Molecular Simulation

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gmos20

Sampling large conformational transitions: adenylate kinase as a testing ground

Sean L. Seyler & Oliver Beckstein
Department of Physics, Center for Biological Physics, Arizona State University, Tempe, AZ 85287-1504, USA
Published online: 29 May 2014.

To cite this article: Sean L. Seyler & Oliver Beckstein (2014): Sampling large conformational transitions: adenylate kinase as a testing ground, Molecular Simulation, DOI: 10.1080/08927022.2014.919497

To link to this article: http://dx.doi.org/10.1080/08927022.2014.919497

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
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 $\rightarrow$ 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 transporters cycle between multiple conformational states in order to fulfil 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 techniques 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. First, 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 conceptualise. 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] Second, there is an abundance of known intermediate crystal structures that span the configuration space between the open and closed crystal structures. Third, 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 focuses on the AdK enzyme of the mesophilic bacterium *Escherichia coli* (AKeco) 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 AKeco 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 (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 catalyses the phosphoryl transfer from an ATP to an AMP molecule in the presence of a catalytic magnesium ion:

\[
\text{Mg}^{2+}\text{ATP} + \text{AMP} \rightleftharpoons \text{Mg}^{2+}\text{ADP} + \text{ADP}. \tag{1}
\]

AdK is a relatively small protein – AKeco contains 214 residues with a molecular weight of 23.6 kDa – and consists of three distinct domains, shown in Figure 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, 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 catalysed 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 of the order of 300 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 characterised structurally, with more than 50 crystal structures of AKeco 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 crystallised 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 characterising conformational transitions, one is often interested in identifying energetic basins of attraction that
define metastable states and the transition state (TS) 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 TS ensemble. Order parameters characterise 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 (TPS)) 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 TS surface is defined for a general dynamical system by its separatrix. The commitor, or commitment probability, $p_i(x, \tau)$, is the probability that trajectories initialised in configuration $x$ will end in state $i$ in a (short) time $\tau$. Stated another way, a trajectory initialised at $x$ will commit to state $i$ in a time $\tau$ with probability $p_i(x, \tau)$. An isocommittor surface is a subspace of phase space such that the commitor is constant. Thus, for a transition system with two stable states A and B, it is clear that the TS surface (and separatrix) must lie on the isocommittor surface corresponding to equal commitment probability, i.e. $p_A = p_B = 1/2$. Isocommittor 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 behaviour 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 $\chi$ should distinguish between states A and B, and a discriminating order parameter requires that, for some intermediate value $\chi^*$. $\chi < \chi^*$ 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 parameterise a free energy function $F(q)$ and $q^*$ maximises $F$ between states A and B, then $q^*$ should coincide with the TS 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 TS surface; an acceptable order parameter is not guaranteed to adequately characterise the TS 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 summarised 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 centres of mass and RMSD with respect to open/closed structure), dihedral or contact space.

<table>
<thead>
<tr>
<th>Collective variables/order parameters</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D LID/NMP–CORE MCD</td>
<td>[24,58,61,69,74]</td>
</tr>
<tr>
<td>RMSD (to an end structure)</td>
<td>[15,61,67,75]</td>
</tr>
<tr>
<td>Indices along Rxn coordinate</td>
<td>[24,47,48,66,76,77]</td>
</tr>
<tr>
<td>2D LID/NMP–CORE MCD</td>
<td>[24,58,61–66]</td>
</tr>
<tr>
<td>LID/NMP–CORE angle</td>
<td>[15,31,59,67,68]</td>
</tr>
<tr>
<td>RMSD (to end structures)</td>
<td>[23,59,62]</td>
</tr>
<tr>
<td>Essential variables</td>
<td>[8,23,43,65,69,70]</td>
</tr>
<tr>
<td>Fraction of native contacts</td>
<td>[16,59]</td>
</tr>
<tr>
<td>Other</td>
<td>[71–73]</td>
</tr>
</tbody>
</table>
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 CMD (between residues 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 10 most significant essential modes (principal components). In characterising the AdK closed ↔ 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 (FESs) or snapshots of simulation conformers is a heuristic measure consisting of the centre of mass distances (CMD) between domains (centre of mass distance between the LID and CORE, and the centre 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 characterise 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, monitors LID–CORE variation differently: LID–CORE CMD, [71] Cα distance between residues 127 and 194 [72] and LID–CORE angle [73] Progress along one dimension (e.g. plotting the potential of mean force (PMF)) was captured using domain CMD. [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 Cα 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 [37]) and A127–A194 (LID–CORE [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 while at the same time making a direct connection to experiments. The FRET pair 152–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 Cα distances, although methods are available (in principle) to post-process 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 with 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.

