Heterogeneous path ensembles for conformational transitions in semi–atomistic models of adenylate kinase

Divesh Bhatt and **Daniel M. Zuckerman**[*]
Department of Computational and Systems Biology, University of Pittsburgh

Abstract

We performed "weighted ensemble" path–sampling simulations of adenylate kinase, using several semi–atomistic protein models. The models have an all–atom backbone with various levels of residue interactions. The primary result is that full statistically rigorous path sampling required only a few weeks of single–processor computing time with these models, indicating the addition of further chemical detail should be readily feasible. Our semi–atomistic path ensembles are consistent with previous biophysical findings: the presence of two distinct pathways, identification of intermediates, and symmetry of forward and reverse pathways.

1 Introduction

Fluctuations and conformational changes are of extreme importance in biomolecules.[1] For example, most enzymes show distinctly different conformations in the apo and the holo forms.[2] Conformational transitions are also typical in non–enzymatic binding proteins,[3] and of course are intrinsic to motor proteins.

The fundamental biophysics of conformational transitions in biomolecules is contained in the ensemble of paths – i.e., trajectories in configurational space – defining the transition. Such path ensembles contain the information about the relevant "mechanisms" for transitions, including possible intermediates (see, for example, discussions by E and vanden Eijnden[4] and by Hummer[5]). In addition, the transition rates can only be calculated accurately from a path ensemble, which implicitly accounts for all barriers and recrossings.[6] From a computational point of view, such path ensembles are difficult to obtain due to rugged energy landscapes and the timescales involved.[6–15] Multiple local minima and/or channels dramatically increase the computational effort required. To put the difficulty of path sampling in perspective, note that equilibrium sampling of fully atomistic models of large biomolecules is not typically feasible.[16] Thus, path sampling using detailed atomistic models for all, but the smallest systems, is difficult – even with potentially efficient methods developed specifically for path sampling. A number of groups have reported atomistic path sampling studies for small systems.[17–19] Radhakrishnan and Schlick[20] reported path sampling studies of atomistic DNA polymerase using initial paths generated by targeted molecular dynamics.[21–23]

Less computationally expensive approaches to determining atomistic paths are available, including targeted and steered molecular dynamics,[21–23] and more rigorous "nudged elastic band",[24–27] and related approaches.[28–31] However, all these methods yield only a single path or a handful, and not the ensemble required for a correct thermal/statistical description.

Specifically, fluctuations in pathways, the possibility of multiple pathways (path heterogeneity), and possible recrossings typically are not accounted for in these approaches.

Coarse–grained (CG) models, on the other hand, permit an alternative strategy for statistical path sampling.[32][34] Although CG models omit chemical detail, they can be sampled significantly faster than fully atomistic models, and, thus, such models are quite attractive for path sampling studies. For example, Zhang *et al*.[35] showed that a simple alpha–carbon model of calmodulin can be fully path sampled using the weighted ensemble path sampling method.[36] Because full path sampling in this model required only a few weeks of single–processor computing, it is evident that better models and/or larger systems could be studied. Network models have also been used to study conformational transitions.[37,38]

In this manuscript, we report weighted–ensemble (WE)[35,36] path sampling studies of adenylate kinase which represent improvements over previous work[35] in several ways. (i) At 214 residues, adenylate kinase is triple the size of the calmodulin domain previously path sampled using WE.[35] (ii) Our models now include significant atomic details, as explained below. (iii) We examine a series of models to test the sensitivity of the path ensemble to the chosen interactions and parameters. (iv) We investigate symmetry, based on our recent formal derivation,[39] between forward and reverse transitions.

Adenylate kinase (Adk) is an enzyme that catalyzes phosphate transfer between AMP and ATP via

$$\text{AMP+ATP} \underset{}{\overset{\text{Mg}^{2+}}{\rightleftharpoons}} \text{2ADP}, \tag{1}$$

and thus helps to regulate the relative amounts of cellular energetic units.[40][42] The crystal structure of Adk for *E. coli* is available in several conformations. Its native apo form (Protein Databank code 4AKE[43]) is shown in Figure 1 (a). In the figure, the blue segments represent the core (CORE), the yellow segment represents the AMP binding domain (BD), and the green segment represents the flexible lid (LID). Upon ligand binding, the enzyme closes over the ligands. The crystal structure (1AKE)[44] of the holo form of the enzyme obtained in complex with a ligand that mimics both AMP and ATP is shown in Figure 1 (b). Clearly, in the apo form, the enzyme shows an Open structure (that we denote as O in this manuscript), and in the holo form, it is Closed (denoted by C throughout).

Adk has been studied previously via computational methods using both coarse–grained models and fully atomistic simulations. Coarse–grained models used to study transition pathways for Adk have, primarily, utilized network models.[37,38,45,46] In these methods, the fluctuations in proteins are represented by harmonic potentials, and the deformations due to these fluctuations are used to estimate the free energy in the basins (end states and/or multiple basins). Subsequently, a minimum energy path is calculated to characterize the transition. Coarse–grained models have also been used with molecular dynamics simulations to determine transition paths.[47,48]

A few groups have also studied conformational fluctuations in Adk using atomistic models. Arora and Brooks[49] performed atomistic (with implicit solvent) umbrella sampling molecular dynamics (MD) simulations along an initial minimum energy path suggested by "nudged elastic band". Kubitzki and de Groot[50] performed replica exchange MD for atomistic Adk to increase conformational sampling of adenylate kinase – and observed both O and C conformers; however, a true path ensemble is not obtained from replica exchange. In other work, fully atomistic MD on the two end structures has been performed to observe fluctuations in the two ensembles[40,51] but direct conformational transitions were not

observed. Beckstein *et al*.52 performed atomistic path sampling studies for conformational transitions in Adk using dynamic importance sampling53 with an approximate biasing scheme.54

In the present study, we use semi–atomistic models to improve chemical accuracy compared to typical coarse–grained models while still performing high quality path sampling. In our models, the backbone is fully atomistic to provide chemically realistic geometry which is often absent from bead–based models. Inter–residue interactions are modeled at a coarse–grained level via the commonly used double–Gō potentials34,46,55,56 that (meta)stabilize two crystal structures. Additionally, one of the models uses residue–specific interactions to probe the effect of such interactions. We use a library–based Monte Carlo (LBMC) scheme to perform sampling.57. LBMC was previously developed in our group and shown to facilitate the use of semi–atomistic models of the type used here.57

Transitions between the Open and the Closed states (both directions) are studied with the weighted–ensemble (WE) path–sampling method36 that has been previously been studied to study folding of proteins,58 protein dimerization,59 and conformational transitions in an alpha–carbon model of calmodulin.35 WE was shown to promote efficient path sampling of conformational transitions in purely alpha–carbon model of calmodulin.35 Additionally, WE is statistically exact: it preserves natural system dynamics, resulting in an unbiased path ensemble.60 Additional strengths of WE include its ability to find multiple pathways,61 the simultaneous determination of the path ensemble and transition rate,35,62 the ease of implementing it at a scripting level, and its natural parallel structure.

Biophysically, we focus on heterogeneity of the path ensemble (multiple pathways) and the forward–reverse "symmetry" of the ensemble. It is possible that evolution has favored the fine–tuned precision of a single pathway in some systems, but the "robustness" of alternative pathways in other cases. Although our semi–atomistic models preclude biochemically precise conclusions, high–quality path sampling permits a complete description of a model system.

The goal of this work, in summary, is to explore the application of statistically rigorous path sampling to conformational transitions in proteins using semi–atomistic models of a moderately large protein. Methodologically, we want to test whether such models can be fully path sampled. Biophysically, we explore several outstanding issues in computational studies of conformational transitions in Adk. Although two different pathways for conformational transitions in Adk have been identified by models of varying level of details, 42,47,48 it has variously been suggested that a different pathway is dominant in the forward and reverse directions47 *versus* a more symmetric scenario in which each pathway occurs with same probability in the two directions.48 With our path sampling study, we can address this issue. We can also validate our path ensembles by comparison to experimental structures.

The manuscript is organized as follows. First, In Section 2 we discuss the models we use to depict the protein. Section 3 then describes the method to generate the ensemble of pathways. In Section 4, we present results for transitions in both the directions for all the three models we used. We discuss the results, efficiency, and future models in Section 5, with conclusions given in Section 6.

2 Semi–atomistic models

We use three semi–atomistic models, expanding on our previous work.57 In all the models, the backbone is represented in full atomistic detail, using the three residues alanine, glycine, and proline.57 All intraresidue interactions are included explicitly, using the OPLSAA all–

atom force field. Both the intra–residue interaction energies and the configurations are stored in libraries as described previously.[57] In brief, we note that libraries of the three types of residues are pre–generated according to the Boltzmann distribution at 300 K, and alanine is used to represent the backbone of all residues besides glycine and proline (a simplification motivated by the similarity of Ramachandran maps for the residues).[63] Ligands are not modeled explicitly in this path sampling study.

The differences in the three models lie in the treatment of inter–residue interactions: two of the models use only double–Gō interactions at backbone alpha carbons, whereas one model uses both double–Gō and residue–specific interactions. Complete information is given below.

