Molecular Simulation
Vol. 40, No. 10–11, May 2014, 1–30

REVIEW

Sampling large conformational transitions: Adenylate kinase as a testing ground

Sean L. Seyler[a] and Oliver Beckstein[a,*]

[a] *Center for Biological Physics, Department of Physics, Arizona State University, Tempe AZ 85287-1504, USA*

(*22 December 2013*)

A fundamental problem in computational biophysics is to deduce the function of a protein from the structure. Many biological macromolecules such as enzymes, molecular motors, or membrane transport proteins perform their function by cycling between multiple conformational states. Understanding such conformational transitions, which typically occur on the millisecond to second time scale, is central to understanding protein function. Molecular dynamics (MD) computer simulations have become an important tool to connect molecular structure to function but equilibrium MD simulations are rarely able to sample on time scales longer than a few microseconds—orders of magnitudes shorter than the time scales of interest. A range of different simulation methods have been proposed to overcome this time-scale limitation. These include calculations of the free energy landscape and path sampling methods to directly sample transitions between known conformations. All these methods solve the problem to sample infrequently occupied but important regions of configuration space. Many path-sampling algorithms have been applied to the closed $\leftrightarrow$ open transition of the enzyme adenylate kinase (AdK), which undergoes a large, clamshell-like conformational transition between an open and a closed state. Here we review approaches to sample macromolecular transitions through the lens of AdK. We focus our main discussion on the current state of knowledge—both from simulations and experiments—about the transition pathways of ligand-free AdK, its energy landscape, transition rates, and interactions with substrates. We conclude with a comparison of the discussed approaches with a view towards quantitative evaluation of path-sampling methods.

Keywords: molecular dynamics simulations; transition path; rare events; sampling; free energy; potential of mean force; adenylate kinase

1 Introduction

Deducing a protein's function from its structure is a central problem in computational biophysics. Many proteins such as enzymes, molecular motors, or membrane transport cycle between multiple conformational states in order to fulfill their function. Thus understanding conformational transitions is central to understanding protein function [1–4]. Such motions typically occur on the millisecond to second time scale and are difficult to study experimentally and computationally. Although equilibrium molecular dynamics (MD) simulations are able to accurately simulate biological macromolecules and have become an important tool in connecting molecular structure to function, equilibrium simulations are rarely able to sample on time scales longer than a few microseconds—orders of magnitudes shorter than the time scales of interest. A key challenge is to efficiently sample realistic macromolecular conformational transitions and to quantitatively relate the simulations to experimentally measurable quantities.

Rare event sampling comprises a set of computational methods used to intelligently sample the important regions in the phase space of a dynamical system. A wide range of path sampling tech-

*Corresponding author. Email: oliver.beckstein@asu.edu

ISSN: 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927022.2014.919497
http://www.informaworld.com

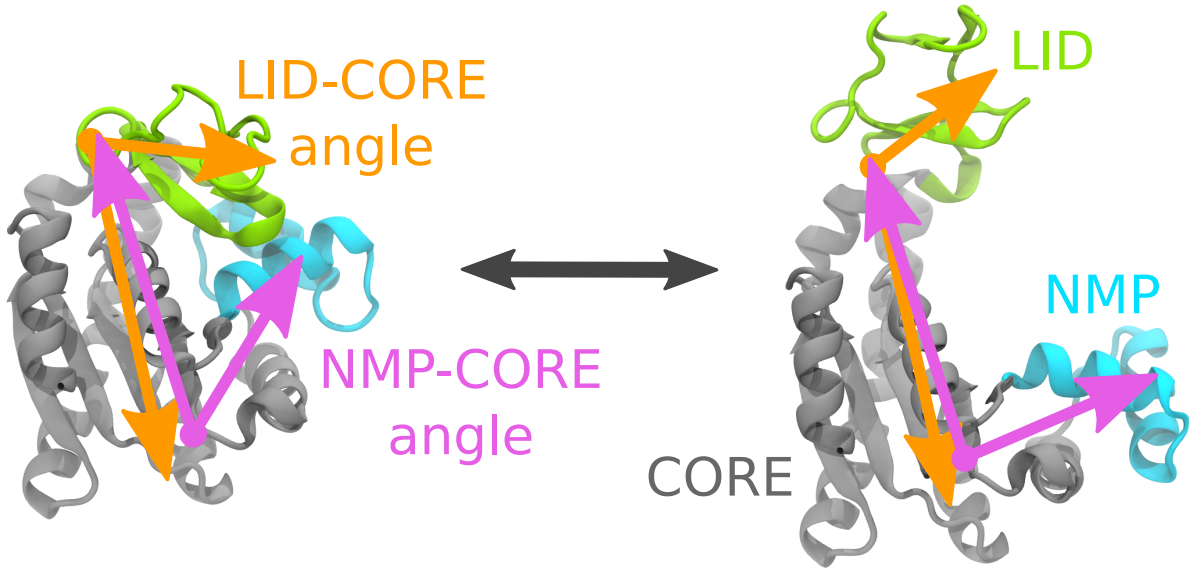


Figure 1. The enzyme AdK undergoes a large conformational change between a closed (left) and open (right) conformation, during which the LID domain (yellow) and NMP domain (blue) move relative to the CORE (gray). One choice of variables for quantifying the two bending motions is the LID-CORE angle (orange) and the NMP-CORE angle (purple) [31].

niques have been proposed to simulate the transition events in regions of phase space that would otherwise be sampled too infrequently by equilibrium MD. The simulation of conformational transitions can be roughly subdivided into three distinct, but interdependent objectives: (1) accurately simulate conformational transition ensembles, (2) determine underlying free energy landscapes, and (3) predict transition probabilities and rates. A wide range of computational approaches to simulate rare transition events have been proposed [5–24], which all, however, necessitate certain assumptions; furthermore, it is not known how accurately these different methods reproduce real transition events. Nevertheless, due to their computational efficiency in comparison to pure equilibrium MD, they remain important tools for generating transition ensembles and gaining intuition for the dynamics of macromolecular systems.

Adenylate kinase (AdK) is an interesting test system in the study of the general mechanisms of conformational transitions. Firstly, it undergoes a large-scale conformational change between open and closed states through hinge-like motions [25] (Figure 1), making the transition relatively easy to conceptualize. As there is still debate about whether the motions are more hinge-like or cracking-like, or possibly a combination of hinging and cracking, we will henceforth refer to the flexible regions (and associated residues) where hinging and/or cracking occurs as the labile regions of the protein [26]. Secondly, there is an abundance of known intermediate crystal structures that span the configuration space between the open and closed crystal structures. Thirdly, its conformational transition has been extensively studied by both biophysical and computational approaches. Finally, the transition mechanism itself is still not fully understood, partly owing to the infrequency and briefness of the transition event, which challenges both experimental and computational approaches.

This review will focus on the AdK enzyme of the mesophilic bacterium *Escherichia coli* (AK_{eco}) since almost all simulations study this specific protein. Thermophilic AdKs have also been computationally studied [27, 28] as have been related kinases such as guanylate kinase [29, 30]. The existence of a large family of related proteins with similar structures but different functionalities provides the opportunity to study, for instance, the molecular determinants for thermostability, tuning of reaction rates for specific temperatures, and selectivity for substrates. However, this review will restrict itself to the more narrow theme of discussing AK_{eco} as a model system to study macromolecular conformational transitions using computer simulation approaches.

1.1 *AdK as a model for macromolecular conformational transitions*

The enzyme adenylate kinase (AdK, EC 2.7.4.3) plays an important role in energy homeostasis by maintaining appropriate relative concentrations of ATP, ADP, and AMP in the cell [32, 33]. It catalyzes the phosphoryl transfer from an ATP to an AMP molecule in the presence of a catalytic magnesium ion:

$$\mathrm{Mg^{2+}ATP} + \mathrm{AMP} \rightleftharpoons \mathrm{Mg^{2+}ADP} + \mathrm{ADP}. \tag{1}$$

AdK is a relatively small protein—$\mathrm{AK_{eco}}$ contains 214 residues with a molecular weight of 23.6 kDa—and consists of three distinct domains, shown in Fig. 1. The *CORE* domain consists of the typical α/β-sandwich fold of a P-loop NTPase [34, 35]. After substrate binding, the ATP-binding *LID* domain (also called the ATPlid domain) and the AMP-binding *NMP* domain ('nucleomonophosphate'; also called the AMPbd domain) close over the substrates. Specifically, $\mathrm{Mg^{2+}ATP}$ is bound both by the P-loop, which is part of the CORE domain, and the LID domain, whereas AMP is held in place by the NMP domain. The phosphoryl transfer reaction is catalyzed by the exact positioning of the reactants, the magnesium ion bound to ATP, and by preventing the diffusion of the inorganic phosphate ion away from AMP, which would result in an incomplete reaction, namely the hydrolysis of ATP [36]. After the completion of the reaction, the enzyme opens up and the products (two ADP molecules) diffuse away [36].

Experimentally, it is known that the opening of the ADP-bound enzyme is the rate limiting step of the reaction with a rate on the order of 300 $\mathrm{s^{-1}}$ [36] and rates of other steps have also been measured [4]. Both experimental [28, 36–40] and computational work (reviewed in more detail below) have contributed to an emerging picture of the conformational changes involved. AdK will also undergo a conformational change in the absence of substrates [28, 39–41], which makes the study of its transition particularly amenable to numerical simulation because it removes the complexity of accurately simulating the protein-ligand interactions. The dynamics of apo enzymes is strongly correlated with the structural changes involved in ligand binding [42–45] and NMR experiments have shown that apo-AdK stochastically samples the same open and closed conformational modes as the ligand-bound enzyme [39, 40].

AdK is also well characterized structurally, with more than 50 crystal structures of $\mathrm{AK_{eco}}$ and homologous proteins that range from closed to open conformations, suggestive of a sequence of intermediate conformations along a transition path [31, 46–48]. Out of these structures only four (PDB IDs 4ake, 2rh5, 3umf, 3gmt) were crystallized in the apo form, i.e. in the absence of a substrate-like ligand, and these structures represent the open conformations of AdK. The majority of structures with an inhibitory ligand or natural substrates adopt a closed or intermediate conformation. Thus, for simulations, there is no shortage of reliable starting closed-state ligand-bound structures or open-state apo structures. The confluence of the aforementioned favorable factors has made AdK an ideal testbed that has—and continues to be—a catalyst for the development and application of novel path sampling techniques, and for generating insight into the mechanistic workings of conformational transitions.

1.2 *Collective variables*

In characterizing conformational transitions, one is often interested in identifying energetic basins of attraction that define metastable states and the transition state region. The $3N$-dimensional configuration space of a system is often projected onto a lower-dimensional space spanned by a small number of 'collective variables,' e.g. 'reaction coordinates' and 'order parameters.' A reaction coordinate should capture the dynamics of a system and can be used to define a transition state ensemble. Order parameters characterize a system's energetic basins of attraction into reactant and product regions, providing a quantitative means for defining the populations of metastable reactant and product states [7, 49, 50]. Appropriate reaction coordinates and order parameters for a given protein can be selected heuristically, through trial and error, by choosing

essential variables (e.g., using normal mode analysis (NMA) or principal component analysis (PCA)), constructing isocommittor surfaces, using path-based (transition path sampling) or state-based (string method) methods, and/or applying machine learning methods [51, 52].

A transition is most simply represented in the context of phase/configuration space trajectories (in the free energy landscape) that begin from a local energetic minimum (state A), proceed through an energetic bottleneck corresponding to a saddle point, and end at another energetic basin (state B). A dynamical system of a sufficiently low number of dimensions will have a transition bottleneck located at a saddle point which corresponds to its phase space separatrix. High-dimensional dynamical systems, however, often contain numerous free energy saddle points that need not correspond to meaningful dynamical bottlenecks, but the *transition state surface* is defined for a general dynamical system by its separatrix. The *commitor*, or commitment probability, $p_i(\mathbf{x}, \tau)$, is the probability that trajectories initialized in configuration $\mathbf{x}$ will end in state i in a (short) time τ. Stated another way, a trajectory initialized at $\mathbf{x}$ will commit to state i in a time τ with probability $p_i(\mathbf{x}, \tau)$. An isocommitor surface is a subspace of phase space such that the committor is constant. Thus, for a transition system with two stable states A and B, it is clear that the transition state surface (and separatrix) must lie on the isocommitor surface corresponding to equal commitment probability, i.e. $p_A = p_B = 1/2$. Isocommitor surfaces are sometimes used as putative reaction coordinates. [7, 49]

The construction of an optimal reaction coordinate that quantitatively describes the (diffusive) dynamics between two states is guaranteed when the behavior of an equilibrium system can be approximated by Markovian dynamics [53]. Qualitatively good reaction coordinates should capture the important underlying energy barrier(s) responsible for the dynamical bottleneck such that the projected motion along the coordinate is consistent with the kinetics of the full system; poor choices, however, do not lead to a separation between transition time scales and characteristic waiting times in metastable equilibria due to memory effects that persist over long time scales. In order for the Markov assumption to hold, which is necessary for computing transition probabilities in transition state theory (TST) or transition tubes in transition path theory (TPT), the characteristic time required to undergo a transition (cross the energetic bottleneck) must be much shorter than the typical time spent in the metastable reactant and product states. While highly diffusive dynamics, such as protein folding funnel dynamics, can be described by a single diffusive reaction coordinate [54], non-Markovian systems and systems with more than two metastable states typically require multidimensional reaction coordinates to adequately describe the dynamics. [55, 56]

A qualitatively good order parameter χ should distinguish between states A and B, and a *discriminating order parameter* requires that, for some intermediate value χ^*, $\chi < \chi^*$ describes a configuration in basin A (B) and $\chi > \chi^*$ describes a configuration in basin B (A). If a well-chosen reaction coordinate q is used to parametrize a free energy function $F(q)$ and q^* maximizes F between states A and B, then q^* should coincide with the transition state surface. It bears mentioning that an order parameter can be chosen such that trajectories initiated at $\chi*$ would end up in A or B with nearly equal probability, but the free energy $F(\chi^*)$ is not coincident with the transition state surface; an acceptable order parameter is not guaranteed to adequately characterize the transition state region. [7, 49, 57]

1.2.1 Collective variables describing the AdK transition

A variety of collective variables have been used to describe the conformational transition of AdK, as described below and summarized in Table 1. A possible concern is that the use of differing order parameters could be responsible for discrepancies in results (or their interpretation). Daily et al. [59], for instance, found that their results were fairly insensitive to the choice of collective variables, and that trajectories could be meaningfully projected onto variables representing a particular motion in Cartesian (distance between domain centers of mass and RMSD with respect to open/closed structure), dihedral, or contact space. The comparison of LID-CORE separation in Figure 2 by Jana et al. [58] demonstrates the similarity in fluctuations between the RMSD of the LID with respect to the LID-closed state, LID-CORE angle, and LID-CORE

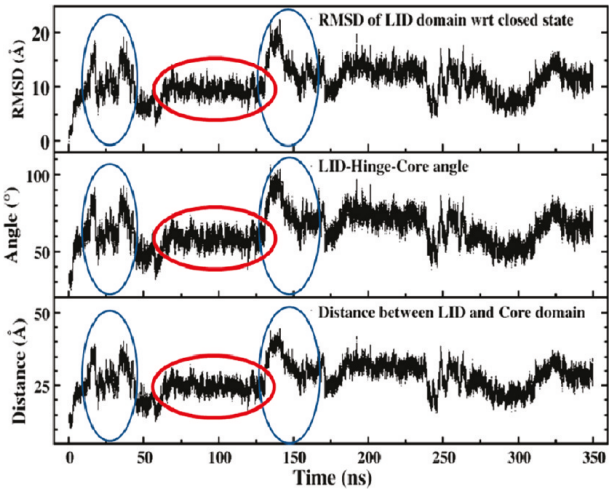


Figure 2. Comparison of the fluctuations of three order parameters in all-atom, explicit solvent equilibrium MD simulations [58]. From top to bottom: RMSD between LID and CORE domains, LID-CORE angle, and LID-CORE mass-center distance. The LID-CORE angle is computed between the C_α atoms in Val148 in the CORE, Pro9 in the labile region, and Glu122 in the LID. The red ellipse highlights fluctuations during the LID-open/NMP-closed state, while the blue ellipses highlight areas of large fluctuation that exhibit remarkable similarity between the three order parameters. [Reprinted with permission from Jana et al. [58]. Copyright 2011, AIP Publishing LLC.]

distance (between residue 122 in the LID and 148 in the CORE), corroborating the findings of Daily et al. [59]. Potestio et al. [60] found that their algorithm optimally decomposed AdK into the three rigid subunits that are consistent with previous studies (LID-CORE-NMP) and could account for 77% of the fluctuation described by the ten most significant essential modes (principal components). In characterizing the AdK closed $\leftrightarrow$ open transition, it is likely that the rigidity of the LID, NMP and CORE domains and simplicity of the bending motions of LID and NMP about the CORE are responsible for the apparently low sensitivity to differing choices of (reasonable) measures of LID-CORE and NMP-CORE fluctuations.

