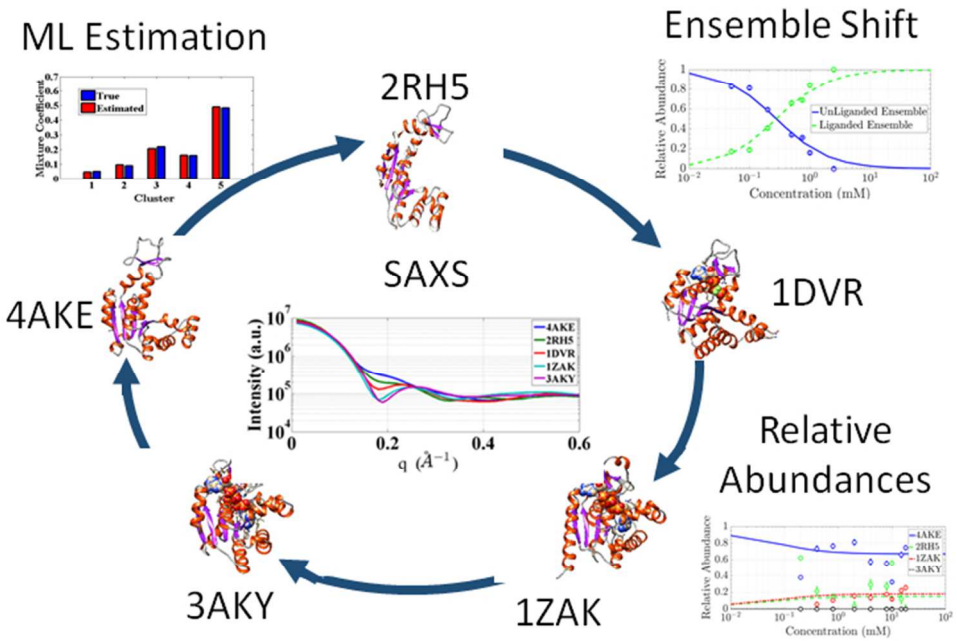
Article

Effects of catalytic action and ligand binding on conformational ensembles of adenylate kinase

Emre Onuk, John Badger, Yujing Wang, Jaydeep Bardhan, Yasmin Chishti, Murat Akcakaya, Dana Brooks, Deniz Erdogmus, David D. L. Minh, and Lee Makowski

Biochemistry, **Just Accepted Manuscript** • DOI: 10.1021/acs.biochem.7b00351 • Publication Date (Web): 02 Aug 2017



TOC Graphic

267x180mm (72 x 72 DPI)

Effects of catalytic action and ligand binding on conformational ensembles of adenylate kinase

Authors: Emre Onuk[1], John Badger[2], Yu Jing Wang[3], Jaydeep Bardhan[4], Yasmin Chishti[3], Murat Akcakaya[5], Dana H. Brooks[6], Deniz Erdogmus[6], David D. L. Minh[7] and Lee Makowski[3,8*]

1. Radiation Oncology Department, University of California, Los Angeles, CA 90095.
2. DeltaG Technologies, San Diego, CA 92122.
3. Department of Bioengineering, Northeastern University, Boston, MA 02115.
4. Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA 02115.
5. Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15261.
6. Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115.
7. Department of Chemistry, Illinois Institute of Technology, Chicago, IL 60616.
8. Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115.

Abbreviations

AdK, Adenylate kinase; AMP, Adenosine monophosphate; AMP-PNP, Adenylyl-imidodiphosphate; ADP, Adenosine diphosphate; ATP, Adenosine triphosphate; Ap5A, P^1,P^5-di(adenosine-5'-)pentaphosphate; SAXS, small angle x-ray scattering.

Keywords

induced fit, conformational transition, ligand binding, SAXS, Adneylate kinase, structural ensembles

Work site: Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115.

Corresponding Author Email: aonuk@mednet.ucla.edu

Short Title: Adenylate Kinase Conformational Ensembles

Abstract

Crystal structures of AdK from *E. coli* capture two states: an 'open' conformation (apo) obtained in the absence of ligands and a 'closed' conformation in which ligands are bound. Other AdK crystal structures suggest intermediate conformations that may lie on the transition pathway between these two states. To characterize the transition from open to closed states in solution, x-ray solution scattering data were collected from AdK in apo form and with progressively increasing concentrations of five different ligands. Scattering data from apo AdK are consistent with scattering predicted from the crystal structure of AdK in the open conformation. In contrast, data from AdK samples saturated with Ap5A do not agree with that calculated from AdK in the closed conformation. Using cluster analysis of available structures, we selected representative structures in five conformational states: open, partially open, intermediate, partially closed, and closed. We used these structures to estimate the relative abundances of these states for each experimental condition. X-ray solution scattering data obtained from AdK with AMP are dominated by scattering from AdK in the open conformation. For AdK in the presence of high concentrations of ATP and ADP the conformational ensemble shifts to a mixture of partially open and closed states. Even when AdK is saturated with Ap5A, a significant proportion of AdK remains in a partially open conformation. These results are consistent with an induced-fit model wherein the transition of AdK from open to a closed state is initiated by ATP binding.

Introduction

Adenylate kinase (AdK) is responsible for catalyzing the reaction ATP + AMP ↔ 2 ADP, which typically occurs with an equilibrium constant close to unity. This is an essential reaction in living cells, and AdK has been identified in yeast, bacteria, and many different organisms throughout the animal kingdom. The AdK structure is comprised of three domains, CORE, LID, and NMP. ATP binding occurs in a pocket involving the LID and CORE domains, whereas AMP binding occurs