All three semi–atomistic models employ double–Gō interactions. Following Ref34, for each of the two crystal structures, residues pairs with alpha carbons less than 8 Å apart are considered native contacts. In the Gō energy of an arbitrary configuration, every native contact from the Open form is assigned an energy of $-\epsilon$, whereas those exclusively found in the Closed form are scaled to be $-e_{\mathrm{scale}}\epsilon$. Gō interactions do not distinguish between different types of residues except in terms of size. This double–Gō potential between two residues i and j with alpha carbon distance r_{ij} is given by

$$u^{\mathrm{G\bar{o}}}\left(r_{ij}\right)=\begin{cases} \infty, & r_{ij}<r_{ij}^{\mathrm{X}}(1-\delta) \\ -\epsilon_{\mathrm{X}}, & r_{ij}^{\mathrm{X}}(1-\delta) \le r_{ij}<r_{ij}^{\mathrm{X}}(1+\delta) \\ 0.3\epsilon, & r_{ij}^{\mathrm{X}}(1+\delta) \le r_{ij}<r_{ij}^{\mathrm{Y}}(1+\delta) \\ -\epsilon_{\mathrm{Y}}, & (*)\, r_{ij}^{\mathrm{Y}}(1-\delta) \le r_{ij}<r_{ij}^{\mathrm{Y}}(1+\delta) \\ 0, & r_{ij} \ge r_{ij}^{\mathrm{Y}}(1+\delta) \end{cases}$$

(2)

where r_{ij}^{X} and r_{ij}^{Y} are the native distances in the two crystal structure ordered such that $r_{ij}^{\mathrm{X}} \langle r_{ij}^{\mathrm{Y}}$ (X and Y equate to Open or Closed), and δ is a well–width parameter chosen to be 0.05. If X equals Open and Y equals Closed, $\epsilon_{\mathrm{X}} = \epsilon$ and $\epsilon_{\mathrm{Y}} = e_{\mathrm{scale}}\epsilon$, and vice–versa. In the case of overlapping square wells, $r_{ij}^{\mathrm{X}}(1+\delta) \rangle r_{ij}^{\mathrm{Y}}(1-\delta)$, the 0.3ϵ barrier in the middle does not exist and the lower limit of the inequality marked with (*) is replaced by $r_{ij}^{\mathrm{X}}(1+\delta)$.

The total Gō interactions are therefore

$$U_{\mathrm{G\bar{o}}}(e_{\mathrm{scale}}) = \sum_{i<j} u^{\mathrm{G\bar{o}}}\left(r_{ij}\right).$$

(3)

For Model 1, we have

$$U_1 = U_{\mathrm{G\bar{o}}}(e_{\mathrm{scale}}=1).$$

(4)

The motivation behind using such a double Gō potential is that the two "end" crystal structures are presumed stable – and the double–Gō protocol guarantees that bistability. Such double Gō interactions have been used to probe the biophysics of several systems.[34,46,55,56]

2.1 Model 1: Pure double Gō with energy symmetry

Previous path sampling studies of proteins were limited to smaller systems and/or simpler models. We, therefore, first study whether the simplest model within the semi–atomistic framework can be fully path sampled. Our Model 1 omits most chemical details and uses only symmetric double–Gō interactions as given in eq 3. That is, native contacts in the Open and the Closed structure are treated identically ($e_{\text{scale}} = 1$).

Because Model 1 is a pure Gō model, the temperature is specified in units of the well depth of Gō interactions, ϵ. We choose the temperature as the highest at which the two experimental crystal structures are stable. We therefore performed a series of Monte Carlo simulations, as described below, at various temperatures. At $T = 0.75\,\epsilon/k_{\text{B}}$, both structures melted, but both remained (meta)stable at $T = 0.7\,\epsilon/k_{\text{B}}$. Thus, for Model 1, all subsequent equilibrium and path sampling simulations were performed at $T = 0.7\,\epsilon/k_{\text{B}}$.

2.2 Model 2: Double Gō with residue–specific interactions

Our second model adds chemical detail, both to improve upon the simplicity of Model 1, and to provide a way to check the sensitivity of our results to modeling choices. Model 2 includes atomistic backbone hydrogen bonding, Ramachandran propensities, and residue–specific contact interactions, as detailed below. Because these interactions are implemented as short ranged, Model 2 is only about 30% slower than Model 1, based on wall–clock time per MC step. Gō interactions are again symmetric, with $e_{\text{scale}} = 1$.

Our semi–atomistic LBMC platform makes the inclusion of additional interactions straightforward. Since the backbone is modeled atomistically, backbone–backbone hydrogen bonding is easily incorporated, as described below. However, due to the absence of explicit side chains in the present implementation, residue–specific chemical interactions can only be incorporated at a coarse–grained level. We use residue–specific contact interactions based on the work of Miyazawa and Jernigan (MJ),[64,65] as discussed below. Specifically, we use the potential energy

$$U_2 = U_{\text{Gō}}\,(e_{\text{scale}}=1) + U_{\text{HB}} + U_{\text{Rama}} + U_{\text{MJ}} \tag{5}$$

where U_{HB} is the hydrogen–bonding potential, U_{Rama} is the potential due to Ramachandran propensities, and U_{MJ} is the residue–specific potential based on MJ interactions. These terms are described below.

Hydrogen bonding for the backbone–backbone interactions is modeled atomistically, but with simplifications appropriate to the otherwise coarse–grained nature of our models. Specifically, we use ordinary Coulomb interactions with OPLSAA charges between the backbone CO and NH groups if the O–H distance less than 2.5 Å. The cutoff was chosen as the distance after which dipole interactions are significantly attenuated. Following previous studies that suggest a dielectric constant of 2–5 inside a protein, we use a value of 3. The use of physical charge and distance units in the hydrogen–bonding interactions allows physical temperature units in the simulation (instead of merely being in relation to the Gō well depth).

Ramachandran propensities were included via the term U_{Rama}, which is based on a potential of mean force obtained by calculating the distribution of φ–ψ dihedral angles in acetaldehyde–alanine–n–methylamide using OPLSAA force field. This distribution was tabulated from a Langevin dynamics simulation at 300 K using the GBSA implicit solvent model in Tinker software package.

The construction of MJ–type interactions required some care. Several variants of the original MJ interaction values have been utilized in the literature (such as scaling the MJ interactions energies, as well as shifting)[66][67] – due to the fact that MJ values are based on folded protein data and are not directly applicable for unfolded states. We follow the suggestion of Jernigan and Bahar[66] to mix MJ values of Table V and Table VI (numbering as in the original MJ paper,[64] with updated values as in Ref[65]) so that the residue specific interactions are modeled as $x \times$(Table V)$+(1 - x) \times$(Table VI). We chose $x = 0.05$ to ensure that the residue–specific interactions are a significant perturbation of the double Gō interactions.

To make the crystal structures (meta)stable, we "titrated in" double Gō interactions (ϵ), until bistability was observed at 300 K. Because, as described, hydrogen bonding introduces physical units into Model 2, the units of Gō well depth, ϵ, are also physical. We found that at $\epsilon/k = 400$ K, both the structures remained (meta)stable.

2.3 Model 3: Pure double Gō without energy symmetry

Finally, to facilitate the generation of large path ensembles in both the Open–to–Closed and Closed–to–Open directions, we also constructed a third model. The new model is designed to overcome the somewhat artifactual over–stabilization of the Closed states in Models 1 and 2 (see results below in Section 4). In brief, our 8 Å cutoff permits significantly more contacts in the Closed state, implicitly but artificially mimicking the presence of ligands in Models 1 and 2. This implicit presence of ligands interferes somewhat with our goal of modeling the ligand–free opening and closing of the enzyme.

In Model 3, therefore, we attempt to make the Open and Closed forms of adenylate kinase more comparable in stability. We decrease the strength of Gō interactions specific to the Closed form to half of Gō interactions (*i.e.*, we set $e_{scale} = 0.5$). Additionally, to focus on the effect of the reduced stability of the Closed form with respect to the Open form, we use only asymmetric double Gō interactions (and no H-bonding, Ramachandran, or MJ interactions). That is, we set

$$U_3 = U_{G\bar{o}} \, (e_{scale} = 0.5) \,. \tag{6}$$

3 Methods

3.1 Dynamical Monte Carlo

We follow many precedents[35][68][69] and use "dynamical" MC for the dynamics of our models. Such an approximation to physical dynamics is consistent with our use of simplified models. Specifically, we use the library–based Monte Carlo (LBMC) algorithm,[57] to propagate the system in both brute–force simulations for generating equilibrium ensembles and path sampling simulations (discussed below in more detail). For both equilibrium and path sampling, the systems always evolve via "natural" LBMC dynamics, and no artificial forces are used to direct conformational transitions, as explained below.

Our LBMC simulations use the same trial moves described in our earlier work.[57] Namely, one flexible peptide plane in the current configuration is swapped with one stored in the library, and a ψ angle is also displaced by a small amount.

