

Allosteric Conformational Transition in Adenylate Kinase: Dynamic Correlations and Implication for Allostery

Ming S. Liu,^{A,B,C} Billy D. Todd,^B and Richard J. Sadus^B

^ACSIRO – Mathematics, Informatics and Statistics, Private Bag 33, Clayton South, Vic. 3169, Australia.

^BCentre for Molecular Simulation, Swinburne University of Technology, P.O. Box 218, Hawthorn, Vic. 3122, Australia.

^CCorresponding author. Email: ming.liu@csiro.au

An essential aspect of protein science is to determine the deductive relationship between structure, dynamics, and various sets of functions. The role of dynamics is currently challenging our understanding of protein functions, both experimentally and theoretically. To verify the internal fluctuations and dynamics correlations in an enzyme protein undergoing conformational transitions, we have applied a coarse-grained dynamics algorithm using the elastic network model for adenylate kinase. Normal mode analysis reveals possible dynamical and allosteric pathways for the transition between the open and the closed states of adenylate kinase. As the ligands binding induces significant flexibility changes of the nucleotides monophosphate (NMP) domain and adenosine triphosphate (ATP) domain, the diagonalized correlation between different structural transition states shows that most correlated motions occur between the NMP domain and the helices surrounding the ATP domain. The simultaneous existence of positive and negative correlations indicates that the conformational changes of adenylate kinase take place in an allosteric manner. Analyses of the cumulated normal mode overlap coefficients and long-range correlated motion provide new insights of operating mechanisms and dynamics of adenylate kinase. They also suggest a quantitative dynamics criterion for determining the allosteric cooperativity, which may be applicable to other proteins.

Manuscript received: 27 August 2009.

Manuscript accepted: 10 February 2010.

Introduction

Conformational dynamics plays essential roles in regulating protein functions.^[1,2] How protein dynamics, arising from protein–protein interactions, protein docking with DNA/RNA, and/or ligand binding with proteins, affect the functions remains largely unknown. For a wide range of protein functions, collective dynamics and large-scale conformational changes have a critical role in cases such as: (a) conformational transition between different states, e.g. induced by ligand binding or inter-protein docking; (b) allostery and cooperativity of multi-site activities in a monomer or of inter-subunit interactions in a multimeric protein (where multiple dynamic equilibrium adjusts regulation mechanisms, and vice versa). An understanding of dynamics collectivity, transitions, and correlations between different domains or conformational states is vital for an understanding of protein functions,^[3] because internal fluctuations and correlated dynamics of proteins intrinsically regulate their biological activities. More generally, it is also important for predicting signalling networks and enzyme activity sites in allostery-based drug designs.^[4]

A typical system showing a defined dynamics-function relationship is adenylate kinase (also known as AdK or myokinase), a phospho-transferase enzyme that catalyzes the inter conversion of nucleotides monophosphate (NMP) and plays an important role in cellular energy homeostasis.^[5–7] AdK undergoes large conformational changes as it catalyzes the reverse reaction

of $\text{Mg}^{2+}\cdot\text{ATP} + \text{AMP} \leftrightarrow \text{Mg}^{2+}\cdot\text{ADP} + \text{ADP}$. In terms of conformational dynamics, AdK is constructed as a three-domain protein (Fig. 1, Table 1), namely the CORE domain, the LID domain (for binding adenosine triphosphate (ATP)/adenosine diphosphate (ADP)/adenosine monophosphate (AMP) and its analogues) and the NMP domain (for binding ADP/AMP and its analogues). Large motions of the LID and NMP domains are associated with nucleotide binding against the CORE domain.^[5–7] Currently in the Protein Data Bank (PDB), there are ~45 free or ligated AdK conformers from bacteria and other organisms. Based on various sets of determined conformations, AdK can transit between an unliganded ‘open’ conformation state and a distinctive bi-substrate-inhibited (e.g. by ATP and AMP) ‘closed’ conformation state (Fig. 1). For the conformational transition of AdK, the LID and NMP domain move largely (Fig. 1 and Table 1) with the CORE domain being relatively rigid. It was also found that fluctuations of the ligand binding domains are rate-limiting for AdK catalysis.^[8] Both the ‘open’ and ‘closed’ conformations of AdK may exist in an equilibrium population, with conformational population shifts.^[1,9,10] Despite extensive experimental^[1,5–7,10,11] and theoretical^[9,12–18] studies on AdK, its catalytic mechanisms, dynamic pathways for structural transition (Fig. 2) and conformational cooperativity remain unclear at the molecular level.

AdK serves as an excellent model system for relating the conformational dynamics and structural transitions to biological

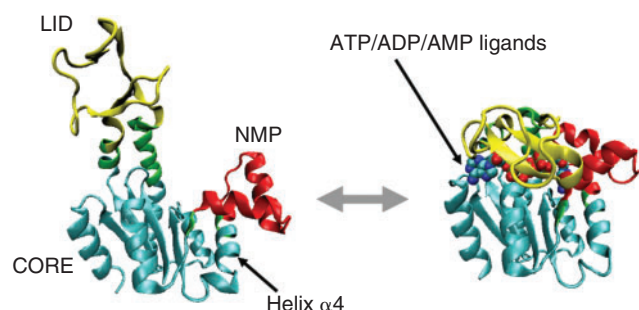


Fig. 1. Structures of AdK and its conformational transitions between the open state (left, e.g. PDB:4AKE) to the closed state (right, e.g. PDB:1AKE). The CORE domain is coloured in cyan, the NMP domain red, and the LID domain yellow. Highly rotational domain motions of LID and NMP results in dynamics hinges at the intersections between LID/NMP and CORE. Green colour indicates the bending residues responsible for the inter-domain rotation (as determined by the DynDym algorithm;^[45–47] also see Table 1).

functions, particularly in determining the long-range dynamic correlations and the role of dynamics in regulating allosteric transition and cooperativity. To determine protein dynamics, experimental methods such as NMR spectroscopy^[19,20] have both size and timescale limitations. Moreover, protein conformational changes usually occur on timescales (e.g. $\mu\text{s}\sim\text{ms}$) currently inaccessible to explicit molecular dynamics simulation. Therefore, new computational approaches are required to overcome the size and timescale difficulties associated with probing protein dynamics. One effective approach is to coarse grain the atomistic details of proteins with an elastic network model to speed up the dynamics simulation.^[21–24] Normal mode analysis (NMA) using elastic network models^[25–28] have been widely accepted to describe the flexibility and collective large-amplitude motions of proteins transiting among different conformations. Through NMA, collective dynamics of the functional domains or motifs can be probed at a timescale and spatial resolution at which either molecular dynamics experiments or explicit simulations encounter limitations.

