PROTEINS: Structure, Function, and Bioinformatics 67:1096–1112 (2007)

# Anatomy of a Structural Pathway for Activation of the Catalytic Domain of Src Kinase Hck

**Nilesh K. Banavali and Benoît Roux***
*Department of Physiology and Biophysics, Weill Medical College of Cornell University, New York, New York 10021*

***ABSTRACT*** **Src kinase activity is implicated in the regulation of downstream signal transduction pathways involved in cell growth processes. Crystallographic studies indicate that activation of Hematopoietic cell kinase (Hck), a member of the Src kinase family, is accompanied structurally by a large conformational change in two specific parts of its catalytic domain: the α-C helix and the activation loop. In the present study, molecular dynamics (MD) simulations are used to characterize the transformation pathway from the inactive to the active state. Four different conditions are considered: the presence or absence of Tyr416 phosphorylation in the activation loop, and the presence or absence of substrate ATP-2$Mg^{+2}$ in the active site. Effective free energy landscapes for local residues are determined using a combination of restrained MD simulations with a Root Mean Square Distance (RMSD) biasing potential to enforce the change followed by free MD simulations to allow relaxation from artificially enforced intermediates. A conceptual subdivision of the kinase catalytic domain into four moving parts: the flexible activation loop segment, the buried activation loop segment, the α-C helix, and the N-terminal end linker, leads to a concise hypothesis in which each of the moving parts are only required to be coupled to their nearest neighbor to ensure bidirectional allostery in the regulation of protein tyrosine kinases. Both Tyr416 phosphorylation and ATP-2$Mg^{+2}$ affect the local backbone torsional free energy landscapes accompanying the structural transition. When these two factors are present together, a metastable coordinated state of ATP-2$Mg^{+2}$ and the phosphorylated Tyr416 is observed that offers a possible explanation for the inhibition of protein kinase activity due to increase in $Mg^{+2}$ ion concentration. Proteins 2007;67:1096–1112.** 

**Key words: molecular dynamics; RMSD restraint; mechanism; Mg-ATP; free energy landscape; allostery; phosphorylation**

## INTRODUCTION

Src kinases are membrane-associated nonreceptor tyrosine kinases that phosphorylate specific tyrosine residues in other proteins, thereby modifying their activity and affecting various downstream signal transduction pathways in which these proteins participate.[1] The overall architecture of all nine members of the Src kinase family is conserved, and mainly consists of the domain that carries out the catalytic activity, and the SH2 and SH3 domains that regulate the activity of the catalytic domain.[2] The SH2 domain recognizes and binds to a phosphorylated tyrosine residue (Tyr 527, chicken c-Src numbering) in the C-terminal tail of the catalytic domain, which is a critical regulatory intramolecular interaction.[3] Dephosphorylation of this residue by phosphatases results in Src kinase activation.[4] The SH3 domain recognizes and binds to the proline-containing linker between the catalytic and SH2 domains.[5] Both these interactions can be disrupted by external peptides with recognition sequences containing a phosphorylated tyrosine or multiple proline residues that bind preferentially to the SH2 and SH3 domains, respectively, allowing the catalytic domain to shift to its activated form.[6] Phosphorylation of Tyr416 in the catalytic domain also results in the activation of the catalytic domain.[7] The downregulation of catalytic domain activity thus primarily depends on a combination of three stabilizing factors: the SH2 domain interacting intramolecularly with the C-terminal tail, the SH3 domain remaining bound to the proline-containing linker between SH2 and catalytic domains, and Tyr416 in the activation loop remaining unphosphorylated. Disruption in any of these three factors causes a shift in equilibrium to the active state of the kinase.[8]

Structural information about the downregulated inactive and upregulated active states of Src kinases is available from X-ray crystallographic studies of three differ-

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: National Institutes of Health; Grant number: CA93577-01.

Nilesh K. Banavali's current address is Division of Molecular Medicine, Wadsworth Center, Empire State Plaza, Albany, NY 12201.