3.2 Path sampling

In systems with rugged energy landscapes, such as proteins, regular brute–force simulations are not efficient for studying transitions. For this reason, we use the statistically rigorous

weighted–ensemble (WE) path sampling method to generate path ensembles of conformational transitions of adenylate kinase between the Open and the Closed states. This method preserves the natural system dynamics[60] and was used previously to study protein folding,[58] protein dimerization,[59] and conformational transitions of calmodulin using an alpha–carbon model.[35] Weighted ensemble studies the probability evolution of trajectories in the configuration space using any underlying system dynamics.[60] In this work, we use the WE method to study transitions in more detailed models to evaluate the effect of increasing chemical detail on the transitions, and to study questions of symmetry in forward and reverse directions.

The procedure to use weighted–ensemble path sampling to study conformational transitions is described in detail elsewhere,[35,36,60] and here we describe our simple implementation briefly. Prior to beginning the simulations, we divide a one dimensional projection of the configurational space (*i.e.*, the DRMS from the target structure in the present study) into a number of bins. The DRMS is a "progress coordinate" or "order parameter" – and is not necessarily the reaction coordinate. The progress coordinate roughly keeps track of the progress to the target state: the DRMS of structures close to the target state is necessarily small. It is also possible to use multidimensional or adaptively changing progress coordinates,[35,60] but was not found necessary here.

In the weighted–ensemble method, an evolving set of trajectories and their probabilities are tracked. Procedurally, several independent trajectories are started in an initial configuration and run for a short time interval τ (consisting of multiple simulation steps) with natural dynamics. At the end of each τ interval, the progress of the trajectories along the progress coordinate is noted (*i.e.*, into which bin along the progress coordinate each trajectory ends). Once bins are tabulated after each τ, trajectories are "split" (replicated with divided probability) and combined. This keeps the same number of trajectories in each occupied bin, prunes low–weight trajectories and splits trajectories with high probability. This splitting and combining of simulations is performed statistically as discussed elsewhere.[35,36,60] The probability remains normalized and all probability flows can be measured.

The full details of our WE simulations are as follows. We employ LBMC to describe the natural system dynamics. We utilize 25 bins between the two states, with 20 simulations (trajectories) in each occupied bin. The end state is defined as being at a DRMS of 1.5 Å from the target crystal structure, a definition used in both directions. Using this definition of the end state, we calculate the probability flux of trajectories entering the target state at the end of each τ.

It should be noted that value of the probability flux into either state – and hence the rate – depends upon the precise definitions of the two states. Although probability flows are good indicators of sampling quality, precise numerical values of the rates are not of great interest in our study of simple models with Monte Carlo dynamics. In this work, we are interested in the path ensembles and not the rates.

4 Results

4.1 Static analysis of conformational differences

For reference, we first analyze the conformational differences between the two end–state static crystal structures to quantify the observed differences in the Open and Closed configurations of Figures 1. Figure 2 shows the α–carbon distance difference map of pairs of residues in the Open and the Closed crystal structures. A large positive value implies that a pair is farther apart in the Open structure than in the Closed structure, whereas a negative value is the opposite. By construction, the figure is symmetric about the diagonal. A few

features of the two structures easily emerge from Figure 2. The inter–residue distances for most of the residue pairs are very similar in the two crystal structures. The major differences are that the distances in the Closed structure between residues labeled LID (114–164) are closer to BD (31–60) and several residues of CORE are smaller than the corresponding distances in the Open structure. Thus, Figure 2 quantifies Figure 1.

From Figures 1 and 2 it is clear that the structural change that characterizes the transition between the Open and the Closed structure is fairly straightforward: the LID and the BD close, and the rest of the protein remains fairly unchanged. Following Figures 1 and 2, for the path sampling studies presented shortly, we monitor inter–residue distances between two pairs of residues: residues 56 (GLY) and 163 (THR), which report on the BD–LID proximity, as well as residues 15 (THR) and 132 (VAL), which report on the CORE–LID proximity. In the Closed structure, $d^c_{56,163}=4.9$Å and $d^c_{15,132}=6.3$Å. On the other hand, in the Open structure, $d^o_{56,163}=23.6$Å and $d^o_{15,132}=17.8$Å. Thus, the relation between the CORE and LID is monitored, along with that of the LID and BD.

4.2 Brute force equilibrium sampling

In order to demarcate the native basins in our analysis of transitions, we first study equilibrium ensembles for the Open and the Closed states of adenylate kinase. Put another way, we want to quantify the size of native–basin fluctuations in our models. Further, we determine whether transition paths can be obtained without the aid of path sampling.

We quantify fluctuations in the equilibrium ensembles in the two basins by using DRMS from the respective crystal structures. Figure 3 (a) shows two sets of DRMS traces for Model 1 for a simulation started from the Open structure: DRMS–from–Open (black line) and DRMS–from–Closed (blue line). Similarly, Figure 3 (b) shows two sets of DRMS traces for Model 1 for a simulation started from the Closed structure: DRMS–from–Closed (black line) and DRMS–from–Open (blue line). Thus, in each panel, the black line represents DRMS from the starting structure, whereas the blue line represents DRMS from the opposing structure.

A comparison of the two panels of Figure 3 shows that the simulation started from the Open structure shows significantly more fluctuations than the simulation started from the Closed structure. Furthermore, the fluctuations drive the simulation started from the Open structure closer to the Closed structure than vice versa. For example, Figure 3 (a) shows that the simulation started from the Open structure gets to within 3 Å of the Closed structure at approximately 70 million MC steps. On the other hand, the simulation started from the Closed structure (Figure 3 (b)) remains farther from the Open structure.

Most importantly, neither simulation show a transition to the opposing structure. The DRMS from the opposing structure for a particular simulation is always significantly larger than DRMS values from the starting structure for the other simulation. To elaborate, let us consider the DRMS–vs–Closed structure for the simulation started from the Closed structure (black line in Figure 3 (b)). The fluctuations in DRMS remain less than 1.5 Å in the native basin for the Closed structure. Comparatively, the largest fluctuations in the simulations started from the Open structure bring it only within at most 3 Å of the Closed structure (blue line in Figure 3 (a)). That is, the opposing native basin is never reached.

We mention that all the DRMS values plotted in Figures 3 (a) and (b) are based on the first 200 residues. This is because the 14 tail residues, which form a helical segment, are very flexible and the helix unravels in either structure at a much lower temperature than the stable part of the protein. Thus, Figure 3 focuses on the rest of the protein. Additionally, although

we show results for $T = 0.7$ here, simulations at lower temperatures also give qualitatively similar results.

We perform an analogous fluctuation analysis for Model 2 which incorporates backbone hydrogen bonding interactions, Ramachandran propensities, and some residue specificity via MJ–type interactions. Figure 4 (a) shows the DRMS (of the first 200 residues) from the Open (black line) and Closed (blue line) structures for a simulation started from the Open structure. Similarly, Figure 4 (b) shows the DRMS traces for a simulation started from the Closed structure. Again, we observe very similar results as for Model 1: the fluctuations in the Open ensemble are larger than in the Closed ensemble, and no transition to the opposing structure is obtained in either simulation.

4.3 Path sampling: Models 1 and 2

Due to the inability of brute–force simulations to show transitions, we use weighted– ensemble path sampling to generate an ensemble of transition pathways with the aim of assessing path heterogeneity. In particular, we examine transitions in both directions for all the three models.

4.3.1 Transition from Open to Closed State—We first check whether our path sampling is sufficient by monitoring the flux into the target state. Figure 5 plots the WE results for probability fluxes obtained into the Closed state for both Models 1 and 2. The "time" axis is merely the number of τ intervals (where one interval contains 2000 LBMC steps). In both models, the fluxes reach linear regimes indicating that the observed transitions are not merely due to initial fast trajectories and the path ensemble is appropriately sampled.

The sensitivity to the models is also apparent in the fluxes shown in Figure 5: Model 2 (which includes hydrogen–bonding, Ramachandran propensities, and MJ–type residue specific interactions) has a smaller flux into the Closed state than Model 1. Residue–specific interactions are expected to roughen the energy landscape, consistent with the observed slowing of transition dynamics. However, the possible change in the Open state basin stability due to addition of these interactions is convoluted with the roughening of the landscape.

We further study the path ensemble by examining individual trajectories. Figure 6 shows, for Model 1, the DRMS from the Closed structure for four typical transitions started in the Open state as a function of time (total number of LBMC steps) obtained via WE path sampling. In contrast with the brute–force simulation in Figure 3 (a), each trajectory in Figure 6 gets to the Closed state (defined to be within a DRMS of 1.5 Å from the Closed structure). Although the trajectories arrive at the target state with different weights, the ones shown in the figures above are obtained after a simple resampling procedure,[70] and, thus, represent trajectories that arrive with relatively large probabilities. Resampling is a statistically rigorous procedure to prune an ensemble.[70] In our resampling scheme, a trajectory arriving at the target with weight w is kept with a probability w/w_{max}.

For trajectories that begin transitions at larger times (such as Trajectory 4 in Figure 6), a significant amount of time is spent in regions with large DRMS values from the Closed structure. Thus, in Figure 7 and beyond, we do not show the "dwell time" in the Open state.