The most commonly used pair of order parameters for plotting 2D free energy surfaces of snapshots of simulation conformers is a heuristic measure consisting of the mass-center distances between domains (center of mass distance between the LID and CORE, and the center of mass distance between NMP and CORE) [24, 58, 61–66]. Several studies use a LID-CORE and NMP-CORE angle pair, as shown in Figure 1, to independently quantify LID and NMP opening motions [15, 31, 59, 67, 68]. Other order parameter pairs to characterize 2D motions include the RMSD to/from the final and/or initial structure(s) [23, 59, 61, 62], essential variables from PCA or NMA [8, 23, 43, 65, 69, 70], and fractions of native contacts relative to the initial and final structures [16, 59]. Lou and Cukier [71], Brokaw and Chu [72], and Pontiggia et al. [73] all measure NMP-CORE separation with the distance between residues 55 and 169, but each, respectively, monitor LID-CORE variation differently: LID-CORE mass-center distance [71], C_α distance between residues 127 and 194 [72], and LID-CORE angle [73]. Progress along one dimension (e.g., plotting the potential of mean force) was captured using domain mass-center distances [24, 58, 69, 74], RMSD to and/or from crystal structures [15, 61, 67, 75], and indices or distances along a computed reaction coordinate pathway [24, 47, 48, 66, 76, 77].

1.2.2 Experimental FRET distances as collective variables

Collective variables that track domain motions have also been directly measured in 1D FRET experiments [37–39]. The distances between the fluorophore labels approximately correspond to C_α distances that can be easily obtained from NMR, crystal structures, or simulations. Comparison of such distances with the thermodynamic free energy landscape derived from MD simulations [31] showed that the distance between residues A55–V169 (NMP–CORE, Sinev et al. [37]) and A127–A194 (LID–CORE, Hanson et al. [39]) track the major conformational changes in apo-AdK in an almost orthogonal manner. A 2D space spanned by these two observables can serve to study the conformational dynamics of AdK at the level of individual domain motions

Table 1. Comparison of collective variables/order parameters used to project AdK motion. See text for details.

Collective variables/Order parameters		References
1D	LID/NMP-CORE MCD	[24, 58, 61, 69, 74]
	RMSD (to an end structure)	[15, 61, 67, 75]
	Indices along Rxn Coordinate	[24, 47, 48, 66, 76, 77]
2D	LID/NMP-CORE MCD	[24, 58, 61–66]
	LID/NMP-CORE angle	[15, 31, 59, 67, 68]
	RMSD (to end structures)	[23, 59, 62]
	Essential Variables	[8, 23, 43, 65, 69, 70]
	Fraction of Native Contacts	[16, 59]
	Other	[71–73]

while at the same time making a direct connection to experiments. The FRET pair I52–K145 [38], on the other hand, reports on the overall opening and closing motion of the enzyme and is therefore a reasonable choice as a 1D collective variable for following the conformational change [24]. It should be noted, however, that it can be difficult to directly assign different thermodynamic states (such as 'open' and 'closed') to multi-modal distance distributions obtained from FRET because these order parameters are not necessarily good reaction coordinates [31, 74]. Furthermore, anisotropy in the fluorophore distribution is not taken into account when using simple C_α distances, although methods are available (in principle) to postprocess simulations in order to obtain the equilibrium label distributions and thus realistic equilibrium distance distributions [78].

1.3 *Rate theories*

Reaction rates, which measure the frequency at which a process such as ligand binding or a conformational change occurs, are essential to describe chemical processes quantitatively. Calculating rates for macromolecular systems from simulations is one of the major goals in computational biophysics. A limited number of studies on AdK discuss rates [12, 59, 75] or quantitatively predict the rates in the enzymatic cycle [79]. To put these important contributions at the forefront of computational method development into context we briefly review rate theories in this section.

Rate calculation is one of the most challenging computational problems because the conformational transitions that need to be measured are rare events. The separation of time scales between the long time spent in local stable states compared to relatively fast transition events necessitates advanced methods that improve sampling of barrier crossings at the expense of sampling stable end states or intermediates. However, it is a very difficult problem to improve sampling in such a way that the underlying dynamics is respected. Indeed, many sampling methods substitute dynamical accuracy (or relinquish the simulation of dynamics altogether) in exchange for improved sampling, making the extraction of rate information either challenging or impossible.

Transition state theory (TST) and Kramers' theory can be used to produce a rate constant expression with exponential dependence on the free energy barrier height (see Eq. 4) under appropriate assumptions [80]. While TST and Kramers' theory can provide reasonable estimates of kinetic rates using equilibrium population densities, they have several shortcomings. Computing rates from simulations using TST requires counting crossing events across an a priori transition state (TS) surface. Since multiple crossing events can occur during a single transition, the predicted reaction rate from TST will be an upper bound on the real rate. A procedure can be used to optimize the TS surface such that the crossing frequency is minimized, while dynamical corrections are applied to account for only the final TS crossing event of a trajectory undergoing a transition [50, 81, 82]. Kramers' theory does not require a TS surface to derive rate constant

expressions in the high and low limits of the friction coefficient, but it still relies on computing equilibrium distributions as in TST.

A reaction coordinate $q(\mathbf{x})$ on the configuration space $\mathbf{x} \in \mathbb{R}^{3N}$ is defined such that the TS surface is located at $q = 0$, dividing the phase space of a system into product ($q < 0$) and reactant ($q > 0$) states. The problem of correctly discounting erroneous crossing events makes TST ill-suited for systems with highly diffusive dynamics (high entropic barriers) where recrossing happens many times during a single reactive transition [83, 84]. Kramers' theory, which is related to TST and can be derived from variational TST in the diffusion limit [85], provides a scheme for deriving rate expressions under the assumption of Langevin dynamics along a suitable reaction coordinate. The original theory was used to derive expressions for reaction rates in the high and low limits of the friction coefficient (collision frequency), corresponding to the positional diffusion and energetic diffusion regimes, respectively [80, 86, 87]. While both TST and Kramers' theory can be generalized to accommodate more than one reaction coordinate [88–90], they cannot offer insight into the mechanistic details of a transition as they do not provide a means for computing the probability densities and currents of reactive transition paths as a function of system configuration. [82]

Transition path theory (TPT) provides a probabilistic framework in which one can determine the probability of finding a trajectory at a particular location in configuration space [91]. Using the concept of an *ensemble of paths* and theoretical structures called *transition tubes*, TPT can be used to compute trajectory probability currents. The string method and transition path sampling (TPS) are common methods for finding transition tubes, although TPS is amenable to other methods [91]. PMF and kinetic rate calculations in the context of TPT requires finding transition tubes, which in turn relies on finding minimum free energy paths or producing trajectories that represent a sample of the real transition path ensemble. Thus, in order to compute rates and free energies accurately, one may ask whether a given simulation method reproduces a realistic transition path ensemble. In Section 7 we outline an approach to quantifying the similarity of transition paths as a means for measuring the quality of transition path ensembles in the context of AdK path sampling.

2 Path sampling methods

2.1 *Molecular dynamics*

Molecular dynamics (MD) is an approach to simulating dynamical systems in which interactions between particles are parametrized by a force field and the equations of motion for all particles are solved with Newton's second law [92, 93]. It is possible to carry out MD with interactions computed at the full quantum mechanical level, but the high computational cost of these *ab-initio* MD methods limits their application to systems of, typically, less than 1000 atoms and a few tens of picoseconds in length. The most common approximation to computing inter-particle interactions is the molecular mechanics approach. Such a classical force field consists of classical representations of covalent bond energetics (bonded forces), which are described using harmonic approximations to capture the energetics of varying bond distances, angles, and dihedrals, and nonbonded forces associated with electrostatic and van der Waals forces [94]. Molecular mechanics force fields are parametrized empirically using both theoretical approximations and experimental data [95].

In the case of biological macromolecules, explicit solvation MD computes the positions, velocities, and force for both the macromolecular solute and the thermal bath of individual water molecules and ions that compose the solvent. This method is considered the most accurate approach, but the large number of water molecules required makes explicit solvent simulations computationally costly. Implicit solvation MD is an alternative approach that treats the solvent as a continuum in order to estimate the free energy of solvation, whereby it attempts to reproduce the mean interaction of, e.g. a protein, with the solvent. Implicit solvation methods can

yield speed-ups beyond an order of magnitude in comparison to explicit solvent simulations, although it has been found that solvation effects can alter the slow dynamics of a macromolecular system Hinsen and Kneller [96] as well as secondary structures and native state populations Roe et al. [97].

There are many other MD-based methods that are designed to reduce computational costs by improving the sampling of configuration space beyond the capabilities of standard equilibrium MD [98], including essential dynamics sampling MD [8], dynamic importance sampling MD (DIMS-MD) [99–101], targeted molecular dynamics [102], accelerated/hyperdynamics [103, 104], replica exchange MD (REMD) [105], temperature-accelerated MD [106, 107], and weighted-ensemble dynamics[12, 108]. Coarse-grained molecular dynamics (CG-MD) simulations can be up to a few orders of magnitude faster than atomistic MD due to the reduction of the number of particles in the system along with the elimination of the fast time scales associated with atomically resolved motions [109, 110] .

2.2 *Elastic networks and coarse-grained models*

Elastic network models (ENMs) are simplified (energetic) representations of proteins that seek to model the functional conformational changes of a protein without resorting to computationally expensive MD [111–113]. Using normal mode analysis (NMA) on a ENM representation of a protein, collective motions of proteins can be captured by slow, low-frequency modes. ENMs are often combined with a coarse-graining (CG) approach (often at the C_α-level) to further reduce computational costs. The normal modes and, hence, the accuracy of the conformational changes produced by a given ENM depend on both the energetic potential and the molecular resolution. Several popular models include the atomic-resolution Tirion model [114], the CG-based Gaussian network model (GNM) [115], anharmonic network models (ANM) [116], and plastic network models (PNM) [47]. Several AdK studies implement an ENM, including the PNM [47], and two other methods utilize normal mode information from an ANM as means to enhance sampling [77, 117]. We also generate and examine AdK closed $\rightarrow$ open transition pathways using three ENM-based transition path servers [16–18] in Section 7 to facilitate comparison.

Whereas many ENMs define energetic basins corresponding to initial and final states under a suitable (e.g. harmonic) representation, multi-basin structure-based models include energetic details between pairs of atoms or residues and merge the energetic terms into a single potential energy landscape using microscopic or macroscopic mixing models [118]. Structure-based models can capture some atomic details, enabling them to capture important global energetic features such as metastable intermediate states and the bumpiness of the real landscape. Several studies of AdK run CG-MD simulations that employ structure-based models [59, 62, 63, 75]. Whitford et al. [118] provide a thorough account of the important aspects of structure-based models, energy landscape theory, and the trade-offs associated with different models.

2.3 *Minimum free energy path methods*

Minimum free energy path (MFEP) and related methods generate single transition paths from a given force field representation by applying local optimization or refinement procedures on the energy landscape. The advantage of this approach partially lies in the choice of representation for the energetic potential, which can range from the full atomistic detail of an MD force field to coarse-grained and/or simplified potentials. Some examples include transition path sampling [7], string methods [10, 119] and related swarming methods [9, 11], nudged elastic band (NEB) methods [6, 120, 121], conjugate peak refinement [5], and the minimum action path approach [16].

2.4 *Motion planning and geometry-based methods*

Another group of related sampling methods that sidestep explicit integration of Newton's equations intelligently search conformation space according to prior information (energetic, conformational/steric) constraints. The class of motion planning algorithms include the rapidly exploring random trees (RRT) algorithm [122, 123], the probabilistic road map (PRM) [124, 125], and the 'mining-minima' algorithm [126]. Other related, geometric-based algorithms include geometry-based essential sampling [127], geometric-based RRT [128], and geometric targeting [14].

2.5 *Free energy sampling*

Free energy sampling methods aim to capture the energetics of a dynamical system as a function of one or more reaction coordinates. Although they are not strictly path sampling algorithms since they do not produce trajectories, it is possible to extract minimum free energy pathways from the free energy landscape using MFEP methods. Umbrella sampling [129] combined with the WHAM [130, 131] or MBAR [132] unbiasing algorithms is a commonly used technique for generating potentials of mean force (PMF), but it requires defining a suitable reaction coordinate (which, when chosen incorrectly, may not necessarily accurately represent the pathway or energetics of a transition) along which energetic sampling is to be performed. For this reason, umbrella sampling is frequently combined with an MFEP method to generate a reaction pathway. Other approaches such as adaptive biasing force methods [133, 134], adaptive reaction coordinate force methods [135], and metadynamics [136] are related methods that can be used to compute free energy surfaces and have the advantage of not requiring an *a priori* reaction coordinate.

3 Energy landscape

The free energy landscape (the PMF) quantifies the free energy required to change the thermodynamic state of the system as described by the collective variables on which the PMF depends, but it does not contain any temporal or dynamical information. Nevertheless, knowledge of the free energy landscape of a system can provide vital clues as to the location of plausible transition pathways, and it is possible to deduce the sequence of domain movements for a given pathway and identify possible intermediate states associated with energetic basins.

3.1 *Conformational equilibrium between open and closed states*

There is general agreement from simulation data that the LID domain of apo-AdK can explore a relatively flat energy landscape [24, 31, 58, 61, 71, 77], which is corroborated by the experimental finding that rigid-body LID motion is the dominant scattering feature in SAXS [137].

The population ratio of the open to closed states can be predicted from the free energy difference between the two states and, importantly, can be compared with experimental measurements. A substantial number of studies indicate that the apo-open state is energetically unfavorable [12, 31, 48, 59, 71, 75, 77]. The early work of Lou and Cukier [71] estimated the free energy of the open state to be about $0.8\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ ($1.37\,k_\mathrm{B}T$) higher than the closed conformation. A number of studies use the free energy difference between conformations as an input parameter for their models. For instance, Lu and Wang [75] adjusted interactions in the closed state so that a free energy difference of $0.75\,k_\mathrm{B}T$ was achieved while Daily et al. [59] chose $\sim 0.5\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ ($\sim 0.84\,k_\mathrm{B}T$) between the native apo-open and apo-closed states in their coarse-grained microscopic double-well model. On the other hand, four studies find the free energy of the open state to be lower than the closed state for unligated AdK [24, 61, 66, 74] with several others indirectly supporting this result, having found their simulations to have a preference for the open confor-

mation. [8, 23, 72] Across all studies, there are large discrepancies in the estimated free energy differences between the two states, and at present the source of the discrepancy is not obvious.