We have demonstrated that, in the cases of signal protein NtrC,^[29] the suppressor of cytokine signalling (SOCS) proteins^[30] and multi-subunit motor proteins,^[31] coarse-grained dynamics is a robust tool to probe dynamics and its regulated functions. In this work, we extended the coarse-grained dynamics to analyse the intrinsic fluctuation and dynamics correlations of AdK and compared the results with those of experiments^[1,5–7,10,11] and molecular dynamics simulations.^[9,14,16–18,32,33] We obtained a quantitative description of residue fluctuations, correlated motions and identified

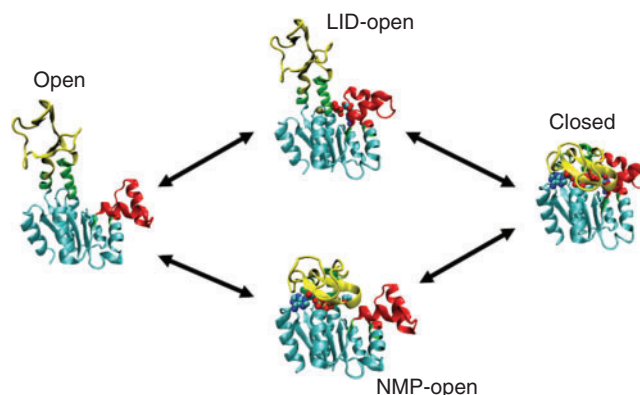


Fig. 2. The competitive order of binding nucleotides at LID or NMP domains will lead to different pathways for a conformational transition in AdK: from the open (with both LID and NMP domains open), via NMP-open (with LID domain closed and NMP domain open) or alternatively via LID-open (the LID domain open with the NMP domain closed), through to the closed (with both domains closed); and vice versa.

the dynamics-oriented pathways for conformational transitions in AdK. Dynamics plays a pivotal role in tracking down the competitive order and allosteric manner of ligands binding in AdK (Fig. 2). The dynamics analyses also lead to new insights on how native fluctuations and long-range correlated motion induce the allosteric transition and cooperativity in different sub-domains of AdK. A general dynamics criterion to elucidate allostery in protein's structural transitions and dynamic changes is also proposed.

Results

Flexibility and Intrinsic Fluctuations of AdK

The residue temperature factor is a measure of fluctuations and flexibility of the protein backbone. Theoretical temperature factors of the open, closed and intermediate transient states of AdK showed that the LID and NMP regions are intrinsically more flexible (Accessory Publication Fig. S2). Major dynamics fluctuations during the conformational transitions differ in the NMP lid and the ATP lid. Except for the termini, significant changes of fluctuations upon ligands binding occur throughout the $\beta\text{L1-L2-L3-L4}$ loop and helix $\alpha 7$ the LID region and helices of $\alpha 2$, $\alpha 3$, and $\alpha 4$ of the NMP region. The overall flexibility of NMP and LID domains decreases significantly upon binding ligands (Accessory Publication Fig. S2). These are the dynamic consequences of structural transitions between the open and the closed conformers. Our dynamics data generally agree

Table 1. Dynamic domain analysis of adenylate kinase (as determined by the DynDym algorithm^[45–47])

The colour scheme of the fixed domain (cyan), moving domain (red or yellow), and bending residues (green) refers to the colours in Fig. 1. 'Domain pairs dynamics' indicates the relative domain motions of the moving domain against the fixed domain. NMP, nucleotides monophosphate; RMSD, root-mean-square deviation

Dynamic domains	Dynamics nature	Size (residues)	Backbone RMSD [Å]	Residues	Domain pairs dynamics	Bending residues (green)
CORE	Fixed (cyan)	135	1.71	1–29 66–116 160–214	–	–
NMP	Moving (red)	36	1.58	30–65	Rotation angle [°] 45.9 ± 0.3 translation [Å] 1.4 closure [%] 99.9	29–30 65–66
LID	Moving (yellow)	43	0.59	117–159	Rotation angle [°] 51.8 translation [Å] 0.8 closure [%] 97.5	116–119 157–172

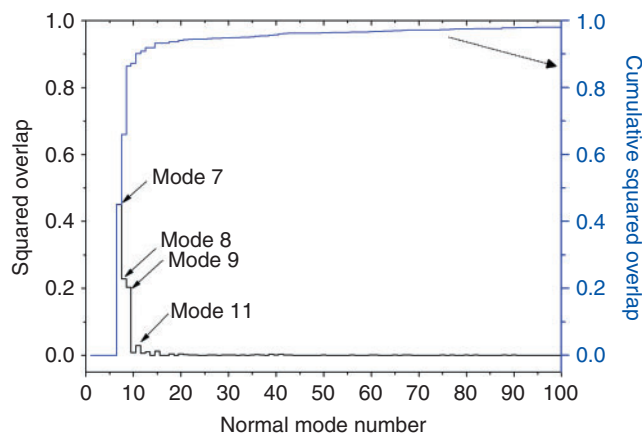


Fig. 3. Squared overlap coefficient for normal modes motions of AdK. Four lowest frequency modes (modes 7, 8, 9, and 11, in a trajectory from the open state) contribute more than 90% conformational changes in the open \rightarrow closed transitions of AdK. Inset (right axis) is the cumulative squared overlap from normal mode 7 up to mode 100.

very well with the backbone dynamics derived from X-ray crystallography^[5-7] and NMR relaxation measurements.^[1,10] This ensures that the coarse-grained dynamics analysis of AdK is accurate and reliable in terms of both elasticity and collective motions. From the coarse-grained dynamics, no difference in fluctuations or correlations was detected between *Escherichia coli* AdK (i.e. PDB:4AKE and PDB:1AKE) and *Aquifex* AdK (i.e. PDB:2RGX and PDB:2RH5). This is consistent with NMR studies^[1] that they basically have the same dynamics properties. Consequently, we present our results and discussion mainly in the format of *E. coli* AdK throughout this paper (key results on *Aquifex* AdK are provided in the Accessory Publication).