To analyze the order of domain closing, and, in particular, to study possible heterogeneity in the path ensemble, we study the four Open–to–Closed trajectories of Figure 6 in more detail. Figure 7 plots the projection of the above four trajectories onto the plane of the BD–LID and LID–CORE distances.

The transition paths traced by the four Model 1 trajectories are significantly different. For Trajectory 1, the BD shuts after the flexible LID gets close to the CORE. For the following discussion, we call this pathway as Open–LID–BD–Closed (first the LID relaxes, and then the BD shuts close). On the other hand, Trajectory 3 shows a dramatically different behavior: the BD snaps shut before the flexible LID gets closer to the CORE (this pathway is labeled as Open–BD–LID–Closed). The other two trajectories are somewhere in between the two extremes.

To quantify heterogeneity in the path ensembles, we compare the ratio of trajectories in the two transition pathways. Specifically, we define a trajectory to follow the Open–LID–BD–Closed (lower right) pathway if it first visits the region $d_{56,163} < 10.0$ Å after last leaving the rectangular Open–state region defined by $d_{56,163} > 10.0$ Å and $d_{15,132} > 10.0$ Å. (As in Section 4.1, d_{ij} is the distance between alpha carbons of residues i and j.) On the other hand, a trajectory follows the Open–BD–LID–Closed (upper left) pathway if it first visits the region $d_{56,163} > 10.0$ Å after last leaving the above Open–state rectangular region. We find that for Open–to–Closed transition using Model 1, approximately 60% of the resampled trajectories follow Open–BD–LID–Closed pathway (akin to Trajectory 3 in Figure 7). The remaining 40% follow the Open–LID–BD–Closed pathway.

The Model 1 transitions exhibit good correspondence with experimental 'intermediates'. Figure 7 also compares the BD–LID and LID–CORE distances of several experimental crystal structures using triangles. Several experimental crystal structures (1DVR, 2AK3, 1AK2, 2BBW – see Supplementary Information) lie along the two transition paths. In particular, structures along the Open–LID–BD–Closed pathway approach 1DVR and 1AK2 structures, whereas structures along the Open–BD–LID–Closed pathway approach 2AK3 crystal structure. The identities of the intermediate structures differ from those found by Beckstein et al.52 There are two possible reasons for this discrepency: we use different models, and the effect of the ratcheting bias used by Beckstein et al.52 on the pathways and free–energy profiles (from observed populations) is not obvious. More details on the additional experimental structures chosen are given in the Supplementary Information.

Further, we look at a few intermediate structures for these two pathways. Figure 8 shows four intermediates along Trajectory 3 of Figure 7. The BD and LID domains near one another before the LID closes. On the other hand, Figure 9 shows four intermediates along Trajectory 1 in Figure 7. The closing of the LID, followed by snapping shut of the BD is clearly visible in the figure. As both Figures 8 and 9 show, the rest of the protein (*i.e.*, the CORE region) maintains a stable shape during the transformation.

To determine the sensitivity of the path ensemble to the model, we similarly analyzed results from Model 2 (which includes hydrogen bonding, Ramachandran propensities, and a level of residue specificity). A similar qualitative picture is obtained for Model 2. Figure 10 plots three of the resampled trajectories from the Open to the Closed structure using Model 2. The "dwell times" in the Open state have been omitted for clarity. Again, the symbols have the same meaning as in Figure 7 (except that open symbols represent the equilibrium fluctuations obtained using Model 2). The transition from the Open–to–Closed structure again takes place by the two pathways: Open–BD–LID–Closed and Open–LID–BD–Closed. The ratio of paths in the two pathways is the same as that for Model 1, and several additional experimental structures show BD–LID and LID–CORE distances similar to structures along the two types of transition pathways.

4.3.2 Transition from Closed to Open State—We also studied "reverse" transitions – from the Closed to the Open state. Figure 11 shows the flux into C as a function of "time" for both Models 1 and 2. Compared to Figure 5 for the transition from the Open to the

Closed state, the flux into state B is several orders of magnitude lower. This observation mirrors the previously described larger fluctuations in the Open state ensemble. Flux into the Open state for Model 2 with residue specific chemistry is higher than for Model 1, despite the expected roughening of the energy landscape. This necessarily reflects a free energy shift, suggesting MJ interactions de–stabilize the Closed state compared to a pure double–Gō model. Such a shift seems appropriate given that we do not model ligands which implicitly lead to more contacts in the Closed state and consequent over–stabilization in the Gō model.

For the Closed–to–Open transition using either model, we obtain pathways which mirror the Open–to–Closed transition: the LID fluctuates in the Closed state, and this is followed by the BD snapping open on a relatively fast time scale. For both Models 1 and 2, successful trajectories appear to follow only the Closed–LID–BD–Open pathway for Closed–to–Open transition (reverse order of the Open–BD–LID–Closed pathway in the Open–to–Closed transition direction). The absence of symmetry is surprising given our recent formal demonstration,39 and there seem to be two possible reasons. First, the transients for the Closed–BD–LID–Open pathway are long–lived. Lengthy transients are consistent with the low reverse reaction rates, shown in Figure 11, for both Models 1 and 2. Second, our state definitions may be flawed as discussed in Section 5.3.

To clarify the issue of the symmetry of path ensembles between forward and reverse directions, we constructed and path sampled Model 3.

4.4 Path ensemble symmetry analysis in Model 3

The slow Closed–to–Open transitions indicates that, for Models 1 and 2, the free energy of the Closed structure is significantly lower than that of the Open structure. As discussed in Section 2.3, this suggested the use of Model 3, which decreases in magnitude the favorable energy for contacts present only in the Closed state. That is, Model 3 reduces the free energy asymmetry between the Open/Closed states.

Model 3 thereby facilitates study of the symmetry between forward and reverse transitions. As shown in Figure 12, although the flux in the Open–to–Closed direction in Model 3 is higher than in the Closed–to–Open direction, the difference between the fluxes in the two directions is much less than that for Models 1 and 2. The increased Closed–to–Open rate implies that the relative stability of the Closed state is reduced compared to Models 1 and 2. Importantly, the relatively linear behavior of fluxes in both the directions implies our path sampling is sufficient – well beyond transients.

For Model 3, we examine the same classification of pathways as above. Both paths are frequently observed in both directions. In Figure 13, we show the ratio of probabilities of the two paths as a functions of simulation time in the two directions. Values in each window are averaged over 500 τ increments. The results for the Open–to–Closed direction (diamonds) are shown for a single simulation, whereas the Closed–to–Open transitions (circles) are shown for 6 independent simulations. Despite large fluctuations, the ratios of paths in the two directions are similar. We discuss the issue of path symmetry further in Section 5.

4.5 Intermediate detection

The path ensembles obtained can also be used to detect intermediates during the transitions in the two directions. A framework for analyzing the transition path ensemble was recently reviewed by E and vanden–Eijnden.4 In particular, transition, or reactive, trajectories can be traced along the configuration space (or, a projection of it) to compute populations of regions of the configuration space: regions with high populations correspond to intermediate

structures that will be observed more frequently during transitions. Note also the analyses proposed by Hummer.5

Figure 14 shows fractional populations in regions of configuration space obtained for Model 3 during transitions in the two directions. Both panels show significant similarity – two distinct channels (or, transition tubes) are visible, corresponding to the two paths. Regions with higher populations suggest metastability – *i.e.*, intermediates.

In both the figures, the region of the transition channels with LID–CORE distance in the 10– 15 Å range and with BD–LID distance similar to that in the Closed state is significantly more populated than other transition regions in the two directions. Such a structure is illustrated by Configuration 3 of Figure 8. This not only reiterates the observation in Figure 13 that the Open–BD–LID–Closed pathway is the more dominant transition pathway, but also suggests that the region noted above is a dominant intermediate in both the directions. A qualitatively similar result is obtained for the Open–to–Closed direction for Models 1 and 2 (as noted above, Closed–BD–LID–Open pathway was not observed for Models 1 and 2).

5 Discussion

5.1 Models

An important issue in any coarse–grained study is the sensitivity of the results to the particular model(s) used. To address this point, we used three different semi–atomistic models of adenylate kinase. For the models used, we find that the transition pathways are not significantly affected by the models we used. In particular, we find two dominant pathways (Open–LID–BD–Closed and Open–BD–LID–Closed) that occur in all the models. Although the rates vary considerably among models, we do not expect realistic kinetics in simplified models.

Our choice of models was governed by the basic requirement of obtaining full path sampling of conformational transitions – in order to study path ensembles, heterogeneity, and symmetry. Two of the models are based purely on structure (Gō model) and the other (Model 2) includes some level of residue specificity via Miyajawa–Jernigan interactions, as well as hydrogen bonding energies and Ramachandran propensities. In Model 2, the chemical energy terms are significant perturbations to the Gō interactions. (as quantified by MJ interactions between residues). This model is designed to be able to capture a minimum level of biochemistry. However, Model 2 still requires significant Gō–type interactions to stabilize the two physical states. In the future, we plan to utilize more detailed and explicit side chain–side chain and side chain–backbone interactions to reduce the dependence on Gō–type interactions.