It is also difficult to definitively answer this question from the experimental data. The main evidence for apo-AdK being able to access open and closed conformations in an equilibrium

$$\text{open} \underset{k_{\text{close}}}{\overset{k_{\text{open}}}{\rightleftharpoons}} \text{closed} \tag{2}$$

comes from 1D FRET data. The opening and closing rates determine the equilibrium constant $K = k_{\text{close}}/k_{\text{open}}$ and the free energy difference between closed and open conformations,

$$\Delta G = G_{\text{closed}} - G_{\text{open}} = -k_{\text{B}}T \ln K = -k_{\text{B}}T \ln \frac{k_{\text{close}}}{k_{\text{open}}}. \tag{3}$$

From the apo opening and closing rates for a thermophilic enzyme $k_{\text{open}} = 6,500 \pm 500\,\text{s}^{-1}$ and $k_{\text{close}} = 2,000 \pm 200\,\text{s}^{-1}$ measured by Henzler-Wildman et al. [38] (see also Section 6.2 below) one arrives at $\Delta G = (1.2 \pm 0.1)\,k_{\text{B}}T$, i.e. the apo open state is marginally more stable than the closed state. However, all simulation studies discussed here consider the mesophilic *E. coli* enzyme AK_{eco}, which displays much higher rates for the substrate-bound enzyme than the thermophilic one [36, 38, 41]. The experimental free energy difference between AK_{eco} apo-closed and apo-open is thus not known. Hanson et al. [39] measured specific LID opening and closing rates for AK_{eco} (see Section 6.2), which translate into $\Delta G = (-0.6 \pm 0.5)\,k_{\text{B}}T$. Thus, on average, the LID is more likely to be found in a closed conformation than in an open one. However, these results average over all states of the NMP domain and thus cannot answer the question about the equilibrium Eq. 2.

The conformational equilibrium of AdK was also studied under the effect of cellular crowding using a postprocessing approach [138]. First the motions of a test protein were simulated in the absence of crowding. Then conformations from the simulation were used to compute the change in chemical potential $\Delta\mu$ upon moving a conformation to crowded solutions of varying concentrations. AdK was found to be susceptible to strong crowding effects on the open-to-closed population ratio where intracellular crowding can significantly shift the apo-AdK equilibrium toward the closed state. Given that the closed $\rightarrow$ open transition is thought to be the rate-limiting step, a shift towards the closed ensemble would reduce the closed $\rightarrow$ open transition rate. These ideas appear to be consistent with a recent NMR measurements (combined with MD simulations) that found that the conformational equilibrium of AK_{eco} can be strongly influenced by the concentration of the osmolyte TMAO in solution [139].

3.2 *Potential of mean force*

The majority of studies produce results supporting the presence of a significant energetic barrier between the open and closed states of apo-AdK [12, 15, 16, 23, 31, 48, 58, 62–64, 69, 71, 72, 74–77, 79, 117]. On the other hand, Snow et al. [8], Arora and Brooks [61], Matsunaga et al. [24], and Song and Zhu [66] produce 1D PMF calculations that show an energetic minimum near the open crystallographic structure and an monotonic [8, 61, 66] or almost monotonic [24] rise in energy toward the closed crystal state. Potoyan et al. [74] obtained an energy difference between the two conformational states on the order of $1.5\,k_{\text{B}}T \pm 0.5\,k_{\text{B}}T$, in contrast to the many tens of $k_{\text{B}}T$ from estimates obtained by Arora and Brooks [61] and suggesting that the use of an implicit solvent and a reaction coordinate that is not local to allosteric states in the earlier work may have lead to artifacts. Matsunaga et al. [24] and Song and Zhu [66] show smaller disagreement with Potoyan et al. [74], predicting free energy differences on the order of 20 $k_{\text{B}}T$, but still differing by an order of magnitude.

The wide range of free energy differences between the open and closed states in 1D PMFs

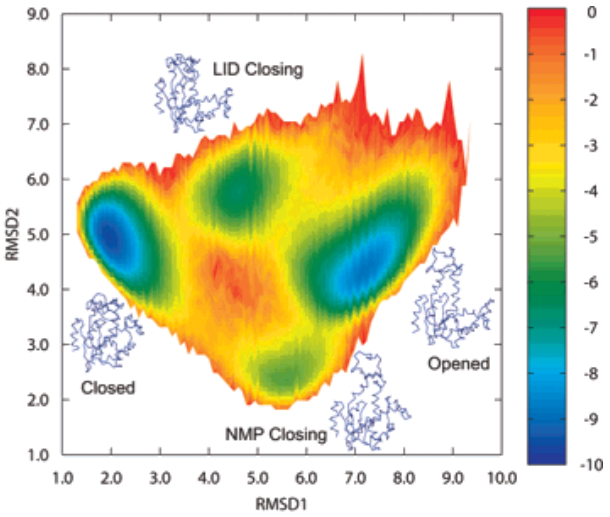


Figure 3. A 2D free energy surface (FES) produced from CG-MD that employs a coarse-grained structure-based potential at the C_α level [75]. The free energy is measured in k_BT. RMSD1 is the RMSD from the closed structure, while RMSD2 is the RMSD from the LID-open/NMP-closed structure depicted at the bottom of the plot. Note that the LID-closed/NMP-open structure at the top is 'far' from the LID-open/NMP-closed structure and corresponds to the other intermediate ensemble. The left and right blue regions coincide with the closed and open state ensembles, respectively; the upper basin corresponds to ensemble basins of LID-closed/NMP-open (intermediate) structures and the lower-right basin, a LID-open/NMP-closed ensemble. Note that NMP domain motion entails slightly higher free energy barriers than LID motion. [Reprinted with permission from Lu and Wang [75]. Copyright (2008) American Chemical Society.]

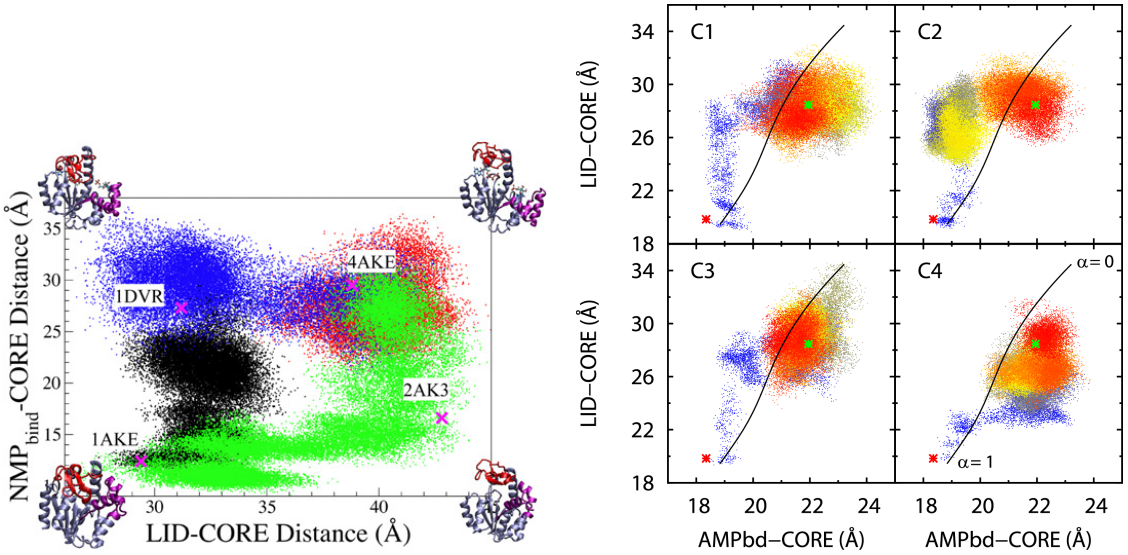
is likely due the use of different reaction coordinates coupled with the evident diversity of the conformational dynamics of AdK. Both Arora and Brooks [61] and Song and Zhu [66] strongly suggest discrepancies are due to differing choices of order parameters, and we have noticed a pattern among the methods and results supporting this intuition. The reaction coordinates used by Arora and Brooks [61] and Matsunaga et al. [24] were obtained using the NEB and string methods, respectively, while Song and Zhu [66] used the principal curve approach [119] in combination with their conformer data from MD simulations (see the black lines in Fig. 4). All three of these methods, however, used the interpolating line between the closed and open crystal structures as the initial path for applying MFEP or principal curve optimization. In contrast, Potoyan et al. [74] employed a structural morphing technique measured by the structural overlap parameter [140] that ensures locality near the allosteric states.

Since umbrella sampling constrains sampling along a chosen pathway that is generally not optimal, the barrier estimates will be be larger than barriers associated with reaction coordinates lying closer to populated transition regions in configuration space. Thus, the lower free energy estimate obtained by Potoyan et al. [74] can be treated as an approximate upper bound.

Song and Zhu [66] acknowledge that the diversity of conformers obtained from their simulations, disagreement with results from FRET [39], and strong evidence of metastable intermediate states from other studies suggest that a single transition tube is inadequate for sampling the diversity of the conformational dynamics of AdK. Many studies, however, have computed a 2D free energy surface (FES) projected onto the LID-CORE and NMP-CORE mass-center distances [24, 58, 61–66] or angle coordinates [15, 31, 59, 67, 68] and directly show the presence of (at least) two local energy minima between the open and closed energy basins; these results corroborate the prediction that two meta-stable intermediate structure ensembles exist and are accessible via two separate pathways that connect the open and closed basins as shown, for example, in Fig. 3 [75].

3.3 *Metastable intermediate states*

A more diverse (multi-pathway) picture of the conformational landscape of AdK than can be gleaned from 1D PMF calculations is bolstered by the work of



(a) Brokaw and Chu [72] (b) Song and Zhu [66]

Figure 4. 2D plots of conformers from simulation snapshots projected onto LID-CORE and NMP-CORE mass-center distance coordinates. (a) Conformers from equilibrium MD show two simulations from the 1ake closed state for the apo (green) and holo (black) enzyme, and apo and holo simulations from the 4ake open state (red and blue, respectively) [72]. The closed-apo and open-holo simulations points show dominant initial LID-opening/closing motions, with the closed-apo simulations suggesting a complete closed → open transition. (b) Four apo-AdK closed → open transitions showing simulation progression from the beginning (blue) to the middle (yellow) to the end (red) [66]. The green and red stars denote the location of the 1ake and 4ake states, respectively. Note that the horizontal and vertical axes are transposed between the two plots. [Figure (a) Reprinted from Brokaw and Chu [72], Copyright (2010), with permission from Elsevier. Figure (b) Copyright 2013 Song and Zhu [66], reprinted under the terms of the Creative Commons Attribution License.]

Bhatt and Zuckerman [12], Whitford et al. [63], Lu and Wang [75], and Beckstein et al. [31], giving strong indications of pathways connecting two local energy minima—corresponding to a LID-open/NMP-closed and a LID-closed/NMP-open conformation—residing between the open and closed ensembles. Jana et al. [58] found an intermediate state where the LID and NMP domains intermediate between their open and closed conformations with the NMP domain slightly closer to the open state; this intermediate state is deemed the half-open/half-closed (HOHC) and corresponds closely to the 1ak2 crystal structure that they believe has direct kinetic involvement in the catalytic cycle. An HOHC-like intermediate state was predicted by experimental results [4, 28, 38] and was explicitly corroborated by the simulations performed by Wang et al. [79]. Other studies either directly reveal energy landscapes that host intermediate states resembling the HOHC state [12, 31, 63, 75] or demonstrate the existence of at least two pathways, one of which may progress through an intermediate state similar to the HOHC state [16, 23, 64, 66, 72, 76, 77, 117]. Pontiggia et al. [73] were able to identify many substates, characterized by $\sim$ 5-10 ns residence times, via clustering all-atom MD simulations using k-medoid structural partitioning. While Brokaw and Chu [72] did not compute energies, projections of conformers onto LID-CORE/NMP-CORE mass-center distance space showed possible transient structures near the 1dvr (LID-closed/NMP-open) and 2ak3 (LID-open/NMP-closed) crystallographic structures (Fig. 4(a)). The presence of two intermediate ensembles was clearly evident in weighted ensemble simulations performed by Bhatt and Zuckerman [12] that exhibited two regions of relatively high fractional conformer populations flanking the densest regions corresponding to the initial and final populations.

4 Transition pathways

Finding a functionally relevant pathway between conformational states of a protein is a primary goal of path sampling. Large macromolecular conformational changes, however, are not necessar-

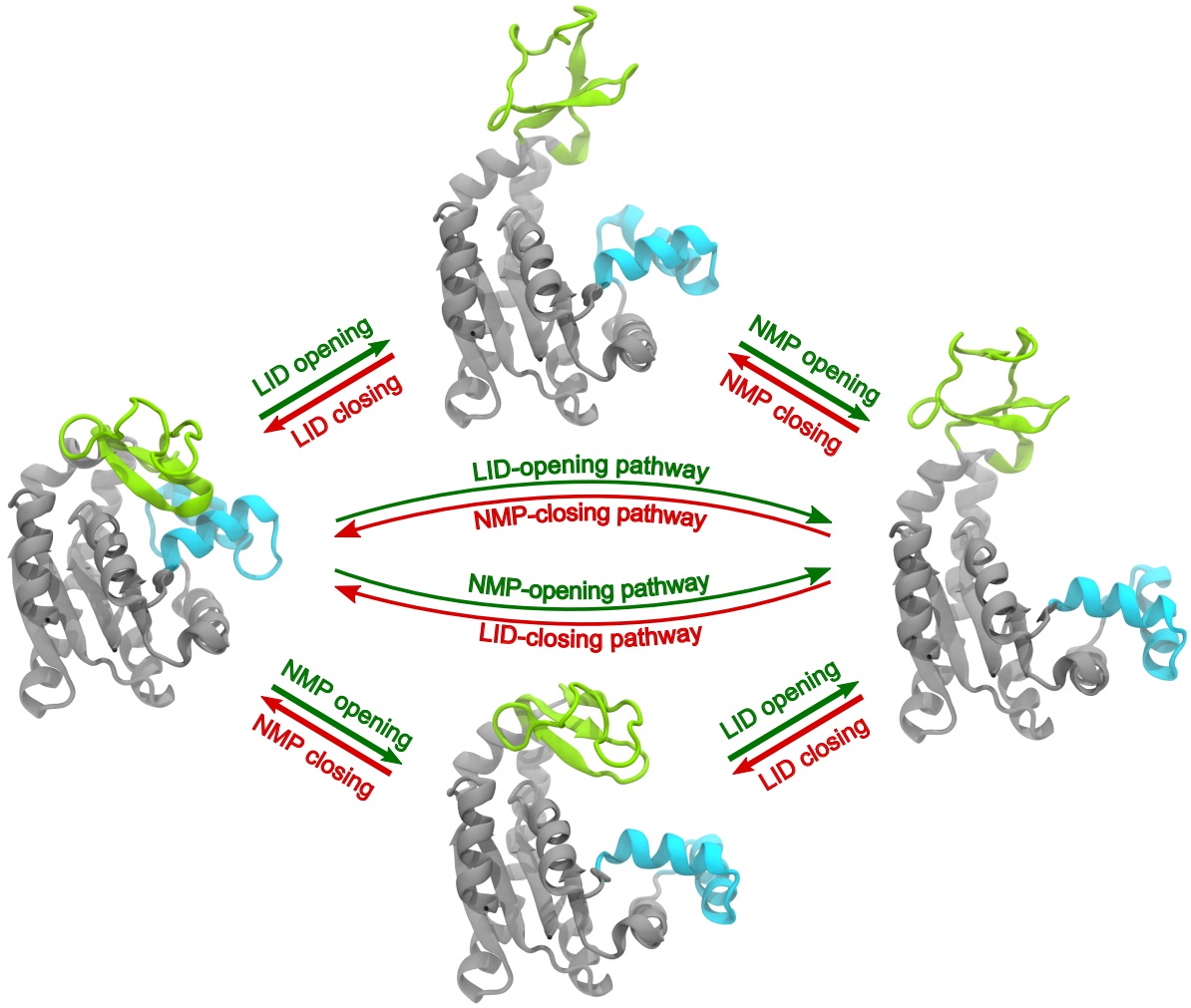


Figure 5. Two plausible pathways for the conformational change between a closed (left - represented by PDB ID 1ake) and open (right - represented by PDB ID 4ake) conformation, during which the LID domain (green) and NMP domain (blue) move sequentially. The top path in the closed → open direction (green arrows) consists of the LID-opening first, placing AdK in one possible intermediate state (top - represented by PDB ID 2ak3), followed by opening of the NMP domain to the open state. The bottom path in the closed → open transition is defined by NMP-opening preceding the LID-opening, with a LID-closed/NMP-open intermediate structure (bottom - represented by PDB ID 1ak2). The red arrows show the two domain closing orders for the open → closed transition.

ily restricted to one well defined pathway, often proceeding through several metastable states that correspond to local free energy minima in the configuration space. In the case of the closed ↔ open AdK transition, the LID and NMP domains both swing open or closed around labile regions of the protein, but whether these two motions are correlated to some extent or relatively independent is an unresolved question. NMR and single-molecule FRET experiments have indicated the presence of two pathways and two corresponding intermediate states [4, 40], and computational studies implementing a variety of sampling methods generally agree that (at least) two pathways are plausible between the open and closed states [12, 16, 23, 31, 58, 63, 64, 66, 72, 75–77, 117]. In Fig. 5, the two proposed transition pathways are depicted along with representative intermediate conformations: the top path corresponds to LID-opening preceding NMP-opening in the closed → open transition (labeled in green) and NMP-closing preceding LID-closing in the open → closed transition (red); the bottom path represents NMP-opening preceding LID-opening in the closed → open direction (green) and LID-closing preceding NMP-closing in the open → closed direction (red). It is also possible that variations of these two pathways occur along with other intermediate ensembles of states, although there is still no consensus about the actual structure and distributions of AdK transition pathways and intermediate states.