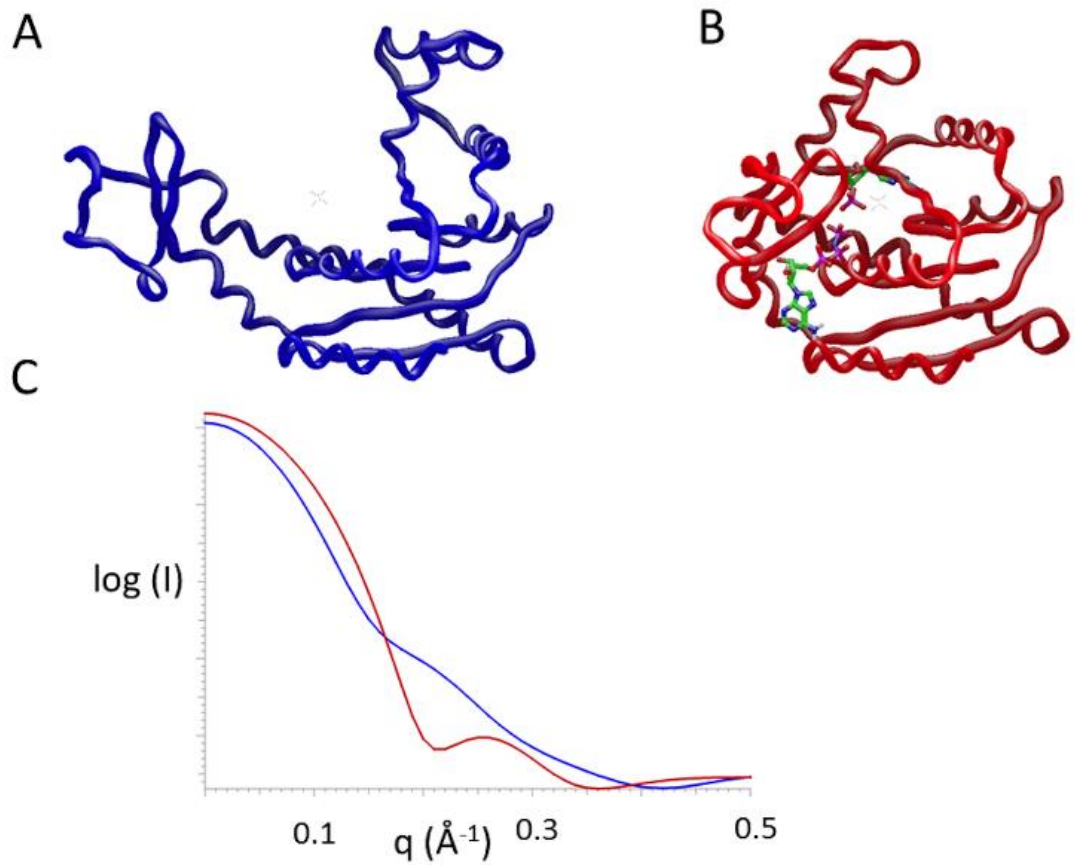


Figure 1. Main chain traces for *E. coli* AdK showing (A) an apo structure in the open conformation (PDB ID: 4AKE) and (B) a structure containing AMP and AMP-PNP in the closed conformation (PDB ID: 1ANK). (C) X-ray scattering patterns calculated using CRYSOL for the open (blue curve) and closed (red curve) structures.

in a pocket involving the NMP and CORE domains. The available crystal structures of AdK from *E. coli* capture two distinct states: an 'open' conformation (apo) obtained in the absence of ligands and a 'closed' conformation in which ligands are bound (Figure 1A and 1B). Key crystal structures of AdK include an apo form (PDB ID: 4AKE)[1], a structure containing AMP and AMP-PNP in the closed conformation (PDB ID: 1ANK)[2], a structure containing ADP and AMP in the closed conformation (PDB ID: 2ECK)[3], and a structure in the closed conformation that contains a tightly binding ligand, Ap5A, that mimics the transition state of AMP with ATP (PDB ID: 1AKE)[4]. The crystal structures of AdK in the closed conformation obtained with these different bound ligands are very similar to each other and NMR spectra obtained from protein in solution saturated with either ADP or Ap5A are very closely related[5]. Most experimental data are consistent with the view that the ligand-free protein is predominantly in the open conformation, but a contrary scenario, where the equilibrium favors a closed state, has also been proposed[6].

The characterization of stable structural states that lie on the pathway between the open and closed conformations has proven elusive, but has been approached by the determination of a large number of crystal structures. These structures have been obtained for ADK from a number of species and in the presence of specifically designed ligands. Beckstein et al.[7] extended our picture of the observed *E. coli* AdK structures by building homology models of the *E. coli* protein based on AdK structures from other species. Sorting the resulting 45 structural models (referred to here as the 'Beckstein set') using a measure of the progress from closed to open structures, they identified several conformations intermediate between open and closed. Operationally, the configuration of LID, CORE, and NMP domains in these 45 structures are distinguished by the distances between the centers of mass of the three domains, and by the variation in these distances (Figure 2). For the available crystal structures, the LID-CORE distance varies only when the NMP is open,

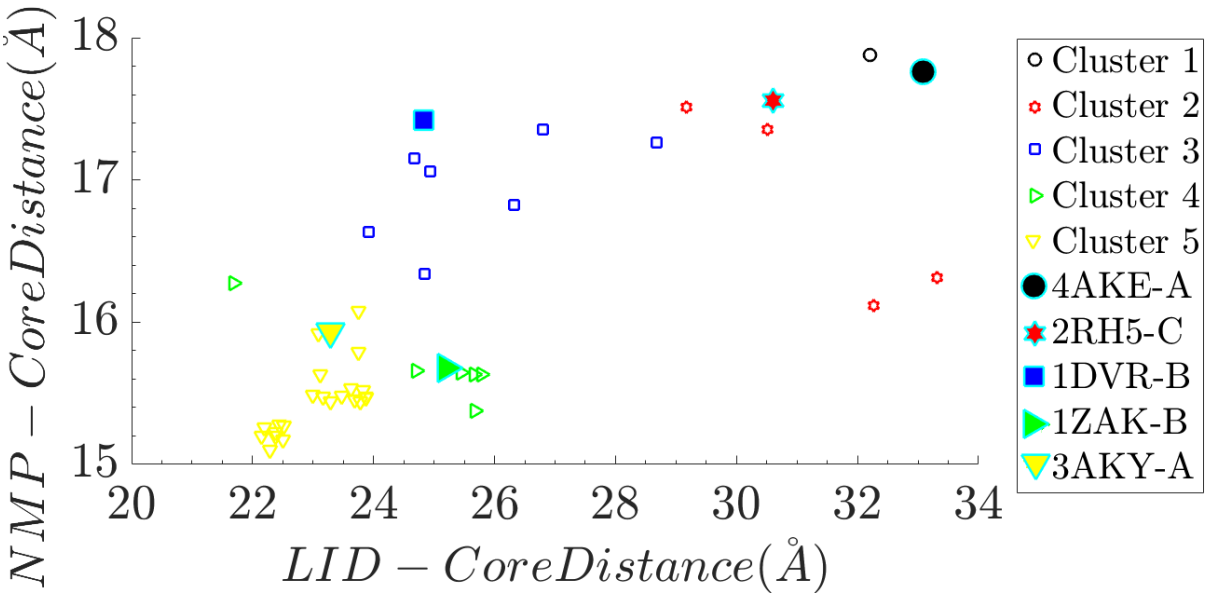


Figure 2. Diagram of the NMP-CORE and LID-CORE distances (calculated from their centers of mass for the 45 *E. coli* models in the Beckstein set. Membership in the five clusters identified by k-means clustering is indicated by color codes. The representative member of each of these clusters is indicated by a filled circle.

whereas the NMP-CORE distance appears more variable when the LID is closed. Conformations with the LID open and NMP closed are rarely observed. These observations indicate that the NMP and LID domains do not move independently of one another.

Although detailed structural information has been provided by these extensive crystallographic analyses and modeling, the nature and order of events for the transition between open and closed conformational endpoints remains unclear. Contrasting conceptual pictures of AdK action include both induced-fit models and population-shift mechanisms. Specifically, an induced-fit transition from open to closed AdK conformations might initially be triggered by the binding of either an AMP or ATP molecule and would be followed by binding of the partner molecule. Alternatively, if the free energy difference between open and closed conformations is sufficiently small, with only minimal energy barriers between them, both open and closed conformations could be

accessed by AdK in the absence of ligands. In this model, preferential ligand binding to a pre-existing closed state would stabilize the closed conformation and shift the population of AdK molecules towards that state. In either scenario, enzymatic activity would follow once ATP and AMP are bound and the protein takes on the closed conformation. A variety of molecular dynamics and modeling techniques have been used to investigate atomic-level aspects of this conformational transition. Most simulation methods report that in the absence of ligands an open conformation is favored over closed conformation with moderate or non-existent energy barriers between the two states but estimates for the free energy difference between these states vary[8-12].

To better understand the structural transitions involved in the enzymatic activity of AdK, we report here x-ray solution scattering data from *E.coli* AdK in the apo form and with ligands that include AMP, ADP, ATP, a non-functional ADP analog ADP-βS, and the tightly binding inhibitor Ap5A. Protein conformational differences of the size observed for open and closed conformations of AdK (Figure 1A and 1B) lead to distinct x-ray scattering patterns (Figure 1C). X-ray solution scattering has traditionally been used for study of samples in which a single protein conformation is predominant but recent studies have demonstrated that it is also possible to deconvolve the relative abundances of multiple conformations from scattering data [13,14]. X-ray scattering data were collected as titration series, in which protein samples contained ligand concentrations that ranged from lower than the protein concentration to large ligand excesses. Estimation of the relative proportions of open, intermediate, and closed conformations during titration provides insight into the distribution of conformations in both apo and ligand-containing ensembles, and the pathway between the open and closed conformations.