Another limitation is that we did not consider the ligand in our path sampling simulations. The inclusion of ligand could influence the observed pathways significantly. We have plans for modeling ligand via "mixed models" that include all–atom ligands and binding sites, with a coarse–grained picture for the rest of the protein. Such an explicit inclusion of ligands, with the corresponding degrees of freedom of the unbound ligands in the Open form should reduce the dependence on arbitrary Gō interactions. A study with explicit ligands could require a higher dimensional progress coordinates to use in weighted ensemble simulations: one coordinate for protein structure (as is done in this work), and a second (or further) coordinates for the distance between ligands and the protein. Note that weighted ensemble can mix real– and configurational–space coordinates: it was originally designed for binding studies.36

5.2 Multiple paths

The two dominant pathways – Open–LID–BD–Closed and Open–BD–LID–Closed (and reverse on opening) – have also been observed by other groups using coarse–grained models with network analysis46 and molecular dynamics47,48. These works utilized double–well potentials. We showed that adding approximate chemical details to the double–well potential preserves the two types of pathways.

Further, Whitford *et al*.47 observed that different pathways were dominant in the two directions. On the other hand, results of Lu and Wang48 suggest that any path is similarly favored in both directions. We discuss this symmetry of pathways below.

5.3 Path symmetry

Recently, we investigated the conditions when there should be symmetry – *i.e.*, when pathways in the forward and the reverse directions occur with the same ratio.39 We show that exact symmetry will hold when a specific (equilibrium–based) steady state is enforced. Approximate symmetry is expected if the initial and final states are well–defined physical basins lacking slow internal timescales, so that trajectories emerging from a state "forget" the path by which they entered. Figure 13 suggests that the ratio of the two different pathways in the two directions is very similar for Model 3, which was fully path sampled in both directions.

Such a symmetry is clearly absent from our results (even after accounting for statistical fluctuations) in Models 1 and 2. Although we observed transitions in both the directions for all the models, Closed–to–Open transitions in all the models (especially in Models 1 and 2) are harder to obtain. In particular, the Closed–BD–LID–Open pathway is not observed in our simulations for Models 1 and 2. This indicates that the conditions required for symmetry39 are not met in one or more ways: transitions in the Closed–to–Open direction may be dominated by transients and/or there may be internal barriers within the Closed state. We are currently working on developing WE path sampling methods that allow steady states to be sampled directly and efficiently62 and we hope to investigate the issue of symmetry further. Related steady–state methods are already available.71⁻73

The above discussion suggests possible reasons behind the lack of path symmetry observed by Whitford *et al*.47 and path symmetry observed by Lu and Wang48. The lack of symmetry observed by Whitford *et al*.47 is, perhaps, due to transitions being studied from one crystal structure conformation (instead of from an ensemble of conformations relevant to that state) to an ensemble of conformations of the opposing state: possible internal barriers within a state may not allow for symmetry to be observed. On the other hand, Lu and Wang48 study transitions from an ensemble of one state to the ensemble of the opposing state.

5.4 CPU time and efficiency

One of the basic goals of this work was to determine the level of detail we can include in a model, while still allowing for full sampling of the path ensemble. Thus, we now discuss the computational effort that was required. All simulations were performed on single 3 GHz Intel processors. The results shown for Model 1 in the Open–to–Closed direction took approximately one week of single CPU time. More simulation was performed in the Closed–to–Open direction, requiring 3-4 weeks of single CPU time. The results for Model 2 were obtained using approximately the same time as Model 1. For Model 3, the Closed–to–Open transition was not much harder to obtain than the Open–to–Closed transition, and a simulation in each direction required approximately two weeks of single CPU time. Due to the low CPU usage for obtaining path ensembles for the models used here, obtaining path ensembles of better models using WE is possible. See Section 5.1.

It is not hard to estimate the efficiency of WE simulation compared to brute–force. The transition rates determined from WE simulations indicate the time required for brute–force simulations to achieve transitions and hence permit estimates of efficiency. For example, the rate obtained for Closed–to–Open transition for Model 3 is $2.5 \times 10^{-6}/\tau$. Thus, one brute–force transition can be estimated to require the reciprocal amount of time. Since 2000τ require approximately one week of computing, BF is estimated to take approximately 4 years for a single transition. In contrast WE yielded 50 transitions after resemapling (*i.e.*, 50 transitions with equal weights), in about two weeks of single–processor computing. (Before resampling, there were about 3000 WE transitions for each simulation). WE is thus significantly more efficient than BF. For transitions in the other direction and/or other models, a qualitatively similar picture for efficiency emerges.

6 Conclusions

We applied weighted ensemble (WE) path sampling to generate ensembles for conformational transitions between Open (apo) and Closed (holo) forms of adenylate kinase using semi–atomistic models of the protein. The models have an all–atom backbone including beta–carbons, along with varying levels of chemical specificity, and represent a significant jump in complexity compared to previous models studied with WE. No additional driving force was used to enable the transitions. We showed that conformational transitions in both directions are possible for such models via WE, while brute–force simulations are less efficient. Given the relatively small computational effort required for observing transitions using WE (weeks of single–CPU time), more detailed models can be used for full path sampling. Models with further reduction in Gō–type interactions are needed, along with ligand modeling, to study the specific enzyme biochemistry.

All the models show significant hereteogeneity in the transition pathways, consistent with previous work and experimental structures. Two dominant pathways are observed, characterized by the order in which the flexible lid and the AMP binding domains close. Although the transition rates (in terms of Monte Carlo steps) varied significantly depending upon the model used, similar dominant pathways are obtained across the models. The model that allows significant transitions in both forward and reverse directions shows an approximate symmetry of pathway populations, consistent with a recently derived symmetry rule.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

[1]. Berg, JM.; Stryer, L.; Tymoczko, JL. Biochemistry. Freeman; New York: 2002. p. 227-258.

[2]. Hammes GG. Biochem. 2002; 41:8221–8228. [PubMed: 12081470]

[3]. Berg, JM.; Stryer, L.; Tymoczko, JL. Biochemistry. Freeman; New York: 2002. p. 269-272.

[4]. E W, vanden Eijnden E. Annu. Rev. Phys. Chem. 2010; 61:391–420. [PubMed: 18999998]

[5]. Hummer G. J. Chem. Phys. 2004; 120:516–523. [PubMed: 15267886]

[6]. Dellago C, Bolhuis PG, Csajka FS, Chandler D. J. Chem. Phys. 1998; 108:1964–1977.

[7]. Bolhuis PG, Chandler D, Dellago C, Geissler PL. Annu. Rev. Phys. Chem. 2002; 53:291–318. [PubMed: 11972010]

[8]. van Erp TS, Moroni D, Bolhuis PG. J. Chem. Phys. 2003; 118:7762.

[9]. van Erp TS, Bolhuis PG. J. Comp. Phys. 2005; 205:157–181.

[10]. Faradjian AK, Elber R. J. Chem. Phys. 2004; 120:10880–10889. [PubMed: 15268118]

[11]. West AMA, Elber R, Shalloway D. J. Chem. Phys. 2007; 126:145104. [PubMed: 17444753]

[12]. vanden Eijnden E, Venturoli M, Ciccotti G, Elber R. J. Chem. Phys. 2008; 129:174102. [PubMed: 19045328]

[13]. Allen RJ, Warren PB, ten Wolde PR. Phys. Rev. Lett. 2005; 94:018104. [PubMed: 15698138]

[14]. Allen RJ, Frenkel D, ten Wolde PR. J. Chem. Phys. 2006; 124:024102. [PubMed: 16422566]

[15]. Borrero EE, Escobedo FA. J. Chem. Phys. 2007; 127:164101. [PubMed: 17979313]

[16]. Lyman E, Zuckerman DM. J. Phys. Chem. B. 2007; 111:12876–12882. [PubMed: 17935314]

[17]. Juraszek J, Bolhuis PG. Biophys. J. 2008; 95:4246–4257. [PubMed: 18676648]

[18]. Elber R. Biophys. J. 2007; 92:L85–L87. [PubMed: 17325010]

[19]. Hu J, Ma A, Dinner AR. J. Chem. Phys. 2006; 125:114101. [PubMed: 16999460]

[20]. Radhakrishnan R, Schlick T. Proc. Natl. Acad. Sci. 2004; 101:5970–5975. [PubMed: 15069184]

[21]. Schlitter J, Engels M, Kruger P, Jacoby E, Wollmer A. Mol. Simul. 1993; 10:291–309.

[22]. Ma J, Karplus M. Proc. Natl. Acad. Sci. 1997; 94:11905–11910. [PubMed: 9342335]

[23]. Apostolakis J, Ferrara P, Caflisch A. J. Chem. Phys. 1999; 110:2099–2108.

[24]. Elber R. Chem. Phys. Lett. 1987; 139:375.

[25]. Ulitsky A. J. Chem. Phys. 1990; 92:1519.

[26]. Fischer S, Karplus M. Chem. Phys. Lett. 1992; 194:252.

[27]. Sevick EM, Bell AT, Theodorou DN. J. Chem. Phys. 1993; 98:3196.