4.1 *Pathways from the closed state*

In the closed → open transition, a majority of studies show evidence for the closed → LID-opening → NMP-opening → open pathway (LID-opening-first) [12, 31, 58, 63, 72, 75, 117, 122], although a substantial number of investigations report different pathways or contrary evidence. Korkut and Hendrickson [15] found a slight opening of both the LID and NMP domains followed by further opening of the LID domain. Interestingly, using a double-well network model, Chu and Voth [76] found that either pathway could be produced, depending on the initial starting path that was optimized in the minimum free energy path (MFEP) calculation. Using linear interpolation between the open and closed structures as the MFEP starting point produced a LID-opening-first pathway, whereas the PNM-based MFEP produced a NMP-opening-first pathway with a similar barrier height that was shifted closer to the closed structure. Liu et al. [68] found that the closed transition could happen through either pathway with close to equal opportunity.

Three studies found contrary evidence where a slight NMP-opening was the leading motion. A temperature replica exchange MD study performed by Kubitzki and de Groot [23] and an equilibrium MD study performed by Peng et al. [64] found that the dominant path corresponds to an initial partial opening of the NMP domain preceding the opening of the LID; Kubitzki and de Groot [23] observed a partially correlated opening of both the LID and NMP domains following the initial NMP motion. Four equilibrium MD simulations of the closed → open transition conducted by Song and Zhu [66] showed a possible slight initial opening of the NMP domain (Fig. 4(b)), which appears to agree with Kubitzki and de Groot [23]; three of the simulations then showed dominant LID-opening followed by dominant NMP opening, and the fourth simulation showed NMP-opening followed by LID-opening. However, the temporal order of the conformers and the starting conformation of each simulation is not definitive from a visual inspection of the plot. Moreover, the four other closed → open and three other open → closed transitions are not shown by Song and Zhu [66] and the authors do not commit to any explicit predictions about the order of domain motions. It should also be noted that the observation of a slight opening ($\lesssim$ 1 Å change in the NMP-CORE center of mass distance) of the NMP domain is likely to be sensitive to the atomistic and energetic resolution of a system and also the choice of order parameters (such as eigenmodes or principal components that hide individual LID/NMP movements in collective displacements) used to project the conformational motion. It is also possible for small NMP motions to have been overlooked or ignored in some studies, especially if the motion is followed, and overshadowed, by large LID-opening motions.

4.2 *Pathways from the open state*

In the closing direction, most have also found evidence for, or support the hypothesis that, LID-closing is the primary step, followed by NMP-closing [24, 47, 48, 58, 61, 63, 68, 69, 77, 117] with Farrell et al. [14] finding simultaneous domain closure. In contrast, coarse-grained simulations performed by Bhatt and Zuckerman [12], Lu and Wang [75], and Daily et al. [59] produced dominant pathways where NMP-closing was the primary step. Bhatt and Zuckerman [12] performed weighted ensemble transitions where forward and reverse transition symmetry was produced by a specific double Gō model , finding 60% of resampled trajectories following the NMP-closing pathway (40% along the LID-closing pathway) and 60% producing the LID-opening step (40% producing the NMP-opening step) first for the closed → open transition (see Fig. 5 for the definition of the pathways). The open → closed transitions performed by Lu and Wang [75] showed the NMP-closing step first 86% of the time versus 14% for a primary LID-closing step, while the closed → open transitions produced the LID-opening step first 87% of the time and the NMP-closing step 13% of the time. These results agree with theoretical arguments for pathway symmetry given by Bhatt and Zuckerman [141] and discussed in more detail in Section 4.3.

4.3 *Forward and reverse path symmetry*

If multiple pathways exist between two end states in a simulation in equilibrium, then the probability ratios between any two pathways should be the same in the forward and reverse direction for an equilibrium system due to detailed balance [141]. Namely in the case of AdK, the ratio between the probabilities to observe LID-opening and NMP-opening transitions should be the same as the ratio between the reverse transitions, NMP-closing and LID-closing (Fig. 5). Approximate symmetry should hold if the initial and final basins lack slow internal time scales such that trajectories emerging from a metastable basin do not depend on the manner in which they entered that basin; in other words, approximate symmetry demands that trajectories in a metastable state emerge from the basin in a quasi-Markovian way [141]. Thus, if multiple pathways are reported in a simulation study and if the pathway ratios exhibit asymmetry between forward and reverse directions then this appears to be indicative of either a non-equilibrium process (such as the presence of a driving force or bias, as for instance in Beckstein et al. [31]) or internal barriers in the end states.

It is interesting to consider simulation results from the different studies in light of the pathway symmetry hypothesis. Bhatt and Zuckerman [12] suggest that the path asymmetry, where the LID-opening (closing) motion precedes the NMP opening (closing) motion for both open $\rightarrow$ closed and closed $\rightarrow$ open transitions, may be due to simulations having been initialized from a single crystal conformation. Starting from an ensemble of conformations corresponding to the initial state and final states, as done by Bhatt and Zuckerman [12], Lu and Wang [75], and Daily et al. [59], should roughly enforce an equilibrium-based steady state that produces forward and reverse path symmetry. Both Bhatt and Zuckerman [12] (for a double Gō model without energy symmetry) and Lu and Wang [75] observed forward and reverse path symmetry along the NMP-opening/LID-closing path depicted in Figure 5.

Studies performed by Peng et al. [64] and Song and Zhu [66] took the approach of generating a collection of short MD simulations (instead of producing fewer long-time trajectories), which could provide an avenue for verifying the path symmetry hypothesis of Bhatt and Zuckerman [141]. Although equilibrium simulations should not be sensitive to initialization bias in theory, initialized trajectories must be consistent with the ratio of fluxes entering equilibrium substates for exact symmetry to hold [141]; however, approximate symmetry holds if initialized trajectories are in metastable equilibrium such that trajectories have sufficient time to explore the metastable basin prior to emerging from the state. At the very least, trajectories emerging from a state must be sufficiently uncorrelated with its initial structure and be able to emerge from the state in a quasi-Markovian matter. Thus, while properly equilibrated MD simulations should meet the conditions for approximate symmetry if there are no slow internal timescales in metastable basins, it seems reasonable to suspect that using an ensemble of carefully equilibrated initial starting configurations will be more conducive to meeting conditions for approximate symmetry since it facilitates a more thorough exploration of the initialization basin.

The setup of an ensemble of many MD simulations required Peng et al. [64] and Song and Zhu [66] to employ an initialization procedure that ensured individual simulations and trajectories were uncorrelated. The general initialization procedure involved (1) extracting a set of frames (conformations) from an equilibration or equilibrium MD simulation over several nanoseconds, (2) splitting each frame into a several simulations having the same initial conformation, and (3) uniquely assigning new velocities according to the Maxwell-Boltzmann distribution and performing further equilibration if necessary. Simulations generated with this procedure are more likely to be representative of equilibrium ensembles since the mixing of both conformations and velocities ensures that each trajectory is unique and thoroughly uncorrelated with the initial 1ake/4ake crystal structures. The majority of other MD studies performed far fewer simulations with more standard equilibration procedures; a smaller sample size, along with the possibility that some trajectories still retained correlations to the initial crystal structures, makes it more difficult to draw conclusions about possible transition mechanisms, which may

require a statistical approach.

Song and Zhu [66] ran eight simulations between 100 ns and 200 ns long from the closed state, finding five simulations that reached the open structure, while only one remained near the closed state for the entire 200 ns simulation. The data from Song and Zhu [66] (Fig. 4(b)) appear to show a slight NMP-opening in the beginning of the transitions, although the motions are ~ 1 Å, and the starting conformations of the simulations and the temporal order of the conformers are not indicated. Seven unrestrained simulations were also initialized from the open structure but all remained in the vicinity of the conformational space of the open state. Peng et al. [64] performed 300 simulations for 5 ns each, with about a tenth of the simulations extended to a total of 20 ns. Both Peng et al. [64] and Song and Zhu [66] observed a slight initial NMP-opening step along with a diversity of pathways leading to the open state. Peng et al. [64], however, found that the NMP-opening pathway was dominant where the NMP domain was stably open prior to LID opening with a far less populated pathway occurring where the NMP domain was partially open or closed while the LID's state was anticorrelated (closed or open, respectively). These findings disagree with the domain motion order predicted by Lu and Wang [75], Daily et al. [59] and Bhatt and Zuckerman [12]. As Peng et al. [64] only performed transitions from the closed ensemble, and since Song and Zhu [66] did not provide predictions regarding the sequence of domain movements (for either direction of the transition) and did not observe open $\rightarrow$ closed transitions during the length of their simulations, there appears to be insufficient information to draw further conclusions about pathway symmetry. Furthermore, it should be noted that three studies discussed here [12, 59, 75] employed CG models; it is clear that further studies using atomistic equilibrium MD would be required to ascertain whether the observed pathway differences are realistic or artifacts of CG potential models.

4.4 *Further considerations*

It is imperative to keep in mind is that sampling methods relying on the construction of a collective variable space using PCA or NMA are biased toward the generation of motions local to the end (crystal) structures. Essential subspaces are constructed by performing PCA on short MD simulations, which biases the formation of principal components favoring motion local to the known conformations. One can also apply NMA to knowledge-based, structure-based, or ENM potentials, which are defined for known structures and are reasonable approximations to local motions about those states. The collective motion of the largest (first) principal component or normal mode will accordingly display a large proportion of LID movement which is possible from both the open and closed states of AdK. However, high LID flexibility does not guarantee that LID-first motion dominates either the forward or reverse transition. Furthermore, the tendency of NMP domain motion to be reflected by slightly faster modes does not guarantee that NMP motion precedes LID motion. Both normal modes and principal components will thus tend to reflect the local flexibility of the LID domain in the vicinity of the end states. Path sampling methods, essential dynamics sampling MD, coarse-grained MD, and ENM-based multiscale MD, which are based on the analysis of the motion the largest (slowest) modes, may favor LID-first motion due to sampling being biased in that direction, even though the pathways associated with this motion may not be the primary pathways.

5 Substrate binding: holo-AdK

The preceding discussions were almost exclusively concerned with the apo enzyme. In fact, one of the advantages of AdK as a model system for conformational changes is that the protein undergoes thermal fluctuations during which it samples conformational states similar to the ones sampled by the substrate-bound (holo) enzyme [40]. Thus, large-scale closed $\leftrightarrow$ open transitions are observable in the absence of substrate (ATP, ADP, AMP) so that one may unam-

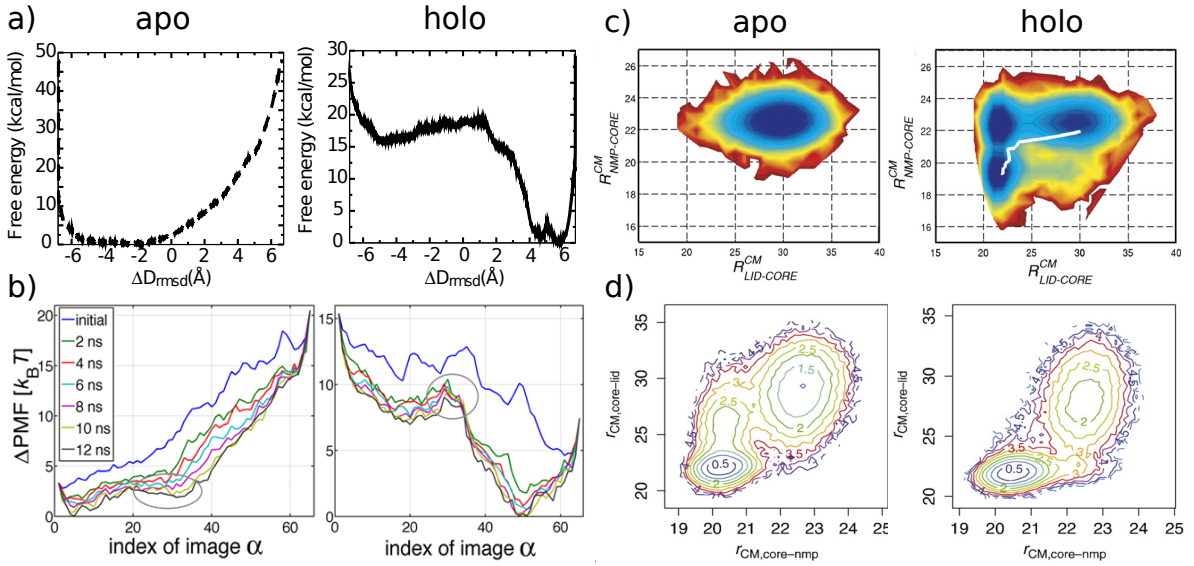


Figure 6. Free energy surfaces of apo and holo AdK. (a) Implicit solvent all atom MD umbrella sampling simulations [61]. (b) Explicit solvent all atom MD MFEP calculations [24]. (c) Structure-based Hamiltonian [62]; energy scale from 0 to $10\,k_BT$ (blue to red). (d) Structure-based Hamiltonian [59]; energies measured in kcal $\cdot$ mol^{-1}. The presence of the ligands was either modelled by explicitly including the inhibitor Ap_5A [24, 61] (a, b) or by incorporating closed-state native contacts into a coarse-grained Hamiltonian based on an interaction model primarily derived from the open state [59, 62] (c, d). In each subfigure, the apo-PMF is on the left and the holo-PMF on the right.[Figure (a) Reprinted from Arora and Brooks [61] with permission, Copyright (2008) National Academy of Sciences, USA. Figure (b) Copyright 2012 Matsunaga et al. [24], reprinted under the terms of the Creative Commons Attribution License. Figure (c) Reprinted from Whitford et al. [62] Copyright (2007), with permission from Elsevier. Figure (d) Reprinted from Daily et al. [59] Copyright (2010), with permission from Elsevier.]

biguously separate effects due to the protein alone from protein-ligand interactions. Experiments and simulations that include ligands often make use of non-hydrolyzable analogs such as Ap_5A (bis(adenosine)-5′-pentaphosphate, which essentially consists of ATP and AMP fused at the γ/α phosphates) instead of the adenosine phosphates. Fundamental questions are how the presence of ligands affects the rates of the conformational transitions, which pathways are populated, and how the free energy difference between closed and open state changes. In particular, a debate continues about whether AdK functions through either (1) an induced fit mechanism by which binding of ligand alters the conformation so that a high-affinity binding site is created [142, 143] or (2) a population shift mechanism [144, 145]. In the latter, also known as conformational selection, the protein samples all conformations with finite probability, even the ones forming an (empty) high-affinity binding site. The presence of a ligand would then shift the conformational equilibrium towards conformations favoring the protein-ligand complex.