Materials and Methods

Materials

The DNA construct used to express wild type *E. coli* adenylate kinase was supplied Dr. Magnus Wolf-Watz (Umeå University, Sweden), and expressed in *E. coli* as previously described [15]. The concentrations of all proteins were determined using an extinction coefficient at 280 nm of 10.000 M^{-1} cm^{-1}. All ligands (AMP, ADP, ATP, ADP-βS, Ap5A) were purchased from Sigma-Aldrich. Ligand concentrations were confirmed by absorbance at 260 nm using an extinction coefficient of 15.0 cm^2 $pmol^{-1}$ for adenine at pH 7.0 [16].The protein buffer contained Mg ions at concentration corresponding to a ~12-fold excess over the protein.

X-ray solution scattering data collection

X-ray solution scattering data were collected at the BioCAT undulator beam line (18ID) at the Advanced Photon Source, Argonne National Laboratory as previously described[17]. Radiation damage was minimized by the use of a flow cell that limited time of exposure of any protein to x-rays to less than 100 ms, leaving the effect of radiation damage on radiosensitive test proteins undetectable [18]. Outlier patterns, generated due to passage of small bubbles through the x-ray beam in the flow cell during exposure, were removed. Two-dimensional scattering patterns were circularly averaged using Fit2D version 14.101[19], applying a polarization correction with polarization factor of 0.99, and binning measured intensities at each pixel into 1000 equally spaced scan bins with a maximum 2θ angle of 28°. Scattering patterns were normalized by the incident photon count.

The sample cell consisted of a thin-walled quartz capillary (1.5 mm in diameter) attached to a programmable pump (Hamilton; Microlab 500 series) that was adjusted to deliver continuous flow through the capillary during data collection. The ambient temperature of the air surrounding the capillary and sample tubing was held at 4° C by attachment of an ethylene glycol bath to the brass capillary holder and through an outer layer of tubing surrounding the inner sample tubing. X-ray scattering patterns were recorded with a MAR165 2k × 2k CCD detector. The specimen-to-detector distance was approximately 170 mm and was calibrated using powder diffraction rings from silver-behenate. The beamline is capable of delivering approximately 2×10^{13} photons/sec/ 100 mA of beam current [18]. As previous experience on the BioCAT beamline has demonstrated that proteins under a variety of physical conditions are damaged after exposure times of a few tenths of a second to a few seconds at these intensity levels, in these experiments 8 to 32 20μm aluminum foils were used (depending on the concentration of protein in the samples) as X-ray beam attenuators to control the incident beam flux. In most cases, the exposures of solution with buffer and ligand were followed by at least seven exposures of solution containing buffer, ligand, and protein, with each exposure lasting about two seconds. Incident beam flux was monitored using nitrogen gas-filled ion chambers placed directly before the sample chamber and an active beam stop to monitor x-ray flux after the sample. No apparent differences in normalization were observed between the two measures of intensity, indicating that changes in ligand concentration made no observable change in absorption. We arbitrarily chose to use the ion chamber measurements for normalization, but could equally well have used data from the active beam stop which would have resulted in indistinguishable measures of scattered intensity.

Solution Scattering Data Sets

Data were collected from the apo protein at a series of concentrations between 5 and 40 mg/ml in order to investigate concentration-dependent conformational changes (Figure 3A). Subsequent X-ray scattering data were collected as titration series, in which protein concentration was fixed at 10mg/ml and ligand concentrations that ranged from lower than the protein concentration to several times the concentration of the protein. X-ray scattering data sets included AdK with AMP (7 samples, 0.02-10 mM ligand), ADP (8 samples, 0.05 – 7.5mM ligand), ATP (9 samples, 0.05 – 10.0mM ligand), Ap5aA (9 samples, 0.1 – 2.0 mM) and ADP-βS (7 samples, 0.05 - 2.5mM ligand). X-ray scattering curves were visualized and manipulated using software from the ATSAS program package [20].

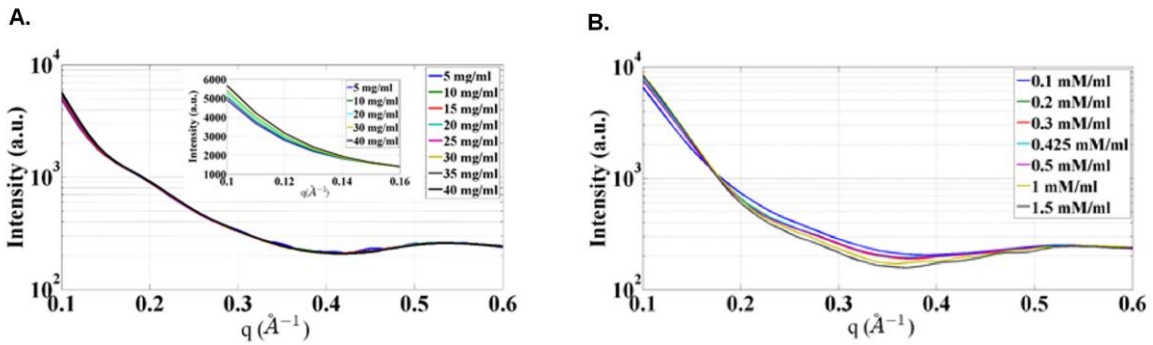


Figure 3. Solution scattering data obtained from *E. coli* AdK (A) in the absence from ligand at concentrations varying from 5 – 40 mg/ml. (B) AdK in the presence of varying concentrations of the tight binding inhibitor Ap5A.

Calculation of scattering patterns

Atomic models used for scattering calculations included several crystal structures of *E. coli* AdK from the Protein Data Bank (4AKE, 1ANK, 2ECK, 1AKE) and the 'Beckstein set': 45 homology

models of *E. coli* AdK in a range of conformations based on AdK structures in the Protein Data Bank obtained from a variety of different species [7].

Calculations of scattering patterns from atomic models were performed using CRYSOL [21] with default parameters and with XS [22]. XS calculates excess scattering intensity by performing a molecular dynamics (MD) simulations on a box of explicit water molecules around the fixed protein and on a box of pure water. The MD calculations were performed using the VMD [23] and NAMD [24] at 277 K using the TIP3P water model [25] and the CHARM22 force field [26]. Following a 20 ps equilibration period followed by a 100 ps MD simulation, 100 snapshots were taken from the simulation runs to obtain scattering curves from an ensemble average.

