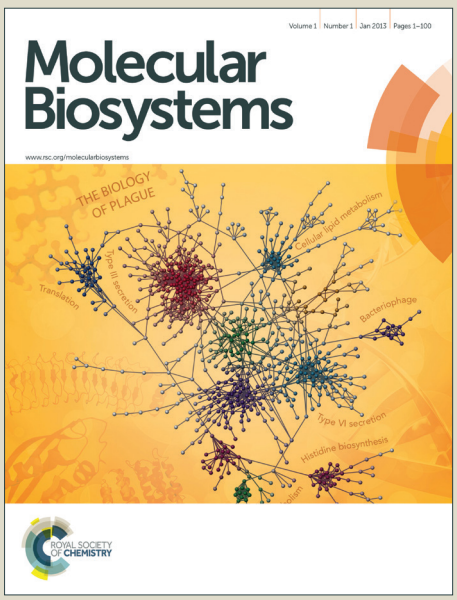
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

A GAUSSIAN NETWORK MODEL STUDY SUGGESTS THAT STRUCTURAL FLUCTUATIONS ARE HIGHER FOR INACTIVE STATES THAN ACTIVE STATES OF PROTEIN KINASES

RAJU KALAIVANI and NARAYANASWAMY SRINIVASAN*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, Karnataka, India

*Corresponding Author : Narayanaswamy Srinivasan

Lab no. 103, Molecular Biophysics Unit

Indian Institute of Science

Bangalore 560012, Karnataka, India

Telephone : +91-80-2293 2837

Fax : +91-80-23600535

Email : ns@mbu.iisc.ernet.in

RUNNING TITLE : Dynamics of kinases in different functional states

No. of manuscript pages : 55

No. of Tables : 4

No. of figures : 9

Supplementary Material : Attached as a separate file named Kalaivani_Srinivasan_2015_supplementary_information.doc.

ABSTRACT

We performed Gaussian Network Model based Normal Mode Analysis of 3-dimensional structures of multiple active and inactive forms of protein kinases. In 14 different kinases, more number of residues (1095) show higher structural fluctuations in inactive states than those in active states (525), suggesting that, in general, mobility of inactive states is higher than active states. This statistically significant difference is consistent with higher crystallographic B-factors and conformational energies for inactive than active states, suggesting lower stability of inactive forms. Only a small number of inactive conformations with the DFG motif in the "in" state were found to have fluctuation magnitudes comparable to the active conformation. Therefore our study reports for the first time, intrinsic higher structural fluctuation for almost all inactive conformations compared to the active forms. Regions with higher fluctuations in the inactive states are often localized to αC-helix, αG-helix and activation loop which are involved in the regulation and/or in structural transitions between active and inactive states. Further analysis of 476 kinase structures involved in interactions with another domain/protein showed that many of the regions with higher inactive-state fluctuation correspond to contact interfaces. We also performed extensive GNM analysis on (i) Insulin Receptor kinase bound to another protein and (ii) holo and apo forms of active and inactive conformations followed by multi-factor analysis of variance. We conclude that binding of small molecules or other domains/proteins reduce the extent of fluctuation irrespective of active or inactive forms. Finally, we show that the perceived fluctuations serve as useful input to predict functional state of a kinase.

KEYWORDS

Active and inactive conformations, Conformational variability, Gaussian Network Model, Normal Mode Analysis, Protein dynamics; Protein kinases

INTRODUCTION

Protein phosphorylation is one of the most influential mechanisms in cellular signalling, affecting up to 30% of cellular proteome[1]. Upon phosphorylation, the protein may be altered in its activity[2], localisation[3] or affinity to bind different proteins[4] and could thus signal information about cell division[5], movement[6], metabolism[7], apoptosis[8] etc. In eukaryotes, protein phosphorylation at serine (S), threonine (T) and tyrosine (Y) residues is brought about by a superfamily of enzymes, henceforth referred as STY kinases. Essentially every biochemical pathway in the cell is directly or indirectly controlled by protein phosphorylation, and thus by STY kinases[9]. The diversity and branching/convergence of signalling pathways is reflected as the large fraction (~2%) of STY kinases encoded in the eukaryotic genomes[10] and over 500 STY kinases in human genome[11]. The importance of STY kinases is further demonstrated by the presence of multitude of regulatory mechanisms that tightly control the activity of each of these kinases[12–17]. Any abnormality in the regulation of their activity, upon mutations, leads to erroneous cellular information processing with wide implications in cancer[18] and autoimmunity[19].

Consistent with the diverse regulatory mechanisms and substrate specificities of the STY kinases, sequence analysis of the kinase catalytic domains indicates a large evolutionary divergence of the superfamily with a meagre ~20% sequence identity across groups[20]. On the other hand, the machinery of all STY kinases involves the covalent transfer of γ-phosphate group from an ATP molecule to a free hydroxyl (-OH) group of S/T/Y residue in a protein substrate. Owing to this common functionality, the catalytic domains of the STY kinases share a conserved structural fold[21], evident from a very large number of available crystal structures of kinases. The common scaffold consists of two lobes: a smaller N-terminal lobe and a larger C-terminal lobe. The N-lobe comprises predominantly of β strands and a

conserved helix, referred as αC-helix, containing a conserved glutamate (E91 - residue numbering followed here and throughout this paper, is that of cAMP-dependent protein kinase, PDB code: 1ATP_E, unless mentioned otherwise). The C-lobe is largely made of helices. Between the two lobes is a cleft within which ATP binds. ATP binding is stabilised by a loop spanning the region 49-57 connecting β1 and β2 strands in the N lobe. This loop is highly flexible, and thus presumably aids the binding of small molecule inhibitors in the cleft. Another flexible segment is the activation loop (residues 145-172), centrally located in the C-lobe, spanning from DFG motif till APE motif. The DFG motif helps position the ATP and cations in appropriate orientation, facilitating phosphoryl transfer[22]. Conformational variability in the αC-helix and activation loop is strongly implicated in the regulation of the kinase between its functional states[23,24].

STY kinases are switched between at least two extreme functional states: active and inactive, by kinase specific regulatory mechanisms. Functionally, the enzyme possesses maximal catalytic competence to phosphorylate protein substrates in the active state, whereas the catalytic rate of the enzyme is minimal in an inactive state. Several crystal structures of STY kinases are available in the active and inactive states. Structurally, in the active conformation, the αC-helix is much closer to the kinase body, stabilising a salt bridge between the glutamate E91 of the αC-helix and a conserved lysine (K72) in the N lobe. This positioning of K72, in turn, positions the ATP during phosphoryl transfer. In addition to the structural positioning of the αC-helix in the active conformation, the activation loop and the P+1 loop are in an open and extended conformation, conducive for substrate binding. Apart from this, the aspartate of DFG motif in the activation loop is near the ATP binding loop, in the “in” conformation and the N lobe and the C lobe are closed with respect to each other. On the other hand, in the inactive conformations, one or more of the above conformational constraints are impeded, with no unique conformation correlating with the inactivity of the enzyme[24]. What is the

feature that positively defines the active and inactive states of the kinase? Although there is a clear correspondence of the structural forms of the STY kinases to the functionally active state, there is no clear understanding of the underlying mechanistic features that facilitate specific functionalities or function-specific regulatory mechanisms. Despite identical sequence and common structural scaffold between the active and inactive states, how do distinct regulatory mechanisms affect specific functional states?

Despite decades of sequence and structural studies on kinases, details of transition through conformational spaces, or mechanistic understanding of regulation of STY kinases remain poorly understood. This is primarily because these studies consider the protein molecule to be static and in isolation. On the contrary, the protein molecules are fluid with functional motions and interactions within and across molecules[25], which are crucial for their functions[26,27] and in understanding their evolution[28]. For instance, a kinase might require a certain conformational flexibility/rigidity to bind to a substrate protein and catalyse the phosphotransfer. We hypothesise that the functionality of an STY kinase is coupled to its dynamics and disposition for collective conformational motions. To address this, we analyse the vibrational dynamics of the known crystal structures of STY kinases in the context of their functionality.

Motions in protein molecules can range from local (e.g., vibrations of atoms) to global (e.g., concerted domain movements). Unlike local (or low collectivity) motions, global motions involving coordinated movement of many atoms have been reported to be functionally important[29–31]. Such global scale motions can be analytically derived using coarse-grained approaches like Elastic Network Model (ENM)[32] based Normal Mode Analysis (NMA)[33,34]. Although minimal and approximate, this approach has been efficient in understanding biologically relevant[35] (e.g., functional hinges, catalytic sites, ligand binding regions) and

116 experimentally consistent[36] (e.g., residues critical for protein folding and function) vibrational
117 motions in protein molecules. In the present study, we use Gaussian Network Model (see
118 Methods) to understand the collective dynamics of STY kinases and corroborate the same
119 with functional relevance.

120 We analysed the dynamics of catalytic domains of 14 STY kinases, whose crystal structures
121 are available in both active and inactive conformations (see Table I) using Gaussian Network
122 Model (GNM). The 14 STY kinases chosen for the study are from diverse groups (AGC,
123 CMGC, CAMK, TK, STE, Other) and sub-families based on the sequence similarity of the
124 kinase domain. It is to be noted that all the entries corresponding to each of the 14 STY
125 kinases (Table I) belong to a unique UniProt ID and are identical in sequence and gross
126 structural fold. They vary only in local backbone conformations corresponding to the levels
127 of catalytic activity of the protein. Are these local conformational variations sufficient to
128 cause changes in the global dynamics of the protein, which can stabilise/destabilise functional
129 states of the protein or enable/disable additional regulatory mechanisms?

130

RESULTS

Fluctuations of active and inactive conformations are not identical

A total of 55 kinase structures, including 14 distinct STY kinases, were studied using GNM based NMA. The entire study, except for one control experiment, was carried out by considering the spatial positions of the Cα atoms only. All small molecules, including water, ATP, ATP analogs, substrate peptides and cations, were stripped off the structures (Influence of bound small molecules on the fluctuations of kinases has been separately dealt with elsewhere in this paper). In essence, only a single chain of Cα atoms belonging to the kinase catalytic domain was used for the construction of network topology for the purpose of GNM. For each of the kinase structures, thus prepared, residue-wise mean square fluctuations in the mode of least frequency were calculated and normalised (See Methods). Global mode Normalised Mean Square (NMS) fluctuations, thus calculated, for the active and inactive conformations of STY kinases were compared with each other.

As an example, for an STY kinase, Insulin receptor kinase (entry no. 5 in Table I) belonging to the Tyrosine Kinase (TK) group, we have plotted the global mode NMS fluctuations of all the residues in the active (PDB ID 1IR3_A, *blue curve*) and inactive (PDB ID 1IRK_A, *red curve*) conformations (Fig. 1*A*). The structure corresponding to the inactive conformation, 1IRK_A, contains missing coordinates for residues 1158, 1162 and 1163 (numbering as in the PDB file) in the activation loop. Here, and throughout the study, the missing residues in a kinase were not modelled and the fluctuations corresponding to the missing residues along with 2 preceding and 2 succeeding residues were ignored. This is because even small errors in modelling the missing residues, could result in altered network topology, and thus could suggest spurious fluctuations. On the other hand, absence of residues in the structure would lead to altered network topology near the missing coordinates; thus, 2 preceding and 2

succeeding residues are also excluded from the analysis. Thus, in 1IRK_A, fluctuations corresponding to residues 1156-1165 were ignored (See the region of missing *blue curve* in Fig. 1*A*).

Fig. 1*A* shows that the NMS fluctuations observed in the active and inactive conformations are not identical. Throughout the NMS fluctuation profile, a trend of higher magnitude of fluctuations is seen in the inactive conformation than in the active conformation. The largest difference between the active and inactive fluctuation profiles is seen in the αC-helix region. For a quantitative analysis, we calculated the areas under each of the curves, after excluding residues 1156-1165 that correspond to the missing residues and two flanking residues on either side. The area under the blue curve, corresponding to the active conformation, is 61.61 units square, while the area under the red curve, corresponding to the inactive conformation, is 76.78 units square (Fig. 1*D*).

Further, a comparison of 11 structures of cAMP-dependent Protein Kinase (PKA), 8 solved in the active conformation (PDB IDs: 1JBP_E, 1APM_E, 1L3R_E, 1RDQ_E, 2ERZ_E, 2CPK_E, 1FMO_E, 1ATP_E, *blue curve*) and 3 in the inactive conformations (PDB IDs: 2QVS_E, 4DFY_A, 1SYK_A, *red curve*) was done (Fig. 1*B*). For simplicity, the residues are numbered from 1 to 255, irrespective of the individual UniProt/PDB numbering. Blue curve in Fig. 1*B* represents the mean NMS fluctuation of each residue in PKA calculated from the 8 independent active structures, while the red curve represents the mean NMS fluctuation of each residue calculated from the 3 independent inactive structures. The plot clearly highlights that there are specific sets of residues, marked by green '*' symbol below the 0 ordinate, that have significantly different distributions of fluctuations in the active and inactive forms (two-tailed unpaired T-test, p-value < 0.05). Specifically for these regions with differential fluctuations in the active and inactive forms, we calculated the area under the curves (Fig.