The prevalent computational strategy to address these questions is to compute apo- and holo free energy landscapes as functions of one or two collective variables (see Section 1.2) although recent studies also directly compute transition rates (discussed in Section 6.2). Comparison of the resulting PMFs directly shows which conformational states are stabilized by the presence of the ligands. Some representative examples of 1D and 2D PMFs are shown in Fig. 6). Most computational studies agree that the presence of substrates will decrease the free energy of the closed conformation relative to the open one so that the closed conformation will become the more favorable state [24, 59, 61, 62, 79] (Fig. 6). All simulations that found that the apo enzyme only samples a single broad and open state report the emergence of a second, lower free energy state when protein-ligand interactions are taken into account [24, 61, 62] (Fig. 6a–c). This finding is commonly interpreted as describing the case where the closure of AdK requires ligand-binding [24, 127]. The alternative view is that ligand-binding shifts the occupancy of pre-existing conformational states [59, 79] (Fig. 6d). Wang et al. [79] calculated rates from their structure-based and experimentally calibrated simulations and showed that the populations of states strongly depend on substrate concentrations. Both structure-based models [59, 62, 63, 79] and explicit all-atom simulations [24, 61, 72] indicate that substrate-protein interactions are

primarily responsible for making the substrate-bound closed conformation the lowest free energy state. For instance, in equilibrium MD simulations starting from the open conformation and with bound substrates (Mg^{2+}ATP and AMP), the LID domain closed due to ligand-LID interactions over the course of 20 ns [72] (see the blue trajectory in Fig. 4(a)). Careful analysis of a holo-MFEP showed that specific binding of the AMP-moiety of Ap_5A to the NMP domain induces NMP-closure, and that this step characterizes the transition state ensemble of the complete closure pathway [24]. The LID domain, on the other hand, only closes after the ATP-moiety of Ap_5A has bound to the dehydrated P-loop [24]. A structure-based model also explicitly supports the view that substrate binding to the NMP domain and its closure is the rate limiting step [79].

The experimental data for AdK are generally interpreted to describe a conformational selection mechanism [4]; simulation results are more ambiguous. Matsunaga et al. [24] conclude that the rate limiting step in their simulations, NMP-closure, is induced by AMP binding, although LID closure resembles a conformational selection mechanism. Even the cases in which the closed state appears as a minimum in the energy landscape when ligand-protein interactions are taken into account is not necessarily proof for the classical induced fit mechanism [142] because even though the protein might not sample a fully closed conformation, the binding site itself can already sample active conformations [61]. Ultimately, mechanisms for ligand binding coupled to conformational change form a continuum between *induced fit* and *conformational selection* [145–147]. Either mechanism is described by a distinct pathway through an energy landscape. How much each pathway contributes to the overall reaction depends on the fluxes along these pathways, which in turn are products of kinetic rates (both of ligand binding and of conformational transitions) and ligand and enzyme concentrations [146–148]. For instance, the mechanism of ligand binding of NADPH to dihdyrdofolate reductase can continuously change from conformational selection to induced fit with increasing ligand concentration [146]. Wang et al. [79] analysed the reaction fluxes of their simplified model of AdK and, because they did not find a significant flux from the closed-unbound state to the closed-bound state, conclude that AdK should be described by an induced-fit model. Although this reasoning is in line with Sullivan and Holyoak [143]'s analysis that enzymes with lids ought to function by induced fit, Hammes et al. [146] point out that the situation may actually be more complicated if broad conformational states of low energy exist that can all bind the ligand.

What seems clear, however, from the studies discussed so far and others [42–45] is that the conformational dynamics of the apo enzyme plays a large role in determining the conformational change that is inducible by ligand binding. In other words, the fold of the protein already encodes the conformations that are populated in the ligand bound enzyme [4].

6 Time scales

NMR measurements suggest that NMP- and LID-opening rates of *E. coli* apo-AdK could be on the order of $1.9 \times 10^7\,\mathrm{s}^{-1}$ at 20°C, corresponding to typical times of only 52 ns [41]. NMR and FRET studies on apo-AdK from the thermophile *Aquifex aeolicus* yielded slower rates of $k_{\mathrm{open}} = 6,500 \pm 500\,\mathrm{s}^{-1}$ and $k_{\mathrm{close}} = 2,000 \pm 200\,\mathrm{s}^{-1}$ [38]. FRET experiments that specifically probed LID dynamics of the *E. coli* apo enzyme arrived at $k_{\mathrm{open}} = 120 \pm 40\,\mathrm{s}^{-1}$ and $k_{\mathrm{close}} = 220 \pm 70\,\mathrm{s}^{-1}$ [39].

In the mesophilic *E. coli* holo-enzyme the LID-opening rate was measured by NMR relaxation at the same temperature as above to be $k_{\mathrm{open}} = 286 \pm 85\,\mathrm{s}^{-1}$ (3.5 ms) while the LID-closing rate was around $k_{\mathrm{close}} = 1,374 \pm 110\,\mathrm{s}^{-1}$ (0.73 ms); for the thermophile the rates were $44 \pm 20\,\mathrm{s}^{-1}$ and $1571 \pm 100\,\mathrm{s}^{-1}$ [36]. 1D FRET measurements of LID dynamics in the presence of the non-hydrolysable ligands AMP-PNP and AMP yielded $k_{\mathrm{open}} = 160 \pm 40\,\mathrm{s}^{-1}$ (6.25 ms) and $k_{\mathrm{close}} = 440 \pm 110\,\mathrm{s}^{-1}$ [39].

The experimental data show the range of time scales from about 50 ns to 5 ms or more that one should expect to observe although one should note that there appears to be a large discrepancy between the apo-AK_{eco} rates derived from NMR [41] and FRET [39] and therefore the lower

bound on conformational times of $\sim$ 50 ns should be reviewed. Nevertheless, an important conclusion from those experiments was that LID-opening comprises the overall rate limiting step [36, 38, 39] of the reaction Eq. 1.

6.1 *Time scale estimate from transition state theory*

Large macromolecular conformational changes such as the transition between open and closed states of AdK are rare events. Nevertheless, there are multiple indications that they might be observable in equilibrium MD simulations. Brokaw and Chu [72] reported almost complete opening of initially closed apo-AdK over one 100-ns trajectory (Fig. 4(a)). The actual transition event only took about 40 ns, which is in line with the general idea that barrier crossing times can be much shorter than the overall average time to switch states. Song and Zhu [66] observed spontaneous transitions from the closed conformation to the open conformation within less than 20 ns in multiple repeat simulations (Fig. 4(b)). Other studies also report opening of apo-AdK in less than 200 ns [149].

Experimentally measured rates for the rate-limiting step of enzyme opening are on the order of 300 s^{-1},[38] suggesting typical transition times of about 3 ms for the opening step of the product-bound enzyme. The non-rate-limiting closing step of the reactant-bound enzyme occurs on even shorter times, about 0.7 ms [38]. These estimates are for the holo-enzyme where substrate-protein interactions increase the free energy barrier from closed to open as indicated by simulations [24, 61]. For the apo transition, simulations suggested a shallower energy landscape in which typical barriers are only on the order of 4–6 $\mathrm{kcal \cdot mol^{-1}}$ [31] or have virtually disappeared from the open state [24, 61, 66].

In order to roughly estimate a time scale for apo transitions and to gain some intuition about the numbers involved, we obtain a simple transition state theory (TST) estimate for the transition rate constant,

$$k = \nu \exp(-\Delta G^{\dagger}/k_{\mathrm{B}}T), \tag{4}$$

with the frequency pre-factor ν and the effective barrier height $\Delta G^{\dagger}$. If we assume that ν does not change very much between the apo and holo process, $\nu_{\mathrm{apo}} \approx \nu_{\mathrm{holo}}$, then the ratio

$$\frac{k_{\mathrm{apo}}}{k_{\mathrm{holo}}} = \frac{\nu_{\mathrm{apo}} \exp(-\Delta G^{\dagger}_{\mathrm{apo}}/k_{\mathrm{B}}T)}{\nu_{\mathrm{holo}} \exp(-\Delta G^{\dagger}_{\mathrm{holo}}/k_{\mathrm{B}}T)} \approx e^{-(\Delta G^{\dagger}_{\mathrm{apo}} - \Delta G^{\dagger}_{\mathrm{holo}})/k_{\mathrm{B}}T} \tag{5}$$

between apo and holo transition rate only depends on the difference in barrier heights. If the apo barrier $\Delta G^{\dagger}_{\mathrm{apo}}$ is lowered by a moderate $5\,k_{\mathrm{B}}T$ (3 $\mathrm{kcal \cdot mol^{-1}}$) compared to the one encountered for the product-bound enzyme, then the apo transition rate would already be faster by a factor of $\sim$ 150, i.e. typical transition times would be on the order of 20 μs, down from 3 ms . For an apo barrier $10\,k_{\mathrm{B}}T$ below the holo barrier, the transition times drop to 140 ns, close to the lowest experimental estimate of $\sim$ 50 ns [41].

To go beyond a relative comparison of rates for the apo and holo enzyme, the exponential prefactor ν, which is the maximum kinetic rate in the absence of a free energy barrier, must be determined. The exponential prefactor must be deduced from experimental or computational measurements of transition times, theoretically or computationally predicted energy barriers, or experimental FRET measurements of the decay of the autocorrelation function for motion in a metastable well [80, 150]. If the exponential prefactor ν is known, then kinetic rate predictions can be extracted given a known free energy landscape using Eq. 4. Potoyan et al. [74] note that, given a rate description provided by Kramers' theory (resulting in an expression like Eq. 4 for the rate constant) and an exponential prefactor typical for polypeptides [150], open $\rightarrow$ closed transitions occurring on time scales shorter than a microsecond imply a free energy difference below several $k_{\mathrm{B}}T$ between the open and closed states. We thus expect the prefactor to be on

the order of (or larger than) 10^{-6} s^{-1} to describe transitions taking place on the microsecond time scale (or shorter).

To estimate the exponential prefactors at room temperature for the opening and closing steps Lu and Wang [75] combined the experimental values measured by Wolf-Watz et al. [36] for the rate constants, respectively k_{open} and k_{closed}, and the free energy barrier heights from their model (F_{open} and F_{closed}) and applied the TST rate relations $k_{\mathrm{open}} = k_{0,\mathrm{open}} \exp(-F_{\mathrm{open}}/k_{\mathrm{B}}T)$ and $k_{\mathrm{closed}} = k_{0,\mathrm{closed}} \exp(-F_{\mathrm{closed}}/k_{\mathrm{B}}T)$. k_{cat} is taken to be the rate constant of the rate-limiting step. (For more details see Fig. 3 of Lu and Wang [75] depicting the 1D barrier heights, k_{open}, k_{closed}, and k_{cat} as functions of temperature for both the NMP-closing and LID-closing pathways.) Since the double-well model used by [75] had two dynamical pathways and four free energy barriers (between the open, closed and two intermediate metastable states): k_{open} and k_{closed} were computed for both the LID-closing and NMP-closing pathways; F_{open} was taken to be the free energy difference between the open state and the larger of the two barriers along each of the LID-closing and NMP-closing pathways; and F_{closed} was taken to be the free energy difference between the closed state and the larger of the two barriers along each of the LID-opening and NMP-opening pathways.

6.2 *AdK kinetic rates*

A competition between entropy and enthalpy is likely to determine opening and closing rates where, in the presence of a ligand, the closing transition is facilitated by the enthalpic interaction of LID-CORE contact formation and is resisted by entropic contributions (decreasing entropy) from LID-CORE closing and backbone dihedral rigidification in the CORE. The opening transition is facilitated by entropic contributions from greater rigid-body and backbone dihedral motion in the NMP domain in the open state, while CORE-NMP contacts are enthalpic inhibitors to opening [31, 59]. Thus, it was hypothesized that a small number of contacts or residues alone is unlikely to determine the transition rate [59]. In the work of Lu and Wang [75], the catalytic (turnover) rate k_{cat} was predicted at different temperatures using the highest barrier height measured in their simulation model. By approximating k_{cat} as the larger of k_{open} and k_{closed}, the transition state theory transition rate estimate from Eq. 4, and the experimental values for the opening and closing rates ($k_{\mathrm{open}} = 286\, s^{-1}$ and $k_{\mathrm{closed}} = 1374, s^{-1}$, from Wolf-Watz et al. [36]), Lu and Wang [75] could determine ν_{open} and ν_{closed} using the room temperature values of ν_{open} and ν_{closed}. These values were extrapolated to find that k_{open} was the rate-limiting step below 323 K.

Wang et al. [79] employed MD simulations of a structure-based model parametrized to reproduce the experimental probability distributions of LID-closed and LID-open conformations [39]. Effects of ligands were explicitly taken into account, in contrast with implicit approaches used by other structure-based models [62, 63]. Due to the highly efficient nature of their model, Wang et al. [79] were able to obtain both a PMF and rates for the interconversion between different conformational states directly from the trajectories. From the low energy basins in the PMF they found the open and closed conformation and two intermediate states, namely an NMP-closed/LID-open state and an HOHC (NMP open/half-open, LID half-open/half-closed) conformation similar to configurations previously found in simulations [12, 31, 58, 63, 75]. Wang et al. [79] determined rates and fractional pathway fluxes between the opening, closing, and two intermediate states, and for apo- and holo-AdK and ATP- and AMP-bound models were computed. The opening and closing rate constants for both the apo and holo enzyme were found to be several times smaller than experimental measurements; however, the relative sizes of the opening and closing rates were reasonably consistent with experimental values. NMP-opening and NMP-closing were found to be flux limiting for both pathways (regardless of LID/NMP closing/opening order) and for both the closed $\rightarrow$ open and open $\rightarrow$ closed transitions.

The findings of Wang et al. [79] agree with the MFEP results of Matsunaga et al. [24] but disagree with the common interpretation of the experimental data that LID-opening of the

holo-enzyme is the rate limiting step [36, 38, 39]. While the rate analysis used by Wang et al. [79] could not determine the rate-limiting step for the closing transition (although they were able to determine the more important flux-limiting step [146]), it was found that their rate and flux descriptions were consistent with each other for the closed → open transition and transition numbers for the transitions between adjacent basins in the forward direction approximated numbers for the reverse transition.

7 Comparison of path sampling methods

The question which method is the most suitable to sample conformational transitions does not have a definitive answer. The answer depends on the priorities of the simulator and is determined by the relative importance of different aspects of the method. Table 2 summarizes the methods discussed in this review and qualitatively assesses their capabilities. The approaches were grouped by what we considered the main method, e.g. a novel method or the method that supported the major conclusions of the work. We further distinguished by sub-method, i.e. supporting methods that were used in the context of the main method. Notes on the level of detail of the molecular representation, the solvation model, what events are sampled (e.g. equilibrium ensembles or directed transition, or only structure along pathways without directional/temporal information), and the corresponding references complete the objective description of each method. We subjectively rated the suitability and performance of each approach in a number of categories such as the suitability for calculation of free energies or sampling of conformational transitions. We adopted a coarse ternary rating system in which the symbols +, 0, − have the following meaning: +: The method is well suited to provide the required information or has a high level of resolution or efficiency; data is considered of high quality (although evaluating the absolute accuracy of most methods is extremely challenging and therefore our ratings should not be taken as a definitive assessment of the quality of any method—see also Section 7.2 below). 0: The method can provide the information but is not exactly geared towards it (for instance, computing a PMF from equilibrium MD via Boltzmann sampling); data are produced at intermediate level of quality or efficiency. −: The method is inherently limited in that it is not possible to meaningfully obtain the desired information (e.g. rates from umbrella sampling or ENM).

7.1 *Quantitative path comparison*

In order to quantitatively compare transition paths generated by different computational methods, we developed a novel set of computational techniques and tools that measure transition path similarity [152]. Similarity measurements are based on the Hausdorff [153] or Fréchet [154] path metrics that measure a distance between curves; smaller distances correspond to greater similarity with 0 indicating identity, while larger distances imply greater dissimilarity. The distances between transition path pairs are summarized in distance matrices and form a foundation upon which qualitative observations are built or further quantitative statistical analysis can be performed. Combined with clustering algorithms and dimensionality reduction methods, these techniques have proven useful for studying transition paths generated by different simulation methods [152]. There is also promise in being able to compare the performance of different dimensionality reduction methods, including formal techniques such as principal component analysis, independent component analysis [155–158], and full correlation analysis [159], along with heuristic projections (e.g., angle-angle and domain mass-center coordinates used for the AdK transition).

As an example, in Fig. 7 we show a quantitative comparison of a number of apo-AdK closed → open transitions that were simulated in our earlier work [31] or were generated by publicly available servers [14, 16–18, 20–22]. For each method, two pathways were generated, utilizing either inherent stochasticity in the method or small parameter variations for fully de-

Table 2. Qualitative comparison of methods used to sample conformational transitions of AdK.