Determination of optimum basis sets of conformations

Because the information content of solution scattering patterns does not allow us to estimate of the relative abundances of all 45 different conformations, it is necessary to cluster similar structures together (Figure 4) and represent each cluster's contribution to scattering by that from the structure

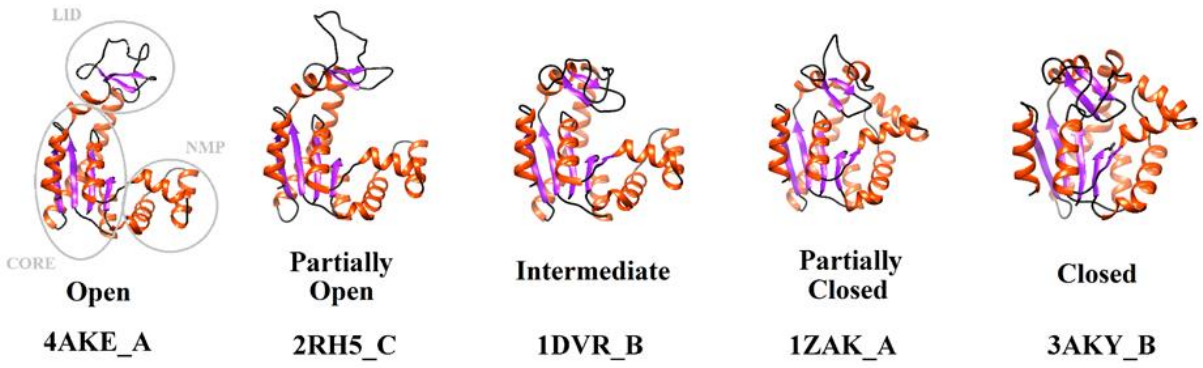


Figure 4. Rendering of the five representative AdK conformations selected from the five clusters identified by k-means clustering.

most representative of the cluster (Figure 5). The 45 scattering patterns were clustered using a subset selection method based on k-means clustering [33] and the Cramer-Rao bound (CRB) [13,27,34]. SAXS patterns were predicted using XS[22] for the 45 structural models of *E. coli* AdK in the Beckstein set[7].

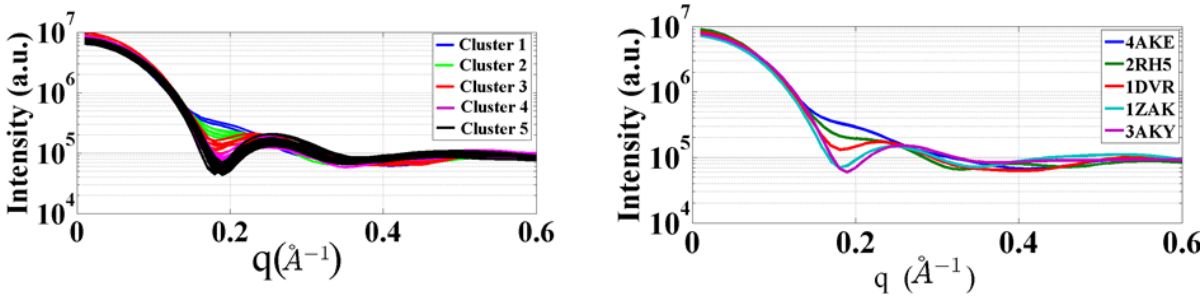


Figure 5. X-ray solution scattering patterns predicted from (left) 45 structurs of *E. coli* AdK from the Beckstein set, and, (right) the five representative structures obtained by clustering. Scattering calculations were performed with XS (45).

The choice of 5 clusters[13] was based on a model order criterion in which was developed a score function based on the total clustering error and used with the Bayesian Information Criterion (BIC) for model order selection. This approach requires computing the criterion over a range of numbers of clusters and then evaluating the BIC as a function of model order (number of clusters). The resulting score function as a function of number of clusters is almost unchanging for all numbers of clusters greater than 5. This observation led to the conclusion that there seem to be 5 groupings of conformations that are distinguishable in the scattering data from AdK. Nevertheless, calculations were carried out with 3-7 clusters and different choices of basis vectors. These calculations confirmed that addition of a 6th basis vector (cluster) did not significantly lower the discrepancy between calculated and observed data when averaged over all diffraction patterns.

Determination of structure populations from solution scattering data

Abundances of conformations were determined using the maximum likelihood estimation (MLE) method described previously [13]. Briefly, measured SAXS intensity data, I_m, is assumed to be a linear combination of known conformation intensities, $\boldsymbol{I}_C = \left[I_C^1, \ldots, I_C^{N_c}\right]^T$ (I_C is the matrix of all representative intensities for every cluster selected after CRB subset selection method), with the relative abundance of each conformation, α_T, and a multiplicative log-normally distributed noise, w. Then we can write the measured intensity data as

$$I_m = \boldsymbol{I}_c^T \alpha_T w$$

Multiplicative noise is assumed because analysis of the SAXS data indicated that the standard deviation of the intensity varies linearly with the mean of intensity[13]. We use MLE to find relative abundances using the previously determined basis SAXS intensity set which gives the minimum mean square intensity error. MLE can be formulated as:

$$\underset{\boldsymbol{\alpha_T}}{\operatorname{argmin}} \frac{1}{2} (\ln I_m - \ln(\boldsymbol{I}_C^T \alpha^T))^T (\ln I_m - \ln(\boldsymbol{I}_C^T \alpha^T))$$

$$subject\ to \quad \alpha_T^T \mathbf{1} = 1\ and\ \mathbf{0} \preccurlyeq \alpha_T \preccurlyeq \mathbf{1}$$

Results

X-ray solution scattering characterization of apo and ligand-bound conformational states

The conformational state of apo AdK was evaluated from a series of eight solution scattering data sets in which the protein concentrations varied from 5 – 40 mg/ml. Variations in the scattered intensities between members of this series are of low significance; for example, for data in the small angle scattering regime ($q < 0.3$ Å^{-1}) there is no pair of data sets where the difference between scattering curves exceeds $\chi^2 = 2.0$. Experimental data over the resolution range $0.064 < q < 0.5$ Å^{-1} were compared to calculated scattering curves obtained with CRYSOL[21] using atomic models

of AdK representing both the open (PDB ID: 4AKE) and closed (PDB ID: 1AKE) conformation using the program OLIGOMER[28]. These comparisons showed that the experimental scattering curves from the ligand-free protein can be almost entirely accounted for by the calculated scattering from crystallographic model of the open crystal structure, with a fit to data of $\chi^2 = 2.0$ (Figure 6A). There was no apo AdK data set in this concentration series for which the fit was