TST and Kramers’ theory can be used to produce a rate constant expression with exponential dependence on the free energy barrier height (see Equation (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 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 optimise the TS surface such that the crossing frequency is minimised, 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 \( q(x) \) on the configuration space \( x \in \mathbb{R}^{3N} \) 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 generalised 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]

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 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 require finding transition tubes, which in turn rely on finding minimum free energy paths (MFEPs) 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
MD is an approach to simulating dynamical systems in which interactions between particles are parametrised 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 10s 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 non-bonded forces associated with electrostatic and van der Waals forces.[94] Molecular mechanics force fields are parametrised 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 with explicit solvent simulations, although it has been found that solvation effects can alter the slow dynamics of a macromolecular system [96] as well as secondary structures and native state populations.[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 MD,[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 CG 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 NMA on an 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 approach (often at the Cα-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,[115] anharmonic network models (ANMs) [116] and plastic network models (PNMs).[47] Several AdK studies implement an ENM, including the PNM,[47] and two other methods utilise normal mode information from an ANM as a means to enhance sampling.[77,117] We also generate and examine AdK closed → 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. MFEP methods

MFEP and related methods generate single transition paths from a given force field representation by applying local optimisation 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 CG and/or simplified potentials. Some examples include TPS,[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] Another group of related sampling methods that sidesteps explicit integration of Newton’s equations intelligently searches conformation space according to prior information (energetic, conformational/steric) constraints. The class of motion-planning algorithms includes the rapidly exploring random trees (RRTs) algorithm,[122,123] the probabilistic road map [124,125] and the ‘mining-minima’ algorithm.[126] Other related, geometric-based algorithms include geometry-based essential sampling,[127] geometric-based RRTs [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 MFEPs from the free energy landscape using MFEP methods. Umbrella sampling [129] combined with the weighted histogram analysis method (WHAM) [130,131] or multistate Bennett acceptance ratio (MBAR) [132] unbiasing algorithms is a commonly used technique for generating PMFs, 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 FESs 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 to 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 small-angle X-ray scattering.[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 unfavourable.[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 kcal·mol⁻¹ (1.37k_BT) 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 0.75k_BT was achieved while Daily et al.[59] chose ∼ 0.5 kcal·mol⁻¹ (∼ 0.84k_BT) between the native apo-open and apo-closed states in their CG 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 conformation.[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

$$\frac{k_{\text{open}}}{k_{\text{close}}} \overset{\text{open}}{=} \overline{\text{closed}} \quad (2)$$

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

$$\Delta G = G_{\text{closed}} - G_{\text{open}} = -k_B T \ln K$$

$$= -k_B T \ln \frac{k_{\text{close}}}{k_{\text{open}}} \quad (3)$$

From the apo-opening and apo-closing rates for a thermophilic enzyme, respectively, $k_{\text{open}} = 6500 \pm 500 \text{s}^{-1}$ and $k_{\text{close}} = 2000 \pm 200 \text{s}^{-1}$, measured by Henzler-Wildman et al.[38] (see also Section 6.2), one arrives at $\Delta G = (1.2 \pm 0.1)k_BT$, 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 AKeco, which displays much higher rates for the substrate-bound enzyme than the thermophilic one.[36,38,41] The experimental free energy difference between AKeco apo closed and apo open is thus not known. Hanson et al.[39] measured specific LID opening and closing rates for AKeco (see Section 6.2), which translate into $\Delta G = (-0.6 \pm 0.5)k_BT$. 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 Equation (2).