Method		Refs	Solvent Interactions	Resolution[a]	Pathways and sampling[b]			Energetics[b]	Time scales[b]	Speed[b]
Main method[c]	Sub-method(s)[d]				Type[e]	Sampling	Quality	Suitability	Suitability	
Eq MD		[43, 69, 73] [64, 72]	explicit	AA	T	–	+	0	+	–
DIMS-MD	2D US-MD/WHAM	[31]	implicit	AA	T	+ –	+ –	0 0	– –	+ –
DREM	1D & 2D WHAM	[71]	explicit	AA	T	+ –	+ –	0 0	0 –	0 –
TEE-REX		[23]	explicit	AA	T	+	+	0	0	0
EDS-MD		[8]	explicit	UA	T	+	+	0	0	0
coMD	TMD, ANM, MC/Metrop	[117]	explicit	AA	T	+	+	0	0	0
WED	double Gō	[12]	N/A	SA	E	+	+	0	+	+
NL-ENM	NMA-driven + strain-en calcs	[67]	N/A	AA	P	–	0	-/0	–	+
PNM	CPR	[47]	N/A	C_α	P	–	0	-/0	–	+
DWNM		[76]	N/A	C_α	P	–	0/+	0	–	+
ANM-MC	NMA-driven	[77]	N/A	SA	P	–	0	-/0	–	+
CG-ENI	CHARMM22 energy min	[48]	N/A	AA	P	–	0	0	–	+
VAMM	NMA-driven	[15]	N/A	C_α	P	–	0/+	0	–	+
Geometric Targeting		[14]	N/A	AA	P	+	0/+	–	–	+
CG-MD	struct-based potential	[62]	N/A	C_α	T	+	+	0/+	0	+
CG-MD	two-well pot	[75]	implicit	C_α	T	+	+	0/+	0	+
CG-MD	double Gō	[59]	N/A	C_α	T	+	+	0/+	0	+
CG-MD, NM-ENM	struct-based, Tirion	[63]	N/A	C_α	T	+, -	+, 0	0/+, -/0	0, -	+, +
CG-MD		[151]	N/A	SA	T	+	+	0/+	0	+
Eq MD	prin curve + 1D US-HREM/WHAM	[66]	explicit	AA	T	+ –	+ –	0 +	– –	– –
Eq MD	2D US-MD/WHAM	[58]	explicit	AA	T	+ –	+ –	0 +	+ –	– –
NEB	$1D^+$ US-MD/WHAM	[61]	implicit	AA	P	0	0	+	–	0
Structural Morphing	1D US-MD	[74]	explicit	AA	P	0/+	0/+	+	–	0
On-the-fly String	$1D^+$/MBAR US-TMD	[24]	explicit	AA	P	0	0/+	0/+	–	0
dMD	importance sampling	[21]	implicit	AA	P	0	0	–	–	+
ENM	OM min	[16]	N/A	C_α	P	0	0	–	–	+
ENM	NMA/geometry	[65]	N/A	C_α/AA	P	0/+	0	–	–	+

[a]Resolution is described as **AA**: all-atom; **UA**: united-atom; **SA**: semiatomistic (residue-level CG); **C_α**: C_α-level CG

[b]Suitability in various categories is qualitatively rated with **+**: very suitable/fast; **0**: adequate/normal; **−**: unsuitable or impossible/slow (see text for more details).

[c]Method name: **Eq MD**: equilibrium molecular dynamics; **TEE-REX**: temperature enhanced essential dynamics replica exchange; **EDS-MD**: enhanced dynamics sampling MD; **coMD**: collective MD; **WED**: weighted ensemble dynamics; **NL-ENM**: non-linear ENM; **PNM**: plastic network model; **DWNM**: double-well network model; **ANM-MC**: anisotropic network model Monte Carlo; **CG-ENI**: coarse-grained elastic network interpolation; **VAMM**: virtual atom molecular mechanics; **CG-MD**: coarse-grained MD; **NEB**: nudged elastic band method; **dMD**: discrete MD

[e]Sub-method name: **US-MD**: umbrella sampling MD; **WHAM**: Weighted Histogram Analysis Method; **TMD**: targeted MD; **ANM**: anisotropic network model; **MC**: Monte-Carlo; **Metrop**: Metropolis algorithm; **CPR**: conjugate peak refinement; **US-HREM**: umbrella sampling Hamiltonian replica exchange MD; **$1D^+$**: 2D free energy surface based on 1D reaction coordinate; **MBAR**: Multistate Bennett Acceptance Ratio; **OM**: Onsager-Machlup action minimization.

[e]Simulation type: **T**: transition; **E**: ensemble; **P**: pathway.

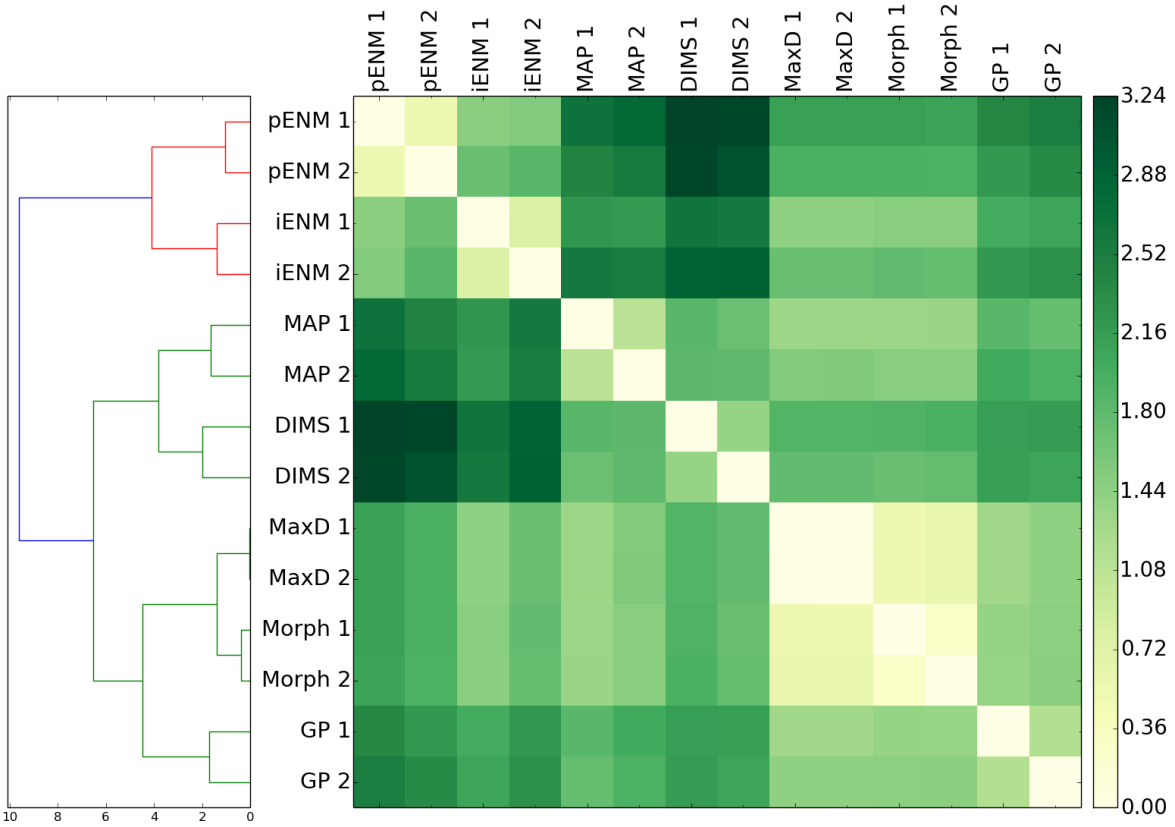
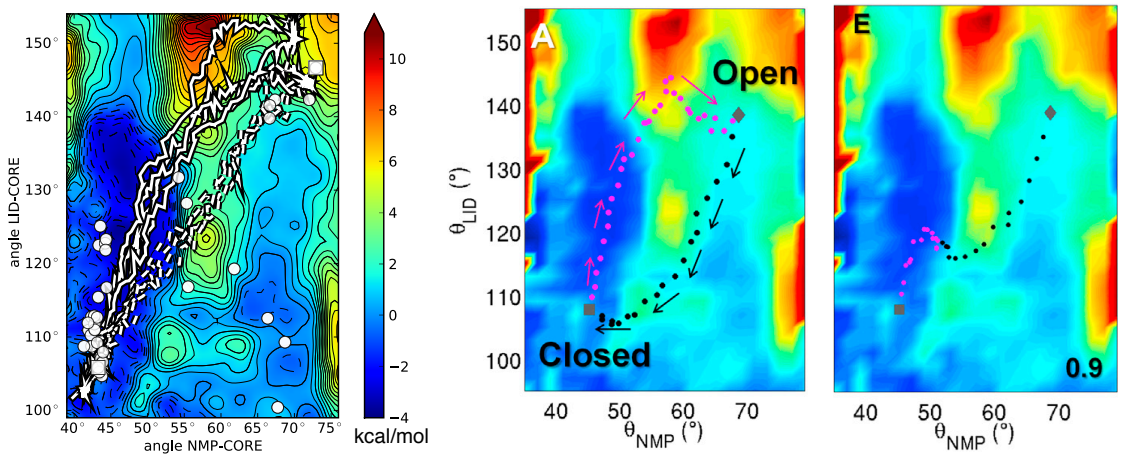


Figure 7. Quantitative comparison of apo-AdK transition paths generated by a range of methods[152]: Two independent trajectories from DIMS [13, 31], MAP (MinActionPath) [16], GeoPath [14], iENM [17], pathENM [18], NIMSim [22], MaxD (Maxwell Daemon discrete MD) [21], MolMovDB [20]. Distances are measured in Å, corresponding to a C_α-RMSD [152]. Note that the matrix is symmetric, reflecting the commutative property of metrics, while squares on the diagonal are 0 and correspond to identity. The dendrogram on the left depicts the result of hierarchical clustering using the Ward linkage criterion to define child cluster distances. The height at the top of each bifurcation is the distance between the two child clusters (according to Ward's criterion).

terministic approaches [152]. The clustered distance matrix (Fig. 7) shows that the two repeats for each method are closest to each other, as indicated by the light 2×2 squares along the diagonal. However, broader patterns also emerge. For instance, the MolMovDB [20] pathways are very similar to the transitions produced by Maxwell Daemon discrete MD [21] and generally equidistant from all other pathways, suggesting that MolMovDB produces in some sense an 'average' pathway. The all-atom DIMS-MD [13, 31] pathways, on the other hand, are typically farthest from any other transitions and also from each other, indicating that the method produces structurally diverse transitions [13]. Two ENM-based methods [17, 18] cluster with each other, while the MinActionPath method [16], which also uses an ENM as the underlying interaction representation, behaves more like DIMS-MD and the geometric pathways method [14] in that their pathways are different from all other ones.

7.2 *Validation of sampled transition paths*

If we want to determine which method most closely predicts realistic transition paths we face the difficult situation that so little is known experimentally of the transient conformations during a transition event. The separation of time scales between very short transition times (barrier crossing times for certain fast folding transitions were estimated from FRET photon counting analysis to be $< 2\mu$s [160] and still obeyed Kramer's rate theory [161]) and long residency times in stable or metastable conformations (> 1 ms for AdK) makes experimental measurements of



(a) DIMS-MD[31] (b) coMD[117] (unidirectional) (c) coMD[117] (shooting)

Figure 8. Comparison of path sampling algorithms against the free energy surface of apo-AdK, projected onto domain angles. The PMF was calculate from all-atom, implicit solvent umbrella sampling simulations [31]. (a) Examples of DIMS-MD transitions in the closed → open (continuous lines) and open → closed (dashed lines) direction [31]. White markers indicate the location of crystallographic structures. (b) Unidirectional coMD transitions [117]. (c) Representative coMD transition trajectories that are constructed from simultaneously evolving two paths from the endpoints towards each other ('shooting'), generated with a Metropolis acceptance ratio $f = 0.9$ [117]. The magenta trajectory was started from the closed conformation while the black one started from the open state. [Figure (a) Reprinted from Beckstein et al. [31] Copyright (2009), with permission from Elsevier. Figure (b), (c) Reprinted from Gur et al. [117] Copyright 2013, with permission from Elsevier.]

properties along the transition path very challenging. Therefore, it is generally very difficult to directly evaluate simulated transition paths against experimental data.

One possibility is to predict transiently formed contacts or close approaches that do not exist in either end state. Cross-linking experiments can then be used to validate these predictions. Comparison to 'intermediate' crystal structures [46] can also be used to validate intermediate structures sampled during transitions [31, 47, 48]. Comparison to crystal structures only constitutes one line of evidence and must be evaluated in conjunction with additional independent data because the factors responsible for differing crystal conformations are unlikely to be present during a conformational transition, namely different crystal contacts or the presence of special co-crystallized ligands.

If multiple rates can be measured experimentally, then the correct prediction of those rates from the simulations would act as a powerful validation. Because rates depend sensitively on both dynamic and energetic properties, they are likely only correct if the simulations provide an accurate model of the underlying physics, including sampling the correct ensemble of pathways.

Finally, in the absence of experimental data we can compare sampled pathways to pathways generated with supposedly higher quality (but computationally more demanding) approaches. For example, transition pathways should be close to the minimum free energy pathways and, as such, should move through the valleys and across the saddle points of the underlying free energy landscape. Assuming one can find reasonable collective variables for sampling the low dimensional PMF, the free energy landscape can be determined to high accuracy. In this manner, Beckstein et al. [31] compared DIMS-MD transitions to the PMF in the domain-angle variables (Fig. 1) and found reasonably good agreement (Fig. 8(a)). Recently, Gur et al. [117] also compared transitions generated with their coMD method to the same PMF. With the right choice of parameters, the pathways respect even secondary features in the PMF, including some backtracking to move around local maxima (Fig. 8(c)) and thus these promising results are fully consistent with the independently generated free energy surface. The comparison between DIMS-MD and coMD suggests an important ingredient for sampling transition paths. DIMS-MD in Ref. [31] uses the RMSD to the target conformation in order to implement a biased random walk in configuration space from the initial conformation to the target. coMD relies on local ANM modes to guide

transitions starting from both endpoints to meet somewhere in the middle, reminiscent of the shooting moves in the transition path sampling method [7]. However, if coMD is performed as unidirectional sampling towards a target, the same pattern as for DIMS-MD emerges in which trajectories depend on the starting conformation and are initially dominated by local features in the energy landscape, but towards the end are mostly driven by a reduction in the RMSD progress variable and less by the mean force (Fig. 8(b)). The use of shooting transitions appears to alleviate this problem and lead to more faithful sampling of the underlying free energy landscape.

Ultimately we are interested in how well a given method produces transitions that are indistinguishable from trajectories drawn from an equilibrium ensemble of transitions. At least in principle, such an ensemble can be produced by long unbiased equilibrium MD simulations, which represent arguably the most realistic computational description of rare events and can thus serve as a 'gold standard'. With the help of a suitable metric [152], it is then possible to quantify the quality of transition paths against these standard trajectories (Seyler and Beckstein, unpublished).

8 Conclusions

AdK (and in particular AK_{eco}) has become a useful test system to study macromolecular transitions because it is a small protein that undergoes a large and relatively easily characterizable conformational change. An abundance of experimental data are available, including multiple structures and rates. At the same time it shares the characteristics of other typical problems of interest such as conformational changes in other enzymes, molecular motors, signalling proteins, and membrane transport proteins. Real fundamental mechanistic questions remain open, including a lively debate over the coupling of conformational dynamics to the enzymatic chemistry [28, 162–164], how experimental 1D FRET measurements of the conformational change should be interpreted [31, 74], whether there is a fixed order in which NMP and LID domain move, and whether AdK is better described by a population shift/conformational selection model or by an induced fit mechanism . Computer simulations have helped in developing a detailed molecular picture of the events and lead to new hypotheses such as 'switching by cracking' [62, 63, 165, 166], which is consistent with experimental mutation studies [167] and the perturbing of energetic barriers associated with active site flexibility via the addition of denaturants [168, 169]. However, different computational methods disagree on a number of fundamental properties such as the order of domain movements or the topography of the underlying free energy landscape. Although some of these disagreements may be due to differences in how transitions are projected into low-dimensional spaces, major differences likely reflect differences in how well physical reality is modeled by different methods. What seems missing to complete an objective comparison and evaluation of the merits of each method is an experimental or computational 'gold standard' for macromolecular transitions in AdK.