Normal Mode Analysis of Conformational Transitions of AdK

Superimposition of AdK structures (Fig. 1 and Accessory Publication Fig. S1) gave a pair-wise root-mean-square deviation (RMSD) of 7.92 Å between the open and closed conformations. The normal modes of up to 200 modes were determined for every modelled structure, and the overlap coefficients of each normal mode for possible structural transitions were also derived. Fig. 3 shows the squared overlap of the open state and the cumulative squared overlap coefficients of different normal modes in the open \rightarrow closed transition. Among these normal modes, the lowest frequency normal mode 7 (modes 1 to 6 are global motions and do not contribute to any conformational transitions) is the most dominant mode of motion, which contributes \sim 70% of overall conformational change. In AdK, modes 7, 8, 9, and 11 have the highest coupled overlaps with the conformational transition of open \rightarrow closed. Of course, no single normal mode (e.g. mode 7) can represent the observed conformational changes, an accurate representation of which requires the combination of the key normal modes, e.g. modes 7, 8, 9, and 11.^[34,35]

Considering the possible pathways for the conformational transition in AdK, normal modes 7, 8, 9, and 11 (Fig. 3) have the largest mode-overlap coefficients and provide the four most probable schemes for the open \leftrightarrow closed transition: Their conformational changes cover \sim 95% of the overall structural transition between the open and closed states. For an understanding of the flexibility and dynamical changes in these possible pathways, Fig. 4 depicts the residues RMSD in these dominant modes

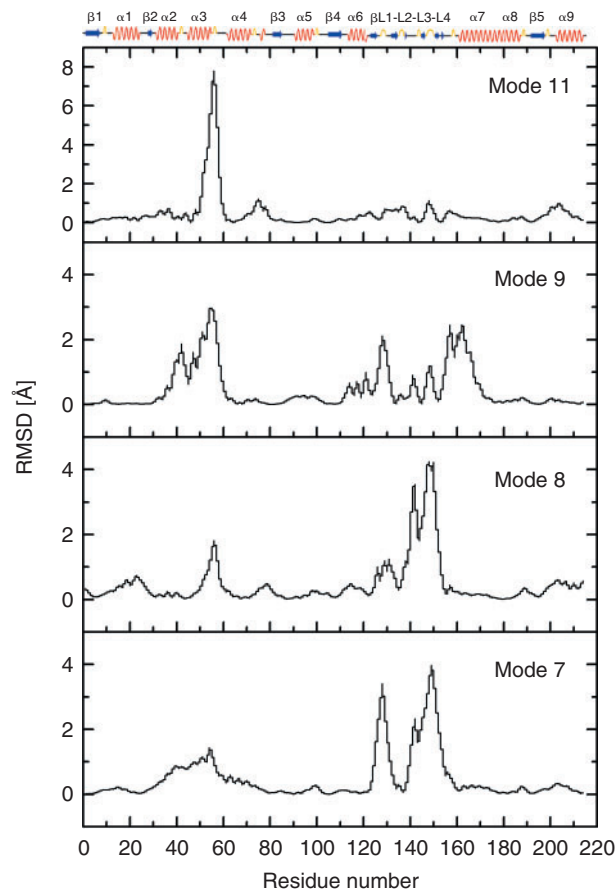


Fig. 4. Residues RMSD of the four normal modes (modes 7, 8, 9, and 11) dominantly responsible for the conformational transition of open \leftrightarrow closed of AdK.

of motion. The RMSD clearly indicate that the binding pocket region does not move very much, rather the LID domain (the β L1-L4 loops) and NMP domain (the α 2- α 4 helices) conduct most dynamics associated with the conformational transition.

In addition, the cumulative squared overlap of normal modes (up to 100) of AdK accounts for the validation and effectiveness of conformational transitions in the open \leftrightarrow closed transition (Fig. 5, top, either via LID-open state or NMP-open state) or the closed \rightarrow open transition (Fig. 5, bottom, either via LID-open state or NMP-open state). Fig. 5 indicates that for the open \rightarrow closed transition, with more agreed overlaps between open \rightarrow closed and open \rightarrow NMP-open transitions, AdK has higher probability transiting through the NMP-open state. In contrast, for the closed \rightarrow open transition, there is no superior intermediate domain motion, and both LID-open and NMP-open have same probability for an AdK transition.

Dynamic Correlations in AdK

From normal mode overlaps, we have identified different possible pathways during this transition (Fig. 5). To further investigate dynamically important residues or segments of AdK responsible for structural transitions, ensembles of conformation were sampled and the dynamics of every residue was determined and interpreted by residue fluctuation and correlation (Eqn 1, Fig. 6, and Accessory Publication Fig. S3). The dynamic correlations in AdK are presented in Fig. 6a by the colour-coded contour diagrams of difference correlation matrix of motions upon

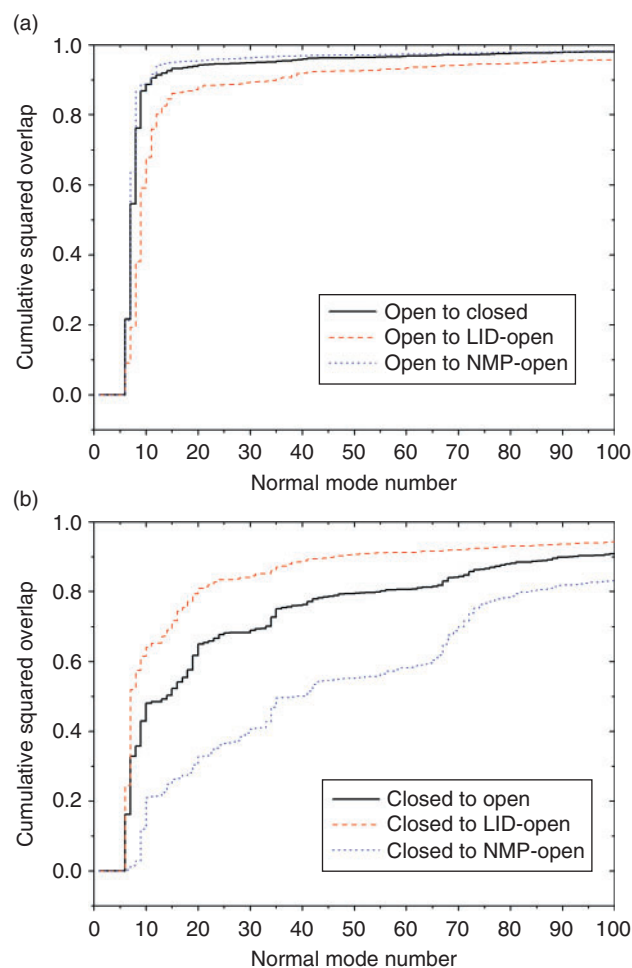


Fig. 5. Cumulative squared overlap of normal modes (up to 100) of AdK indicates different overlap and probability in pathways of the open \leftrightarrow closed conformational transition: (top) for open \rightarrow closed transition via the LID-open or NMP-open state; (bottom) for closed \rightarrow open transition via the LID-open or NMP-open state.

phosphorylation (with the red colours indicating anti-correlated motion against the blue colours). The open and closed states of AdK inherit great differences in flexibility and fluctuation properties from their intrinsic network.