The conformational equilibrium of AdK was also studied under the effect of cellular crowding using a post-processing 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 towards the closed state. Given that the closed → open transition is thought to be the rate-limiting step, a shift towards the closed ensemble would reduce the closed → open transition rate. These ideas appear to be consistent with recent NMR measurements (combined with MD simulations) that found that the conformational equilibrium of AKeco can be strongly
influenced by the concentration of the osmolyte trimethylamine oxide 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.[12,15,16,23,31,48,58,62–64,69,71,72,74–77,99,117] 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 a monotonic [8,61,66] or almost monotonic [24] rise in energy towards the closed crystal state. Potoyan et al. [74] and Zhu [66] produce 1D PMF calculations that show an energetic minimum near the open crystallographic structure and a monotonic [8,61,66] or almost monotonic [24] rise in energy towards the closed crystal state. Potoyan et al. [74] obtained an energy difference between the two conformational states of the order of $1.5k_B T \pm 0.5k_B T$, in contrast to the many $10$s of $k_B T$ 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 led to artefacts. Matsunaga et al. [24] and Song and Zhu [66] show smaller disagreement with Potoyan et al. [74], predicting free energy differences of the order of $20k_B T$, but still differing by an order of magnitude.

The range of free energy differences between the open and closed states in 1D PMFs is likely due to the use of different reaction coordinates coupled with the evident diversity of the conformational dynamics of AdK (Figure 3). Both Arora and Brooks [61] and Song and Zhu [66] strongly suggest that 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 Figure 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 optimisation. 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 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, suggests that a single transition tube is inadequate for sampling the diversity of the conformational dynamics of AdK. Many studies, however, have computed a 2D FES projected onto the LID–CORE and NMP–CORE CMDs[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 Figure 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 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 are intermediate between their open and closed conformations with the NMP domain being 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, characterised by \( 5 \text{–} 10 \text{ns} \) 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 CMD space showed possible transient structures near the 1dvr (LID-closed/NMP-open) and 2ak3 (LID-open/NMP-closed) crystallographic structures (Figure 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 necessarily 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 \( \rightarrow \) 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 Figure 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 \( \rightarrow \) open transition (labelled in green) and NMP-closing preceding LID-closing in the open \( \rightarrow \) closed transition.
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), 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 (DWNM), Chu and Voth [76] found that either pathway could be produced, depending on the initial starting path that was optimised in the MFEP calculation. Using linear interpolation between the open and closed structures in configuration space (as the initial path for MFEP) produced a LID-opening-first pathway; initialising the MFEP calculation from a PNM-derived initial path produced an 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 REMD 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 $\rightarrow$ open transition conducted by Song and Zhu [66] showed a possible slight initial opening of the NMP domain (Figure 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 opening followed by LID-opening. Nevertheless, the observation of a slight A˚ change in the NMP–CORE CMD of the NMP domain is likely to be sensitive to the atomicistic 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, CG 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 $\rightarrow$ open transition (see Figure 5 for the definition of the pathways). The open $\rightarrow$ 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 $\rightarrow$ 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 are discussed in more detail in Section 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 directions 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 (Figure 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-closing (opening) motion precedes the NMP-opening (closing) motion for both open $\rightarrow$ closed and closed $\rightarrow$ open transitions, may be due to simulations having been initialised 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 initialisation bias in theory, initialised trajectories must be consistent with the ratio of fluxes entering equilibrium substates for exact symmetry to hold [141]; however, approximate symmetry holds if initialised 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 their initial structure and be able to emerge from the state in a quasi-Markovian manner. Thus, while properly equilibrated MD simulations should meet the conditions for approximate symmetry if there are no slow internal time scales 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 initialisation basin.