References

[1] J. Yon, D. Perahia, and C. Ghélis, *Conformational dynamics and enzyme activity*, Biochimie 80 (1998) pp. 33–42.
[2] D. Kern and E. R. Zuiderweg, *The role of dynamics in allosteric regulation*, Curr Opin Struct Biol 13 (2003) pp. 748–757.
[3] M. Karplus, Y. Q. Gao, J. Ma, A. Van Der Vaart, W. Yang, and A. H. Zewail, *Protein structural transitions and their functional role*, Philos. Trans. Math. Phys. Eng. Sci. Ser. A 363 (2005) pp. 331–356.
[4] K. Henzler-Wildman and D. Kern, *Dynamic personalities of proteins*, Nature 450 (2007) pp. 964–72.
[5] S. Fischer and M. Karplus, *Conjugate peak refinement - an algorithm for finding reaction paths and accurate transition-states in systems with many degrees of freedom*, Chem. Phys. Lett. 194 (1992) pp. 252–261.
[6] G. Henkelman and H. Jónsson, *Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points*, J. Chem. Phys. 113 (2000) pp. 9978–9985.
[7] P. Bolhuis, D. Chandler, C. Dellago, and P. Geissler, *Transition path sampling: Throwing ropes over rough mountain passes, in the dark*, Ann. Rev. Phys. Chem. 53 (2002) pp. 291–318.
[8] C. Snow, G. Qi, and S. Hayward, *Essential dynamics sampling study of adenylate kinase: comparison to citrate*

synthase and implication for the hinge and shear mechanisms of domain motions., Proteins: Struct. Funct. Genet. 67 (2007) pp. 325–337.

[9] A. C. Pan, D. Sezer, and B. Roux, *Finding transition pathways using the string method with swarms of trajectories.*, J. Phys. Chem. B 112 (2008) pp. 3432–3440.

[10] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *String method in collective variables: minimum free energy paths and isocommittor surfaces.*, J. Chem. Phys. 125 (2006) p. 24106.

[11] A. van der Vaart and M. Karplus, *Minimum free energy pathways and free energy profiles for conformational transitions based on atomistic molecular dynamics simulations.*, J. Chem. Phys. 126 (2007) p. 164106.

[12] D. Bhatt and D. M. Zuckerman, *Heterogeneous path ensembles for conformational transitions in semiatomistic models of adenylate kinase*, J Chem Theory Comput 6 (2010) pp. 3527–3539.

[13] J. R. Perilla, O. Beckstein, E. J. Denning, and T. Woolf, *Computing ensembles of transitions from stable states: Dynamic importance sampling*, J. Comp. Chem. 32 (2011) pp. 186–209.

[14] D. W. Farrell, K. Speranskiy, and M. F. Thorpe, *Generating stereochemically acceptable protein pathways*, Proteins 78 (2010) pp. 2908–21.

[15] A. Korkut and W. A. Hendrickson, *Computation of conformational transitions in proteins by virtual atom molecular mechanics as validated in application to adenylate kinase*, Proc. Natl. Acad. Sci. U.S.A. EarlyEd (2009).

[16] J. Franklin, P. Koehl, S. Doniach, and M. Delarue, *MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape*, Nucleic Acids Res. 35 (2007) pp. W477–482.

[17] M. Tekpinar and W. Zheng, *Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model*, Proteins 78 (2010) pp. 2469–2481.

[18] W. Zheng, B. R. Brooks, and G. Hummer, *Protein conformational transitions explored by mixed elastic network models*, Proteins 69 (2007) pp. 43–57.

[19] D. Weiss and M. Levitt, *Can morphing methods predict intermediate structures?*, J. Mol. Biol. 385 (2009) pp. 665–674.

[20] S. Flores, N. Echols, D. Milburn, B. Hespenheide, K. Keating, J. Lu, S. Wells, E. Z. Yu, M. Thorpe, and M. Gerstein, *The database of macromolecular motions: new features added at the decade mark*, Nucleic Acids Res 34 (2006) pp. D296–301.

[21] P. Sfriso, A. Emperador, L. Orellana, A. Hospital, J. L. Gelpí, and M. Orozco, *Finding conformational transition pathways from discrete molecular dynamics simulations*, J. Chem. Theory Comput. 8 (2012) pp. 4707–4718.

[22] D. M. Krüger, A. Ahmed, and H. Gohlke, *NMSim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins*, Nucleic Acids Res 40 (2012) pp. W310–316.

[23] M. B. Kubitzki and B. L. de Groot, *The atomistic mechanism of conformational transition in adenylate kinase: a TEE-REX molecular dynamics study.*, Structure 16 (2008) pp. 1175–1182.

[24] Y. Matsunaga, H. Fujisaki, T. Terada, T. Furuta, K. Moritsugu, and A. Kidera, *Minimum free energy path of ligand-induced transition in adenylate kinase*, PLoS Comput Biol 8 (2012) p. e1002555.

[25] M. Gerstein, G. Schulz, and C. Chothia, *Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers.*, J. Mol. Biol. 229 (1993) pp. 494–501.

[26] P. C. Whitford, J. N. Onuchic, and P. G. Wolynes, *Energy landscape along an enzymatic reaction trajectory: hinges or cracks?*, HFSP J 2 (2008) pp. 61–4.

[27] E. Bae and G. N. Phillips, Jr, *Identifying and engineering ion pairs in adenylate kinases. insights from molecular dynamics simulations of thermophilic and mesophilic homologues*, J Biol Chem 280 (2005) pp. 30943–8.

[28] K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus, and D. Kern, *A hierarchy of timescales in protein dynamics is linked to enzyme catalysis*, Nature 450 (2007) pp. 913–916.

[29] O. Delalande, N. Férey, G. Grasseau, and M. Baaden, *Complex molecular assemblies at hand via interactive simulations*, J. Comput. Chem. 30 (2009) pp. 2375–2387.

[30] O. Delalande, S. Sacquin-Mora, and M. Baaden, *Enzyme closure and nucleotide binding structurally lock guanylate kinase*, Biophys. J. 101 (2011) pp. 1440–1449.

[31] O. Beckstein, E. J. Denning, J. R. Perilla, and T. B. Woolf, *Zipping and unzipping of adenylate kinase: Atomistic insights into the ensemble of open↔closed transitions*, J. Mol. Biol. 394 (2009) pp. 160–176.

[32] D. E. Atkinson, *The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers*, Biochemistry 7 (1968) pp. 4030–4.

[33] P. P. Dzeja, R. J. Zeleznikar, and N. D. Goldberg, *Adenylate kinase: kinetic behavior in intact cells indicates it is integral to multiple cellular processes*, Mol Cell Biochem 184 (1998) pp. 169–82.

[34] M. Saraste, P. R. Sibbald, and A. Wittinghofer, *The P-loop — a common motif in ATP- and GTP-binding proteins*, Trends in Biochemical Sciences 15 (1990) pp. 430 – 434.

[35] D. D. Leipe, E. V. Koonin, and L. Aravind, *Evolution and classification of P-loop kinases and related proteins*, Journal of Molecular Biology 333 (2003) pp. 781 – 815.

[36] M. Wolf-Watz, V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser, and D. Kern, *Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair*, Nature Struct. Mol. Biol. 11 (2004) pp. 945–949.

[37] M. A. Sinev, E. V. Sineva, V. Ittah, and E. Haas, *Domain closure in adenylate kinase.*, Biochemistry 35 (1996) pp. 6425–6437.

[38] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, and D. Kern, *Intrinsic motions along an enzymatic reaction trajectory*, Nature 450 (2007) pp. 838–844.

[39] J. A. Hanson, K. Duderstadt, L. P. Watkins, S. Bhattacharyya, J. Brokaw, J.-W. Chu, and H. Yang, *Illuminating the mechanistic roles of enzyme conformational dynamics*, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) pp. 18055–18060.

[40] J. Adén and M. Wolf-Watz, *NMR identification of transient complexes critical to adenylate kinase catalysis*, J. Am. Chem. Soc. 129 (2007) pp. 14003–14012.

[41] Y. E. Shapiro and E. Meirovitch, *Activation energy of catalysis-related domain motion in E. coli adenylate kinase.*, J. Phys. Chem. B 110 (2006) pp. 11519–11524.

[42] D. Tobi and I. Bahar, *Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state*, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) pp. 18908–18913.

[43] R. I. Cukier, *Apo adenylate kinase encodes its holo form: A principal component and varimax analysis*, J. Phys. Chem. B ASAP (2009).

[44] A. Bakan and I. Bahar, *The intrinsic dynamics of enzymes plays a dominant role in determining the structural*

changes induced upon inhibitor binding, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) pp. 14349–14354.
[45] D. Seeliger and B. L. de Groot, *Conformational transitions upon ligand binding: Holo-structure prediction from apo conformations*, PLoS Comp. Biol. 6 (2010) p. e1000634.
[46] C. Vonrhein, G. J. Schlauderer, and G. E. Schulz, *Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases*, Structure 3 (1995) pp. 483–490.
[47] P. Maragakis and M. Karplus, *Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase.*, J. Mol. Biol. 352 (2005) pp. 807–822.
[48] Y. Feng, L. Yang, A. Kloczkowski, and R. L. Jernigan, *The energy profiles of atomic conformational transition intermediates of adenylate kinase*, Proteins: Struct. Funct. Genet. 77 (2009) pp. 551–558.
[49] W. E, W. Ren, and E. Vanden-Eijnden, *Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes*, Chemical Physics Letters 413 (2005) pp. 242 – 247.
[50] R. B. Best and G. Hummer, *Reaction coordinates and rates from transition paths*, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) pp. 6732–6737.
[51] J. R. Perilla and T. B. Woolf, *Towards the prediction of order parameters from molecular dynamics simulations in proteins*, J Chem Phys 136 (2012) p. 164101.
[52] M. A. Rohrdanz, W. Zheng, and C. Clementi, *Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions*, Annu. Rev. Phys. Chem. 64 (2013) pp. 295–316.
[53] S. V. Krivov, *On reaction coordinate optimality*, J. Chem. Theory Comput. 9 (2013) pp. 135–146.
[54] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, *Diffusive dynamics of the reaction coordinate for protein folding funnels*, J. Chem. Phys. 104 (1996) pp. 5860–5868.
[55] P. Bolhuis, C. Dellago, and D. Chandler, *Reaction coordinates of biomolecular isomerization*, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) pp. 5877–5882.
[56] W. Min, X. S. Xie, and B. Bagchi, *Two-dimensional reaction free energy surfaces of catalytic reaction: effects of protein conformational dynamics on enzyme catalysis*, J. Phys. Chem. B 112 (2007) pp. 454–466.
[57] S. Krivov and M. Karplus, *One-dimensional free-energy profiles of complex systems: Progress variables that preserve the barriers*, J. Phys. Chem. B 110 (2006) pp. 12689–12698.
[58] B. Jana, B. V. Adkar, R. Biswas, and B. Bagchi, *Dynamic coupling between the LID and NMP domain motions in the catalytic conversion of atp and amp to adp by adenylate kinase*, J Chem Phys 134 (2011) p. 035101.
[59] M. D. Daily, G. N. Phillips Jr., and Q. Cui, *Many local motions cooperate to produce the adenylate kinase conformational transition*, J Mol Biol 400 (2010) pp. 618–631.
[60] R. Potestio, F. Pontiggia, and C. Micheletti, *Coarse-grained description of protein internal dynamics: an optimal strategy for decomposing proteins in rigid subunits*, Biophys. J. 96 (2009) pp. 4993–5002.
[61] K. Arora and C. L. Brooks, III, *Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism*, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) pp. 18496–18501.
[62] P. C. Whitford, O. Miyashita, Y. Levy, and J. N. Onuchic, *Conformational transitions of adenylate kinase: switching by cracking*, J. Mol. Biol. 366 (2007) pp. 1661–1671.
[63] P. C. Whitford, S. Gosavi, and J. N. Onuchic, *Conformational transitions in adenylate kinase:allosteric communication reduces misligation*, J. Biol. Chem. 283 (2008) pp. 2042–2048.
[64] C. Peng, L. Zhang, and T. Head-Gordon, *Instantaneous normal modes as an unforced reaction coordinate for protein conformational transitions*, Biophys. J. 98 (2010) pp. 2356–2364.
[65] A. Ahmed, F. Rippmann, G. Barnickel, and H. Gohlke, *A normal mode-based geometric simulation approach for exploring biologically relevant conformational transitions in proteins*, J. Chem. Inf. Model. 51 (2011) pp. 1604–1622.
[66] H. D. Song and F. Zhu, *Conformational dynamics of a ligand-free adenylate kinase*, PLoS One 8 (2013) p. e68023.
[67] O. Miyashita, P. G. Wolynes, and J. N. Onuchic, *Simple energy landscape model for the kinetics of functional transitions in proteins.*, J. Phys. Chem. B 109 (2005) pp. 1959–1969.
[68] M. S. Liu, B. D. Todd, and R. J. Sadus, *Allosteric conformational transition in adenylate kinase: Dynamic correlations and implication for allostery*, Aust. J. Chem. 63 (2010) pp. 405–412.
[69] H. Lou and R. Cukier, *Molecular dynamics of apo-adenylate kinase: A principal component analysis*, J. Phys. Chem. B 110 (2006) pp. 12796–12808.
[70] D. Seeliger, J. Haas, and B. L. de Groot, *Geometry-based sampling of conformational transitions in proteins.*, Structure 15 (2007) pp. 1482–1492.
[71] H. Lou and R. I. Cukier, *Molecular dynamics of apo-adenylate kinase: A distance replica exchange method for the free energy of conformational fluctuations*, J. Phys. Chem. B 110 (2006) pp. 24121–24137.
[72] J. B. Brokaw and J.-W. Chu, *On the roles of substrate binding and hinge unfolding in conformational changes of adenylate kinase*, Biophys J 99 (2010) pp. 3420–9.
[73] F. Pontiggia, A. Zen, and C. Micheletti, *Small- and large-scale conformational changes of adenylate kinase: a molecular dynamics study of the subdomain motion and mechanics.*, Biophys. J. 95 (2008) pp. 5901–5912.
[74] D. A. Potoyan, P. I. Zhuravlev, and G. A. Papoian, *Computing free energy of a large-scale allosteric transition in adenylate kinase using all atom explicit solvent simulations*, J. Phys. Chem. B 116 (2012) pp. 1709–1715.
[75] Q. Lu and J. Wang, *Single molecule conformational dynamics of adenylate kinase: Energy landscape, structural correlations, and transition state ensembles*, J. Am. Chem. Soc. 130 (2008) pp. 4772–4783.
[76] J.-W. Chu and G. A. Voth, *Coarse-grained free energy functions for studying protein conformational changes: a double-well network model.*, Biophys. J. 93 (2007) pp. 3860–3871.
[77] N. Kantarci-Carsibasi, T. Haliloglu, and P. Doruker, *Conformational transition pathways explored by Monte Carlo simulation integrated with collective modes*, Biophys. J. 95 (2008) pp. 5862–5873.
[78] L. S. Stelzl, P. W. Fowler, M. S. Sansom, and O. Beckstein, *Flexible gates generate occluded intermediates in the transport cycle of LacY*, J Mol Biol e-pub (2013).
[79] Y. Wang, L. Gan, E. Wang, and J. Wang, *Exploring the dynamic functional landscape of adenylate kinase modulated by substrates*, J. Chem. Theory Comput. 9 (2013) pp. 84–95.
[80] H.-X. Zhou, *Rate theories for biologists*, Q. Rev. Biophys. 43 (2010) pp. 219–293.
[81] D. G. Truhlar, B. C. Garrett, and S. J. Klippenstein, *Current status of Transition-State theory*, J. Phys. Chem. 100 (1996) pp. 12771–12800.
[82] E. Vanden-Eijnden and F. A. Tal, *Transition state theory: variational formulation, dynamical corrections, and error estimates*, J. Chem. Phys. 123 (2005) p. 184103.