1*E*). As with the Insulin receptor kinase example, PKA structures also exhibit higher fluctuations in the inactive conformation (area under the curve = 16.08 units square) than in the active conformation (area under the curve = 14.16 units square).

To compare the residue-wise fluctuations in all of the 55 STY kinases (Table I), information of topologically equivalent sets of residues across kinases is required. To this end, multiple sequence alignment of the catalytic domains of the 55 STY kinases was done using T-Coffee[37]. We plotted the NMS fluctuations, perceived from GNM based global Normal Mode, for each of the topologically equivalent residues (Fig. 1*C*) for the active (*blue curve*) and inactive (*red curve*) conformations separately. For simplicity, the residues are commonly numbered from 1 to 361, irrespective of the individual UniProt/PDB numbering. The blue curve represents the mean NMS fluctuations calculated from 31 independent active conformations; the red curve represents that of 24 independent inactive conformations. We observe that there are specific regions, marked by green '*' symbol below the 0 ordinate, that have significantly different distributions of fluctuations in the active and inactive conformations (two-tailed unpaired T-test, p-value < 0.05). For these regions of differential fluctuations, area under the curve for active conformations is 7.15 units square and the area under the curve for inactive conformations is 11.58 units square. These analyses indicate that the dynamics of certain residues are different in the active and inactive conformations. The relative propensity of such residues to fluctuate, with respect to other residues in the structure, varies in magnitude depending on whether the structure is in active or inactive conformation.

Fluctuations in the inactive forms are higher

To further characterise the residue-wise fluctuations, the magnitudes of global mode NMS fluctuations in the active and inactive forms were compared. To this end, we randomly chose

one PDB entry per STY kinase corresponding to the active conformation (Table I) and compared it with each of the inactive conformations of the same kinase, resulting in 24 active-inactive conformation pairs (See Table II). In a scatter plot in Fig. 2*A*, the global mode NMS fluctuations of each of the residues in the 24 inactive conformations are plotted against that of the topologically equivalent residues in the corresponding active counterparts. The black diagonal line is of unity slope, upon which the points would lie if the fluctuations of active and inactive conformations were identical. But in observation, more number of residues (n = 1095) have greater fluctuations in the inactive conformation than those (n = 525) that have greater fluctuations in the active conformation. This distribution of data points across the unity slope is statistically significant (Chi-Square-test, *p*-value = 1.58e-45). This indicates that there is a systematic difference between the fluctuations in the active and inactive conformations, in that the residues fluctuate more when the kinase is in the inactive conformation than when it is in the active conformation. We argue that if the global mode fluctuations indicate higher flexibility for more number of residues in the inactive conformation, then we would expect the kinases to have lesser stability in the inactive conformation than in its active conformation. This would be reflected as higher B-factors and conformational energies of the crystal structures of inactive conformations.

B-factors of the Cα atoms of the kinase catalytic domain were retrieved from each of the structures in the 24 active-inactive conformation pairs and normalised (See Methods). B-factors of each of the Cα atoms in the kinase catalytic domains of 24 inactive conformations were plotted against that of the topologically equivalent Cα atoms in the corresponding active conformations in Fig. 2*B*. In corroboration with our argument, more number of residues (3693) have a higher normalised B-factor in the inactive conformation than those (2232) that have a higher normalised B-factor in the active conformation. This distribution of data points is statistically significant (Chi-Square-test, *p*-value = 2.68e-73). Although the B-factors are

influenced by effects of crystal packing and dynamics of higher modes, it is important to note that it nevertheless reflects the trend observed in the Normal Mode fluctuation analysis. Further, FoldX[38] program was used to calculate the conformational energy of the active and inactive structural forms of each of the 14 STY kinases (Table I) in terms of difference in free energy of folding, $\Delta G_{Folding}$ (kcal/mol). If for an active or inactive state of an STY kinase, multiple structures were available (for e.g., entry no. 4 in Table I), mean of $\Delta G_{Folding}$ (kcal/mol) of all the available structures in a specific state was calculated. A plot of the mean conformational energies of the inactive conformations against that of the active conformations is shown in Fig. 2*C*. 10 out of the 14 STY kinases considered for analysis have higher energies (lower stability) in the inactive conformation, in line with our results of NMA.

Functional regions contribute to higher fluctuations in the inactive forms

What are the regions that contribute to the difference in fluctuations between the active and inactive states? To assess this, for each of the 24 active-inactive conformation pairs, the difference in the global mode NMS fluctuations of residues between the active and inactive states was calculated (see Methods). Residue-wise mean difference in fluctuations between the active and inactive states $\langle NMS\_Fluct_i^{inactive} - NMS\_Fluct_i^{active} \rangle$ was mapped on to the kinase fold in a blue-white-red scheme as shown in Fig. 3*A*. In the figure, regions that fluctuate more in the active state, thus having negative mean difference, are coloured blue and those that fluctuate more in the inactive state, having positive mean difference, are coloured red. This plot highlights that the residues with fluctuation difference are not randomly distributed throughout the structure, but are concentrated in specific regions. The only region that exhibited a trend of higher fluctuations in the active state is the ATP binding loop (*coloured blue* in Fig. 3*A*). On the contrary, αC-helix, activation loop and αG-helix (see

Table S1 for definition of the functional regions) exhibit larger fluctuations in the inactive state (*coloured red* in Fig. 3*A*).

This is in line with a previous study that reported higher flexibility of functional residues in yeast enolase in the inactive forms, when compared to the active forms[39], adding evidence to the hypothesis that functional regions have differentiated dynamics, in addition to differentiation in structural conformations.

Activation loop segments[23,24] and αC-helix[40] have been known to be important conformational switches during the transition between active and inactive states of STY kinases. αC-helix forms important interactions both in the N lobe and with the activation loop in the C-lobe, which is critical for the functional state of the kinase. It should also be noted that the αC-helix regulates the kinase by binding to external binding proteins like cyclin[41] (in the case of CDKs) and is possibly involved in allosteric modulation of the kinase[42]. The role of αG-helix in the switch is being elucidated lately. Growing evidence suggests that αG-helix is involved in the activation of kinase[43] and docking substrates[44–46]. Given this scenario, what is the functional consequence of differential structural fluctuations in these regions between the active and inactive states? Due to phosphorylation being one of the most predominant signalling mechanisms in cells, specificity in signal transduction is achieved by regulating STY kinases extensively. Most STY kinases interact with other molecules, resulting in activation/regulation of the kinase. In this context, we hypothesise that the regions having differential dynamics in the active and inactive states are crucial for the regulation of the kinase and/or important in the active-inactive switch. To test this, 476 known crystal structures of STY kinases solved in complex with one or more interacting partners were collected from iPfam (see List S1)[47]. The inter-chain interactions at each residue in the STY kinase catalytic domain from all the 476 structures were investigated (see Methods). This

residue-wise interaction has been mapped on to the kinase fold in a blue-white-red scheme as shown in Fig. 3*B*. Residues that participate in the most number of inter-chain interactions are coloured red and those that form the least number of interactions are coloured blue. It is clear from this plot that αC-helix, activation loop and αG-helix are the regions that are most commonly involved in protein-protein interactions.

Interestingly, the regions which show difference in fluctuations between the active and inactive states are also commonly involved in protein-protein interactions. This is in corroboration with other studies that report regions with differential flexibility between the active and inactive forms to be involved in protein-protein interactions[48]. From the above observations, it could be reflected that during the active-inactive switch of the kinase, dynamics of regions important for protein-protein interactions change. It is to be noted that using only the Cα coordinates of the STY kinase catalytic domain in the active and inactive states, we were able to deduce the regions of functional relevance, protein-protein interaction and regulation, viz., αC-helix, activation loop and αG-helix. We point out for the first time that αG-helix plays a crucial role in protein-protein interactions in congruence with altered dynamics during the active-inactive switch.

From the above analyses, we interpret that the residues most commonly involved in protein-protein interactions in STY kinases have decreased dynamics and higher stability when the kinase switches from an inactive to active state. This is clearly seen as a correlation (cc = 0.44, *p*-value < e-14) between the mean fluctuation difference (Fig. 3*A*) and the number of inter-chain interactions (Fig. 3*B*) of the residues. It is well known that the binding of a protein to another affects (and/or is affected by) the dynamics and stability of the proteins involved[49,50]. It should be noted that the active-inactive switch of some STY kinases is regulated by interaction with another protein. This protein-protein interaction could help

stabilise the kinase in a specific conformation by altering the kinase conformation around the interface region, which indirectly affects the flexibility and dynamics of the kinase around the interface region. According to this argument, binding of a protein to an STY kinase should bring about change in the fluctuation of the kinase irrespective of the functional state of the kinase. To this end, we studied the GNM based normal mode of an example STY protein kinase, solved in complex with a protein regulator.

Effect of bound protein on NMS Fluctuations: A case study with Insulin Receptor kinase:

Insulin Receptor kinase (entry no. 5 in Table I), crystallised in the active (PDB ID 1IR3_A) and inactive (PDB ID 1IRK_A) states, was used in the previous analyses. While the active state conformation is phosphorylated and is bound to small molecules, Phosphoaminophosphonic Acid-Adenylate Ester (ANP) and Mg ions, both the active and inactive forms are crystallised with no bound proteins to the kinase. By virtue of the difference in conformations alone, we observe intrinsic higher fluctuations of residues in the inactive conformation than in the active conformation (Fig. 1*A*). The difference in the area under the two curves in Fig. 1*A* is 14.85 square units.

Depetris et al.[51] solved the crystal structure (PDB ID 2AUH) of Insulin Receptor (InsR) kinase in complex with Grb 14 BPS (Growth factor receptor bound protein's InsR binding region), a fragment of an adapter protein that negatively regulates the kinase. In this bound state structure (2AUH_A, Fig. 4*A*), the kinase is phosphorylated and assumes an active conformation. The interface between InsR and Grb 14 BPS is formed by residues in the αC-helix, activation loop and αG-helix of the kinase (Fig. 4*B*). We performed GNM based NMA of the bound InsR kinase structure (2AUH) in 2 ways: (i) Bound_isolated - considering only the Cα atoms of the kinase catalytic domain in isolation, after stripping the bound Grb 14

BPS (Fig. 4*A*), (ii) Bound_complex - considering the Cα atoms of both the kinase catalytic domain and the bound Grb 14 BPS (Fig. 4*B*). It is to be noted that in both the cases, the conformations of the kinase domain are identical, and correspond to the active state conformation. We compared the fluctuations of topologically equivalent residues of the Insulin Reseptor kinase in the (i) bound_isolated (PDB ID 2AUH_A, *blue curve*) and (ii) bound_complex (PDB ID 2AUH_A and 2AUH_B, *red curve*) forms in Fig. 4*C*. The area under the two curves is plotted in Fig. 4*D*, and it is clearly seen that the bound_complex form indeed has much lower fluctuations than the bound isolated structure. The difference in the areas under the two curves is 16.88 units square. Further inspection of the NMS fluctuation profiles and the structure of the Insulin Receptor kinase complex clearly shows that the higher fluctuations are localised in the αC-helix, activation loop and αG-helix regions (Fig. 4*C*), which form the interface of the protein-protein complex (Fig. 4*B*).

Combining this result with the results of comparative analyses of active and inactive conformations suggests that the fluctuations of the kinase scaffold can be influenced by two independent means: (i) by virtue of 3D conformations of the active and inactive states and (ii) by virtue of bound proteins. Irrespective of the functional form, the protein-bound form has a lower fluctuation than the protein-unbound form. Therefore, we contend that the fluctuations in the active and inactive forms of kinases can be influenced, over and above, by interactions with other proteins/domains.

Specific structural regions fluctuate more in the inactive conformations

Within the kinase domain, we quantitatively assessed the dynamics of specific structural segments of known function: αC-helix, catalytic loop, activation loop, αG-helix, hinge residues and ATP binding loop residues (Fig. 5*A*-*F* respectively). When all of the 24 active-inactive conformation pairs were considered together, the hinge residues (see Table S2 for the

list of residues) did not show significant difference in fluctuations between the active and inactive states (Fig. 5*E*). 80 residues have higher magnitude of NMS fluctuations in the inactive state and 60 residues have higher magnitude of NMS fluctuations in the active state (Chi-Square-test, p-value > 0.05). Similarly, although we observe mild differences, ATP binding loop also has similar fluctuation magnitudes in the active and inactive states (Fig. 5*F*, Chi-Square-test, p-value > 0.001). On the contrary, αC-helix, catalytic loop, activation loop and αG-helix showed significantly higher fluctuations in the inactive state than in the active state (Fig. 5, *A*-*D*, Chi-Square-test, p-value < e-6).

All, but "DFG-in", inactive conformations exhibit equivalent high fluctuations

As has already been noted, STY kinases exist in inactive states by virtue of many different conformations. In fact, a given STY kinase may assume more than one structural form during its inactivity. In this regard, it is of interest to understand whether there are fluctuation differences between the different inactive state conformations. To this end, we classified the inactive STY kinase structures into three forms depending on : (i) conformation of the DFG motif, (ii) conformation of the αC-helix and (iii) the relative orientation of the N and C lobes (Table III).