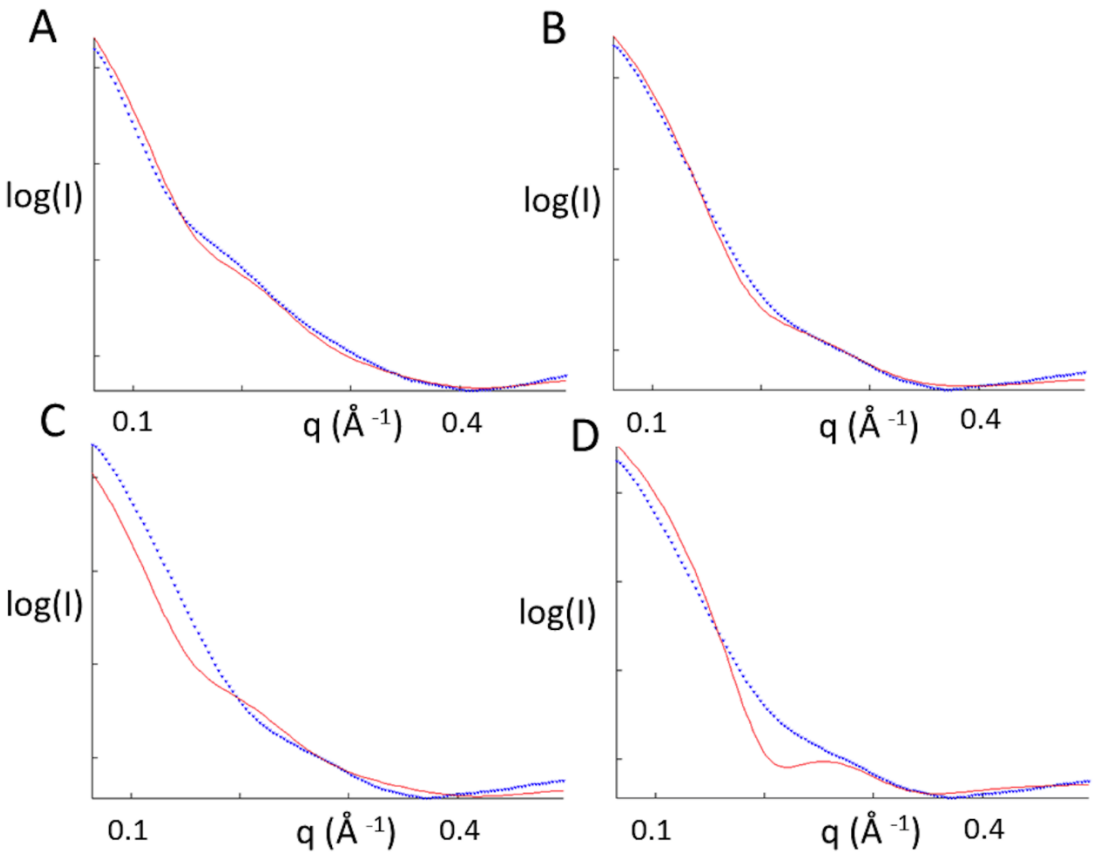


Figure 6. Experimental x-ray solution scattering intensity curves (blue dots) and the best fits of calculated patterns from combinations of open (PDB ID: 4AKE) and closed structures (PDB ID: 1ANK) as calculated by OLIGOMER. (A) Data from apo AdK at protein concentration of 10mg/ml is best fit by an entirely open scattering curve ($\chi^2 = 2.0$). (B) Data from AdK with ligand Ap5A at slightly over-saturated ligand concentration (0.5mM) is best fit with at 0.42:0.58 ratio of open and closed conformations ($\chi^2 = 2.4$). The fits of the experimental data for this Ap5A-containing sample to scattering patterns calculated from (C) the purely open conformation ($\chi^2 = 5.3$) and, (D) the purely closed conformation ($\chi^2 = 5.3$) are much poorer than the combined fit.

significantly improved by including a calculated scattering contribution from AdK in the fully closed conformation. As we will see (below), consideration of intermediate states suggests that up to 5% of the unliganded proteins are in a partially closed or partially open conformation. We conclude from these calculations that the crystallographic structure of the open conformation of AdK is a reliable representation of the dominant conformation in the solution state.

To characterize the closed conformational state in terms of experimental x-ray solution scattering data, we measured scattering from AdK in the presence of the strongly binding ligand Ap5A (K_d ~ 2.5 nM [29]) at concentrations ranging from 1/4 the protein concentration to a several-fold excess, *i.e.* extending into a concentration regime where the protein is heavily saturated by ligand. Comparing these data to the scattering curves obtained from the apo AdK samples, it is evident that the presence of Ap5A makes a significant impact to the x-ray scattering curve across the entire range of concentrations studied. As expected, calculated scattering from AdK in the open conformation gives a poor fit (χ^2 = 5.3) to the experimental data (Fig 6D). Surprisingly, the calculated scattering from *E. coli* AdK in the ligand-bound, closed conformation gives an equally poor fit to the experimental data (χ^2 = 5.3, Fig 6C). The scattering data may be more adequately (but still imperfectly) matched (χ^2 = 2.4) by scattering contributions that include both the open and closed conformations of the protein obtained from relevant crystal structures (Figure 6B, C, D).

Mapping of the ligand-induced shifts in conformational ensembles

As a first step towards determining the conformational ensemble at each experimental condition, we formulated the variation in scattering as a transition from the 'unliganded ensemble', comprising the relative abundances of the set of conformations in the absence of ligand, and 'liganded ensembles'

which comprise the relative abundances of the set of conformations under saturation (assuming that all proteins have bound ligand). Each titration series was fit by MLE using a basis set consisting of the apo AdK scattering pattern and the scattering pattern for the saturated ligand ensemble specific to each series. The proportion of proteins in the liganded ensemble and the unliganded ensemble is estimated by expressing each scattering pattern as a linear combination of scattering in the absence of ligand and scattering at saturation of each ligand. The transition from unliganded to liganded ensembles exhibits sigmoidal behavior indistinguishable from a simple 1:1 ligand binding (Figure 7).

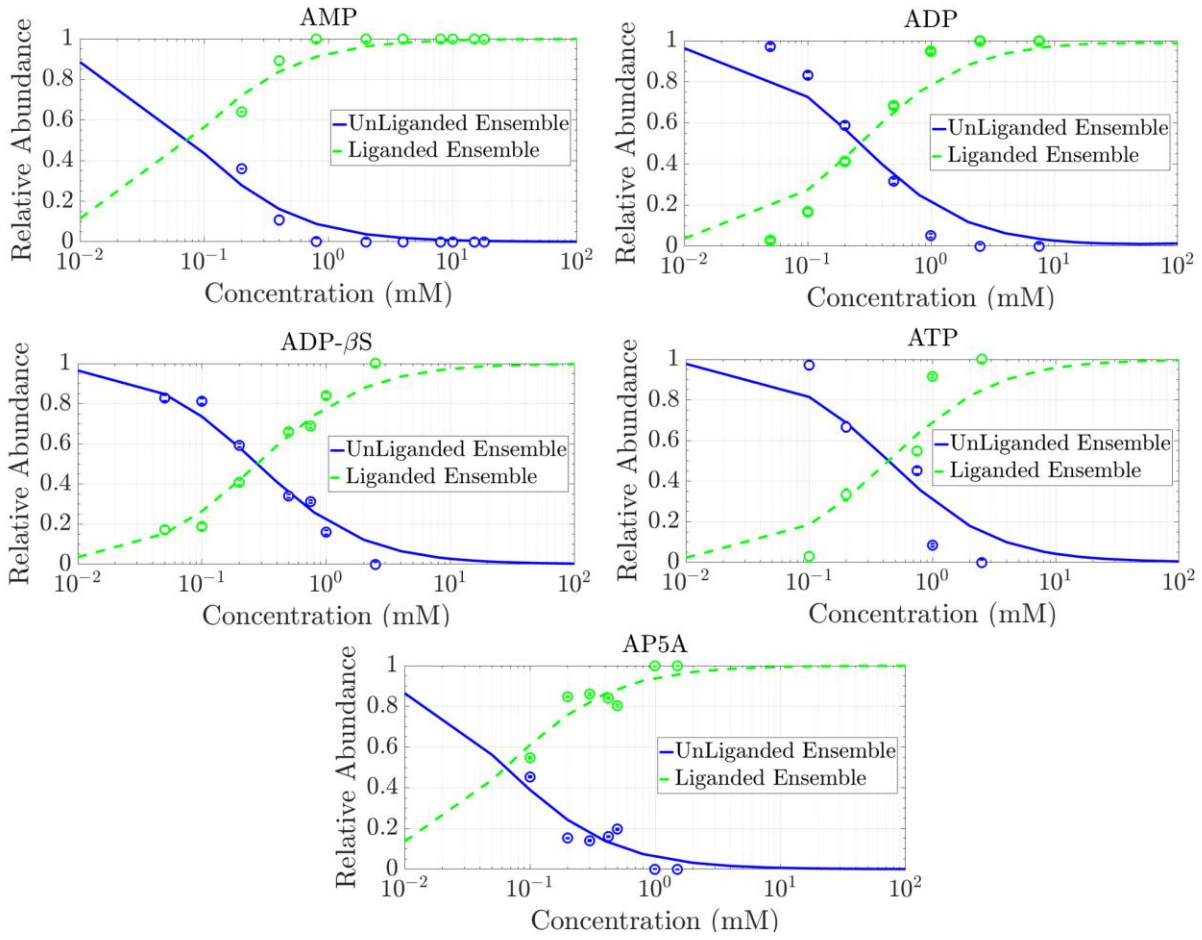


Figure 7. Ensemble shifts as a function of ligand concentration in terms of 'unliganded' and 'liganded' ensembles. The population shifts are shown as a function of ligand concentration for (A) AMP, (B) ADP, (C) ADP-βS, (D) ATP, and (E) Ap5A.