Significantly, major dynamical changes in AdK, in terms of residues fluctuation and correlation upon phosphorylation (see Fig. 6a), are observed in the LID domain (especially the β L1-L4 loop) and NMP domain (the α 2- α 4 helices). The remarkable changes of the correlation coefficient C_{ij} matrix indicate the largest correlated motion (red regions in Fig. 6a, where residues move in the same scale of magnitude and the same direction) of helices α 2- α 3 and helix α 4 occurs with an anti-correlated motion of the segments upstream and downstream of the active site region (blue spots in Fig. 6a, where residues move in the same scale of magnitude but opposite direction). Apart from the fact that the loop β L1-L2-L3-L4 and helices α 2- α 3 move the most in the structural re-arrangement of AdK (Fig. 1 and Accessory Publication Figs S1, S2), the positive and negative correlations indicate that the concerted changes between the open and closed states of AdK actually occur via strongly correlated motions of the LID domain and NMP domain, particularly the correlated motion among helices α 2- α 3, α 4, helices α 6- α 7, and the loop of β L1-L2-L3-L4.

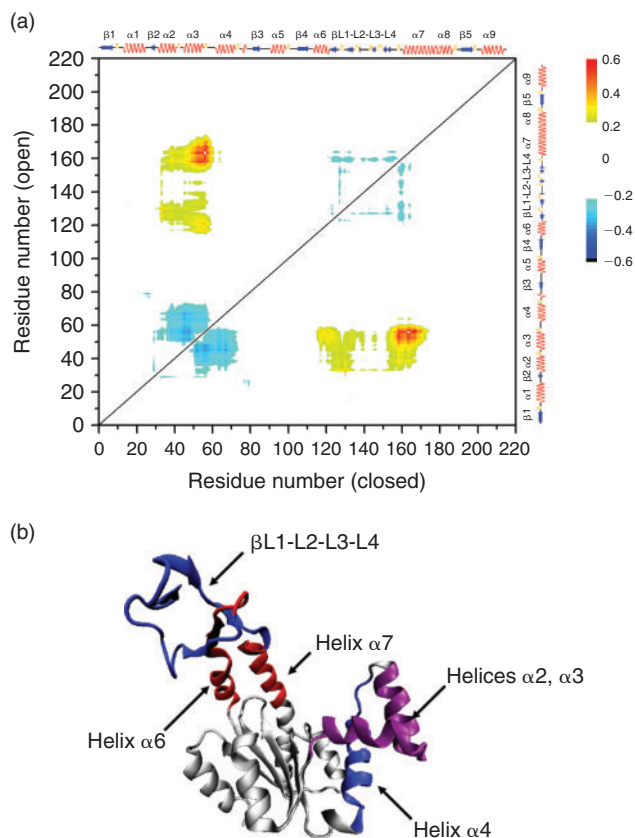


Fig. 6. (a) Difference correlation matrix of the intra-domain motions of AdK, with colour coding in 2D contour map mode (the colour scheme: red, a correlated motion; blue, an anti-correlated motion; and red (blue) regions correspond to same (opposite) direction distortions), between the open and closed conformational transitions. (b) Correlations difference leads to distinctive dynamics of different motifs of AdK. A correlated motion of helix- α 6 and helix- α 7 is coupled to helix- α 2, α 3 motion but anti-correlated with β L1-L2-L3-L4, while helices of α 2, α 3 are anti-correlated with helix α 4. The simultaneous existence of both positive and negative correlations at helices of α 2, α 3 is highlighted by the colour purple (i.e. a mix of red and blue).

However, phosphorylation (e.g. ATP/AMP \leftrightarrow ADP) does not produce significant dynamical changes of the ATP or AMP binding pockets, as well as the LID and NMP domains. For the remaining segments of AdK, only minor correlated movements of other secondary structures are detected from this correlation map. These dynamic correlations have not been determined previously either by NMR experiments^[1,5-7,10,11] or by theoretical simulations.^[9,14,16-18,32,33]

Discussion

Competitive Order and Allosteric Coordinative Ligands Binding in AdK

Our coarse-grained dynamics results for AdK demonstrate the important role of structural topology in regulating intrinsic dynamics and their associated functions. The key unresolved issues for AdK are: (a) how conformational changes of the LID domain or the NMP domain are propagated, coordinating each other in a dynamically correlated way, to the counterparts and rest of AdK; and (b) the competitive order and pathways for ligands (e.g. AMP and ATP) binding or unbinding into AdK. These issues would have fundamental significance when we attempt to

define dynamic criteria to elucidate allostery and cooperativity of catalysis activities in kinase and other proteins.

For AdK binding ATP or AMP, NMR relaxation dynamics^[10] showed that the open and closed states coexist in a dynamic equilibrium with almost equal population. Therefore, the motions of LID-open and NMP-open have a competitive order and cooperativity, while the order would have direct impacts on the kinetics of AdK catalysis.^[8] Cumulative squared overlap of normal modes (Fig. 5, top) shows that for the open \rightarrow closed transition, there are highly agreed overlaps between the open \rightarrow closed and open \rightarrow NMP-open transitions. This means that AdK has a higher probability of transiting via a pathway through the NMP-open state, which verifies the prediction on unidirectional LID domain closure in advance to the NMP domain.^[18] However, the sequence or order of unbinding of ligands (for the closed \rightarrow open transition, Fig. 5) could not be determined exclusively by NMA, and AdK shows no preferred unbinding order.