In Fig. 6*A*, the mean global mode NMS fluctuations of all the 31 active structures (*blue curve*) and the 6 "DFG-in" inactive structures (*red curve*) is plotted. Interestingly, we observe that the difference in fluctuations between the active and "DFG-in" inactive conformations is not significant. This suggests that the conformation of the DFG motif influences the flexibility of the kinase to a large extent. When present in the "DFG-in" conformation, the inactive kinase exhibits fluctuations similar to that of an active kinase. On the other hand, the other reported conformations of the DFG motif in the inactive state[52], viz., "flip", "Src-like" and "out" (Fig. 6*B*-*D*) do not show much variation among themselves, but, show higher

fluctuations compared to active states. The fluctuations in different categories are quantitatively plotted as the area under the curves in Fig. 6*E*, which clearly shows that the areas under the curve for the active and "DFG-in" inactive are quite similar, and the areas under the curves for the DFG “flip”, “Src-like” and “out” are similar to that of an average inactive conformation. Also, we analysed if there exists variations in fluctuations between the “in” and “out” conformations of αC-helix (Fig. S1) and “open” and “closed” orientations of the N and C lobes (Fig. S2). There were not any detectable variations in the fluctuations of these different inactive conformations (Fig. 6*F*). Thus, we conclude that, all the different inactive conformations of STY kinases, except for those that have the DFG motif in the “in” conformation, exhibit equivalently higher fluctuations, when compared to that of the active conformations.

Observed difference in fluctuations is not entirely a consequence of small molecule binding or mutations

In this work, Gaussian Network Model analysis of a protein kinase domain structure uses the atomic coordinates of Cα atoms only. All the external binding factors like ATP, Mg, Mn, substrate peptide, etc. were neglected. One could argue that if a bound factor was present in the active state structure and absent in the inactive state structure, a void space would be created in the binding state during the removal of the bound factor before the construction of the topological network. This could lead to differential GNM network topologies. By this argument, the difference in the fluctuations between the active and inactive states is merely a consequence of systematic presence of bound factors in the kinase structures, and is not a true property of the functional states. To assess this, we selected all possible pairs of active conformations of an STY kinase (with a unique UniProt ID) crystallised with and without one/more of the small molecules listed above. These are called Active-Holo and Active-Apo

structures respectively. Similarly, all possible pairs of inactive conformations of an STY kinase (with a unique UniProt ID) crystallised with and without one/more of the small molecules are called Inactive-Holo and Inactive-Apo structures respectively. We identified 34 such active conformation pairs and 11 such inactive conformation pairs. In Fig. 7*A*, we have plotted the global mode NMS fluctuations of the active conformation pairs when bound (*blue curve*) and unbound (*red curve*) to small molecules. It is clear from this plot that the global mode fluctuations are not affected significantly by the binding of small molecules like ATP, Mg, Mn, substrate peptide, etc. when the kinase is in the active conformation. Similarly, the global mode NMS fluctuations of the inactive conformations when bound (*blue* curve) and unbound (*red* curve) to small molecules is plotted in Fig. 7*B*. Although we observe lower fluctuations of the Inactive-Holo (*blue curve*) conformations when compared to the Inactive-Apo (*red curve*) forms, the magnitude of difference seen is not large enough to completely explain the entire fluctuation difference seen between the active and inactive conformations. In Fig. 8, the quantitative areas under the curves of different forms are plotted. It can be seen that the fluctuations of the Active-Holo and Active-Apo conformations are very similar and the difference between their areas under the curve is negligible (0.35 units square). Similarly, the fluctuations of the Inactive-Holo and Inactive-Apo conformations are very similar to each other, giving rise to a mean difference of 4.6 units square. This shows that the holo forms are indeed more stable and less dynamic than the corresponding apo forms. However, these differences do not fully account for the high difference seen between the active and inactive states. The distribution of areas under the curve of the Active-Apo and the Inactive-Holo are statistically different from each other (two-tailed unpaired T-test, p-value < 0.01), demonstrating that the bound factors alone cannot explain the entire difference in the fluctuation magnitude seen between the active and inactive forms. Further, we also verified that this result holds true in the case of all-atom GNM based global normal mode for an

example STY kinase PKA (Fig. S3)

In addition, if the presence of bound small molecules were to cause the lower fluctuations in the active state, then there should be a systematic presence and absence of bound small molecules between the active and inactive structures. To this end, we analysed whether the following factors could systematically explain the functional state (active or inactive) of the kinase structure: (i) presence or absence of ATP/ATP analog, (ii) presence or absence of substrate peptide, (iii) presence or absence of mutations in the kinase catalytic domain, (v) presence or absence of cations, like Mg or Mn, near the DFG loop, (v) presence or absence of phosphorylated residues in the kinase catalytic domain and (vi) kinase type. We performed a multi-factor Analysis of Variance (ANOVA, see Table S2 for the factors matrix) to explain the functional state of the structure. We found no main effect of any of the factors (see Table IV, $p > 0.286$ for all factors), thus ascertaining that there is no systematic binding of small molecules between the active and inactive state structures.

Taken together, our analyses suggest that although the binding of small molecules reduces the flexibility of the kinase, it cannot entirely explain the fluctuation difference seen between the active and inactive forms. Therefore, the higher fluctuations of the inactive states compared to the active states seem to be intrinsic features of those structural forms.

Functional state can be predicted using fluctuations alone

Can we reliably predict the functional state of a structure (active or inactive) using only the global mode NMS fluctuations? In this regard, a linear classifier was trained and tested on the data set in a leave-one-out method (see Methods). The input for the training was the global mode NMS fluctuations of specific segments, viz., αC-helix, αG-helix and activation loop. The classifier could predict significantly better than random (Chi-Square-test, *p*-value <

444 0.05). All the three structural segments showed reliable prediction accuracy as compared to
445 that of random prediction (Fig. 9). This further demonstrates that the global mode of the STY
446 kinase structures contains sufficient information about the functional conformation of the
447 protein kinase.

448

DISCUSSION

Despite the knowledge of a large number of crystal structures of STY kinases, we do not completely understand how and why only certain conformations enable the catalytic activity of the kinase. Detailed structural and domain motion[53,54] analyses on the different active and inactive conformations were carried out (See Fig. S4-S8), but no specific differentiation could be attributable to the functional regions unless fluctuations are considered. Additionally, kinases can be regulated by the binding of other proteins during specific functional states[55]. Using GNM based NMA, we have, for the first time, identified and quantified the difference in fluctuations between the functional states of STY kinases. Our results suggest that although the fundamental modes of motion are conserved (high overlap values between the global modes, indicated in Table II) during different functional states, residues tend to fluctuate more when in the inactive states than in the active states. This can have several implications in the function of kinases. First, relative increase in vibrational dynamics of the kinase can directly affect its binding to different partners. This can regulate binding of the kinase to a substrate or a regulatory protein. Some regions in the STY kinase catalytic domain have been shown to bind to regulatory proteins that directly inhibit/regulate substrate binding[56]. Binding to differential partners in the cellular environment can be favoured or impeded by the differential dynamics of the structure. Second, it is known that local packing density around a residue in a given structure affects the vibrational fluctuations of the residue perceived through GNM[57]. The fluctuation analyses in the current study suggest that the overall kinase structure could become more compact during the transition to an active state. This can result in potential increase in the local packing densities around the functionally sensitive residues in the active form. This is in corroboration with previous structural studies that have documented the closure of the C-lobe and N-lobe during the inactive-to-active switch[24]. Third, it is well known that although the STY kinases adopt a very

similar active conformation, the inactive conformations are varied[24]. This observation can be explained in the light of the findings of the current study. Accentuated fluctuations in the inactive conformations can, in principle, provide the means for the kinase to move through the conformational space and sample many conformations. On the other hand, lower fluctuation amplitudes would keep the active structures relatively stable. This is in agreement with a previous study[58] that investigated the structural fluctuations of wild-type and mutant RET and MET receptor tyrosine kinases. Upon conducting structural and simulation studies, it was reported that the cancer mutants and the inactive conformations of the kinases have higher local mobility at functionally sensitive regions and sampled larger conformation space, in comparison to the wild-type active conformations. It was concluded that the inactive forms of the RET and MET kinases enjoyed much lesser thermodynamic stability than the active forms.

We have identified specific regions within the kinase structure that show high difference in fluctuations between the functional states. Some of these regions, αC-helix and activation loop, were implicated to be crucial in the conformational transition between active and inactive states[23,24]. Other regions, like αG-helix and GH-loop, do not show structural deviations during the switch. Nevertheless, these are important regions of regulation that are involved in protein-protein interactions, crucial to the activity of the kinase[43–46,59]. Effect of mutations in αC-helix[60,61], mutations and phosphorylation in activation loop[62–64], mutations and binding partners αG-helix[45,46,65,66] on the activity of the kinase are well documented[67]. In addition, binding of a small protein/molecule in these regions could bring about change in the flexibility of the kinase irrespective of its active or inactive conformation. This further stresses the relationship between protein-protein interactions and functional switch of the kinase, through alteration of dynamics. Thus, GNM based global modes are sensitive to the functional and regulatory aspects of the kinase, although such differences cannot be

understood from the structural analyses alone. We have also analysed the NMS fluctuations of inactive STY kinases solved in different conformations. Our study shows that the "DFG-in" conformations of the inactive kinases could be as stable as the active conformations. However, other conformational variations do not affect the flexibility of the inactive conformation significantly.

Reliable prediction of the functional state of the kinase based on the fluctuations in specific regions further supports the sensitivity and information in the GNM modes, indicating that the Cα spatial positions contain enough information to infer structural fluctuations specific to functional states.

METHODS

Gaussian Network Model based Normal Mode Analysis

In the Elastic Network (EN) based Normal Mode Analysis (NMA), a protein molecule is represented as a 3-dimensional mass-spring system with the masses/nodes at the Cα atoms and the springs connecting the proximate Cα nodes. This is a simplified model of Cα-Cα virtually bonded and non-bonded interactions between the proximate residues, modelled as harmonic potentials. Based purely on the topology of constraints imposed by the springs, which is prescribed by the 3-D structure, ENM analytically determines the vibrational modes around the equilibrium state[68]. In Gaussian Network Model (GNM), a simplification of ENM, isotropic Gaussian-distributed vibrations are calculated using an identical spring constant γ for all the interactions defined by a cut-off distance for proximity[36]. GNM determines the relative amplitudes of fluctuations of each of the nodes or Cα atoms without any directionality. It should be noted that GNM assumes the native conformation of the protein to be in an equilibrium state, and samples only the microstates around this folded state at equilibrium. It has been extensively verified that despite the simplicity, the fluctuations perceived from GNM correlate with experimentally determined crystallographic temperature factors[36] and functional motions of proteins[69].

Positional coordinates of the Cα atoms in the backbone of the kinase catalytic domain of the STY kinase structures were alone used in the GNM based Normal Modes calculations. If any ligands or substrates or inhibitors are bound at the kinase domain, they were deleted before using the structure in calculations. All pairs of Cα atoms that lie within a cut-off distance r_c were to be connected to each other. A 7Å r_c cut-off was used for all the analyses throughout, as determined to be the optimal interaction distance by previous studies[70,71]. Nevertheless, we verified that the results reported in the study are not dependent on a specific cut-off value.

Using a range of cut-off values, we were able to reproduce similar trends and results (See Fig. S9).

For each structure, a matrix that defines the inter-Cα connections, called the Kirchhoff connectivity matrix **Γ**, was formulated as follows:

For any node i, let $\mathbf{R}_i^0$ be the equilibrium position vector and $\mathbf{R}_i$ be the instantaneous position vector. Deformation from the mean position or fluctuation at node i is given by the vector $\Delta\mathbf{R}_i$:

$$\Delta\mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0$$

Similarly, if $\mathbf{R}_{ij}$ is the distance vector from node i to j, then the fluctuation in the distance vector $\mathbf{R}_{ij}$ is given by:

$$\Delta\mathbf{R}_{ij} = \mathbf{R}_{ij} - \mathbf{R}_{ij}^0 = \Delta\mathbf{R}_j - \Delta\mathbf{R}_i$$

In the Kirchhoff matrix, the element in i^{th} row and j^{th} column, Γ_{ij}, is given by:

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0, & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{j,j\neq i} \Gamma_{ij}, & \text{if } i = j \end{cases}$$

Once the network topology of linear springs and nodes is constructed in the form of Kirchhoff matrix, the normal modes are analytically solved as described below.

If the distribution of the distance vector fluctuations $\Delta\mathbf{R}_{ij}$ is isotropic and Gaussian in *X*, *Y* and *Z* directions, the network potential of *N* residues can be calculated using the components of $\Delta\mathbf{R}_i$: ΔX_i, ΔY_i and ΔZ_i, where (X_i, Y_i, Z_i) is the position vector of the i^{th} Cα atom.

$$V_{GNM} = \frac{\gamma}{2}\left[\sum_{i,j}^{N} \Gamma_{ij}\left[\left(\Delta X_i - \Delta X_j\right)^2 + \left(\Delta Y_i - \Delta Y_j\right)^2 + \left(\Delta Z_i - \Delta Z_j\right)^2\right]\right]$$

where γ is the uniform spring constant for all the interactions defined by the Kirchhoff matrix.