[83] P. Metzner, C. Schütte, and E. Vanden-Eijnden, *Illustration of transition path theory on a collection of simple examples*, J. Chem. Phys. 125 (2006) p. 084110.
[84] Z. Zhang and H. S. Chan, *Transition paths, diffusive processes, and preequilibria of protein folding*, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) pp. 20919–20924.
[85] E. Pollak and J. Ankerhold, *Improvements to kramers turnover theory*, J. Chem. Phys. 138 (2013) p. 164116.
[86] H. Kramers, *Brownian motion in a field of force and the diffusion model of chemical reactions*, Physica 7 (1940) pp. 284–304.
[87] P. Hänggi, P. Talkner, and M. Borkovec, *Reaction-rate theory: fifty years after kramers*, Rev. Mod. Phys. 62 (1990) pp. 251–341.
[88] A. Ma, A. Nag, and A. R. Dinner, *Dynamic coupling between coordinates in a model for biomolecular isomerization*, J. Chem. Phys. 124 (2006) p. 144911.
[89] C. D. Snow, Y. M. Rhee, and V. S. Pande, *Kinetic definition of protein folding transition state ensembles and reaction coordinates*, Biophys. J. 91 (2006) pp. 14–24.
[90] M. Sega, P. Faccioli, F. Pederiva, and H. Orland, *Kramers theory for conformational transitions of macromolecules*, 2008.
[91] W. E and E. Vanden-Eijnden, *Transition-path theory and path-finding algorithms for the study of rare events*, Annu Rev Phys Chem 61 (2010) pp. 391–420.
[92] D. Frenkel and B. Smit, *Understanding Molecular Simulations*, Academic Press, San Diego, 2nd ed., 2002.
[93] S. A. Adcock and J. A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins*, Chem Rev 106 (2006) pp. 1589–615.
[94] A. D. Mackerell, *Empirical force fields for biological macromolecules: Overview and issues*, J. Comp. Chem. 25 (2004) pp. 1584–1604.
[95] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, *Systematic validation of protein force fields against experimental data*, PLoS ONE 7 (2012) p. e32131.
[96] K. Hinsen and G. R. Kneller, *Solvent effects in the slow dynamics of proteins*, Proteins: Struct. Funct. Bioinf. 70 (2008) pp. 1235–1242.
[97] D. R. Roe, A. Okur, L. Wickstrom, V. Hornak, and C. Simmerling, *Secondary structure bias in generalized born solvent models: comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation*, J. Phys. Chem. B 111 (2007) pp. 1846–1857.
[98] D. M. Zuckerman, *Equilibrium sampling in biomolecular simulations*, Annu. Rev. Biophys. 40 (2011) pp. 41–62.
[99] D. M. Zuckerman1 and T. B. Woolf, *Efficient dynamic importance sampling of rare events in one dimension*, Physical Review E 63 (2000) pp. 1–10.
[100] D. M. Zuckerman and T. B. Woolf, *Dynamic reaction paths and rates through importance-sampled stochastic dynamic reaction paths and rates through importance-sampled stochastic dynamic reaction paths and rates through importance-sampled stochastic dynamics*, J. Chem. Phys. 111 (1999) pp. 9475–9484.
[101] T. B. Woolf, *Path corrected functionals of stochastic trajectories: towards path corrected functionals of stochastic trajectories: towards relative free energy and reaction coordinate calculations*, Chemical Physics Letters 289 (1998) pp. 433–441.
[102] J. Schlitter, M. Engels, and P. Krüger, *Targeted molecular-dynamics - a new approach for searching pathways of conformational transitions*, J. Mol. Graph. 12 (1994) pp. 84–89.
[103] A. F. Voter, *Hyperdynamics: Accelerated molecular dynamics of infrequent events*, Phys. Rev. Lett. 78 (1997) pp. 3908–3911.
[104] D. Hamelberg, J. Mongan, and J. A. McCammon, *Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules.*, J. Chem. Phys. 120 (2004) pp. 11919–11929.
[105] Y. Sugita and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*, Chem. Phys. Lett. 314 (1999) pp. 141–151.
[106] M. R. So/rensen and A. F. Voter, *Temperature-accelerated dynamics for simulation of infrequent events*, J. Chem. Phys. 112 (2000) pp. 9599–9606.
[107] C. F. Abrams and E. Vanden-Eijnden, *Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics*, Proceedings of the National Academy of Sciences 107 (2010) pp. 4961–4966.
[108] G. Huber and S. Kim, *Weighted-ensemble brownian dynamics simulations for protein association reactions*, Biophys. J. 70 (1996) pp. 97–110.
[109] C.-P. Chng and L.-W. Yang, *Coarse-grained models reveal functional dynamics–II. molecular dynamics simulation at the coarse-grained level–theories and biological applications*, Bioinform. Biol. Insights 2 (2008) pp. 171–185.
[110] W. G. Noid, *Perspective: Coarse-grained models for biomolecular systems*, J Chem Phys 139 (2013) p. 090901.
[111] L.-W. Yang and C.-P. Chng, *Coarse-grained models reveal functional dynamics–i. elastic network models–theories, comparisons and perspectives*, Bioinform. Biol. Insights 2 (2008) pp. 25–45.
[112] I. Bahar, T. R. Lezon, L.-W. Yang, and E. Eyal, *Global dynamics of proteins: bridging between structure and function*, Annu Rev Biophys 39 (2010) pp. 23–42.
[113] Y.-H. Sanejouand, *Elastic network models: theoretical and empirical foundations*, Methods Mol Biol 924 (2013) pp. 601–616.
[114] M. M. Tirion, *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis*, Phys. Rev. Lett. 77 (1996) pp. 1905–1908.
[115] I. Bahar, A. R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential*, Fold. Des. 2 (1997) pp. 173–181.
[116] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, *Anisotropy of fluctuation dynamics of proteins with an elastic network model*, Biophys. J. 80 (2001) pp. 505–515.
[117] M. Gur, J. D. Madura, and I. Bahar, *Global transitions of proteins explored by a multiscale hybrid methodology: Application to adenylate kinase*, Biophysical Journal 105 (2013) pp. 1643 – 1652.
[118] P. C. Whitford, K. Y. Sanbonmatsu, and J. N. Onuchic, *Biomolecular dynamics: order-disorder transitions and energy landscapes*, Rep. Prog. Phys. 75 (2012) p. 076601.
[119] W. E, W. Ren, and E. Vanden-Eijnden, *Finite temperature string method for the study of rare events*, J Phys Chem B 109 (2005) pp. 6688–6693, pMID: 16851751.
[120] H. Jónsson, G. Mills, and K. W. Jacobsen, *Nudged elastic band method for finding minimum energy paths of transitions*, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, eds. B. J. Berne, G. Ciccoti, and D. F.

Coker, chap. 16, pp. 385–394, World Scientific, 1998.
[121] G. Henkelman, B. P. Uberuaga, and H. Jónsson, *A climbing image nudged elastic band method for finding saddle points and minimum energy paths*, J. Chem. Phys. 113 (2000) pp. 9901–9904.
[122] S. Kirillova, J. Cortés, A. Stefaniu, and T. Siméon, *An NMA-guided path planning approach for computing large-amplitude conformational changes in proteins*, Proteins 70 (2008) pp. 131–143.
[123] B. Raveh, A. Enosh, O. Schueler-Furman, and D. Halperin, *Rapid sampling of molecular motions with prior information constraints*, PLoS Comput Biol 5 (2009) p. e1000295.
[124] L. Kavraki, P. Svestka, J.-C. Latombe, and M. Overmars, *Probabilistic roadmaps for path planning in high-dimensional configuration spaces*, Robotics and Automation, IEEE Transactions on 12 (1996) pp. 566–580.
[125] M. Apaydin, A. Singh, D. Brutlag, and J.-C. Latombe, *Capturing molecular energy landscapes with probabilistic conformational roadmaps*, in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, vol. 1, pp. 932–939 vol.1, 2001.
[126] M. S. Head, J. A. Given, and M. K. Gilson, *"mining minima": Direct computation of conformational free energy*, J. Phys. Chem. A 101 (1997) pp. 1609–1618.
[127] D. Seeliger, J. Haas, and B. L. de Groot, *Geometry-based sampling of conformational transitions in proteins*, Structure 15 (2007) pp. 1482–92.
[128] J. Cortés, T. Siméon, V. Ruiz de Angulo, D. Guieysse, M. Remaud-Siméon, and V. Tran, *A path planning approach for computing large-amplitude motions of flexible molecules*, Bioinformatics 21 Suppl 1 (2005) pp. i116–25.
[129] J. P. Valleau and G. M. Torrie, *A guide to Monte Carlo for statistical mechanics: 2. Byways*, in *Statistical Mechanics. Part A: Equilibrium Techniques*, ed. B. J. Berne, vol. 5 of *Modern Theoretical Chemistry*, chap. 5, pp. 169–194, Plenum Press, New York, 1977.
[130] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, *The Weighted Histogram Analysis Method for free-energy calculations on biomolecules. I. The method*, J. Comp. Chem. 13 (1992) pp. 1011–1021.
[131] B. Roux, *The calculation of the potential of mean force using computer simulations*, Comp. Phys. Comm. 91 (1995) pp. 275–282.
[132] M. R. Shirts and J. D. Chodera, *Statistically optimal analysis of samples from multiple equilibrium states*, J Chem Phys 129 (2008) p. 124105.
[133] E. Darve, D. Rodríguez-Gómez, and A. Pohorille, *Adaptive biasing force method for scalar and vector free energy calculations*, J. Chem. Phys. 128 (2008) p. 144120.
[134] J. Hénin, G. Fiorin, C. Chipot, and M. L. Klein, *Exploring multidimensional free energy landscapes using time-dependent biases on collective variables*, Journal of Chemical Theory and Computation 6 (2010) pp. 35–47.
[135] C. B. Barnett and K. J. Naidoo, *Free energies from adaptive reaction coordinate forces (FEARCF): an application to ring puckering*, Mol. Phys. 107 (2009) pp. 1243–1250.
[136] A. Laio and M. Parrinello, *Escaping free-energy minima.*, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) pp. 12562–12566.
[137] M. D. Daily, L. Makowski, G. N. Phillips, and Q. Cui, *Large-scale motions in the adenylate kinase solution ensemble: coarse-grained simulations and comparison with solution x-ray scattering*, Chem. Phys. 396 (2012) pp. 84–91.
[138] H. Dong, S. Qin, and H.-X. Zhou, *Effects of macromolecular crowding on protein conformational changes*, PLoS Comput. Biol. 6 (2010) p. e1000833.
[139] J. Åden, A. Verma, A. Schug, and M. Wolf-Watz, *Modulation of a pre-existing conformational equilibrium tunes adenylate kinase activity*, J. Am. Chem. Soc. 134 (2012) pp. 16562–16570.
[140] Plotkin, Wang, and Wolynes, *Correlated energy landscape model for finite, random heteropolymers*, Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics 53 (1996) pp. 6271–6296.
[141] D. Bhatt and D. M. Zuckerman, *Beyond microscopic reversibility: Are observable non-equilibrium processes precisely reversible?*, J Chem Theory Comput 7 (2011) pp. 2520–2527.
[142] D. E. Koshland, *Application of a theory of enzyme specificity to protein synthesis*, Proc Natl Acad Sci U S A 44 (1958) pp. 98–104.
[143] S. M. Sullivan and T. Holyoak, *Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection*, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) pp. 13829–13834.
[144] K. U. Linderstrøm-Lang and J. A. Schellman, *Protein structure and enzyme activity*, in *The Enzymes*, eds. P. D. Boyer, H. A. Lardy, and K. Myrbäk, Academic Press, New York, 1959.
[145] P. Csermely, R. Palotai, and R. Nussinov, *Induced fit, conformational selection and independent dynamic segments: an extended view of binding events*, Trends Biochem Sci 35 (2010) pp. 539–46.
[146] G. G. Hammes, Y.-C. Chang, and T. G. Oas, *Conformational selection or induced fit: A flux description of reaction mechanism*, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) pp. 13737–13741.
[147] H.-X. Zhou, *From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions*, Biophys J 98 (2010) pp. L15–7.
[148] A. D. Vogt and E. Di Cera, *Conformational selection or induced fit? a critical appraisal of the kinetic mechanism*, Biochemistry 51 (2012) pp. 5894–5902.
[149] J. Ping, P. Hao, Y.-X. Li, and J.-F. Wang, *Molecular dynamics studies on the conformational transitions of adenylate kinase: a computational evidence for the conformational selection mechanism*, Biomed Res. Int. 2013 (2013) p. 628536.
[150] J. Kubelka, J. Hofrichter, and W. A. Eaton, *The protein folding 'speed limit'*, Curr. Opin. Struct. Biol. 14 (2004) pp. 76–88.
[151] R. D. H. Jr., L. Lu, and G. A. Voth, *Multiscale coarse-graining of the protein energy landscape*, PLoS Computational Biology 6 (2010) pp. 1–12.
[152] S. L. Seyler, A. Kumar, M. Thorpe, and O. Beckstein, *Quantifying macromolecular conformational transition paths*, 2013, in preparation.
[153] H. Alt and L. Scharf, *Computing the Hausdorff distance between curved objects*, Int. J. Comput. Geom. Ap. 18 (2008) pp. 307–320.
[154] H. Alt and M. Godau, *Computing the fréchet distance between two polygonal curves*, J. Comput. Geom. Appl. 5 (1993) pp. 75–91.
[155] A. Hyvärinen and E. Oja, *Independent component analysis: algorithms and applications*, Neural Netw. 13 (2000) pp. 411–430.
[156] P. Comon, *Independent component analysis, a new concept? - higher order statistics*, Signal Processing 36 (1994) pp. 287–314.
[157] Y. Naritomi and S. Fuchigami, *Slow dynamics in protein fluctuations revealed by time-structure based independent*

component analysis: the case of domain motions, J. Chem. Phys. 134 (2011) p. 065101.

[158] Y. Naritomi and S. Fuchigami, *Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis*, J. Chem. Phys. 139 (2013) p. 215102.

[159] O. F. Lange and H. Grubmüller, *Full correlation analysis of conformational protein dynamics*, Proteins 70 (2008) pp. 1294–1312.

[160] H. S. Chung, K. McHale, J. M. Louis, and W. A. Eaton, *Single-molecule fluorescence experiments determine protein folding transition path times*, Science 335 (2012) pp. 981–984.

[161] H. S. Chung and W. A. Eaton, *Single-molecule fluorescence probes dynamics of barrier crossing*, Nature 502 (2013) pp. 685–688.

[162] A. V. Pisliakov, J. Cao, S. C. L. Kamerlin, and A. Warshel, *Enzyme millisecond conformational dynamics do not catalyze the chemical step*, Proceedings of the National Academy of Sciences 106 (2009) pp. 17359–17364.

[163] M. Karplus, *Role of conformation transitions in adenylate kinase*, Proceedings of the National Academy of Sciences 107 (2010) p. E71.

[164] S. C. L. Kamerlin and A. Warshel, *Reply to karplus: Conformational dynamics have no role in the chemical step*, Proceedings of the National Academy of Sciences 107 (2010) p. E72.

[165] O. Miyashita, J. N. Onuchic, and P. G. Wolynes, *Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins.*, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) pp. 12570–12575.

[166] U. Olsson and M. Wolf-Watz, *Overlap between folding and functional energy landscapes for adenylate kinase conformational change*, Nat Commun 1 (2010) p. 111.

[167] T. P. Schrank, D. W. Bolen, and V. J. Hilser, *Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins*, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) pp. 16984–16989.

[168] H. J. Zhang, X. R. Sheng, X. M. Pan, and J. M. Zhou, *Activation of adenylate kinase by denaturants is due to the increasing conformational flexibility at its active sites*, Biochem. Biophys. Res. Commun. 238 (1997) pp. 382–386.

[169] L. Rundqvist, J. Adén, T. Sparrman, M. Wallgren, U. Olsson, and M. Wolf-Watz, *Noncooperative folding of subdomains in adenylate kinase*, Biochemistry 48 (2009) pp. 1911–1927.