The presence of both positive and negative correlations accompanying the conformational changes of AdK (Fig. 6a) indicates that, as the ligands bind and phosphorylation proceeds, the LID and NMP domain have a kind of tightly correlated 'breathing' motion (Fig. 6b). That is, a correlated motion of helix- α 6 and helix- α 7 is coupled to helix- α 2, α 3 motion but anti-correlated with β L1-L2-L3-L4, while helices of α 2 and α 3 are anti-correlated with helix α 4. This clearly depicts how different regions of AdK facilitate conformational changes upon ligand binding via certain correlated dynamical pathways. The simultaneous existence of both positive (against helix α 6 and helix α 7) and negative (against helix α 4) correlations at helices of α 2, α 3 strongly suggest that dynamics in coupling to structural transitions have an allosteric cooperativity (i.e. binding at one site affects the conformation of the second and/or other sites.^[3,4]). It is likely that the LID-NMP interface, helices of α 4- α 5, loop β 4, and helix α 8, may translate and propagate the allosteric dynamics changes between LID and NMP. This interface has a unique structural mechanism, as seen from Fig. 1 and Accessory Publication Fig. S1. The helices α 4- α 5, loop β 4, and helix α 8 are physically packed to each other, and they are down stream motifs of helix- α 2, α 3 and helix- α 6, α 7. It is apparent that the most benefit of allosteric cooperativity is to reduce mis-ligation^[18] and sustain catalytic cycle at an equal rate of LID motion, NMP motion and ligands turnover.^[8]

Transitional Pathways in AdK

The binding of ligands allows an ordered closure of the LID and NMP domains in AdK. We have a clear dynamics picture (see above) of a mechanical 'breathing' mechanism of LID and NMP during the structural transition. The open \leftrightarrow closed transition of AdK is accomplished by distant tertiary motifs (as indicated by Fig. 1 and the distance-difference matrix in the Accessory Publication Fig. S1) with strongly correlated and anti-correlated dynamics over large distances (across 10.4 to 47.7 Å). Clearly it is NMP binding that induces both correlated motions and anti-correlated motions between helices α 2- α 3 and LID and the α 2- α 3 and α 4 helices. These dynamical correlations are driven over a large distance and likely controlled by entropy equilibria where enzymatic activation stabilizes the conformation in favour of the fully closed state.^[10] As a result, allosteric coupled dynamics and cooperativity within AdK lead to allosteric pathways for conformational transitions, e.g. closure and opening of the LID and NMP domains.

Based on coarse-grained dynamics results, we can propose an updated mechanism for AdK catalysis and transitional pathways.

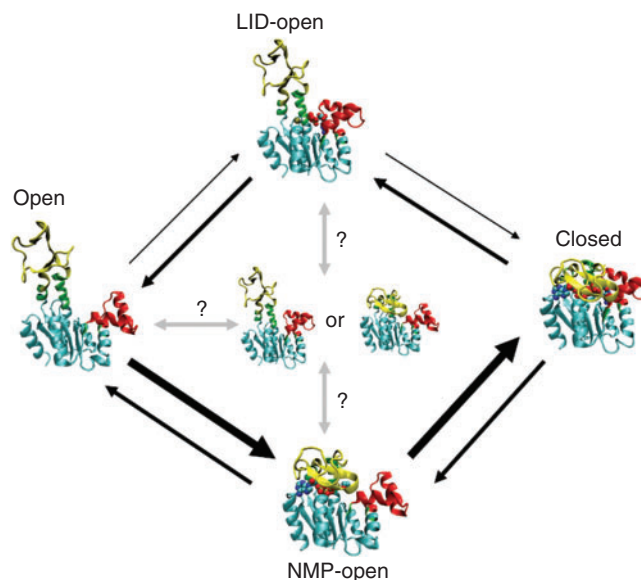


Fig. 7. Proposed dynamic pathways in AdK as determined from NMA and dynamics correlations – the different width of arrows indicates different probability of transition, with thicker width meaning higher probability. The question marks denote the possible conformational equilibrium between the open and the LID-open and/or NMP-open states.

As shown in Fig. 7, the open state is inclined to go through the NMP-open state as transiting into the closed state, with higher probability through the pathway of open \rightarrow NMP-open than open \rightarrow LID-open (where as LID domain always close preceding the NMP domain); on the other side, the closed state has almost equal opportunity following the pathways via NMP-open and LID-open as transiting into the open state. For the later case, pre-existence of a conformational equilibrium between the open and LID-open and/or NMP-open states is possible (as denoted by the grey coloured arrows and question marks in Fig. 7). Transition within this equilibrium scheme is not known yet, at least from coarse-grained dynamics studies.

As demonstrated by NMR dynamics experiments,^[10] addition of ADP to AdK can produce an equilibrium mixture of ATP, AMP, and ADP. Also the ATP-bound (NMP-open) and AMP-bound state (LID-open) are more stable than the non-bound states. This may explain why both LID-open and NMP-open have same probability and there is no dominant transient state in the closed \rightarrow open transition.

A Dynamic Criterion to Determine the Allostery

Apart from AdK, dynamics coupling, correlated motion, and allosteric cooperativity appear to be conserved in the long range communication and conformational transitions of globular proteins.^[3,36–38] For example, a simple mutation can produce marked effects at distal sites via undefined pathways for a conventionally non-allosteric protein.^[37] CheY, a kinase protein like AdK, shows an allosteric mechanism^[38] reconciling the 'induced-fit' scheme^[39,40] with the 'population-shift' theory^[2,41] for allostery.

Given that pre-existing pathways and folding cooperativity seem to be intrinsic for globular proteins,^[37,38,41] collective motions and fluctuations of a protein can initiate its allostery upon external impacts (e.g. ligand binding). Dynamics certainly plays an essential role in allosteric regulations.^[2,38] Our coarse-grained dynamics analysis of AdK suggests a possible

dynamic criterion to determine the allostery in general. Given two distinctive conformational states, dynamical fluctuations and correlations, either among the distant functional motifs (e.g. in AdK) or different subunits,^[42] can be determined to account for the conformational transitions between them. If these correlations result in both correlated and anti-correlated modes of motions, allosteric cooperativity will occur simultaneously. Of this mechanism, allostery prevails upon dynamics,^[3,38,43] and allosteric communications and cooperativity can be mediated solely by changes in motions.^[3]

It is also important to note that NMA handles large-scale motions of proteins with a linear elastic approximation, but for some biological functions (such as catalytic sites mutations) non-linearity is also essential. These non-linear events can only be determined precisely by coarse-grained dynamical correlations with two distinct states available,^[12,28] e.g. the open, closed, and intermediate states of AdK (more generally the active and inactive states of a protein). In practice, allostery that normally involves large-scale motion and possible partial unfolding can be interpreted by structural fluctuations (by Eqn 2) and dynamical correlation differences (by Eqn 1) via normal modes of the starting and reference structures. Allostery can be verified by the existence of both correlated and anti-correlated dynamics as a protein switches from one conformational state to the other.