If $\Delta\mathbf{X}^T$, $\Delta\mathbf{Y}^T$ and $\Delta\mathbf{Z}^T$ are $[\Delta X_1, \Delta X_2, \ldots, \Delta X_N]$, $[\Delta Y_1, \Delta Y_2, \ldots, \Delta Y_N]$ and $[\Delta Z_1, \Delta Z_2, \ldots, \Delta Z_N]$ respectively, then

$$V_{GNM} = \frac{\gamma}{2}\left[\Delta\mathbf{X}^T\mathbf{\Gamma}\Delta\mathbf{X} + \Delta\mathbf{Y}^T\mathbf{\Gamma}\Delta\mathbf{Y} + \Delta\mathbf{Z}^T\mathbf{\Gamma}\Delta\mathbf{Z}\right]$$

$$V_{GNM} = \frac{\gamma}{2}\left[\Delta\mathbf{R}^T\left(\mathbf{\Gamma} \otimes \mathbf{E}\right)\Delta\mathbf{R}\right]$$

where $\Delta\mathbf{R}^T$ is $[\Delta X_1, \Delta Y_2, \ldots, \Delta Z_N]$, $\Delta\mathbf{R}$ is its transpose, a column vector of 3N dimension and $\mathbf{E}$ is a 3 x 3 identity matrix. Since the vibrations are assumed to be isotropic, the Eigen values of the $\mathbf{\Gamma} \otimes \mathbf{E}$ matrix of order 3N are threefold degenerate.

The normal modes, or the modes of motion theoretically available to the protein structure, are obtained by transforming the Kirchhoff matrix $\mathbf{\Gamma}$ into the product of 3 matrices: (i) unitary matrix $\mathbf{U}$ of Eigen vectors $\mathbf{u}_i$ of $\mathbf{\Gamma}$, (ii) diagonal matrix $\mathbf{\Lambda}$ of Eigen values λ_i of $\mathbf{\Gamma}$ and (iii) $\mathbf{U}^T$

$$\mathbf{\Gamma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$$

Here, the Eigen value λ_i denotes the frequency of oscillation of the i^{th} mode; the Eigen vector $\mathbf{u}_i$ denotes the shape of the i^{th} mode. There are N-1 modes in all, excluding the first mode that corresponds to the translation motion of the entire protein molecule. The mode with the lowest frequency, involving the slowest motion, is known to have the highest degree of collectivity, in that they involve changes distributed over a large number of residues. These,

563 in general, correspond to conformational changes or biologically relevant motions in the
564 protein. All the analyses in this study have been done using the first slowest mode (global
565 mode) of the protein after verification of correspondence of the global modes of the active
566 and inactive conformations (overlap values in Table II). Additionally, a few of the low
567 frequency modes of the inactive STY conformations were also cumulatively assessed for
568 overlap with the global mode of the active conformation (Table II).

569 The mean-square fluctuations of a residue or node *i* is given by

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle = \langle \Delta X_i^2 \rangle + \langle \Delta Y_i^2 \rangle + \langle \Delta Z_i^2 \rangle$$

570 It can be verified[72] that the expected value of the mean-square residue fluctuations can be
571 determined from the inverse of Kirchhoff matrix or a sum of contributions of *N*-1 modes
572 $2 \leq k \leq N$:

$$\langle \Delta \mathbf{R}_i^2 \rangle = \frac{3k_B T}{\gamma} [\mathbf{\Gamma}^{-1}]_{ii} = \frac{3k_B T}{\gamma} [\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T]_{ii}$$

$$= \frac{3k_B T}{\gamma} \sum_k [\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T]_{ii}$$

573 where k_B is the Boltzmann constant and T is the absolute temperature. The fluctuations of
574 residues due to the k^{th} mode is given by:

$$\langle \Delta \mathbf{R}_i^2 \rangle = \frac{3k_B T}{\gamma} \left[\frac{\mathbf{u}_k \mathbf{u}_k^T}{\lambda_k} \right]_{ii}$$

575 For every structure, mean-square fluctuations of each of the residues (in Å^2) in the k^{th} mode
576 were calculated and normalised by division with the mean-square fluctuation of the most
577 dynamic residue in the same mode of the same structure. This resulted in Normalised Mean-

578 Square (NMS) fluctuations, with values utmost 1, of residues for each structure. For example,
579 the fluctuation of the i^{th} residue in the k^{th} mode is normalised as:

$$\text{NMS fluctuations} = \langle \Delta \mathbf{R}_i^2 \rangle_{normalised} = \frac{\langle \Delta \mathbf{R}_i^2 \rangle_k}{\max \{ \langle \Delta \mathbf{R}_i^2 \rangle_k \}}$$

580 Fluctuations of the residues missing in the structures, along with 2 sequentially preceding and
581 2 sequentially succeeding residues, were not considered for analysis. This is to extrude any
582 spurious interaction or topology constraint arising due to the missing residues.

583 **Normalisation of B-factors and validation**

584 For every structure, B-factors of each of the residues (in $Å^2$) were extracted from the PDB
585 file and normalised by division with the B-factor of the most dynamic residue (maximum B-
586 factor) in the same structure. This resulted in Normalised B-factors, with values utmost 1, of
587 residues for each structure.

588 Residue-wise NMS fluctuations calculated from the global mode of the protein structures
589 were compared against the Normalised B-factors of equivalent residues of the same structure.
590 As expected, we observe a significant correlation of global mode NMS fluctuations with the
591 experimentally determined B-factors (See Fig. S10).

592 **Mean Fluctuation difference between the active and inactive forms**

593 For each pair of STY kinases, global mode NMS Fluctuations of residues in the active state
594 were subtracted from the corresponding NMS fluctuations in the inactive state. Mean
595 fluctuation difference was calculated by averaging the difference across the 11 pairs of STY
596 kinases:

$$\text{Mean Fluctuation Difference} = \langle NMS\_Fluct_i^{inactive} - NMS\_Fluct_i^{active} \rangle$$

Inter-chain interactions count from iPfam

So far, the calculations for GNM based NMA were carried out using only the Cα atomic positions of the kinase domain. The fluctuations perceived from such an analysis were validated in the light of all the known inter-chain interactions effected by the residues. To this end, we collated all the kinase structures solved in complex with interacting proteins from iPfam. From this, Kinase-inhibitor peptide complexes and intra-chain interactions were filtered out. In the remaining 476 structures, we considered the number of inter-chain interactions at each residue. The sequences of the 476 structures were aligned to find the topologically equivalent sets of residues using ClustalW2[73]. Finally, at each topologically equivalent residue position, the summation of the number of inter-chain interactions was calculated and mapped on to the kinase fold.

Classifier Analysis

A linear classifier was trained and tested on a data set of 11 pairs of STY kinases using leave-one-out method (training set contained 21 observations and test set contained the remaining one observation; such training and testing were carried out for each of the 22 kinases in the data set). GNM perceived global mode NMS Fluctuations of specific regions were used for training and testing. The accuracy with which the functionality (active/inactive) of the test observation was predicted by the classifier was calculated as an average across the 22 structures.

CONCLUSIONS

In summary, we have analysed the global vibrational mode of the active and inactive conformations of STY kinases. We establish that there are systematic differences between the two forms in terms of relative magnitudes of flexibility. We have found significantly high flexibility of residues in all except "DFG-in" inactive conformations, and have also pointed out that such differences are contributed by the activation loop, αG-helix and the αC-helix which are involved in structural alterations between active and inactive states or in the regulation. Binding of small molecules or proteins to the kinase reduces the flexibility of the kinase irrespective of its functional state.

627 **ACKNOWLEDGEMENTS**

628 The authors would like to thank Dr. R.M.Bhaskara and other members of the group for useful
629 inputs and suggestions.

630 This research is supported by Indo-French Collaborative Grant (CEFPRA), Mathematical
631 Biology programme sponsored by Department of Science and Technology and Department of
632 Biotechnology, Government of India. RK is supported by a fellowship from University
633 Grants Commission, Government of India.

634

REFERENCES

1. S. B. Ficarro, M. L. McCleland, P. T. Stukenberg, D. J. Burke, M. M. Ross, J. Shabanowitz, D. F. Hunt and F. M. White, *Nat. Biotechnol.*, 2002, **20**, 301–5.

2. J. D. Corbin, I. V Turko, A. Beasley and S. H. Francis, *Eur. J. Biochem.*, 2000, **267**, 2760–7.

3. W. H. Biggs, J. Meisenhelder, T. Hunter, W. K. Cavenee and K. C. Arden, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 7421–6.

4. B. P. Bullock and J. F. Habener, *Biochemistry*, 1998, **37**, 3795–809.

5. F. Hans and S. Dimitrov, *Oncogene*, 2001, **20**, 3021–7.

6. S. E. Hernández, M. Krishnaswami, A. L. Miller and A. J. Koleske, *Trends Cell Biol.*, 2004, **14**, 36–44.

7. R. Acin-Perez, D. L. Gatti, Y. Bai and G. Manfredi, *Cell Metab.*, 2011, **13**, 712–9.

8. M. M. Dix, G. M. Simon, C. Wang, E. Okerberg, M. P. Patricelli and B. F. Cravatt, *Cell*, 2012, **150**, 426–40.

9. T. Pawson and J. D. Scott, *Trends Biochem. Sci.*, 2005, **30**, 286–90.

10. G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, *Science*, 2002, **298**, 1912–34.

11. S. Hanks, *Genome Biol.*, 2003, **4**, 111.

12. C. Kim, C. Y. Cheng, S. A. Saldanha and S. S. Taylor, *Cell*, 2007, **130**, 1032–43.

654 13. S. M. Gisler, S. Pribanic, D. Bacic, P. Forrer, A. Gantenbein, L. A. Sabourin, A. Tsuji,
655 Z.-S. Zhao, E. Manser, J. Biber and H. Murer, *Kidney Int.*, 2003, **64**, 1733–45.

656 14. C. T. Eggers, J. C. Schafer, J. R. Goldenring and S. S. Taylor, *J. Biol. Chem.*, 2009,
657 **284**, 32869–80.

658 15. L. R. Masterson, A. Mascioni, N. J. Traaseth, S. S. Taylor and G. Veglia, *Proc. Natl.*
659 *Acad. Sci. U. S. A.*, 2008, **105**, 506–11.

660 16. F. S. Kinderman, C. Kim, S. von Daake, Y. Ma, B. Q. Pham, G. Spraggon, N.-H.
661 Xuong, P. A. Jennings and S. S. Taylor, *Mol. Cell*, 2006, **24**, 397–408.

662 17. A. J. Boettcher, J. Wu, C. Kim, J. Yang, J. Bruystens, N. Cheung, J. K. Pennypacker,
663 D. A. Blumenthal, A. P. Kornev and S. S. Taylor, *Structure*, 2011, **19**, 265–76.

664 18. I. Shchemelinin, L. Sefc and E. Necas, *Folia Biol. (Praha).*, 2006, **52**, 81–100.

665 19. H. Patterson, R. Nibbs, I. McInnes and S. Siebert, *Clin. Exp. Immunol.*, 2014, **176**, 1–
666 10.

667 20. S. K. Hanks, A. M. Quinn and T. Hunter, *Science*, 1988, **241**, 42–52.

668 21. S. K. Hanks and T. Hunter, *FASEB J.*, 1995, **9**, 576–96.

669 22. L. N. Johnson, M. E. M. Noble and D. J. Owen, 1996, **85**, 149–158.

670 23. B. Nolen, S. Taylor and G. Ghosh, *Mol. Cell*, 2004, **15**, 661–75.

671 24. M. Huse and J. Kuriyan, *Cell*, 2002, **109**, 275–82.

672 25. W. Wong and N. R. Gough, *Sci. Signal.*, 2009, **2**, eg4.

673 26. R. G. Smock and L. M. Gierasch, *Science*, 2009, **324**, 198–203.

674 27. A. J. Engler, P. O. Humbert, B. Wehrle-Haller and V. M. Weaver, *Science*, 2009, **324**,
675 208–12.

676 28. N. Tokuriki and D. S. Tawfik, *Science*, 2009, **324**, 203–7.

677 29. R. L. Jernigan, M. C. Demirel and I. Bahar, *Int. J. Quantum Chem.*, 1999, **75**, 301–
678 312.

679 30. I. Bahar, *Rev. Chem. Eng.*, 1999, **15**, 319–348.

680 31. I. Bahar, B. Erman, R. L. Jernigan, A. R. Atilgan and D. G. Covell, *J. Mol. Biol.*, 1999,
681 **285**, 1023–37.

682 32. M. Tirion, *Phys. Rev. Lett.*, 1996, **77**, 1905–1908.