The set-up of an ensemble of many MD simulations required Peng et al. [64] and Song and Zhu [66] to employ an initialisation procedure that ensured individual simulations, and trajectories, were uncorrelated. The general initialisation procedure involved (1) extracting a set of frames (conformations) from an equilibration or equilibrium MD simulation over several nanoseconds, (2) splitting each frame into 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 all remained in the vicinity of 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 10th 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; the pathways where the NMP domain was partially open or closed while the LID’s state was anticorrelated were far less populated. (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 → 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 the three other studies mentioned above [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 artefacts of CG potential models.

4.4. Further considerations

It is imperative to keep in mind that sampling methods relying on the construction of a collective variable space using PCA or NMA are biased towards 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, CG-MD and ENM-based multiscale MD, which are based on the analysis of the motion of 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 those sampled by the substrate-bound (holo) enzyme.[40]

Thus, large-scale closed ↔ open-transitions are observable in the absence of substrates (ATP, ADP and AMP) so that one may unambiguously separate dynamics due to the protein alone from protein–ligand interactions. Experiments and simulations that include ligands often make use of non-hydrolysable analogues such as Ap5A (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 those 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 stabilised by the presence of the ligands. Some representative examples of 1D and 2D PMFs are shown in Figure 6. Most computational studies agree that the presence of substrates will decrease the free energy of the closed conformation relative to the open conformation so that the closed conformation will become the more favorable state [24,59,61,62,79] (Figure 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] (Figure 6(a)–(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] (Figure 6(d)).

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

Figure 6. (Colour online) FESs 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 10kT (blue to red). (d) Structure-based Hamiltonian [59]; energies measured in kcal·mol⁻¹. The presence of the ligands was either modelled by explicitly including the inhibitor Ap5A [24,61] (a and b) or by incorporating closed-state native contacts into a CG Hamiltonian based on an interaction model primarily derived from the open state [59,62] (c and 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.]
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 Figure 4(a)). Careful analysis of a holo-MFEP showed that specific binding of the AMP-moiety of Ap$_5$A to the NMP domain induces NMP-closure, and that this step characterises the TS ensemble of the complete closure pathway.[24] The LID domain, on the other hand, only closes after the ATP-moiety of Ap$_5$A 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 are 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 nicotinamide adenine dinucleotide phosphate dihydrofolate 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, concluded that AdK should be described by an induced-fit model. Although this reasoning is in line with Sullivan and Holyoak’s [143] 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 \textit{E. coli} apo-AdK could be of the order of $10^{7}$ s$^{-1}$ at 20$^\circ$C, corresponding to typical times of only 52 ns.[41] NMR and FRET studies on apo-AdK from the thermophile \textit{Aquifex aeolicus} yielded slower rates of $k_{\text{open}} = 6500 \pm 500$ s$^{-1}$, and $k_{\text{close}} = 2000 \pm 200$ s$^{-1}$.[38] FRET experiments that specifically probed LID dynamics of the \textit{E. coli} apo enzyme arrived at $k_{\text{open}} = 120 \pm 40$ s$^{-1}$ and $k_{\text{close}} = 220 \pm 70$ s$^{-1}$.[39]

In the mesophilic \textit{E. coli} holo-enzyme the LID-opening rate was measured by NMR relaxation at the same temperature as above to be $k_{\text{open}} = 286 \pm 85$ s$^{-1}$ (3.5 ms) while the LID-closing rate was around $k_{\text{close}} = 1374 \pm 110$ s$^{-1}$ (0.73 ms); for the thermophile the rates were 44 $\pm$ 20 and 1571 $\pm$ 100 s$^{-1}$.[36] 1D FRET measurements of LID dynamics in the presence of the non-hydrolysable ligands AMP–PNP and AMP yielded $k_{\text{open}} = 160 \pm 40$ s$^{-1}$ (6.25 ms) and $k_{\text{close}} = 440 \pm 110$ 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$_{\text{eco}}$ 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 Equation (1).