Benoît Roux current address is Institute of Molecular Pediatric Science, Gordon Center of Integrative Science, The University of Chicago, 929 East 57th Street, Chicago, IL 60637, E-mail: roux@uchicago.edu

*Correspondence to: Benoît Roux, Department of Physiology and Biophysics, Weill Medical College of Cornell University, New York, NY 10021. E-mail: roux@uchicago.edu

Received 1 May 2006; Revised 19 October 2006; Accepted 8 November 2006

Published online 22 March 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21334

ent members of the Src family: Hematopoietic cell kinase (Hck), c-Src, and Lck. Hck is found in myeloid and lymphoid cells[9] that play a role in development of cancer.[10,11] The crystal structures of Hck in complex with two different inhibitors[12,13] show the structural details of its inactive state, which are also very similar to those for the highly homologous c-Src.[14,15] The molecular basis of the regulatory interactions between SH2 and C-terminal tail and SH3 and proline-containing linker is clarified in these fully assembled structures of the downregulated inactive form of the kinase. The catalytic domain itself is divided into two lobes, the N-lobe and the C-lobe, between which lies the catalytic site where ATP and peptide substrates bind. The overall structural architecture of the catalytic domain is very similar to that of other protein kinases, such as c-AMP activated protein kinase A (PKA)[16] and the insulin receptor kinase (IRK).[17] The inactive state of the catalytic domain of Hck has the activation loop with Tyr416 (approximately residues 410–430, chicken Src numbering) folded back onto the catalytic site, and the α-C helix (residues 304–316) rotated outward away from the catalytic site. The hydrogen bond between Lys295 and Glu310, which is known to be critical for catalysis,[12] is completely dissociated in this inactive conformation. These features, absent in the active state structures of both PKA and IRK, are represented in the unphosphorylated inactive IRK structure.[18] Structural studies on other protein kinases have shown that their catalytic domains adopt active state conformations similar to PKA, but may vary widely in their inactive state conformations.[19]

The crystal structure of the lymphocyte kinase Lck, another closely related member of the Src kinase family found in white blood cells, reveals an active state conformation reminiscent of that of PKA.[20] In this conformation, the phosphorylated Tyr416 containing activation loop is located away from the catalytic site and the α-C helix is rotated inwards such that the hydrogen bond between Lys295 and Glu310 is formed. Phosphorylation of the tyrosine residue in the activation loop of the catalytic domain clearly results in substantial conformational changes.

These changes are communicated to the spatially remote regulatory domains, as is characteristic of a long-range allosteric coupling mechanism.[21] The peptide linkers between the SH2 and SH3 domains (residues 140–147), and the SH2 and catalytic domains (residues 246–259) are both shown to play a part in this long-range allosteric coupling. Three separate modifications related to these linkers are able to supersede the inhibitory effect of SH2 domain interaction with the phosphorylated Tyr527: (a) introduction of structurally destabilizing multiple glycine mutations in the short linker between the SH2 and SH3 domains in c-Src[22]; (b) introduction of polyproline-rich HIV 1-Nef protein disrupting the intramolecular binding of the proline-rich region (residues 250–254 in the linker between the SH2 and catalytic domains) to the SH3 domain[6]; and (c) alanine mutation of the strictly conserved Trp260 residue at the end of the linker between the SH2 and catalytic domains.[23] This Trp260 residue at the N-terminal end of the catalytic domain is also in close contact with the α-C helix, which in turn is closely associated with several residues of a buried segment of the activation loop (residues 407–411), in close proximity to the critical Tyr416 residue in the activation loop.

All these local interactions trace an allosteric pathway through the kinase that seems to functionally connect the distant parts of the enzyme. The bidirectional nature of this allosteric pathway is suggested by biochemical experiments showing that Tyr416 phosphorylation causes greater availability of the SH2 and SH3 domains to interact with the phosphotyrosine containing p-YEEI peptide and polyproline-containing HIV 1-Nef protein, respectively.[8] However, the mechanism by which information about conformational state of one part of the kinase is allosterically transmitted to the other parts, and whether this process is concerted or not, remains unclear.