683 33. K. Hinsen, *Proteins*, 1998, **33**, 417–29.

684 34. I. Bahar and A. J. Rader, *Curr. Opin. Struct. Biol.*, 2005, **15**, 586–92.

685 35. I. Bahar and R. L. Jernigan, *J. Mol. Biol.*, 1998, **281**, 871–84.

686 36. I. Bahar, A. R. Atilgan and B. Erman, *Fold. Des.*, 1997, **2**, 173–81.

687 37. C. Notredame, D. G. Higgins and J. Heringa, *J. Mol. Biol.*, 2000, **302**, 205–17.

688 38. R. Guerois, J. E. Nielsen and L. Serrano, *J. Mol. Biol.*, 2002, **320**, 369–87.

689 39. *Coarse-Graining of Condensed Phase and Biomolecular Systems*, CRC Press, 2008.

690 40. H. Huang, R. Zhao, B. M. Dickson, R. D. Skeel and C. B. Post, *J. Phys. Chem. B*,
691 2012, **116**, 4465–75.

692 41. *Handbook of Cell Signaling, Three-Volume Set*, Academic Press, 2003.

693 42. L. Palmieri and G. Rastelli, *Drug Discov. Today*, 2013, **18**, 407–14.

694 43. R. Scholz, M. Suter, T. Weimann, C. Polge, P. V Konarev, R. F. Thali, R. D. Tuerk, B.
695 Viollet, T. Wallimann, U. Schlattner and D. Neumann, *J. Biol. Chem.*, 2009, **284**,
696 27425–37.

697 44. A. C. Dar, T. E. Dever and F. Sicheri, *Cell*, 2005, **122**, 887–900.

698 45. S. S. Taylor, N. M. Haste and G. Ghosh, *Cell*, 2005, **122**, 823–5.

699 46. *Quantitative Biology: From Molecular to Cellular Systems*, CRC Press, 2012.

700 47. R. D. Finn, B. L. Miller, J. Clements and A. Bateman, *Nucleic Acids Res.*, 2014, **42**,
701 D364–73.

702 48. S. M. Hanson, D. J. Francis, S. A. Vishnivetskiy, E. A. Kolobova, W. L. Hubbell, C. S.
703 Klug and V. V Gurevich, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 4900–5.

704 49. R. Grünberg, M. Nilges and J. Leckner, *Structure*, 2006, **14**, 683–93.

705 50. G. R. Smith, M. J. E. Sternberg and P. A. Bates, *J. Mol. Biol.*, 2005, **347**, 1077–101.

706 51. R. S. Depetris, J. Hu, I. Gimpelevich, L. J. Holt, R. J. Daly and S. R. Hubbard, *Mol.*
707 *Cell*, 2005, **20**, 325–33.

52. *From Molecules to Medicines: Structure of Biological Macromolecules and Its Relevance in Combating New Diseases and Bioterrorism*, Springer Science & Business Media, 2009.

53. S. Hayward, A. Kitao and H. J. Berendsen, *Proteins*, 1997, **27**, 425–37.

54. G. P. Poornam, A. Matsumoto, H. Ishida and S. Hayward, *Proteins*, 2009, **76**, 201–12.

55. B. Okram, A. Nagle, F. J. Adrián, C. Lee, P. Ren, X. Wang, T. Sim, Y. Xie, X. Wang, G. Xia, G. Spraggon, M. Warmuth, Y. Liu and N. S. Gray, *Chem. Biol.*, 2006, **13**, 779–86.

56. Y. Bourne, M. H. Watson, M. J. Hickey, W. Holmes, W. Rocque, S. I. Reed and J. A. Tainer, *Cell*, 1996, **84**, 863–74.

57. B. Halle, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 1274–9.

58. A. Dixit, A. Torkamani, N. J. Schork and G. Verkhivker, *Biophys. J.*, 2009, **96**, 858–74.

59. S. J. Deminoff, V. Ramachandran and P. K. Herman, *Genetics*, 2009, **182**, 529–39.

60. A. P. Kornev, S. S. Taylor and L. F. Ten Eyck, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 14377–82.

61. G. H. Iyer, M. J. Moore and S. S. Taylor, *J. Biol. Chem.*, 2005, **280**, 8800–7.

62. J. M. Steichen, G. H. Iyer, S. Li, S. A. Saldanha, M. S. Deal, V. L. Woods and S. S. Taylor, *J. Biol. Chem.*, 2010, **285**, 3825–32.

63. M. Cao, D. Meng, L. Wang, S. Bei, W. J. Snell and J. Pan, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 12337–42.

64. J. A. Adams, *Biochemistry*, 2003, **42**, 601–7.

65. L. R. Pearce, D. Komander and D. R. Alessi, *Nat. Rev. Mol. Cell Biol.*, 2010, **11**, 9–22.

66. V. Gelev, H. Aktas, A. Marintchev, T. Ito, D. Frueh, M. Hemond, D. Rovnyak, M. Debus, S. Hyberts, A. Usheva, J. Halperin and G. Wagner, *J. Mol. Biol.*, 2006, **364**, 352–63.

67. R. Bayliss, A. Fry, T. Haq and S. Yeoh, *Open Biol.*, 2012, **2**, 120136.

68. T. Haliloglu, I. Bahar and B. Erman, *Phys. Rev. Lett.*, 1997, **79**, 3090–3093.

69. I. Bahar, A. Atilgan, M. Demirel and B. Erman, *Phys. Rev. Lett.*, 1998, **80**, 2733–2736.

70. I. Bahar and R. L. Jernigan, *J. Mol. Biol.*, 1997, **266**, 195–214.

71. S. Miyazawa and R. L. Jernigan, *Macromolecules*, 1985, **18**, 534–552.

72. A. Rader, C. Chennubhotla, L.-W. Yang and I. Bahar, *The Gaussian Network Model: Theory and Applications, in "Normal Mode Analysis. Theory and Applications to Biological and Chemical Systems,"* Chapman & Hall / CRC Mathematical and Computational Biology Series, CRC Press, Taylor & Francis Group, 2006.

73. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, *Bioinformatics*, 2007, **23**, 2947–8.

74. Y. Zhang and J. Skolnick, *Nucleic Acids Res.*, 2005, **33**, 2302–9.

747

Table I Data set of 14 STY kinases, in the active and inactive forms, used in the present study. UniProt ID and classification of the STY kinases are listed along with the identities of the used structures (in Protein Data Bank - PDB ID) in the active and inactive conformations. Span of ATOM residues (numbered according to the data deposited in PDB structures) comprising the kinase catalytic domain and their crystallographic resolution (Res) are also tabulated. Sequence identity is calculated based on the alignment of kinase domain sequences extracted from the PDB files using T-Coffee[37]. Alignments of all possible active-inactive pairs (e.g., for entry no. 2 of UniProt ID P00519 in the table, each of the 4 active sequences were aligned with each of the 3 inactive sequences, giving a total of 12 alignment scores) were done and the mean alignment score is noted. Although the active and inactive forms of the same STY kinase are compared with each other, score of <1 may frequently result from point mutations in one of the structures and/or missing residues in one/both of the structures. Likewise, mean Root Mean Square Deviation (RMSD) is calculated from all possible pair wise structural superpositions between the active and inactive conformations using TM-align[74].

no.	UniProt ID and classification	Functional conformation	PDB ID	Kinase domain ATOM residues	Res (Å)	Mean Seq alignment score	Mean RMSD (Å)
1	O14965 PK : Ser/Thr protein kinase : Aurora	Active	1OL5_A	133-383	2.5	1.0	0.90
		Inactive	1OL7_A	133-383	2.75		
2	P00519 PK : Tyr protein kinase : ABL	Active	1OPL_A 2F4J_A 2G2I_A 2GQG_A	261-512 242-493 242-493 242-493	3.42 1.91 3.12 2.4	0.999	1.62
		Inactive	2HYY_A 2FO0_A 2G1T_A	242-493 261-512 242-493	2.4 2.27 1.8		
3	P00533 PK : Tyr protein kinase : EGF receptor	Active	2GS2_A 1M14_A 2ITP_A	688-955 688-955 712-979	2.8 2.6 2.74	0.996	2.37
		Inactive	3GOP_A	688-955	2.8		

			3GT8_A 2GS7_A 1XKK_A	688-955 688-955 712-979	2.95 2.6 2.4		
4	P05132 PK : AGC Ser/Thr protein kinase : cAMP	Active	1JBP_E 1APM_E 1L3R_E 1RDQ_E 2ERZ_E 2CPK_E 1FMO_E 1ATP_E	43-297 43-297 43-297 43-297 43-297 43-297 43-297 43-297	2.2 2.0 2 1.26 2.2 2.7 2.2 2.2	0.999	2.27
		Inactive	2QVS_E 4DFY_A 1SYK_A	43-297 43-297 43-297	2.5 3 2.8		
5	P06213 PK : Tyr protein kinase : Insulin receptor	Active	1IR3_A	996-1271	1.9	0.997	2.65
		Inactive	1IRK_A	996-1271	2.1		
6	P08069 PK : Tyr protein kinase : Insulin receptor	Active	1K3A_A	969-1244	2.1	0.998	1.3
		Inactive	1M7N_A	996-1271	2.7		
7	P10721 PK : Tyr protein kinase : CSF-1/PDGF receptor	Active	1PKG_A	589-927	2.9	0.998	2.57
		Inactive	1T45_A	589-935	1.9		
8	P12931 PK : Tyr protein kinase : SRC	Active	1Y57_A	267-520	1.91	0.998	2.11
		Inactive	2SRC_A 1FMK_A	267-520 267-520	1.5 1.5		
9	P24941 PK : CMGC Ser/Thr protein kinase : CDC2/CDKX	Active	1W98_A 1JST_A 1QMZ_A 1FIN_A	4-286 4-286 4-286 4-286	2.15 2.6 2.2 2.3	0.999	2.42
		Inactive	1HCL_A 1HCK_A	4-286 4-286	1.8 1.9		
10	P47811 PK : CMGC Ser/Thr protein kinase : MAP kinase	Active	3PY3_A	24-308	2.1	1.0	1.54
		Inactive	1P38_A	24-308	2.1		
11	P49137 PK : CAMK Ser/Thr protein kinase	Active	1NXK_A	64-325	2.7	0.996	1.78
		Inactive	1NY3_A 1KWP_A	64-325 64-325	3 2.8		
12	P63086 PK : CMGC Ser/Thr protein kinase : MAP kinase	Active	2ERK_A	23-311	2.4	1.0	1.51
		Inactive	1ERK_A	23-311	2.3		
13	Q13153 PK : STE Ser/Thr protein kinase :	Active	1YHV_A 1YHW_A 3Q52_A	270-521 270-521 270-521	1.8 1.8 1.8	0.998	2.19

	STE20	Inactive	1F3M_C	270-521	2.3		
14	P00523 PK : Tyr protein kinase : SRC	Active	3DQW_A	267-520	2.02	0.998	1.32
		Inactive	2OIQ_A	267-520	2.07		

748

749

Table II Data set of 24 pairs of active-inactive conformations of kinases considered for comparison. From the 14 unique STY kinases listed in Table I, one active conformation was chosen from each entry and compared with each of the inactive conformations. The active-inactive conformation pairs used for quantitative comparison analyses are listed. In order to assess the similarity between the global mode of the active conformation and the low modes of the inactive conformations, the overlap between the modes was calculated. Overlap between the global active mode and global inactive mode is indicated in column 6. Cumulative overlaps of the lowest 5 inactive modes with the global active mode are also tabulated. High overlap values close to 1 suggest that the global mode of the inactive conformations corresponds to the global mode of the active conformations.