6.1. Time scale estimate from TST

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 (Figure 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 of apo-AdK in less than 200 ns.[149]

Experimentally measured rates for the rate-limiting step of enzyme opening are of 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 in 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 of the order of 4–6 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 TST estimate for the transition rate constant,

\[ k = \nu \exp\left(-\frac{\Delta G^\dagger}{k_BT}\right), \tag{4} \]

with the frequency pre-factor \(\nu\) and the effective barrier height \(\Delta G^\dagger\). If we assume that \(\nu\) does not change very much between the apo and holo processes, \(\nu_{\text{apo}} \approx \nu_{\text{holo}}\), then the ratio

\[
\frac{k_{\text{apo}}}{k_{\text{holo}}} = \frac{\nu_{\text{apo}} \exp\left(-\frac{\Delta G^\dagger_{\text{apo}}}{k_BT}\right)}{\nu_{\text{holo}} \exp\left(-\frac{\Delta G^\dagger_{\text{holo}}}{k_BT}\right)}
\approx e^{-\left(\Delta G^\dagger_{\text{apo}} - \Delta G^\dagger_{\text{holo}}\right)/k_BT} \tag{5}
\]

between apo and holo transition rates only depends on the difference in barrier heights. If the apo barrier \(\Delta G^\text{apo}\) is lowered by a moderate 5\(k_BT\) (3 kcal \(\cdot\) mol\(^{-1}\)) compared with that 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 of the order of 20 \(\mu\)s, down from 3 ms. For an apo barrier 10\(k_BT\) 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 \(\nu\), 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 \(\nu\) is known, then kinetic rate predictions can be extracted given a known free energy landscape using Equation (4). Potoyan et al. [74] note that, given a rate description provided by Kramers’ theory (resulting in an expression like Equation (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 \(k_BT\) between the open and closed states. We thus expect the prefactor to be of 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, \(k_{\text{open}}\) and \(k_{\text{closed}}\), and the free energy barrier heights from their model \((F_{\text{open}}\) and \(F_{\text{closed}}\)) and applied the TST rate relations \(k_{\text{open}} = k_{0,\text{open}} \exp\left(-F_{\text{open}}/k_BT\right)\) and \(k_{\text{closed}} = k_{0,\text{closed}} \exp\left(-F_{\text{closed}}/k_BT\right)\). \(k_{\text{cat}}\) is taken to be the rate constant of the rate-limiting step. (For more details see Figure 3 of Lu and Wang [75] depicting the 1D barrier heights, \(k_{\text{open}}, k_{\text{closed}}\) and \(k_{\text{cat}}\) as functions of temperature for both the NMP-closing and LID-closing pathways.) Since the double-well model used by Lu and Wang [75] had two dynamical pathways and four free energy barriers (between the open, closed and two intermediate metastable states), \(k_{\text{open}}\) and \(k_{\text{closed}}\) were computed for both the LID-closing and NMP-closing pathways, \(F_{\text{open}}\) 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 \(F_{\text{closed}}\) 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 hypothesised 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 \(k_{\text{cat}}\) was predicted at different temperatures using the highest barrier height measured in their simulation model. By approximating \(k_{\text{cat}}\) as the larger of \(k_{\text{open}}\) and \(k_{\text{closed}}\), while using the TST transition rate estimate from Equation (4) and the experimental values for the opening and closing rates \((k_{\text{open}} = 286\) s\(^{-1}\) and \(k_{\text{closed}} = 1374\) s\(^{-1}\)), from Wolf-
Watz et al. [36], Lu and Wang [75] could determine \( v_{\text{open}} \) and \( v_{\text{closed}} \) for a range of temperatures using the room temperature values of \( v_{\text{open}} \) and \( v_{\text{closed}} \). These values were extrapolated to find that \( k_{\text{open}} \) was the rate-limiting step below 323K.

Wang et al. [79] employed MD simulations of a structure-based model parametrised to reproduce the experimental probability distributions of LID-closed and LID-open conformations.[39] Effects of ligands were explicitly taken into account, in contrast to 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 HOHC) 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 open, closed and two intermediate states for apo- and holo-AdK as well as ATP- and AMP-bound models. The opening and closing rate constants for both the apo and holo enzymes were found to be several times smaller than experimental values; 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 \( \rightarrow \) open transition, while transition numbers (for the transitions between adjacent basins) in the forward direction approximated numbers for the reverse transition, demonstrating path symmetry.