Computer simulations based on atomic models can help in understanding the complex dynamics of Src kinases that governs such bidirectional allostery. A previous computational study, focused on identifying the basis of long-range allostery in Hck activation through restrained Targeted Molecular Dynamics (TMD),[22] suggested that the structural effects of activation on the N-lobe were transmitted as "instantaneous correlations" to the regulatory domains. A more recent TMD study focused on the hinge-bending separation of the two internally restrained lobes of the Src kinase and the response of the activation loop to this hinge-bending motion.[24] With the lobes artificially restrained with dihedral constraints, the activation loop showed spontaneous opening only in the presence of ADP with two $Mg^{+2}$ ions and phosphorylated Tyr416, which was attributed to phosphate–phosphate repulsion and hydration requirements. The results were also interpreted to suggest the possibility of an intramolecular ("cis") phosphorylation of Tyr416 by its own catalytic site. In spite of the qualitative insight provided by all these experimental and theoretical studies, there is clearly a need to progress towards a more detailed understanding of the conformational pathways linking Tyr416 phosphorylation to structural activation of Src kinases.

In the present study, structural pathways for the activation of the Hck catalytic domain are obtained using restrained MD simulations in explicit solvent. The process is examined under four different conditions concerning the presence of neutral ligand ATP-$2Mg^{+2}$ and Tyr416 phosphorylation. To reduce the main drawback of restrained TMD simulations, unbiased (free) MD simulations are subsequently used to allow spontaneous relaxation of the initially biased pathways. Conformational statistics gathered are characterized through residue-level backbone torsion conformational analysis and sidechain interaction analysis to quantitatively dissect the large-scale structural transformation into its local components. A metastable coordination state between ATP-$2Mg^{+2}$ and the phosphorylated Tyr416 is also observed and the

consequences of its existence are explored. The implications of the details of the structural activation pathways in the catalytic function and bidirectional allosteric regulation of Hck are discussed.

## MATERIALS AND METHODS

The c30a1 academic version of the CHARMM program[25] and the all-atom PARAM22 CHARMM force field for ATP-2$Mg^{+2}$ and proteins[26,27] was used for all calculations. The Hematopoietic cell kinase (Hck) catalytic domain inactive state structure was obtained from the 2.0 Å resolution crystal structure (PDB ID: 1QCF) in the presence of the inhibitor PP1; the initial active state structure of the Hck catalytic domain was obtained from the 1.7 Å resolution Lck active form structure (PDB ID: 3LCK)[20] by sequence alignment using CLUSTALW[28] and homology modeling using the program MODELLER.[29] ATP was modeled into the active site by placing the adenine base moiety of ATP in the same orientation as the adenine base moiety of the bound inhibitor PP1 in the inactive state Hck crystal structure. The initial internal conformation of ATP and its associated $Mg^{+2}$ ions was obtained from the Mn-ATP structure in the PKA crystal structure (PDB ID: 1ATP).[30] The phosphorylated form of the catalytic domain was generated by adding a phosphate on the Tyr416 side chain with a chemically reasonable initial orientation with respect to the six membered tyrosine phenyl ring. A 150 m*M* KCl solution box with dimensions 77.6 × 62.1 × 52.4 $Å^3$ was used to solvate all structures (catalytic domain, catalytic domain with Tyr416 phosphorylated, catalytic domain with ATP-2$Mg^{+2}$, catalytic domain with $Mg^{+2}$-ATP and Tyr416 phosphorylated) such that the edge of the box from any protein non-hydrogen atom was at least 8 Å. The systems were then minimized with 5 kcal/mol harmonic restraints on non-hydrogen solute atoms for 500 Steepest Descent steps.