Conclusion

The residue fluctuations, domain collectivity, detailed internal correlations, and even dynamical pathways associated with structural transitions of AdK are well captured by coarse-grained dynamics using the elastic network model. The coarse-grained dynamics of AdK predicted the dynamics and allostery that generally compare well with experimental relaxation dynamics, yet with less limitation in computing timescale and/or system size. The diagonalized correlations between the open and closed states indicate that most dynamics correlations in AdK occur around the LID and NMP domains, and ligands binding induce flexibility changes and allosteric cooperativity within different regions of AdK. These structural and dynamical analyses suggest that allostery and cooperativity of proteins are intrinsically related to dynamical fluctuations and correlations. The studies of AdK also demonstrate that coarse-grained dynamics represents a robust and powerful tool for probing protein dynamics and its roles in biological functions.

Materials and Methods

Structural Models and Conformational Transition of AdK

AdK is homologous to a large family of signal transduction kinase.^[5-7] This superfamily has highly conserved ligand binding residues. Structures of *E. coli* and *Aquifex* AdK in both the unliganded 'open' state (with both LID and NMP domains open, e.g. PDB:4AKE, PDB:2RH5^[1,6,7]) and the fully liganded 'closed state' (with both domains closed, e.g. PDB:1AKE, PDB:2AKY, PDB:1ANK, PDB:2RGX^[1,6,7]) were determined in the crystalline and solution forms, and they were taken as the modelling structures for conformational transitions of AdK. While the rigid CORE domain essentially contains a twisted five strand β -sheet structure ($\beta 1$ – $\beta 5$), key structural changes from the 'open' to 'closed' state are the rotational inter-domain motions of the LID and NMP domains. This is illustrated in Table 1 and also in the Accessory Publication Fig. S1 (these structural re-arrangements are represented by a distance-difference matrix between the two static conformational states, which highlights

the existence of inter-domain motions). For conformational transitions, structures of PDB:1AK2 and PDB:2AK3 were also used as models for possible intermediate states of AdK. In order to mimic the transient structures during the 'open \leftrightarrow closed' transition, we manually designed the partially-liganded 'NMP-open' state (with LID domain closed and NMP domain open, modelled by superposing the LID domain of PDB:1AKE against the rest of PDB:4AKE) and the partially-liganded 'LID-open' state (the LID domain open with the NMP domain closed, modelled by superposing the NMP domain PDB:1AKE against the rest of PDB:4AKE). Fig. 2 shows a schematic pathway for the structural transition of AdK from the 'open' to the 'closed' state via certain transient states, and vice versa.

For AdK, the competitive order of binding nucleotides at LID or NMP domains will lead to different pathways in the open \leftrightarrow closed transition, namely via the path of either open \leftrightarrow LID-open \leftrightarrow closed or open \leftrightarrow NMP-open \leftrightarrow closed (Fig. 2). Currently, the exact order of binding (as well the exact sequence in the conformational changes of LID or NMP domains) is not known.

Dynamic Correlations in AdK

Dynamic correlations in proteins can be interpreted by the dynamics of cross-correlations among residue fluctuations when the protein undergoes conformational changes.^[29] As the covariance matrix calculates the correlated fluctuations of atoms throughout the protein, to determine coupling between intra- and inter-domain changes, we define a difference correlation matrix of motions between two discrete conformations of protein, e.g. upon ligand binding, as^[29]

$$C_{AB} = (C_{ij}^A - C_{ij}^B)/2 = (\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle^A - \langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle^B)/2, \quad (1)$$

where A and B refer to two dynamically distinguished conformations. In the case of AdK, they refer to the 'open', 'closed', and transient conformers between them. The diagonalized and normalized C_{AB} should have values varying from -1 , 0 to $+1$ showing the strongly-correlated, non-correlated, and strongly-anti-correlated motions, respectively, between conformations of A and B . The term $\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle$ refers to the variance-covariance matrix of the atomic fluctuations between atom i and atom j . This matrix can be derived from dynamics data from either experimental measurements or molecular simulations. However, both experimental methods and molecular dynamics simulations currently have great difficulty in handling large biomolecules over long timescales. Here, an alternative approach is used for an accurate, and computationally fast, determination of $\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle$. Using the dynamics algorithm of an improved Gaussian network model (GNM),^[25] the atomic details of a protein can be coarse-grained by a Gaussian elastic network of its backbone C^α atoms. Mean-square fluctuations of residues, or the cross-correlations, can then be determined via^[28]

$$\langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_j \rangle = \frac{3k_B T}{\gamma} [\Gamma^{-1}]_{ij}, \quad (2)$$

where γ is a theoretical parameter mimicking the strength of the harmonic potential of the elastic network of C^α atoms. $[\Gamma^{-1}]_{ij}$ is the ij th element of the inverse Kirchhoff matrix Γ , a symmetric matrix also known as the valency adjacency matrix in graph topology theory. In the GNM approximation, the elements of Γ

are given by^[25]

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \text{ and } \mathbf{r}_{ij} \leq r_C \\ 0, & \text{if } i \neq j \text{ and } \mathbf{r}_{ij} > r_C \\ -\sum_{i \neq j} \Gamma_{ij} & \text{if } i = j \end{cases} \quad (3)$$

The summation for evaluating Γ_{ij} is performed over all off-diagonal elements on the i th residue at a cut-off distance of r_C , which defines the range of interactions between residues. The mean-square fluctuations of C^α atoms from Eqn 2 also lead to theoretical temperature factor B -values, i.e. $B_i = 8\pi^2 \cdot \langle \Delta \mathbf{r}_i \cdot \Delta \mathbf{r}_i \rangle / 3$, which compare well with X-ray crystallography measurements and other methods.^[25]

Normal Mode Analysis using Elastic Network Model

NMA provides an approximation of protein motions by expanding the motion into a superposition of different normal modes of vibrations, where each normal mode is characterized by a frequency of vibration of overdamped dynamics. Low frequency NMA modes using elastic network models have been applied widely to globular proteins^[26,27] to describe their collective dynamics, dynamic domains, hinge-bending motion, and other functionally important motions. As proteins perform most biological functions dynamically starting near the native states, NMA using the elastic network model of proteins is based on a harmonic perturbation approximation of the potential energy function around a global minimum conformational state. Therefore, NMA captures the intrinsic modes of motions encoded in the protein structures. Conformational transitions in proteins, such as open or closed conformers upon ligand binding or oligomerization due to protein-protein interaction, can be well captured via NMA.