The 1:1 binding model is not entirely appropriate for AMP, ADP, or ADP-βS: AMP can bind to both sites, although with much higher affinity to the site formed by the CORE and NMP domains[15]. ADP-βS can also bind to AdK with a 2:1 molar ration. ADP will bind to both sites and then interconvert with AMP and ATP so all three species are present in the binding sites. A two site model fits these data sets slightly better than a 1:1 binding model, but the improvement in fit is not statistically significant.

Assuming a 1:1 model is an adequate representation of binding, these SAXS data make it possible, in some cases, to make a rough estimate of K_d for specific ligand binding, from the MLE-estimated fraction of the protein that adopts the liganded ensemble. The free ligand concentration is obtained from the experimental ligand concentration after accounting for the number of ligands removed by binding to the protein. Fitting for K_d with a sigmoidal function to the proportions estimated from the scattering data gives a determination of the equilibrium binding constant in the same way as for other biophysical methods[30]. For the simple case of single binding, this analysis predicts K_ds of ~0.1 and ~0.4mM for AMP and ATP respectively. In practice, the very high binding constant for Ap5A relative to the experimentally accessible regime (mM ligand concentrations compared to expected binding affinity in the nanomolar range) precludes an accurate determination of K_d.

Optimal basis set of structural states

In order to interpret the x-ray solution scattering data in terms of a more comprehensive set of structural states than the open and closed AdK crystal structures, we employed the 45 structures of *E. coli* AdK derived from crystallography and homology modeling by Beckstein et al[7]. The analysis of crystallography-based models resulted in definition of five clusters and the structure most representative of each cluster. These clusters are designated here as 'open', 'partially open',

'intermediate', 'partially closed', and 'closed' (Figure 4). The relationships between the structural domains (LID, CORE, and NMP) in the five representative conformations can be visualized by plots of the NMP-CORE distance versus the LID-CORE distance (Figure 2). The partially open conformation differs from the fully open one largely in the position of the LID, which has moved towards the closed position in this conformation. Similarly, the partially closed conformation differs from the fully closed one largely in the position of the LID. The intermediate conformation has a closed LID and open NMP, but under the experimental conditions we have considered, there is little evidence from scattering for the existence of this conformation. The predicted scattering patterns from each of these representative structures (Figure 5) provide a basis set for defining the conformational ensembles of AdK in terms of the relative abundances of these five structures.

Relative populations of basis set structures

The relative abundances of the representative structures under each experimental condition were determined using the maximum likelihood estimation (MLE) approach as used for the liganded and unliganded ensembles. To estimate the populations of each conformation, each scattering pattern was represented as a linear combination of scattering from the representative structures with the scaling factor (relative abundance of each conformation) chosen to optimize the correspondence of calculated to observed scattering under a noise model that accurately reflected the experimental errors[13].

MLE analysis of the unliganded ensemble indicated that roughly 95% of the protein is in the 'open' conformation, 2.5% in the partially open state, and 2.5% in the partially closed conformation. Fully closed conformations, if present, were below the detection limit for these experiments.

The best fits of the five-member basis set of structures to the scattering data for AdK with various concentrations of the different ligands is shown in Figure 8. These plots show contributions from four

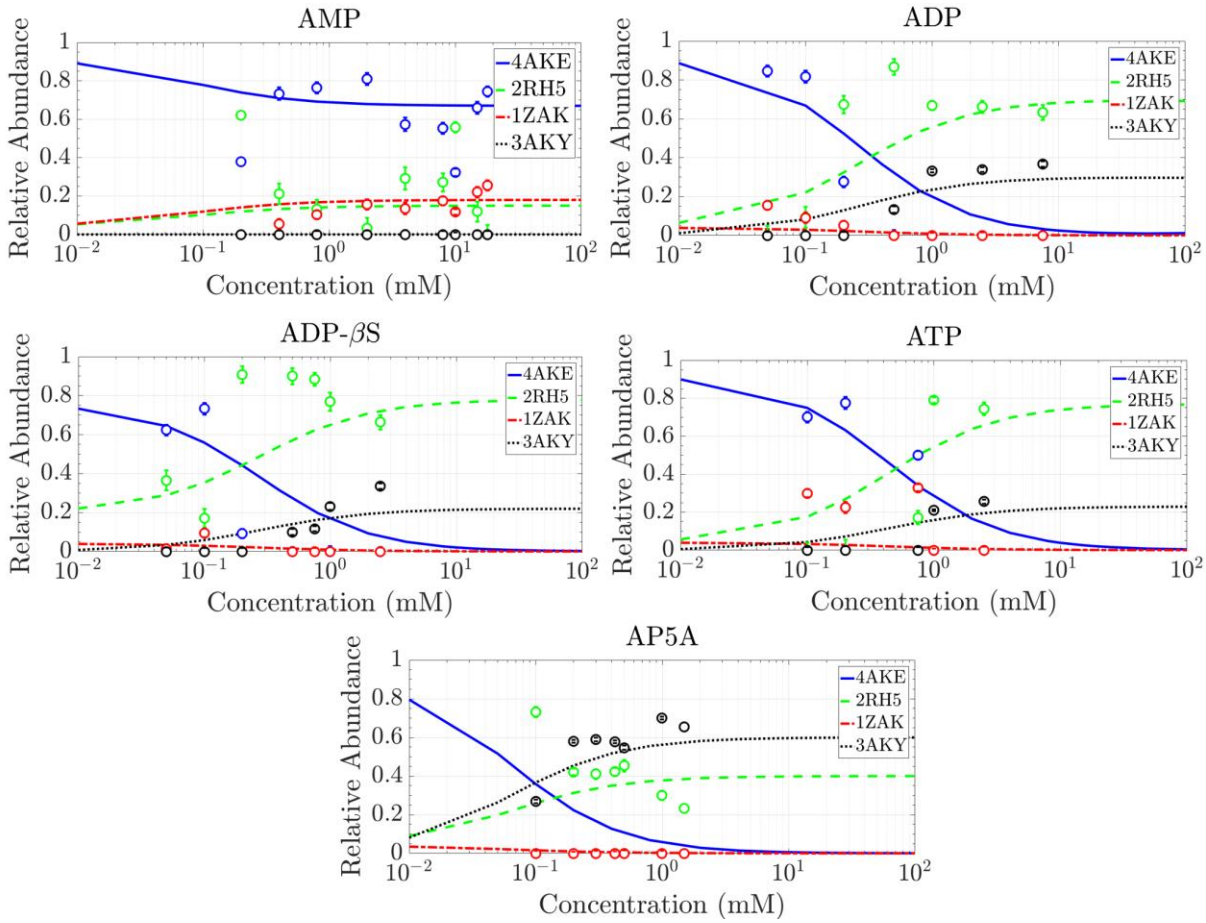


Figure 8. Relative abundances of conformations calculated with MLE demonstrating the predicted conformational shifts for representative conformational states, as a function of ligand concentration, for (A) AMP, (B) ADP, (C) ADP-βS, (D) ATP, and (E) Ap5A.

structures because no contribution from the structure representing the intermediate form (PDB ID 1DVR) was detected under any experimental condition. For comparison, smooth sigmoidal functions representing the population of each representative structure were calculated from linear combinations of unliganded and liganded ensembles (where the liganded ensemble is different for each ligand) and the K_d values determined from the analysis diagrammed in Figure 7. In other words, the curves in Figure 8 are not fits to the data presented in Figure 8. Rather they have been calculated directly from

the abundances of constituents in the unliganded and liganded ensembles (endpoints) and the dissociation constants inferred through the fitting exhibited in Figure 7. That the apparent discrepancies in Figure 8 are relatively large is a reflection of the uncertainties intrinsic to separation of the abundances of 5 components using data from individual scattering curves. Nevertheless, the trends and relative abundances remain largely self-consistent throughout each titration series.