[28]. Gillilan RE, Wilson KR. J. Chem. Phys. 1992; 97:1757.

[29]. Elber R, Ghosh A, Cardenas A, Stern H. Adv. Chem. Phys. 2004; 126:123.

[30]. Passerone D, Ceccarelli M, Parrinello M. J. Chem. Phys. 2003; 118:2025.

[31]. Olender R, Elber R. J. Chem. Phys. 1996; 105:9299.

[32]. Dill KA, Bromberg S, Yue K, Fichig KM, Yee DP, Thomas PD, Chan HS. Protein Sci. 1995; 4:561–602. [PubMed: 7613459]

[33]. Onuchic JN, Wolynes PG. Curr. Opin. Struct. Biol. 2004; 14:70–75. [PubMed: 15102452]

[34]. Zuckerman DM. J. Phys. Chem. B. 2004; 108:5127–5137.

[35]. Zhang BW, Jasnow D, Zuckerman DM. Proc. Natl. Acad. Sci. 2007; 104:18043–18048. [PubMed: 17984047]

[36]. Huber GA, Kim S. Biophys. J. 1996; 70:97–110. [PubMed: 8770190]

[37]. Maragakis P, Karplus M. J. Mol. Biol. 2005; 352:807–822. [PubMed: 16139299]

[38]. Chennubhotla C, Bahar I. PLoS Comput. Biol. 2007; 3:1716–1726. [PubMed: 17892319]

[39]. Bhatt, D.; Zuckerman, D. arXiv. [accessed February 11, 2010]. http://arxiv.org/abs/ 1002.2402http://arxiv.org/abs/1002.2402

[40]. Pontiggia F, Zen A, Micheletti C. Biophys. J. 2008; 95:5901–5912. [PubMed: 18931260]

[41]. Shapiro YE, Sinev MA, Sineva EV, Tugarinov V, Meirovitch E. Biochem. 2000; 39:6634–6644. [PubMed: 10828981]

[42]. Hanson JA, Duderstadt K, Watkins LP, Bhattacharyya S, Brokaw J, Chu J-W. Proc. Natl. Acad. Sci. 104(46):18055–18060.

[43]. Muller CW, Schlauderer G, Reinstein J, Schultz GE. Structure. 1996; 4:147–156. [PubMed: 8805521]

[44]. Muller CW, Schultz GE. J. Mol. Biol. 1992; 224:159–177. [PubMed: 1548697]

[45]. Whitford PC, Miyashita O, Levy Y, Onuchic JN. J. Mol. Biol. 2007; 366:1661–1671. [PubMed: 17217965]

[46]. Chu JW, Voth GA. Biophys. J. 2007; 93:3860–3871. [PubMed: 17704151]

[47]. Whitford PC, Gosavi S, Onuchic JN. J. Biol. Chem. 2008; 283:2042–2048. [PubMed: 17998210]

[48]. Lu Q, Wang J. J. Am. Chem. Soc. 2008; 130:4772–4783. [PubMed: 18338887]

[49]. Arora K, Brooks CL. Proc. Natl. Acad. Sci. 2007; 104:18496–18501. [PubMed: 18000050]

[50]. Kubitzki MB, de Groot BL. Structure. 2008; 16:1175–1182. [PubMed: 18682219]

[51]. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Karplus M, Hubner CG, Kern D. Nat. 2007; 450:838–844.

[52]. Beckstein O, Denning EJ, Perilla JR, Woolf TB. J. Mol. Biol. 2009; 394:160–176. [PubMed: 19751742]

[53]. Zuckerman DM, Woolf TB. J. Chem. Phys. 1999; 111:9475–9484.

[54]. Zuckerman, DM.; Woolf, TB. arXiv. [accessed September 27, 2002]. http://arxiv.org/abs/ physics/0209098http://arxiv.org/abs/physics/0209098

[55]. Best RB, Chen Y-G, Hummer G. Struct. 2005; 13:1755–1763.

[56]. Levy Y, Cho SS, Shen T, Onuchic JN, Wolynes PG. Proc. Natl. Acad. Sci. 2005; 102:2373– 2378. [PubMed: 15701699]

[57]. Mamonov AB, Bhatt D, Cashman DJ, Ding Y, Zuckerman DM. J. Phys. Chem. B. 2009; 113:10891–10904. [PubMed: 19594147]

[58]. Rojnuckarin A, Kim S, Subramanian S. Proc. Natl. Acad. Sci. 1998; 95:4288–4292. [PubMed: 9539729]

[59]. Fisher EW, Rojnuckarin A, Kim S. J. Molec. Struct.–Themochem. 2000; 529:183–191.

[60]. Zhang BW, Jasnow D, Zuckerman DM. J. Chem. Phys. 2010; 132:054107. [PubMed: 20136305]

[61]. Zhang, BW.; Jasnow, D.; Zuckerman, D. arXiv. [accessed February 16, 2009]. http://arxiv.org/ abs/0902.2772http://arxiv.org/abs/0902.2772

[62]. Bhatt D, Zhang BW, Zuckerman D. J. Chem. Phys. 2010; 133:014110. [PubMed: 20614962]

[63]. Lovell SC, Word JM, Richardson JS, Richardson DC. Proteins: Structure Function and Genetics. 2000; 40:389–408.

[64]. Miyazawa S, Jernigan RL. Macromolecules. 1985; 18:534–552.

[65]. Miyazawa S, Jernigan RL. J. Mol. Biol. 1996; 256:623–644. [PubMed: 8604144]

[66]. Jernigan RL, Bahar I. Curr. Op. Struct. Biol. 1996; 6:195–209.

[67]. Gan HH, Tropsha A, Schlick T. J. Chem. Phys. 2000; 113

[68]. Shimada J, Kussell EL, Shakhnovich EI. J. Mol. Biol. 2001; 308:79–95. [PubMed: 11302709]

[69]. Shimada J, Shakhnovich EI. Proc. Natl. Acad. Sci. 2002; 99:11175–11180. [PubMed: 12165568]

[70]. Liu, JS. Monte Carlo Strategies in Scientific Computing. Springer; New York: 2004.

[71]. Warmflash A, Bhimalapuram P, Dinner AR. J. Chem. Phys. 2007; 127:154112. [PubMed: 17949137]

[72]. Dickson A, Warmflash A, Dinner AR. J. Chem. Phys. 2009; 130:074104. [PubMed: 19239281]

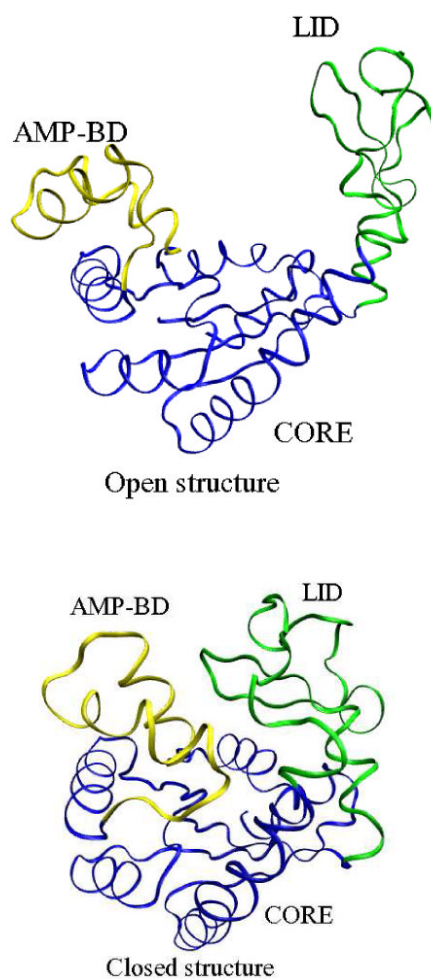[73]. vanden Eijnden E, Venturoli M. J. Chem. Phys. 2009; 131:044120. [PubMed: 19655850]