STY kinase no. (from Table I)	UniProt/ Swiss-Prot ID	s.no.	PDB_ID Active conformation	PDB_ID Inactive conformation	Overlap between the global modes	Cumulative overlap of inactive modes with the global active mode			
						Lowest 2	Lowest 3	Lowest 4	Lowest 5
1	O14965	1	1OL5_A	1OL7_A	0.993	0.996	0.997	0.997	0.998
2	P00519	2	2G2I_A	2FO0_A	0.927	0.927	0.927	0.928	0.928
		3	2G2I_A	2G1T_A	0.868	0.908	0.908	0.911	0.912
		4	2G2I_A	2HYY_A	0.867	0.870	0.872	0.894	0.894
3	P00533	5	1M14_A	3GOP_A	0.661	0.662	0.663	0.678	0.679
		6	1M14_A	3GT8_A	0.969	0.969	0.970	0.970	0.970
		7	1M14_A	2GS7_A	0.939	0.944	0.960	0.970	0.971
		8	1M14_A	1XKK_A	0.776	0.789	0.789	0.790	0.790
4	P05132	9	1ATP_E	4DFY_A	0.957	0.958	0.970	0.980	0.982
		10	1ATP_E	1SYK_A	0.997	0.997	0.997	0.998	0.998
		11	1ATP_E	2QVS_E	0.997	0.998	0.998	0.998	0.998
5	P06213	12	1IR3_A	1IRK_A	0.963	0.964	0.967	0.971	0.973
6	P08069	13	1K3A_A	1M7N_A	0.977	0.977	0.982	0.984	0.987
7	P10721	14	1PKG_A	1T45_A	0.943	0.947	0.973	0.973	0.975
8	P12931	15	1Y57_A	1FMK_A	0.959	0.960	0.963	0.964	0.964
		16	1Y57_A	2SRC_A	0.932	0.944	0.961	0.962	0.973
9	P24941	17	1JST_A	1HCL_A	0.897	0.897	0.915	0.916	0.936
		18	1JST_A	1HCK_A	0.885	0.886	0.900	0.900	0.925
10	P47811	19	3PY3_A	1P38_A	0.993	0.995	0.995	0.995	0.995
11	P49137	20	1NXK_	1NY3_A	0.757	0.798	0.806	0.807	0.808

			A						
		21	1NXK_A	1KWP_A	0.852	0.855	0.856	0.856	0.856
12	P63086	22	2ERK_A	1ERK_A	0.998	0.998	0.998	0.998	0.999
13	Q13153	23	1YHV_A	1F3M_C	0.952	0.964	0.968	0.973	0.974
14	P00523	24	3DQW_A	2OIQ_A	0.962	0.964	0.971	0.982	0.982

750

751

Table III Conformational details of the inactive STY kinase structures. STY kinases, when inactive, assume a variety of backbone conformations. The aspartate in the DFG motif could assume one of the following conformations: “in”, “flip”, “Src-like inactive” and “out”. αC-helix could be in the “in” position, thus stabilising a salt bridge interaction with a conserved lysine in the N lobe, or in the swung “out” conformation. The orientation of the N and C lobes could be “closed” or “open” with respect to each other. The orientation of N-C lobes and the conformations of DFG motif and αC-helix of the inactive structures used in the study are listed.

STY kinase no.	PDB ID	DFG conformation	αC-helix conformation	N and C lobe orientation
1	1OL7_A	In	In	Closed
2	2HYY_A	Out	In	Open
	2FO0_A	Flip	In	Closed
	2G1T_A	Src-like inactive	Out	Open
3	3GOP_A	Src-like inactive	Out	Open
	3GT8_A	Src-like inactive	Out	Open
	2GS7_A	Src-like inactive	Out	Open
	1XKK_A	Src-like inactive	Out	Open
4	2QVS_E	In	In	Closed
	4DFY_A	Src-like inactive	In	Open
	1SYK_A	In	In	Closed
5	1IRK_A	Out	Out	Open
6	1M7N_A	Flip	Out	Open
7	1T45_A	Out	In	Open
8	2SRC_A	Src-like inactive	Out	Open
	1FMK_A	Src-like inactive	Out	Open
9	1HCL_A	Src-like inactive	Out	Open
	1HCK_A	Src-like inactive	Out	Open
10	1P38_A	Src-like inactive	In	Open
11	1NY3_A	In	Out	Closed
	1KWP_A	In	Out	Closed
12	1ERK_A	In	In	Closed
13	1F3M_C	Src-like inactive	Out	Open
14	2OIQ_A	Out	In	Open

Table IV Multi-factor Analysis of Variance (ANOVA) testing different factors that might vary systematically between the active and inactive forms, and thus contribute to the fluctuation difference between the forms. The results for testing the main effect of different factors on the functional state of the kinase (active/inactive) are tabulated. Sum of squares describes the variance in the kinase functional state contributed by variance in the corresponding factors. F statistic score is the ratio of variance observed between the factor groups to variance observed within the factor groups. For a confidence of ≥95%, if the F score is lesser than the critical F-value, then the corresponding factor plays a significant role in the functional state of the kinase. All p-values ≥0.29 indicate that none of the factors tested has a main effect on whether a kinase is in the active or inactive form.

Source factor	**Sum of squares**	**Degrees of freedom**	**Mean-square**	**F statistic score**	**Prob >F**
ATP binding site occupied	0.04	1	0.04	0.1	0.77
Substrate peptide bound factor	0	1	0	0	1
Mutation in the kinase domain factor	0.25	1	0.25	0.6	0.47
Cation binding near DFG-loop factor	0	1	0	0	1
Phosphorylation in the kinase domain factor	0.57	1	0.57	1.37	0.29
Kinase factor	1.14	10	0.11	0.27	0.96
Error	2.5	6	0.42		
Total	5.5	21			

754

755

Figure 1 Normalised Mean Square (NMS) Fluctuations of residues of kinases calculated from GNM based global Normal Mode. (A) Global mode NMS fluctuations of active (PDB ID 1IR3_A, *blue curve*) and inactive (PDB ID 1IRK_A, *red curve*) conformations of Insulin Receptor kinase. Numberings are according to the residue numbers in the PDB structures supplied by the authors. Missing coordinates' and 2 preceding and 2 succeeding residues' fluctuations are not considered (region of missing *blue curve*) for analyses. Structural and functional motifs in the kinase catalytic domain are marked below the zero ordinate. (B) Mean Global mode NMS fluctuations of 8 active (PDB IDs: 1JBP_E, 1APM_E, 1L3R_E, 1RDQ_E, 2ERZ_E, 2CPK_E, 1FMO_E, 1ATP_E, *blue curve*) and 3 inactive (PDB IDs: 2QVS_E, 4DFY_A, 1SYK_A, *red curve)* conformations of PKA. The standard error of mean in the NMS fluctuations at each residue position is represented as shaded regions (*shaded blue* for active and *shaded red* for inactive). Residues are numbered arbitrarily from 1-255 irrespective of UniProt/PDB numbering. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of active and inactive fluctuations of that residue. (C) Mean Global mode NMS fluctuations of 31 active (*blue curve*) and 23 inactive (*red curve*) conformations of STY kinase structures listed in Table I. The standard error of mean in the NMS fluctuations at each residue position is represented as shaded regions (*shaded blue* for active and *shaded red* for inactive). Residues are numbered arbitrarily from 1-361 irrespective of UniProt/PDB numbering. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of active and inactive fluctuations of that residue. (D) Area under the two curves in Fig. 1*A* is plotted and the values are noted for the active and inactive conformations. (E) For residues having significant difference in fluctuation distributions between the active and inactive conformations in Fig. 1*B*, *green '*'*, area under the two curves are plotted and the corresponding mean values are noted. (F) For residues having significant difference in fluctuation distributions between the active and

inactive conformations in Fig. 1*C*, *green '*'*, area under the two curves are plotted and the corresponding mean values are noted.

756

Figure 2 Comparison of structural fluctuations, B-factors of residues and conformational energies of STY kinases in the active and inactive conformations (A) For all the 24 active-inactive conformation pairs listed in Table II, a scatter plot of the global mode NMS fluctuations of residues in the inactive conformation is plotted against that of topologically equivalent residues in the active counterparts. The black diagonal line represents unity slope. 1095 points lie above and 525 points lie below the line. Distribution of points above and below the unity slope is significantly different; and is marked by '*' symbol. (B) For all the 24 active-inactive conformation pairs listed in Table II, a scatter plot of the normalised B-factors of residues in the inactive conformation is plotted against that of topologically equivalent residues in the active counterparts. The black diagonal line represents unity slope. 3693 points lie above and 2232 points lie below the line. Distribution of points above and below the unity slope is significantly different; and is marked by '*'. (C) Scatter plot of mean conformational energies of all the inactive structures of an STY kinase against that of all the active structures of the same kinase is made. The black diagonal line represents the unity slope. 10 points lie above and 4 points lie below the line.

Figure 3 Comparison of regions of structural fluctuation difference and regions of known protein-protein interface (A) Kinase fold is colour coded with residue-wise mean difference in fluctuations between the inactive and active forms. The mean difference is negative (*blue*) for residues that fluctuate more in the active form and positive (*red*) for residues that fluctuate more in the inactive form. αC-helix, activation loop, αG-helix and loop connecting the αG and αH helices are the residues with high positive mean difference. (B) Kinase fold is colour coded with residue-wise number of inter-chain interactions in 476 kinase complex structures. The residues that make the least number of inter-chain interactions are coloured *blue*; those that make the maximum number of inter-chain interactions are coloured *red*. αC-helix, activation loop and αG-helix residues participate in the most number of protein-protein interactions.

757

Figure 4 Protein-protein interactions can sufficiently alter the global mode dynamics of the kinase. (A) Cartoon representation of InsR kinase chain, when solved in complex with Grb 14 BPS region (PDB ID 2AUH_A). Only the Cα positions of chain A (*blue*) is used to construct the network topology for the purpose of GNM in the bound_isolated condition. (B) Cartoon representation of chain A (*blue*) and chain B (*red*) of the InsR-BPS complex. Cα positions of both chain A and chain B were used to construct the network topology for the purpose of GNM in the bound_complex condition. BPS interacts with the αC-helix, activation loop and αG-helix residues of the kinase. (C) Global mode NMS fluctuations of the bound_isolated (*blue curve*) and the bound_complex (*red curve*) variants of the Insulin Receptor kinase. Numberings are arbitrarily assigned from 1-275 irrespective of the UniProt/PDB numbering. Structural and functional motifs in the kinase catalytic domain are marked below the zero ordinate. (D) Area under the two curves in Fig. 4*C* is plotted and the values are noted for the bound-isolated and bound_complex variants.

758

Figure 5 Comparison of magnitude of global mode NMS fluctuations of structural segments of known functional relevance. (A) αC-helix (B) catalytic residues (C) activation loop (D) αG-helix (E) hinge (F) ATP binding loop. The black diagonal line in the subplots is of unity slope. The number of points above and below the unity line in each of the subplots is marked. '*' symbol indicates that the distribution of points above and below the unity line are statistically significant (p-value < 0.001).

Figure 6 Investigation of effect of different inactive conformations on the fluctuations of the kinases. Mean global mode NMS fluctuations of 31 active conformations listed in Table I (*blue curve*) and (A) 6 DFG-"in" inactive (*red curve*), (B) 2 DFG-"flip" inactive (*red curve*), (C) 12 DFG-"Src-like" inactive (*red curve*), (D) 4 DFG-"out" inactive (*red curve*) conformations of STY kinase structures listed in Table III. The standard error of mean in the NMS fluctuations at each residue position is represented as shaded regions (*shaded blue* for active and *shaded red* for inactive). Residues are numbered arbitrarily from 1-361 irrespective of UniProt/PDB numbering. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of active and inactive fluctuations of that residue. Structural and functional motifs in the kinase catalytic domain are marked below the zero ordinate. For regions of significant differential fluctuations, areas under the curves are plotted for (E) DFG-"in" inactive, DFG-"flip" inactive, DFG-"Src-like" inactive, DFG-"out" inactive and (F) αC-helix-"in" inactive, αC-helix-"out" inactive, lobes-"open" inactive and lobes-"closed" inactive conformations along with those for all the active and inactive conformations for comparison. The mean values for each of the distributions are also indicated.

759

Figure 7 Control analysis to understand the effect of bound small molecules on the fluctuation dynamics. From the active structures enlisted in Table I, all possible active-conformation pairs belonging to a unique UniProt ID, solved with and without one/more bound small molecules (substrate peptide, ATP, ATP analog, Phosphate moiety, cations) were selected and their respective fluctuations were compared. (A) Mean global mode NMS fluctuations of the active conformations of a kinase with bound small molecule(s) (*blue curve*) and that of corresponding active conformations of the same kinase with the small molecule unbound (*red curve*) are plotted. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of active-holo (bound-to-small-molecules) and active-apo (unbound-to-small-molecules) fluctuations of that residue (two-tailed paired T-test, p-value < 0.05). (B) Mean global mode NMS fluctuations of the inactive conformations of a kinase with bound small molecule(s) (*blue curve*) and that of corresponding inactive conformations of the same kinase with the small molecule unbound (*red curve*) are plotted. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of inactive-holo (bound-to-small-molecules) and inactive-apo (unbound-to-small-molecules) fluctuations of that residue (two-tailed paired T-test, p-value < 0.05).

760

Figure 8 Quantitative effect of bound small molecules on the structural fluctuations. For the regions with significant differential fluctuations, the areas under the curves are plotted for active-holo, active-apo, inactive-holo and inactive-apo conformations. The mean of each of the distributions is indicated.

Figure 9 Classifier prediction accuracy. Global mode NMS fluctuations of different structural segments (indicated residue numbers correspond to PDB numbering of CDK2: 1JST_A) were used to train and test the classifier. A random classification would result in 50% accuracy. Significant improvement from random classification is indicated with '*' symbol.