Analysis of data for AdK titrated with Ap5A show a rapid decline of the open conformation and a corresponding rise in partially open and fully closed conformations in a ~40:60 ratio as Ap5A concentration increases. However, even at heavily saturating concentrations of Ap5A, only about 60% of the molecules are in a completely closed conformation with as many as 40% of NMP domains remaining open. The existence of many crystal structures of fully closed conformation suggests that either the proportion of fully closed structures increases with protein concentration (as may occur during crystallization) or that crystallization preferentially captures AdK in a stable, closed conformation. This implies that, in solution, AdK in complex with Ap5A apparently exhibits dynamic transitions between closed and partially open conformations that would not be accommodated within the confines of a crystal lattice.

Our analysis shows that the presence of AMP has relatively little impact on the AdK conformational equilibrium. We estimate that, even when heavily saturated, ~80% of AdK remains in the fully open conformation and the remaining AdK population is evenly divided between partially open and partially closed forms. AMP binds between the CORE and NMP domains but not only does AMP binding have little impact on the LID conformation but the vast majority of NMP domains also remain open in the presence of this ligand.

ATP more strongly impacts the equilibrium between AdK conformational states. Our results are consistent with the elimination of the open conformation as ATP concentration increases (Figures 7 and 8). When heavily saturated by ligand, about 80% of AdK appears to be in the partially open conformation, indicating the impact of ATP interaction with the LID domain. In this conformation, the NMP domain remains open, poised for binding of incoming AMP molecules. The remaining 20% of the population shifts to the fully closed state, with both LID and NMP closed. It is possible that this conformation precludes subsequent binding of AMP.

X-ray scattering data collected from AdK with increasing concentrations of both ADP and the ADP-βS also demonstrate a pronounced shift in the population distribution of conformational forms. In both cases the population of the open conformation appears to decline sharply with increasing ligand concentration, and is replaced by the partially open and closed conformational forms. For ADP, the population of partially open conformations plateaus at slightly more than twice the closed conformation, whereas for the ADP-βS data the partially open conformation comprises ~90% of the sample. As with the ATP sample (but in contrast to the AMP sample), there is little evidence for a significant population of partially closed conformational states.

Impact of choosing different numbers of clusters on relative abundances

Comparable calculations were carried out using 3-7 clusters. When using less than 4 clusters, the discrepancy between calculated and observed was always greater than with 5 clusters; whereas increasing the number of clusters beyond 5 did little to lower the discrepancy independent of the choice of added structure (vector). The specifics of the calculation differed depending on how appropriate the choice of representative structure was. For instance, removing 1DVR from the set of 5 representative structures and re-doing the calculation did little to increase the discrepancy between

calculated and observed because our results have shown that structures similar to 1DVR are poorly represented in the solutions studied. However, if any other structure was removed, the fit to the data was much worse. When additional structures were added to our prototypical 5, the results changed in predictable ways. If 1AKE was added (a closed conformation similar to 3AKY), the total abundance of closed structures did not change, but was distributed between 1AKE and 3AKY. However, changing the relative abundances of the two open structures did little to change the predicted scattering, supporting the assertion that the data sets cannot distinguish between these two similar conformations. Comparable results were obtained if the added structure was closely related to the open 4AKE conformation.

Discussion

Population distribution in conformational ensemble of unliganded and liganded ADK

Interpretation of the SAXS data reported here requires a distinction to be made between the conventional labels of 'bound' and 'unbound' from 'closed' and 'open'. It is clear from our results that 'bound' does not always equal 'closed' and that 'unbound' does not equal an ensemble with all proteins 'open'. NMR experiments have indicated the presence of 'closed' conformations of AdK in the absence of ligands [31]. An important capability enabled by our SAXS analysis is the means to calculate the proportions of 'open', 'partially open', 'partially closed' and 'closed' conformations, which is difficult to accomplish by other techniques.

For AdK in the absence of ligand, the solution scattering analysis is consistent with the standard view that the population of conformational states is heavily dominated by an open conformation that is well approximated by the open structure identified crystallographically. The observed data is relatively consistent with the predictions of this model alone (Figure 1A), and the MLE analysis predicts that

~95% of the conformational ensemble in the open conformation, with the remainder of the population split between the partially open and partially closed conformations.

Our data on AdK with Ap5A are consistent with the expectation that the binding constant for this compound is orders of magnitude higher than the mM concentrations of protein and compound used in these experiments. Beyond saturating concentrations of ligand, the fraction of protein identified in non-open conformations remains nearly constant (Figure 8). The impact of Ap5A on the AdK conformational equilibrium differs somewhat from impact of the biologically relevant ligands AMP, ADP, and ATP. In those cases, a partially open state dominates the set of non-closed states at higher compound concentrations. But with Ap5A there is a 40:60 split between partially open and fully closed conformations (Figure 8). Ap5A contains five bridging phosphate groups and binds at two adenine binding sites whereas AMP and ATP each only bind strongly at a specific site and contain a total of four phosphate groups. These differences in ligand structure may lead to a different energetic balance between conformational states.

Conformational transition between open and closed states

A noteworthy observation is the complete absence of evidence for the presence of intermediate conformations with the LID nearly closed and the NMP domain open. Although crystal structures of AdK in this configuration exist (e.g. 1DVR_B) we have found no evidence for the presence of these conformations in solution. 1DVR[32] is the structure of an inactive mutant of the yeast AdK bound to the ATP analogue AMPPCF$_2$P. Although Schlauderer et al.[32] suggest that the observation of a lid closed (NMP open) state demonstrates that the domain motions occur largely independent from each other [32], our observations indicate that for WT *E. coli* AdK, NMP closing occurs primarily when the LID is closed.

The x-ray scattering curves obtained for samples containing AMP are similar to the curves obtained from the apo AdK samples (Figure 3), and even at the highest AMP concentrations studied, the open conformation is dominant. That is, the conformational analysis suggests that only minor populations exist of the partially open and partially closed conformations, and that there is no significant population of the closed conformation. In contrast, the presence of ATP resulted in a change in x-ray scattering curves that could be modeled by the introduction of a significant population of AdK molecules in the partially open and closed conformations, with a corresponding decrease in the number of AdK in the open conformation. These observations are consistent with a process in which closure of AdK results from binding of ATP rather than AMP. This contrasting impact of ATP and AMP on the equilibrium between open and closed conformations is consistent with the observation of a unidirectional energetic coupling between ATP and AMP binding sites from NMR data [5]. As observed by NMR, ATP binding induced chemical shift perturbations in the relatively distant AMP binding site that did not appear related to non-specific binding of ATP at the AMP site. In the NMR experiments, the presence of ATP was considered to result in a closed state AdK population of 23%, which is consistent with the maximum value obtained in the ATP titration described here.