Figure 1.
Crystal structures of adenylate kinase. The blue segment represents the core of the protein (CORE), the yellow segment is the AMP binding domain (BD), and the green segments is the flexible lid (LID). The left figure is the Apo (or Open) form, and the right figure is holo (Closed) form.
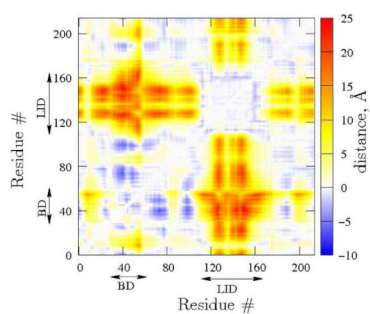
Figure 2.
Map of the difference in the inter–residue distances of the Open from the Closed crystal structure. The map is, by definition, symmetric about the diagonal. The white space on the diagonal just implies that we do not calculate distances between residue pairs less than 2 residues apart along the chain. The residues corresponding to the LID and BD are labeled and the unlabeled residues form the CORE.
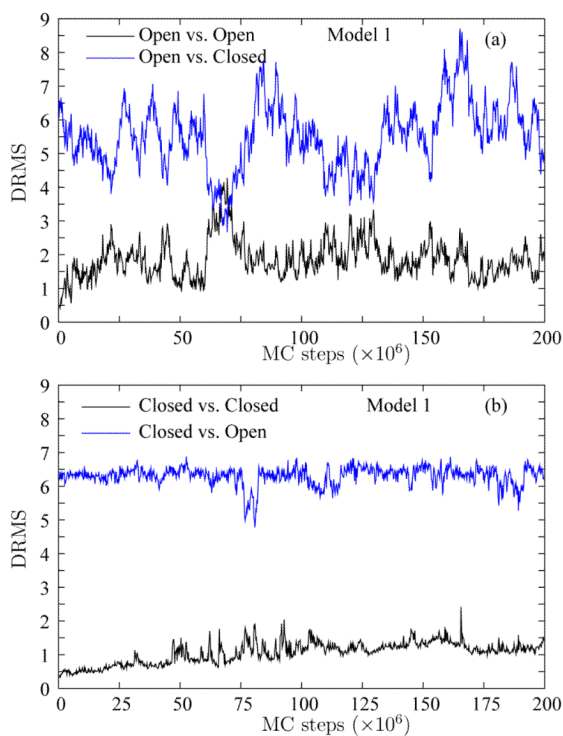
Figure 3.
Stability of the native basins in Model 1. The DRMS in Model 1 (pure Gō) is shown for two simulations (a) starting from the Open structure and (b) starting from the Closed structure. For each simulation, we show the DRMS from the starting structure (black line) and the opposing structure (blue line). Neither simulation show a transition to the opposing structure, but the scales of fluctuations are very different.
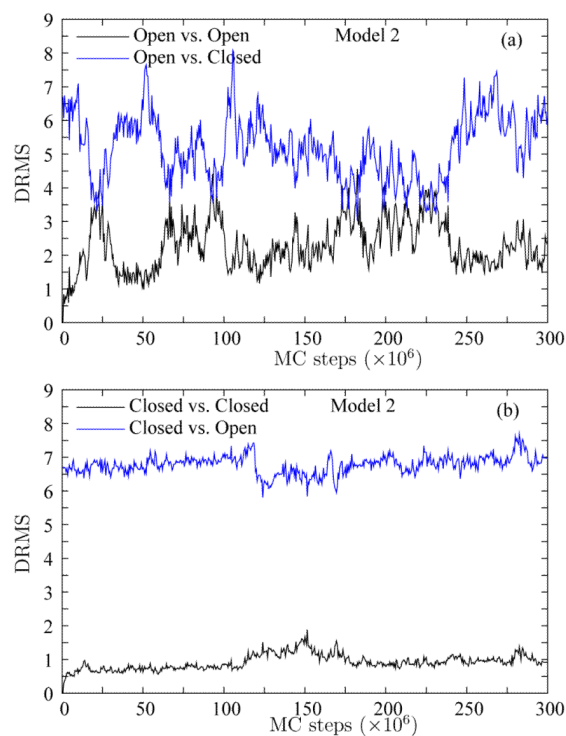
Figure 4.
Stability of the native basins in Model 2. The DRMS in Model 2 (Gō, H-bonding, Ramachandran propensities, and MJ–type residue specific interactions) is shown for two simulations (a) starting from the Open structure and (b) starting from the Closed structure. For each simulation, we show the DRMS from the starting structure (black line) and the opposing structure (blue line). Neither simulation show a transition to the opposing structure, but the scales of fluctuations are very different.
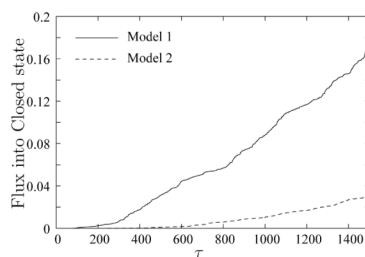
Figure 5.
Comparison of the probability fluxes into the Closed state for two models. The fluxes from a pure double–Gō system (Model 1, solid line) and a system with considerable chemical specificity (Model 2, dashed line) are plotted as functions of WE time intervals (each τ interval is 2000 LBMC steps). In both cases, the fluxes reach approximately linear regimes, suggesting transient effects have attenuated.
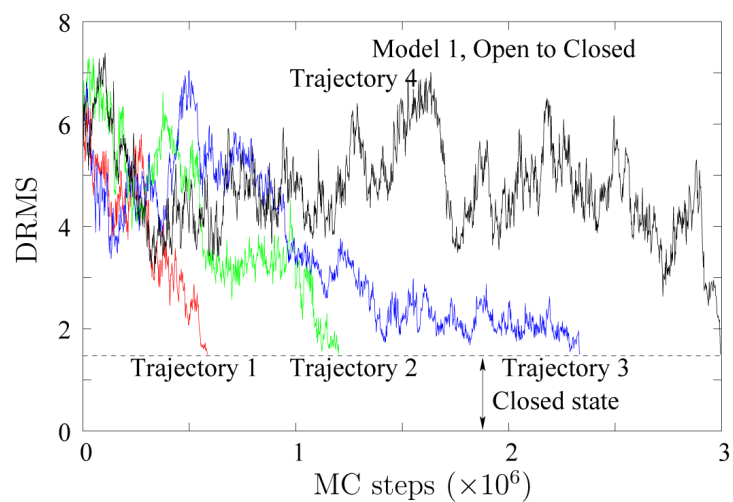
Figure 6.
Open–to–Closed transitions observed for Model 1. Four typical DRMS traces are shown, measured from the Closed crystal structure. The Closed basin is delimited by a DRMS of 1.5 Å from the Closed crystal structure.
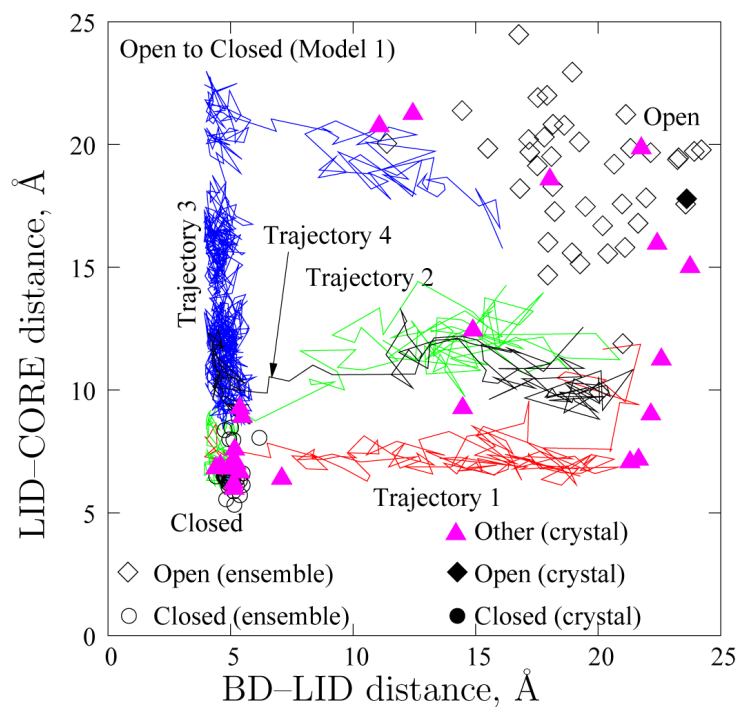
Figure 7.
Path heterogeneity in Model 1 transitions. Four typical trajectories for the Open–to–Closed transition (cf. Figure 6) are shown via distances between the CORE and LID (ordinate) and between the BD and the LID (abscissa). Also depicted are experimental configurations, including those from which our double–Gō models were constructed (black filled symbols) and a number of others (pink triangles) which are tabulated in the Supplementary Information. Open symbols depict the equilibrium ensembles obtained from simulation. The "dwell" times for the trajectories in the Open state have been excluded for clarity.
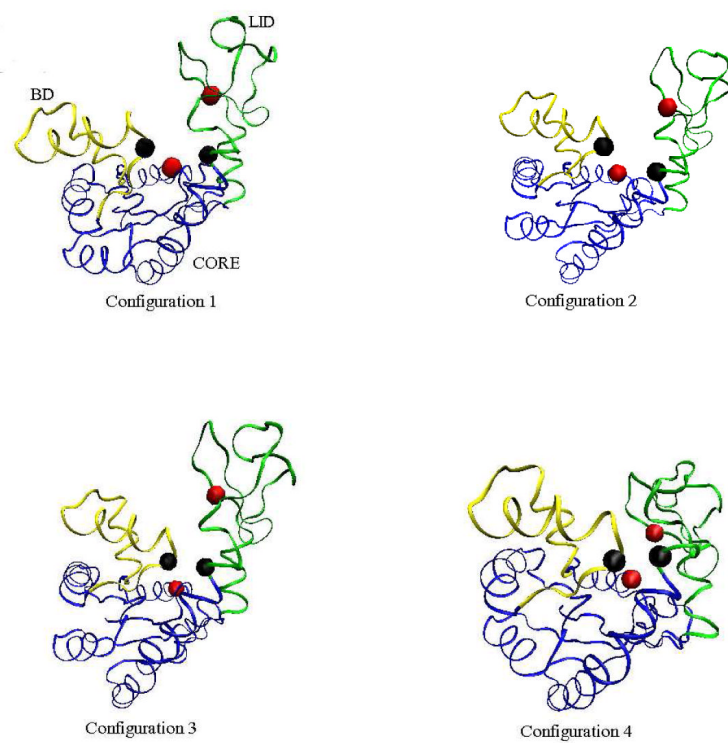
Figure 8.
Four time–ordered, representative structures along the Open–BD–LID–Closed path obtained for Trajectory 3 in Figure 7 (Model 1). The CORE domain is blue, the BD is yellow, and the LID is green. The configurations correspond to (296000, 802000, 822000, 1496000) MC steps, respectively, in Trajectory 3 of Figure 6.
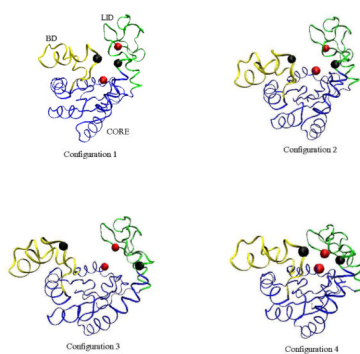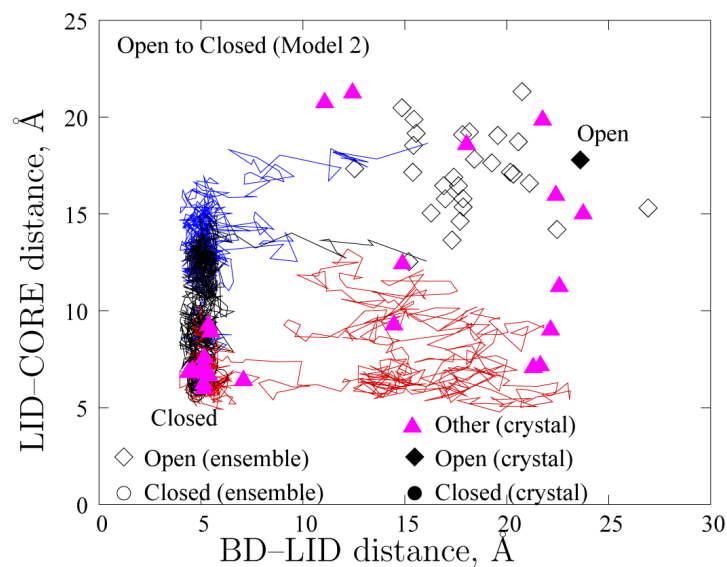
Figure 9.
Four time–ordered, representative structures along the Open–LID–BD–Closed path obtained for Trajectory 1 in Figure 7 (Model 1). The CORE domain is blue, the BD is yellow, and the LID is green. The configurations correspond to (128000, 206000, 306000, 480000) MC steps, respectively, in Trajectory 1 of Figure 6.

