nat. Commun.* **2010**, *1*, 111.
- (23) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12570–5.
- (24) Lu, Q.; Wang, J. *J. Am. Chem. Soc.* **2008**, *130*, 4772–83.
- (25) Whitford, P. C.; Miyashita, O.; Levy, Y.; Onuchic, J. N. *J. Mol. Biol.* **2007**, *366*, 1661–71.
- (26) Maragakis, P.; Spichty, M.; Karplus, M. *Phys. Rev. Lett.* **2006**, *96*, 100602.
- (27) Daily, M. D.; Phillips, G. N.; Cui, Q. *J. Mol. Biol.* **2010**, *400*, 618–31.
- (28) Jana, B.; Adkar, B. V.; Biswas, R.; Bagchi, B. *J. Chem. Phys.* **2011**, *134*, 035101.
- (29) Arora, K.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18496–501.
- (30) Zhuravlev, P. I.; Wu, S.; Potoyan, D. A.; Rubinstein, M.; Papoian, G. A. *Methods* **2010**, *52*, 115–21.
- (31) Banavali, N. K.; Roux, B. *J. Am. Chem. Soc.* **2005**, *127*, 6866–6876.
- (32) Kollman, P. A.; et al. *Acc. Chem. Res.* **2000**, *33*, 889–97.
- (33) Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. *J. Med. Chem.* **2006**, *49*, 6596–606.
- (34) Vorobjev, Y. N.; Hermans, J. *Protein Sci.* **2001**, *10*, 2498–2506.
- (35) Park, S.; Lau, A. Y.; Roux, B. *J. Chem. Phys.* **2008**, *129*, 134102.
- (36) Adkar, B. V.; Jana, B.; Bagchi, B. *J. Phys. Chem. A* **2011**, *115*, 3691–7.
- (37) Plotkin, S. S.; Wang, J.; Wolynes, P. G. *Phys. Rev. E* **1996**, *53*, 6271–6296.
- (38) Phillips, J. C.; et al. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (39) MacKerell, A. D.; Banavali, N. K.; Foloppe, N. *Biopolymers* **2000**, *56*, 257–265.
- (40) Plimpton, S. J. *Comp. Phys.* **1995**, *117*, 1.
- (41) Portman, J.; S Takada, P. G. W. *Phys. Rev. Lett.* **1998**, *23*, 5237.
- (42) Roe, D. R.; Okur, A.; Wickstrom, L.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 1846–57.
- (43) Zhou, H.-X. *Q. Rev. Biophys.* **2010**, *43*, 219–93.
- (44) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opt. Struct. Biol.* **2004**, *14*, 76–88.
- (45) Henzler-Wildman, K. A.; et al. *Nature* **2007**, *450*, 838–44.
- (46) Kabsch, W.; S., C. *Biopolymers* **1983**, *22*, 2577D637.
- (47) Hammond, G. J. *Am. Chem. Soc.* **1955**, *77*, 334.
- (48) Matthews, J. M.; Fersht, A. R. *Biochemistry* **1995**, *34*, 6805–14.