7. Comparison of path sampling methods

The question of 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 summarises the methods discussed in this review and qualitatively assesses their capabilities. The approaches were grouped by what we considered the main method of each study, 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 and – have the following meaning: (+) The method is well suited to provide the required information or has a high level of resolution or efficiency; data are 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). (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 summarised 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 PCA, independent component analysis [155–158] and full correlation analysis, [159] along with heuristic projections (e.g. angle–angle and domain centre of mass coordinates used for the AdK transition).
<table>
<thead>
<tr>
<th>Main method&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Sub-method(s)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Refs</th>
<th>Solvent interactions</th>
<th>Resolution&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Pathways and sampling&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Energetics&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Time scales&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Suitability</th>
<th>Speed&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eq MD</td>
<td></td>
<td>[43,69,73] [64,72]</td>
<td>Explicit AA</td>
<td>T − + +</td>
<td>0 + − −</td>
<td>+ − −</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIMS-MD</td>
<td></td>
<td>[31]</td>
<td>Implicit AA</td>
<td>T + + +</td>
<td>0 − − +</td>
<td>+ − −</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2D US-MD/WHAM</td>
<td></td>
<td>[71]</td>
<td>Explicit AA</td>
<td>T + + +</td>
<td>0 0 0 0</td>
<td>− − −</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DREM</td>
<td></td>
<td>1D &amp; 2D WHAM</td>
<td>Explicit AA</td>
<td>T + + +</td>
<td>0 0 0 0</td>
<td>− − −</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEE-REX</td>
<td></td>
<td>[23]</td>
<td>Explicit AA</td>
<td>T + + +</td>
<td>0 0 0 0</td>
<td>− − −</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDS-MD</td>
<td></td>
<td>[8]</td>
<td>Explicit UA</td>
<td>T + + +</td>
<td>0 0 0 0</td>
<td>− − −</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coMD</td>
<td>TMD, ANM, MC/Metrop</td>
<td>[117]</td>
<td>Explicit AA</td>
<td>T + + +</td>
<td>0 0 0 0</td>
<td>− − −</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WED</td>
<td>Double Gō</td>
<td>[12]</td>
<td>N/A AA</td>
<td>P − 0 −−/−0 − + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL-ENM</td>
<td>NMA-driven + strain-en calcs</td>
<td>[67]</td>
<td>N/A AA</td>
<td>P − 0 −−/−0 − + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNM</td>
<td>CPR</td>
<td>[47]</td>
<td>N/A Cα</td>
<td>P − 0 −−/−0 − + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DREM</td>
<td></td>
<td>[76]</td>
<td>N/A Cα</td>
<td>P − 0−/−0 − + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANM-MC</td>
<td>NMA-driven</td>
<td>[77]</td>
<td>N/A SA</td>
<td>P − 0 −−/−0 − + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG-ENI</td>
<td>CHARMM22 energy min</td>
<td>[48]</td>
<td>N/A AA</td>
<td>P − 0 −−/−0 − + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VAMM</td>
<td>NMA-driven</td>
<td>[15]</td>
<td>N/A Cα</td>
<td>P − 0−/−0 − + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geometric targeting</td>
<td></td>
<td>[14]</td>
<td>N/A AA</td>
<td>P + 0−/−0 − + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG-MD</td>
<td>Structure-based potential</td>
<td>[62]</td>
<td>N/A Cα</td>
<td>T + + + 0+/+ + 0 + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG-MD</td>
<td>Two-well pot</td>
<td>[75]</td>
<td>Implicit Cα</td>
<td>T + + + 0+/+ + 0 + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG-MD</td>
<td>Double Gō</td>
<td>[59]</td>
<td>N/A Cα</td>
<td>T + + + 0+/+ + 0 + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG-MD, NM-ENM</td>
<td>Structure-based, Tirion</td>
<td>[63]</td>
<td>N/A Cα</td>
<td>T +,− +, + 0+/−0 − /−0 − + + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eq MD</td>
<td></td>
<td>[58]</td>
<td>Explicit AA</td>
<td>T + + + 0 − + −</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2D US-MD/WHAM</td>
<td>Principal curve + 1D US-HRE/WHAM</td>