NMA can discriminate conformation changes between large-scale, long-range amplitude motions (via low frequency modes) and confined, local motions (via high frequency modes). The relative importance of normal modes in a particular conformational change is evaluated in terms of collectivity and overlap coefficients.^[26] Collectivity measures the collective degree of protein motion in a certain mode, i.e. the number of residues significantly involved in that mode. The 'overlap coefficient', as defined O_k in Eqn (4),^[26,34] quantifies the degree of agreement between the movement predicted from a particular normal mode sampling (e.g. \mathbf{u}_k) and the 'observed' movement (in this case, the transition between inactive state and active state of AdK), through the dot product of the difference vector $\Delta \mathbf{r}$ between two known structures of the same macromolecule and each normal mode \mathbf{u}_k

$$O_k = \sum_n \Delta \mathbf{r}_n \cdot \mathbf{u}_k^n / \left[\sum_n (\Delta \mathbf{r}_n)^2 \cdot \sum_n (\mathbf{u}_k^n)^2 \right]^{1/2}, \quad (4)$$

where $\Delta \mathbf{r}_n = \Delta \mathbf{r}_n^A - \Delta \mathbf{r}_n^{IA}$, $\Delta \mathbf{r}_n^A$ and $\Delta \mathbf{r}_n^{IA}$ are the n th atomic coordinate of the protein in the active and inactive structures, respectively. A value of 1 for the overlap coefficient means that the direction given by normal mode k is identical to $\Delta \mathbf{r}$, whereas a value of 0 means the direction is uncorrelated with $\Delta \mathbf{r}$. Therefore, high overlap coefficients help lead to identification and prediction of the modes that can be used to describe the possible conformational transition between different states of a protein.^[1,14,26,44] The cumulative overlap coefficients can be used to judge the possible preferred pathways of certain conformational transitions.^[27,43,44]

In this work, NMA was performed mainly by the GNM algorithm^[25] where the protein structure details were coarse-grained by a Gaussian elastic network of the backbone C^α atoms. Fluctuations (interpreted as the residue temperature factor) were derived by Eqn (2). Dynamic correlations in conformational transitions were determined by Eqn (1).

Dynamical Domain Analysis

Dynamical domain analysis was validated using the DynDom algorithm.^[45-47] In the DynDom algorithm, conformational changes of protein, either from X-ray structures, NMA or molecular dynamics simulation, are coarse-grained as quasi-rigid bodies at mechanically large scale. The analysis of a conformational change in terms of domain movements is validated only when the inter-domain deformation is comparable to the intra-domain deformation. The structural basis for defining a dynamic 'domain' is based on the fact that any rigid body displacement can be decomposed by a screw motion about a certain screw axis.^[45,47]

The determination of dynamical domains, hinge axes, and bending residues in AdK was performed with various pairs of structural transitions between different conformational states, as well as from the lowest mode of elastic network computation. In the DynDom analysis, the window length was set initially at five residues until successful domain decomposition (≤ 15 residues). RMSD of the moving domains' best-fit was in the range of 0.9~1.5 Å, and was 0.4~0.9 Å for the fixed domains. The sub-domain matches and mismatches (e.g. hinge) were determined with pair-wise sequence identities above 80%. The domain similarity cut-off was set as the RMSD of $C^\alpha \leq 2.0$ Å, and the dissimilarity cut-off should be RMSD of $C^\alpha \geq 3.0$ Å (while at least seven residues are provided).

Modelling, Simulation, and Visualizations

The modelled structures for different conformational states of AdK, namely 'open', 'closed', 'LID-open', and 'NMP-open', were first relaxed by molecular mechanics (MM) minimization before coarse-graining and the NMA/GNM and DynDom computations. The MM minimization was performed on all-atom structures (with no solvent) for ~2000 steps by NAMD^[48] (with threshold energy less than 10^{-4} kcal mol⁻¹ Å⁻¹). This treatment removes any dynamically uncertain or irregular atoms in the given structure. Further, to ensure our coarse-grained simulations were started with the stable AdK conformations, we have done various sets of all-atom MD simulation of different AdK states in explicit and implicit solvent. We found that in short timescale (<5 ns) and at room temperature (300 K), the structure is very stable, and possible structural transition only occurs at the time scale of $\times 10$ ns and larger with higher temperature (398~498 K).

The dynamics minimization and the normal modes analysis were carried out on a Linux cluster of 145×2 Quad-core Intel CPUs. Typically it takes only a few CPU minutes to accomplish a NMA/GNM calculation of AdK (~214 residues) or MM computing. Structural analysis and graphics visualization were processed by the Virtual Molecular Dynamics VMD package.^[49]

Accessory Publication

Five figures are provided electronically as the Accessory Publication to describe the distance difference matrix, intrinsic fluctuations (B -factors), dynamics cross correlation in static

states, the dominant normal modes for the closed \rightarrow open transitions, and dynamics correlation in *Aquifex* AdK, respectively. The material is available on the Journal's website.

Acknowledgements

This work is partly supported by the Australian Research Council (ARC Discovery Project: DP0557991). Generous allocations of computing resources from the Australian and Victorian Partnerships for Advanced Computing are also acknowledged.