Figure 10.
Path heterogeneity in Model 2. Three typical trajectories for the Open–to–Closed transition are shown via distance between the CORE and LID (ordinate) versus the distance between the BD and the LID (abscissa). Also depicted are experimental configurations, including those from which our double–Gō models were constructed (black filled symbols) and a number of others (pink triangles) which are tabulated in the Supplementary Information. Open symbols depict the equilibrium ensembles obtained from simulation. The "dwell" times for the trajectories in the Open state have been excluded for clarity.
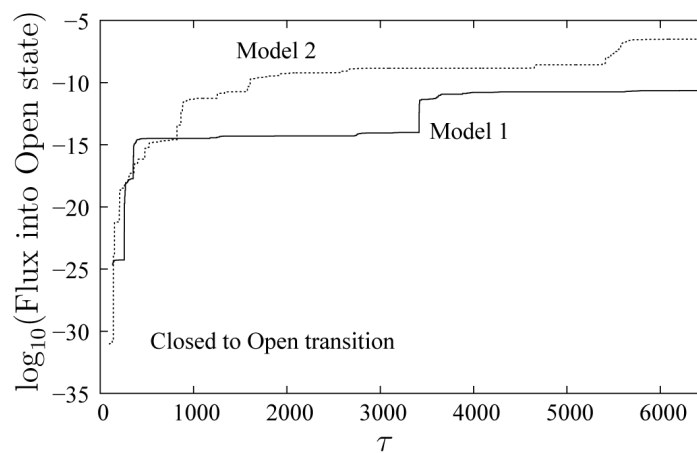
Figure 11.
A comparison of probability fluxes into the Open state for two models as a function of WE time increment (each τ is 2000 LBMC steps). The solid line is for Model 1, whereas the dashed line is for Model 2. The log scale emphasizes the small amount of flux in Closed–to–Open direction for both the models, as compared to fluxes in the reverse direction.
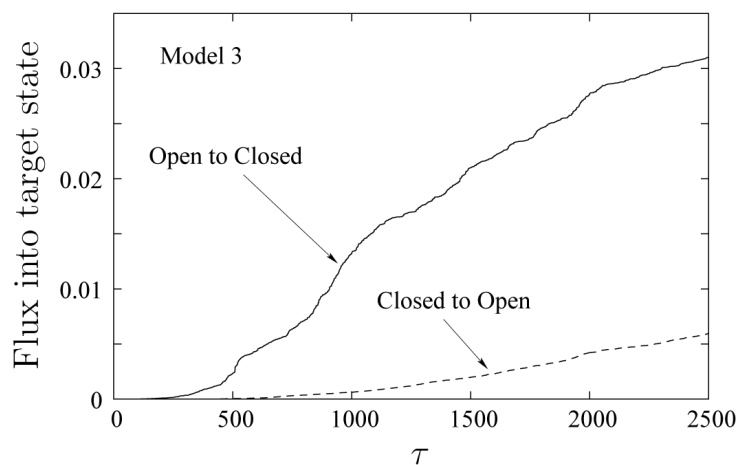
Figure 12.
Probability flux in either direction for Model 3 as a function of WE time increment (each τ is 2000 LBMC steps). For this model, the Closed–to–Open flux is of the same order of magnitude as in the reverse direction.
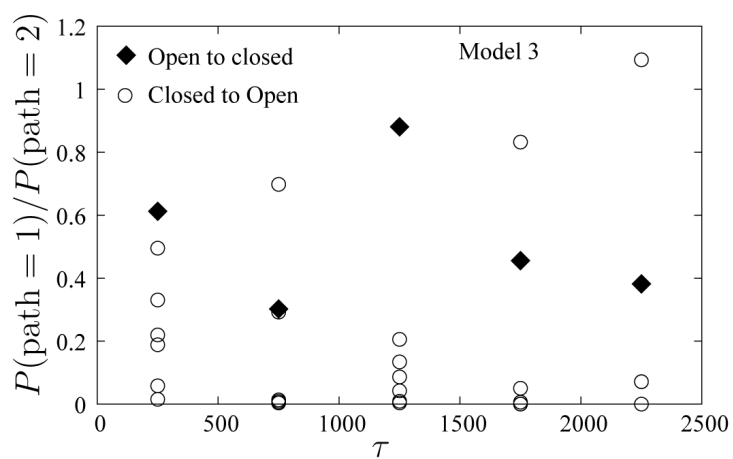
Figure 13.
Ratio of probabilities of the two paths, 1 and 2, in the Open–to–Closed (diamonds) and Closed–to–Open directions (circles). Results from six independent WE simulations are shown in the Closed–to–Open direction to highlight the fluctuations. All data points are window averaged for 500 τ.
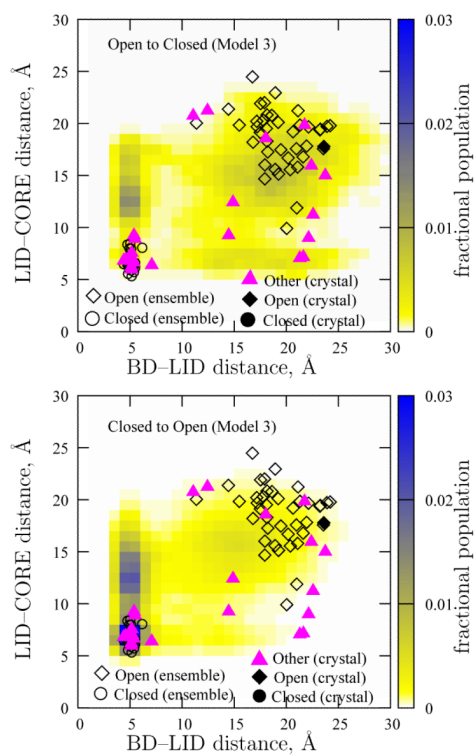
Figure 14.
Fractional populations observed in the transition path ensemmble. The top panel shows fractional populations for the Open–to–Closed direction, whereas the bottom panel is for the Closed–to–Open direction. Also depicted are experimental configurations, including those from which our double–Gō models were constructed (black filled symbols) and a number of others (pink triangles) which are tabulated in the Supplementary Information. Open symbols depict the equilibrium ensembles obtained from simulation.