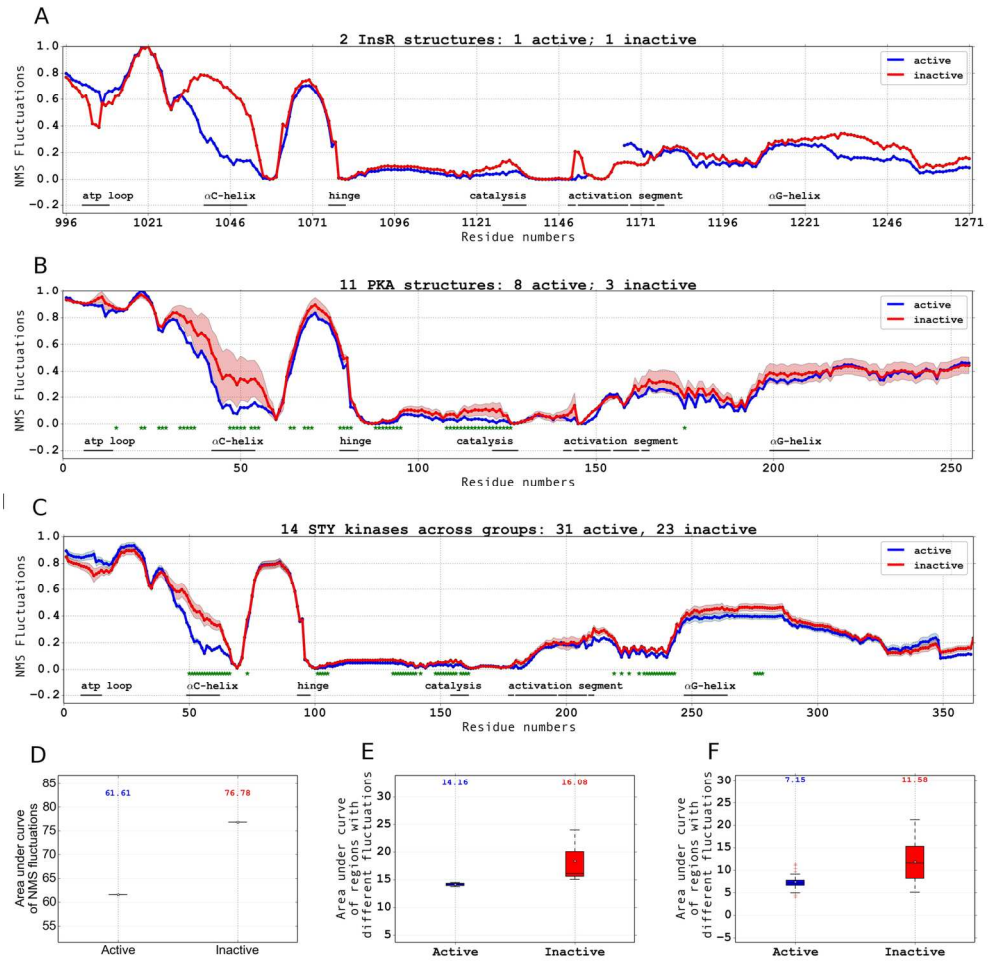


Figure 1 Normalised Mean Square (NMS) Fluctuations of residues of kinases calculated from GNM based global Normal Mode. (A) Global mode NMS fluctuations of active (PDB ID 1IR3_A, blue curve) and inactive (PDB ID 1IRK_A, red curve) conformations of Insulin Receptor kinase. Numberings are according to the residue numbers in the PDB structures supplied by the authors. Missing coordinates' and 2 preceding and 2 succeeding residues' fluctuations are not considered (region of missing blue curve) for analyses. Structural and functional motifs in the kinase catalytic domain are marked below the zero ordinate. (B) Mean Global mode NMS fluctuations of 8 active (PDB IDs: 1JBP_E, 1APM_E, 1L3R_E, 1RDQ_E, 2ERZ_E, 2CPK_E, 1FMO_E, 1ATP_E, blue curve) and 3 inactive (PDB IDs: 2QVS_E, 4DFY_A, 1SYK_A, red curve) conformations of PKA. The standard error of mean in the NMS fluctuations at each residue position is represented as shaded regions (shaded blue for active and shaded red for inactive). Residues are numbered arbitrarily from 1-255 irrespective of UniProt/PDB numbering. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of active and inactive fluctuations of that residue. (C) Mean Global mode NMS fluctuations of 31 active (blue curve) and 23 inactive (red curve) conformations of STY kinase structures listed in Table I. The standard error of mean in the NMS fluctuations at each residue position is represented as shaded regions (shaded blue for active and shaded red for inactive). Residues are numbered arbitrarily from 1-361 irrespective of UniProt/PDB numbering. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of active and inactive fluctuations of that residue. (D) Area under the two curves in Fig. 1A is plotted and the values are noted for the active and inactive conformations. (E) For residues having significant difference in fluctuation distributions between the active and inactive conformations in Fig. 1B, green '*', area under the two curves are plotted and the corresponding mean values are noted. (F) For residues having significant difference in fluctuation

distributions between the active and inactive conformations in Fig. 1C, green '*', area under the two curves are plotted and the corresponding mean values are noted.

149x145mm (300 x 300 DPI)

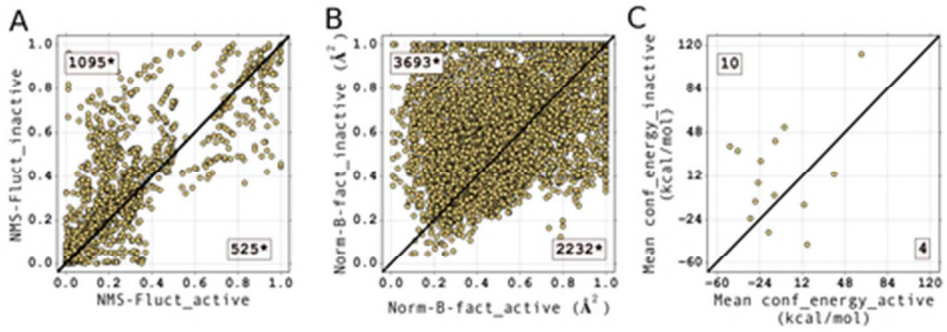


Figure 2 Comparison of structural fluctuations, B-factors of residues and conformational energies of STY kinases in the active and inactive conformations (A) For all the 24 active-inactive conformation pairs listed in Table II, a scatter plot of the global mode NMS fluctuations of residues in the inactive conformation is plotted against that of topologically equivalent residues in the active counterparts. The black diagonal line represents unity slope. 1095 points lie above and 525 points lie below the line. Distribution of points above and below the unity slope is significantly different; and is marked by '*' symbol. (B) For all the 24 active-inactive conformation pairs listed in Table II, a scatter plot of the normalised B-factors of residues in the inactive conformation is plotted against that of topologically equivalent residues in the active counterparts. The black diagonal line represents unity slope. 3693 points lie above and 2232 points lie below the line. Distribution of points above and below the unity slope is significantly different; and is marked by '*'. (C) Scatter plot of mean conformational energies of all the inactive structures of an STY kinase against that of all the active structures of the same kinase is made. The black diagonal line represents the unity slope. 10 points lie above and 4 points lie below the line.
40x14mm (300 x 300 DPI)

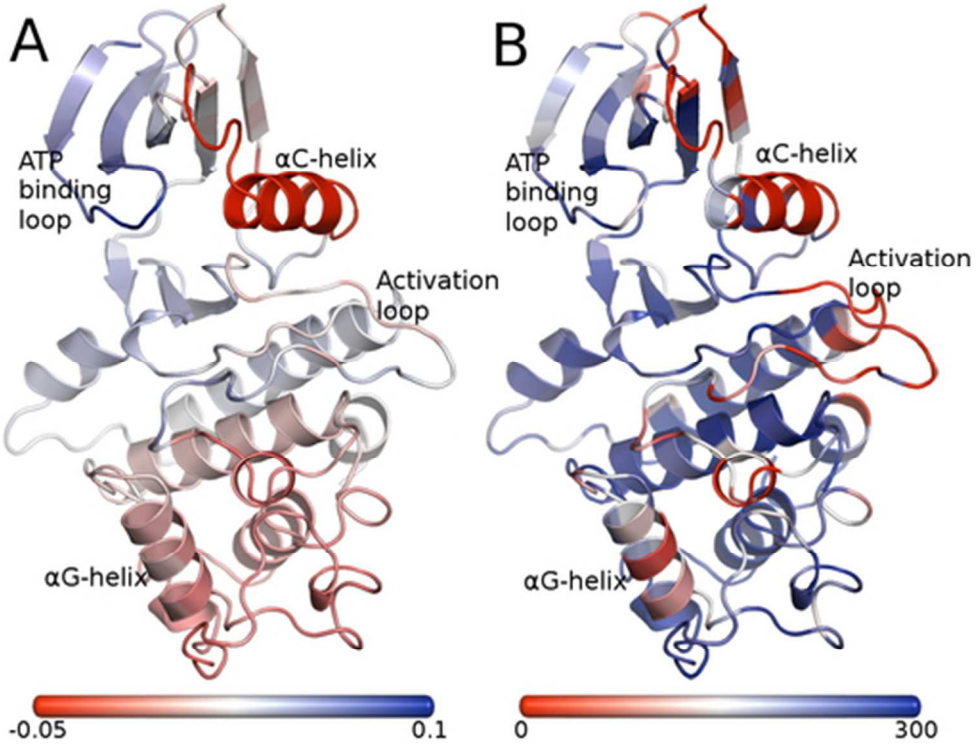


Figure 3 Comparison of regions of structural fluctuation difference and regions of known protein-protein interface (A) Kinase fold is colour coded with residue-wise mean difference in fluctuations between the inactive and active forms. The mean difference is negative (blue) for residues that fluctuate more in the active form and positive (red) for residues that fluctuate more in the inactive form. αC-helix, activation loop, αG-helix and loop connecting the αG and αH helices are the residues with high positive mean difference. (B) Kinase fold is colour coded with residue-wise number of inter-chain interactions in 476 kinase complex structures. The residues that make the least number of inter-chain interactions are coloured blue; those that make the maximum number of inter-chain interactions are coloured red. αC-helix, activation loop and αG-helix residues participate in the most number of protein-protein interactions.

44x33mm (300 x 300 DPI)

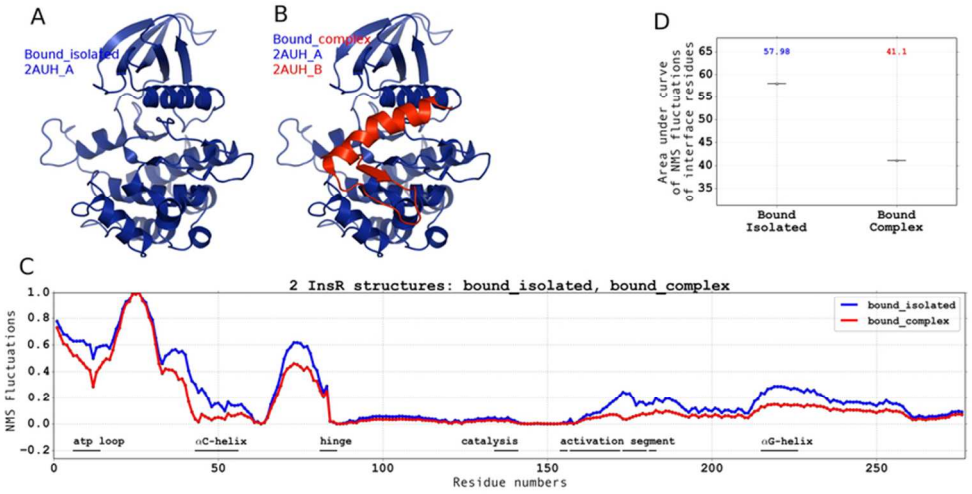


Figure 4 Protein-protein interactions can sufficiently alter the global mode dynamics of the kinase. (A) Cartoon representation of InsR kinase chain, when solved in complex with Grb 14 BPS region (PDB ID 2AUH_A). Only the Cα positions of chain A (blue) is used to construct the network topology for the purpose of GNM in the bound_isolated condition. (B) Cartoon representation of chain A (blue) and chain B (red) of the InsR-BPS complex. Cα positions of both chain A and chain B were used to construct the network topology for the purpose of GNM in the bound_complex condition. BPS interacts with the αC-helix, activation loop and αG-helix residues of the kinase. (C) Global mode NMS fluctuations of the bound_isolated (blue curve) and the bound_complex (red curve) variants of the Insulin Receptor kinase. Numberings are arbitrarily assigned from 1-275 irrespective of the UniProt/PDB numbering. Structural and functional motifs in the kinase catalytic domain are marked below the zero ordinate. (D) Area under the two curves in Fig. 4C is plotted and the values are noted for the bound-isolated and bound_complex variants.