<td>[58]</td>
<td>Explicit AA</td>
<td>T + + + 0 + − +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEB</td>
<td>1D+ US-MD/WHAM</td>
<td>[61]</td>
<td>Implicit AA</td>
<td>P 0 0 + − + −</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural morphing</td>
<td>1D US-MD</td>
<td>[74]</td>
<td>Explicit AA</td>
<td>P 0+/− 0+/− + − 0</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-the-fly string</td>
<td>1D+/MBAR US-TMD</td>
<td>[24]</td>
<td>Explicit AA</td>
<td>P 0 0+/+ 0+/− − 0</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dMD</td>
<td>Importance sampling</td>
<td>[21]</td>
<td>N/A AA</td>
<td>P 0 0 − − + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENM</td>
<td>OM min</td>
<td>[16]</td>
<td>N/A Cα</td>
<td>P 0 0 − − + +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENM</td>
<td>NMA/geometry</td>
<td>[65]</td>
<td>N/A Cα</td>
<td>P 0+/− 0 + − − +</td>
<td>+ + +</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Suitability in various categories is qualitatively rated with: +, very suitable/fast; 0, adequate/normal; −, unsuitable or impossible/slow (see text for more details).
<sup>b</sup> Method name: Eq MD, equilibrium molecular dynamics; TEE-REX, temperature enhanced essential dynamics replica exchange; ED-MD, enhanced dynamics sampling MD; coMD, collective MD; WED, weighted ensemble dynamics; NL-ENM, nonlinear ENM; PNM, plastic network model; DREM, double-well network model; Mont 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
<sup>c</sup> Sub-method name: US-MD, umbrella sampling MD; WHAM, weighted histogram analysis method; TMD, targeted MD; ANM, anisotropic network model; MC, Monte Carlo; Metropolis algorithm; CPR, conjugate peak refinement; US-HRE, umbrella sampling Hamiltonian replica exchange MD; 1D −, 2D FES based on 1D reaction coordinate; MBAR, multistate Bennett acceptance ratio; OM, Onsager-Machlup action minimisation.
<sup>d</sup> Resolution is described as: AA, all-atom; UA, united-atom; SA, semiatomistic (residue-level CG); Cα, Cα-level CG.
<sup>e</sup> Simulation type: T, transition; E, ensemble; P, pathway.
As an example, in Figure 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, utilising either inherent stochasticity in the method or small parameter variations for fully deterministic approaches. [152] The clustered distance matrix (Figure 7) shows that the two repeats for each method are closest to each other, as indicated by the light \(2 \times 2\) 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 are 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 relatively more distant from all other pathways.

### 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 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 with ‘intermediate’ crystal structures [46] can also be used to validate intermediate structures sampled during transitions;[31,47,48] comparison with 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-crystallised 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 with pathways generated with supposedly higher quality (but computationally more demanding) approaches. For example, transition pathways should be close to the MFEPs 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 (Figure 1) and found reasonably good agreement (Figure 8(a)). Recently, Gur et al. [117] also compared transitions generated with their coMD method with 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 (Figure 8(c)) and thus these promising results are fully consistent with the independently generated FES. 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 TPS 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 (Figure 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\textsubscript{eco}) has become a useful test system to study macromolecular transitions because it is a small protein that undergoes a large and relatively easily characterisable conformational change. An abundance of experimental data is 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 modelled 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
[27] Bae E, Phillips GN Jr. Identifying and engineering ion pairs in adenylate kinases. Insights from molecular dynamics simulations


[168] Zhang HJ, Sheng XR, Pan XM, Zhou M. Activation of adenylate kinase by denaturants is due to the increasing conformational flexibility at its active sites. Biochem Biophys Res Commun. 1997;238:382–386.