References

- [1] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübnner, D. Kern, *Nature* **2007**, *450*, 838. doi:10.1038/NATURE06410
- [2] D. Kern, E. R. Zuiderweg, *Curr. Opin. Struct. Biol.* **2003**, *13*, 748. doi:10.1016/J.SBI.2003.10.008
- [3] C.-J. Tsai, A. del Sol, R. Nussinov, *J. Mol. Biol.* **2008**, *378*, 1. doi:10.1016/J.JMB.2008.02.034
- [4] J. A. Hardy, J. A. Wells, *Curr. Opin. Struct. Biol.* **2004**, *14*, 706. doi:10.1016/J.SBI.2004.10.009
- [5] C. W. Muller, G. E. Schulz, *J. Mol. Biol.* **1992**, *224*, 159. doi:10.1016/0022-2836(92)90582-5
- [6] U. Abele, G. E. Schulz, *Protein Sci.* **1995**, *4*, 1262. doi:10.1002/PRO.5560040702
- [7] C. W. Müller, G. J. Schlauderer, J. Reinstein, G. E. Schulz, *Structure* **1996**, *4*, 147. doi:10.1016/S0969-2126(96)00018-4
- [8] M. Wolf-Watz, *Nat. Struct. Mol. Biol.* **2004**, *11*, 945. doi:10.1038/NSMB821
- [9] K. Arora, C. L. Brooks, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18496. doi:10.1073/PNAS.0706443104
- [10] J. Ådén, M. Wolf-Watz, *J. Am. Chem. Soc.* **2007**, *129*, 14003. doi:10.1021/JA075055G
- [11] M. Bellinzoni, A. Haouz, M. Grana, H. Munier-Lehmann, W. Shepard, P. M. Alzari, *Protein Sci.* **2006**, *15*, 1489. doi:10.1110/PS.062163406
- [12] O. Miyashita, J. N. Onuchic, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12570. doi:10.1073/PNAS.2135471100
- [13] N. A. Temiz, E. Meirovitch, I. Bahar, *Proteins* **2004**, *57*, 468. doi:10.1002/PROT.20226
- [14] P. Maragakis, M. Karplus, *J. Mol. Biol.* **2005**, *352*, 807. doi:10.1016/J.JMB.2005.07.031
- [15] H. Lou, R. I. Cukier, *J. Phys. Chem. B* **2006**, *110*, 24121. doi:10.1021/JP064303C
- [16] P. C. Whitford, O. Miyashita, Y. Levy, J. N. Onuchic, *J. Mol. Biol.* **2007**, *366*, 1661. doi:10.1016/J.JMB.2006.11.085
- [17] Q. Lu, J. Wang, *J. Am. Chem. Soc.* **2008**, *130*, 4772. doi:10.1021/JA0780481
- [18] P. C. Whitford, S. Gosavi, J. N. Onuchic, *J. Biol. Chem.* **2008**, *283*, 2042. doi:10.1074/JBC.M707632200
- [19] A. J. Wand, *Nat. Struct. Mol. Biol.* **2001**, *8*, 926. doi:10.1038/NSB1101-926
- [20] A. Mittermaier, L. E. Kay, *Science* **2006**, *312*, 224. doi:10.1126/SCIENCE.1124964
- [21] C. Clementi, H. Nymeyer, J. N. Onuchic, *J. Mol. Biol.* **2000**, *298*, 937. doi:10.1006/JMBI.2000.3693
- [22] F. Ding, N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, E. I. Shakhnovich, *Biophys. J.* **2002**, *83*, 3525. doi:10.1016/S0006-3495(02)75352-6
- [23] K.-i. Okazaki, N. Koga, S. Takada, J. N. Onuchic, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11844. doi:10.1073/PNAS.0604375103
- [24] S. Sharma, F. Ding, N. V. Dokholyan, *Biophys. J.* **2007**, *92*, 1457. doi:10.1529/BIOPHYSJ.106.094805
- [25] I. Bahar, A. R. Atilgan, B. Erman, *Fold. Des.* **1997**, *2*, 173. doi:10.1016/S1359-0278(97)00024-2
- [26] M. Delarue, P. Dumas, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6957. doi:10.1073/PNAS.0400301101
- [27] J. Ma, *Structure* **2005**, *13*, 373. doi:10.1016/J.STR.2005.02.002
- [28] D. Tobi, I. Bahar, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18908. doi:10.1073/PNAS.0507603102
- [29] M. S. Liu, B. D. Todd, S. Yao, Z.-P. Feng, R. S. Norton, R. J. Sadus, *PROTEINS: Structure, Function, and Bioinformatics* **2008**, *73*, 218. doi:10.1002/PROT.22056
- [30] S. Yao, M. S. Liu, S. L. Masters, J.-G. Zhang, J. J. Babon, N. A. Nicola, S. E. Nicholson, R. S. Norton, *Protein Sci.* **2006**, *15*, 2761. doi:10.1110/PS.062477806
- [31] M. S. Liu, B. D. Todd, R. J. Sadus, *Biochim. Biophys. Acta* **2006**, *1764*, 1553.
- [32] H. Lou, R. I. Cukier, *J. Phys. Chem. B* **2006**, *110*, 12796. doi:10.1021/JP061976M
- [33] O. Miyashita, J. N. Onuchic, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12570. doi:10.1073/PNAS.2135471100
- [34] F. Tama, Y. H. Sanejouand, *Protein Eng.* **2001**, *14*, 1. doi:10.1093/PROTEIN/14.1.1
- [35] A. W. Van Wynsberghe, Q. Cui, *Structure* **2006**, *14*, 1647. doi:10.1016/J.STR.2006.09.003
- [36] A. M. Stock, V. L. Robinson, P. N. Goudreau, *Annu. Rev. Biochem.* **2000**, *69*, 183. doi:10.1146/ANNUREV.BIOCHEM.69.1.183
- [37] M. W. Clarkson, S. A. Gilmore, M. H. Edgell, A. L. Lee, *Biochemistry* **2006**, *45*, 7693. doi:10.1021/BI060652L
- [38] M. S. Formanek, L. Ma, Q. Cui, *Proteins* **2006**, *63*, 846. doi:10.1002/PROT.20893
- [39] D. E. J. Koshland, K. Hamadani, *J. Biol. Chem.* **2002**, *277*, 46841. doi:10.1074/JBC.R200014200
- [40] J.-P. Changeux, S. J. Edelstein, *Science* **2005**, *308*, 1424. doi:10.1126/SCIENCE.1108595
- [41] K. Gunasekaran, B. Ma, R. Nussinov, *Proteins* **2004**, *57*, 433. doi:10.1002/PROT.20232
- [42] M. S. Liu, B. D. Todd, R. J. Sadus, *Biochim. Biophys. Acta* **2005**, *1752*, 111.
- [43] W. Zheng, B. R. Brooks, D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 7664. doi:10.1073/PNAS.0510426103
- [44] W. Zheng, B. Brooks, *J. Mol. Biol.* **2005**, *346*, 745. doi:10.1016/J.JMB.2004.12.020
- [45] S. Hayward, A. Kitao, H. J. C. Berendsen, *PROTEINS: Structure, Function, and Genetics* **1997**, *27*, 425. doi:10.1002/(SICI)1097-0134(199703)27:3<425::AID-PROT10>3.0.CO;2-N
- [46] S. Hayward, *PROTEINS: Structure, Function, and Genetics* **1999**, *36*, 425. doi:10.1002/(SICI)1097-0134(19990901)36:4<425::AID-PROT6>3.0.CO;2-S
- [47] S. Hayward, H. J. C. Berendsen, *PROTEINS: Structure, Function, and Genetics* **1998**, *30*, 144. doi:10.1002/(SICI)1097-0134(19980201)30:2<144::AID-PROT4>3.0.CO;2-N
- [48] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, *J. Comput. Phys.* **1999**, *151*, 283. doi:10.1006/JCPH.1999.6201
- [49] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33. doi:10.1016/0263-7855(96)00018-5