When ADP is mixed with AdK, the enzyme undergoes continuous catalytic interconversion of ADP to AMP and ATP (and reverse) for several hours. Our experiments were carried out under conditions where this cycling was taking place as confirmed by functional assay[5,15]. At low ADP concentrations the proportion of open conformations drops as ADP concentration increases, replaced largely by partially open conformations with some partially closed conformations (Figures 7 and 8). These conformations correspond to the singly-bound ensemble of AdK in the presence of ADP. As ADP concentration continues to increase, the partially open conformations are replaced to some extent by fully closed conformations as the second ADP molecule binds, triggering closure of the NMP domain.

At the highest concentrations of ADP studied, about 40% of the ADK are in a fully closed conformation and 60% in the partially open conformation (*i.e*. LID partially open, NMP closed). These conditions correspond to maximum enzymatic activity[15] of ADP, and suggest that during catalytic cycling the protein spends slightly less than half the time in a closed conformation, slightly more than half of the time in a partially open conformation, and very little (if any) time in a fully open conformation. One implication of this conclusion is that the AMP site releases product in the partially open conformation and that the product in the ATP site is only released during very transient opening of the lid[5], suggesting that ATP release may represent a rate limiting step[15].

Overall, these results appear more consistent with proposals[5] that AdK belongs to a group of enzymes in which substrate binding is the predominant mechanism for driving the protein into a catalytically active state, rather than the population-shift model.

Acknowledgement

We thank Prof. Magnus Wolf-Watz for a DNA construct used to produce adenylate kinase. This work was supported by NSF (MSB-1158340) and NIH (R01-GM85648). A paper package that contains all code and data can be found at: http://hdl.handle.net/2047/d20193588.

References

1. Muller C.W., Schlauderer G.J., Reinstein J., Shultz G.E. (1996) Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding, Structure 4, 147-156.

2. Berry M.B., Meador B., Bilderback T., Liang P., Glaser M., Phillips G.N. (1994) The closed conformation of a highly flexible protein: The structure of *E. coli* adenylate kinase with bound AMP and AMPPNP, Proteins 19, 183-198.

3. Berry, M.B., Bae E., Bilderback T.R., Glaser M., Phillips G.N. (2006) Crystal structure of ADP/AMP complex of *Escherichia coli* adenylate kinase, Proteins 62, 555-556.

4. Muller C.W., Shultz G.E. (1992) Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 A resolution. A model for a catalytic transition state, J. Mol. Biol. 224,159-177.

5. Ådén J., Wolf-Watz M. (2007)NMR Identification of Transient Complexes Critical to Adenylate Kinase Catalysis, J. Am. Chem. Soc. 129,14003-14012.

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

7. Beckstein O., Denning E.J., Perilla J.R., Woolf T.B. (2009) Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open<-->closed transitions, J. Mol. Biol. 394,160-176.

8. Kubitzki M.B., de Groot B.L. (2008) The Atomistic Mechanism of Conformational Transition in Adenylate Kinase: A TEE-REX Molecular Dynamics Study, Structure 16,1175-1182.

9. Song H.D., Zhu F. (2013) Conformational Dynamics of a Ligand-Free Adenylate Kinase, PLoS ONE 8, e68023.

10. Whitford P.C., Miyashita O., Levy Y., Onuchic J.N. (2007) Conformational transitions of adenylate kinase: Switching by cracking. J. Mol. Biol. 366,1661-1671.

11. Arora K., Brooks C.L. III. (2007) Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism, Proc. Natl. Acad. Sci. 104,18496-18501.

12. Potoyan D.A., Zhuravlev P.I., Papoian G.A. (2012) Computing free energy of a large-scale allosteric transition in adenylate kinase using all atom explicit solvent simulations, J. Chem. Phys. B 116,1709-1715.

13. Onuk A.E., Akcakaya M., Bardhan J.P., Erdogmus D., Brooks D.H., Makowski L. (2015) Constrained maximum likelihood estimation of relative abundances of protein conformation in a heterogeneous mixture from small angle x-ray intensity measurements, Signal Processing IEEE Transactions 63, 5383-5394.

14. Yang S., Park S., Makowski L., Roux B. (2009) A rapid coarse residue-based computational method for x-ray solution scattering and multiple conformational states of large protein complexes, Biophys. J. 96, 4449-4463.

15. Wolf-Watz W., Thai V., Henzler-Wildman K., Hadjipavlou G., Eisenmesser E.Z., Kern D. (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair, Nature Structural & Molecular Biology 11, 945–949.

16. Holzbaur, E.L.F., Johnson, K.A. (2004) The rate of ATP synthesis by dynein, Biochemistry 25, 428–434.

17. Makowski L., Rodi D.J., Mandava S., Minh D.D.L., Gore D.B., Fischetti R.F. (2008) Molecular crowding inhibits intramolecular breathing motions in proteins, J. Mol. Biol. 375, 529–546.

18. Fischetti, R., Stepanov, S., Rosenbaum, G., Barrea, R., Black, E., Gore, D., Heurich, R., Kondrashkina, E., Kropf, A.J., Wang, S. and Zhang, K. (2004) The BioCAT undulator beamline 18ID: a facility for biological non-crystalline diffraction and x-ray absorption spectroscopy at the Advanced Photon Source, J. Synch. Rad. 11, 399–405.

19. Hammersley, A. (2004) FIT2D V12.012 Reference Manual V6.0.

20. Konarev P.V., Petoukhov, M.V., Volkov V.V., Svergun D.I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis, J. Appl. Cryst. 39, 277−286.

21. Svergun D., Barberato C., Koch M.H.J. (1995) CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, J. Appl. Cryst. 28, 768–773.

22. Park S., Bardhan J.P., Roux B., Makowski L. (2009) Simulated X-ray scattering of protein solutions using explicit-solvent molecular dynamics, J. Chem. Phys. 130,134114.

23. Humphrey W., Dalke A., Schulten K. (1996) VMD: visual molecular dynamics. J. Mol. Graphics 14, 33–38.

24. Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kale L., Schulten K. (2005) Scalable molecular dynamics with NAMD,J. Comp. Chem. 26, 1781–1802.

25. Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W., Klein M.L. (1983) Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79, 926–935.

26. MacKerell A.D., Bashford D., Bellott M., Dunbrack R.L., Evanseck J.D., Field M.J., Fischer S., Gao J., Guo H., Ha S., Joseph-McCarthy D., Kuchnir L., Kuczera K., Lau F.T.K., Mattos C., Michnick S., Ngo T., Nguyen D.T., Prodhom B., Reiher, R.,Roux B., Schlenkrich M., Smith J.C., Stote R., Straub J., Watanabe M., Wiórkiewicz-Kuczera J., Yin D., Karplus M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem. B 102, 3586–3616.

27. Stoica P., Selen Y. (2004) Model-order selection: a review of information criterion rules, Signal Processing Magazine, IEEE 21, 36–47.

28. Konarev P.V., Volkov V.V., Sokolova, A.V., Koch M.H.J., Svergun D.I. (2003) PRIMUS – a Windows-PC based system for small-angle scattering data analysis. J. Appl. Cryst. 36, 1277-1282.

29. Lienhard G.E., Secemski I.I. (1973) P^1, P^5–Di(adensosine-5')pentaphosphate, a Potent Multisubstrate Inhibitor of Adenylate Kinase, J. Biol. Chem. 248, 1121-1123.

30. Pollard T.D. (2010) A Guide to Simple and Informative Binding Assays, Molecular Biology of the Cell 21, 4061-4067.

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

32. Schlauderer, G.J., Proba K., Schulz, G.E. (1996) Structure of a mutant adenylate kinase ligated with an ATP-analogue showing domain closure over ATP, J. Mol. Biol. 256, 223-227.

33. Bishop C. (2007) Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. corr. 2nd printing edn. Springer, New York.

34. Kay S.M. (1993) Statistical signal processing. Estimation Theory 1.