76x38mm (300 x 300 DPI)

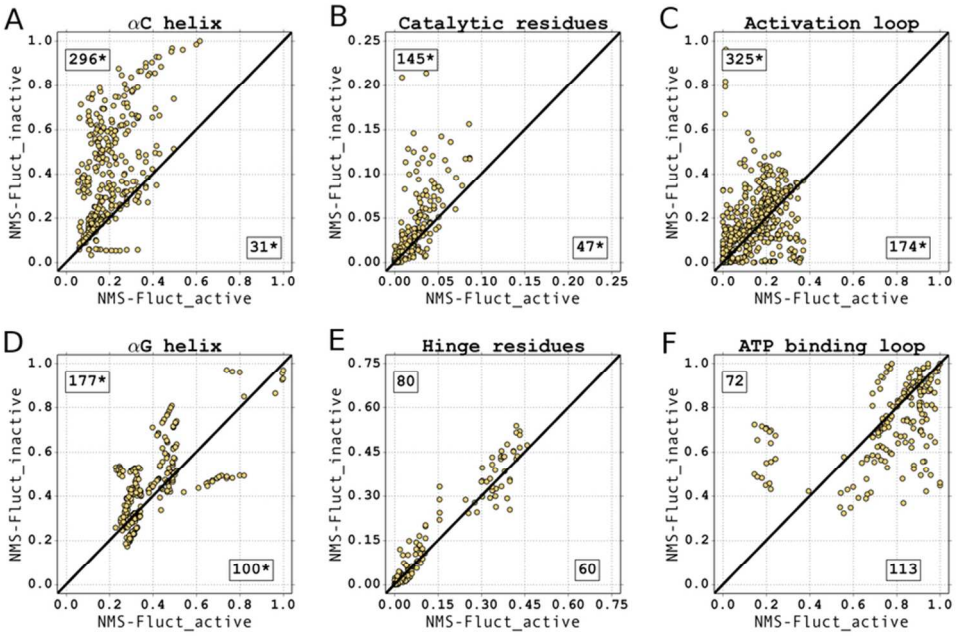


Figure 5 Comparison of magnitude of global mode NMS fluctuations of structural segments of known functional relevance. (A) αC-helix (B) catalytic residues (C) activation loop (D) αG-helix (E) hinge (F) ATP binding loop. The black diagonal line in the subplots is of unity slope. The number of points above and below the unity line in each of the subplots is marked. '*' symbol indicates that the distribution of points above and below the unity line are statistically significant (p-value < 0.001).

80x53mm (300 x 300 DPI)

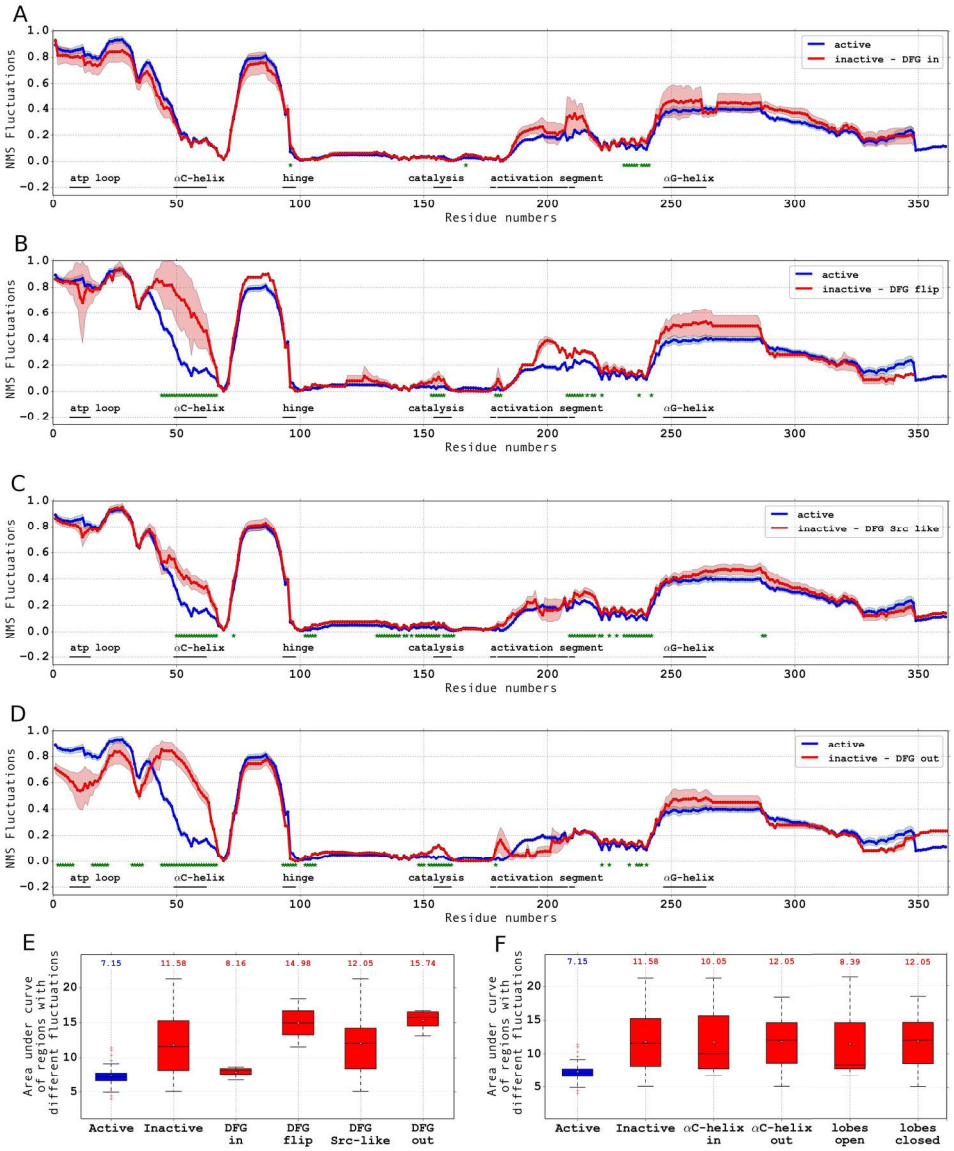


Figure 6 Investigation of effect of different inactive conformations on the fluctuations of the kinases. Mean global mode NMS fluctuations of 31 active conformations listed in Table I (blue curve) and (A) 6 DFG-"in" inactive (red curve), (B) 2 DFG-"flip" inactive (red curve), (C) 12 DFG-"Src-like" inactive (red curve), (D) 4 DFG-"out" inactive (red curve) conformations of STY kinase structures listed in Table III. The standard error of mean in the NMS fluctuations at each residue position is represented as shaded regions (shaded blue for active and shaded red for inactive). Residues are numbered arbitrarily from 1-361 irrespective of UniProt/PDB numbering. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of active and inactive fluctuations of that residue. Structural and functional motifs in the kinase catalytic domain are marked below the zero ordinate. For regions of significant differential fluctuations, areas under the curves are plotted for (E) DFG-"in" inactive, DFG-"flip" inactive, DFG-"Src-like" inactive, DFG-"out" inactive and (F) αC-helix-"in" inactive, αC-helix-"out" inactive, lobes-"open" inactive and lobes-"closed" inactive conformations along with those for all the active and inactive conformations for comparison. The mean values for each of the distributions are also indicated.

186x227mm (300 x 300 DPI)

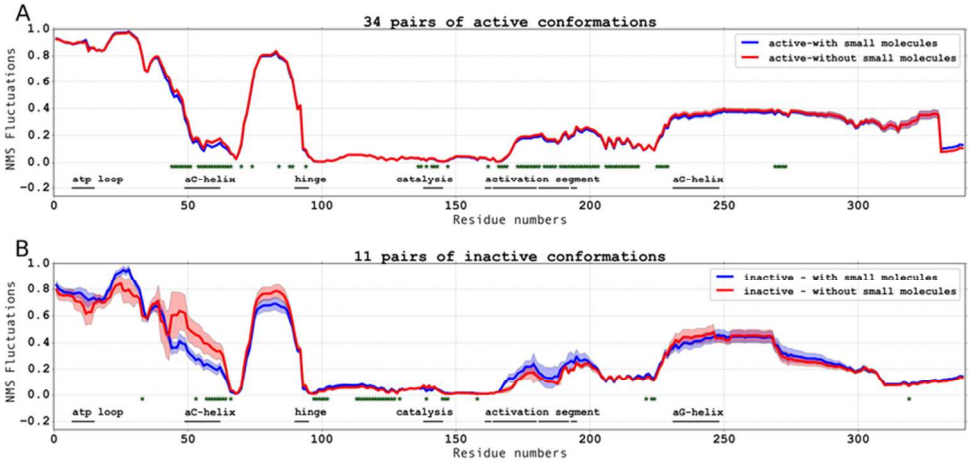


Figure 7 Control analysis to understand the effect of bound small molecules on the fluctuation dynamics. From the active structures enlisted in Table I, all possible active-conformation pairs belonging to a unique UniProt ID, solved with and without one/more bound small molecules (substrate peptide, ATP, ATP analog, Phosphate moiety, cations) were selected and their respective fluctuations were compared. (A) Mean global mode NMS fluctuations of the active conformations of a kinase with bound small molecule(s) (blue curve) and that of corresponding active conformations of the same kinase with the small molecule unbound (red curve) are plotted. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of active-holo (bound-to-small-molecules) and active-apo (unbound-to-small-molecules) fluctuations of that residue (two-tailed paired T-test, p-value < 0.05). (B) Mean global mode NMS fluctuations of the inactive conformations of a kinase with bound small molecule(s) (blue curve) and that of corresponding inactive conformations of the same kinase with the small molecule unbound (red curve) are plotted. Green '*' symbol below the zero ordinate indicates significant difference between the distributions of inactive-holo (bound-to-small-molecules) and inactive-apo (unbound-to-small-molecules) fluctuations of that residue (two-tailed paired T-test, p-value < 0.05).

76x38mm (300 x 300 DPI)

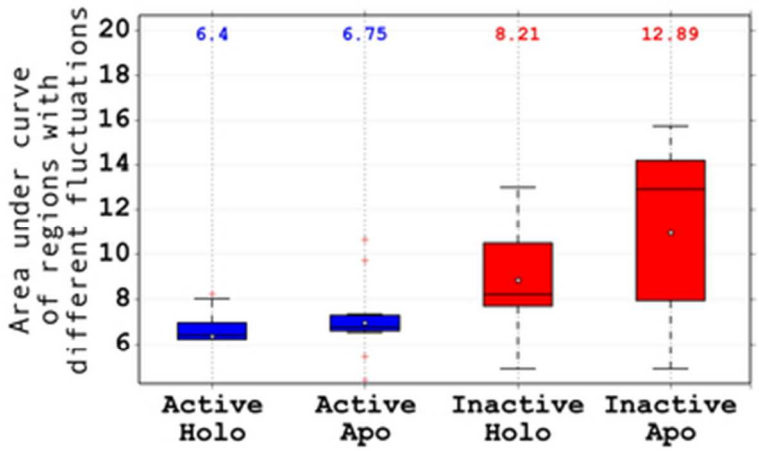


Figure 8 Quantitative effect of bound small molecules on the structural fluctuations. For the regions with significant differential fluctuations, the areas under the curves are plotted for active-holo, active-apo, inactive-holo and inactive-apo conformations. The mean of each of the distributions is indicated.

33x20mm (300 x 300 DPI)

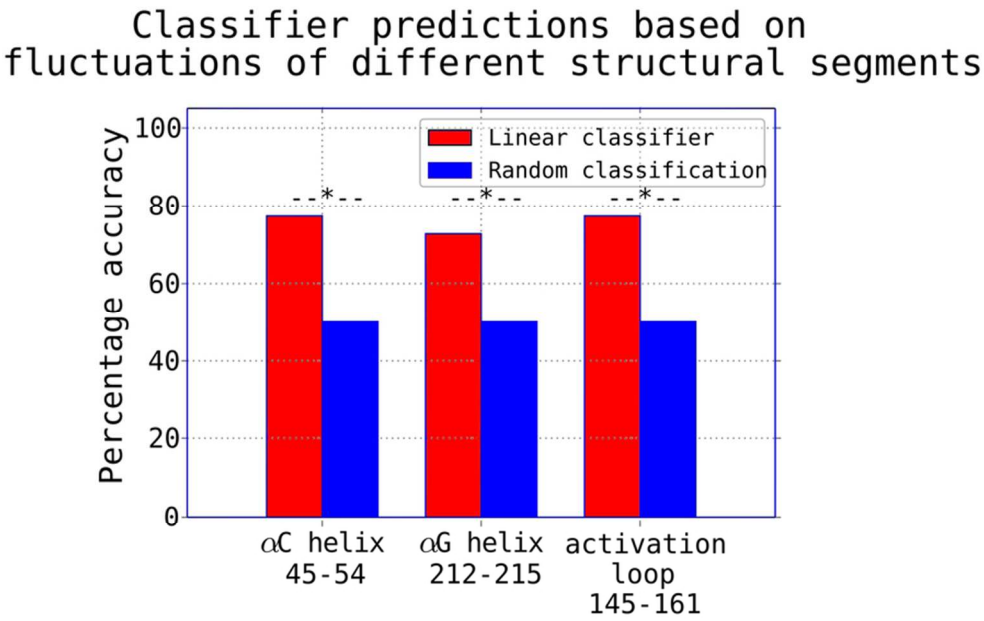


Figure 9 Classifier prediction accuracy. Global mode NMS fluctuations of different structural segments (indicated residue numbers correspond to PDB numbering of CDK2: 1JST_A) were used to train and test the classifier. A random classification would result in 50% accuracy. Significant improvement from random classification is indicated with '*' symbol.

77x48mm (300 x 300 DPI)