The CRYSTAL module in CHARMM[31] was used to generate images for periodic boundary conditions. To enable an integration time step of 2 fs, covalent bonds involving hydrogen were restrained using SHAKE.[32] Long-range electrostatic interactions were treated using the Particle Mesh Ewald (PME) approach[33] with a B-spline order of 4 and a Fast Fourier Transform grid of one point per Å and a real-space gaussian-width kappa of 0.3 $Å^{-1}$. Real space and Lennard-Jones (LJ) interaction cutoffs of 10 Å were used with nonbond interaction lists maintained and heuristically updated out to 15 Å. The migration of the solute protein outside the primary solvent box was discouraged by weak (0.5 kcal/mol) center-of-mass translational and rotational restraints using the MMFP module of CHARMM.[34] All four structures were equilibrated using constant pressure, temperature (NPT) molecular dynamics (MD) simulation[35] with harmonic restraints on nonhydrogen atoms released gradually over the first 100 ps for a total of 2 ns each till the Root Mean Square Distance (RMSD) from the original solute structure stabilized in each system.

The initial structures for the intermediate windows for the transition between the inactive and active conformations of the catalytic domain were generated using a 1D RMSD restraint pulling the structure gradually from the inactive to the active conformation. The region included in the restraint was chosen by initially orienting all nonhydrogen atoms in the catalytic domain in the inactive and active reference structures. The restrained region was chosen using Monte Carlo criteria such that, for each continuous restrained region, the residues in the center had maximal pair wise RMSD and residues at both ends had minimum pair wise RMSD. A total of 82 residues were included in the restraint (residues 289–329 including the α-C helix and residues 397–437 including the activation loop). Each intermediate was generated at 0.1 Å RMSD intervals with 10 ps of NVT sampling before pulling it to the next intermediate value −0.1 Å RMSD away. The final structure was restrained to a RMSD value of 0 Å from the active reference structure. Thus up to 76 intermediate windows were generated spanning the separation of up to 7.5 Å RMSD separating the two reference structures for all four systems. These initial windows were then used as starting points for umbrella sampling MD simulations with the restraint $w_j$ on the $\Delta D_{\mathrm{RMSD}}$ order parameter implemented in CHARMM.[36]

$$w_j = K_{\mathrm{rms}}(\Delta D_{\mathrm{RMSD}} - \Delta D_{\mathrm{min}})^2 \quad (1)$$

where $D_{\mathrm{min}}$ specifies the minimum value for the harmonic potential. The $\Delta D_{\mathrm{RMSD}}$ order parameter is the difference between the RMSD values of each structure from the two reference states:

$$\Delta D_{\mathrm{RMSD}} = \mathrm{RMSD}(X_t, X_{\mathrm{inactive}}) - \mathrm{RMSD}(X_t, X_{\mathrm{inactive}}) \quad (2)$$

where $X_t$ is the instantaneous structure, $X_{\mathrm{inactive}}$ and $X_{\mathrm{active}}$ are the inactive and active reference structures, respectively. The 0.1 Å intervals in the restraint used for pulling to generate initial structures in each window correspond to 0.2 Å intervals for the same structures in the $\Delta D_{\mathrm{RMSD}}$ order parameter. The force constant for this harmonic restraint was gradually reduced from 500 to 20 kcal/mol/$Å^2$ over a period of 70 ps constant volume and temperature (NVT) MD simulation. The final system in each window was then allowed to evolve with the final weak relative 1D RMSD restraint of 20 kcal/mol/$Å^2$ using an NVT MD simulation for a further 140 ps. Each window was subsequently allowed to evolve without any restraint for 300–800 ps, depending on the specific system. The total amount of sampling for the four systems combined (each about 25000 atoms) was > 0.2 μs.

The initial structures for the phosphorus–oxygen distance PMF quantifying the stability of the $Mg^{+2}$-coordinated state between Tyr416 and the γ-phosphate of ATP were obtained by pulling the inactive state conformation of the nonphosphorylated activation loop and its surrounding regions (residues 399–439), in the presence of ATP-2$Mg^{+2}$, to its active state conformation. This pulling

was enforced using a 1D RMSD restraint with 10 ps equilibration at each 0.1 Å interval as described above. Random structures from this pulling run distributed all along the distance coordinate were chosen as starting structures for individual umbrella sampling windows in the distance PMF calculation. Each window was restrained using a 2 kcal/mol/Å$^2$ distance restraint using the MMFP module in CHARMM and allowed to evolve under this restraint. A total of 42 windows were used for this calculation, which was carried out in two conditions for the same starting structures: with and without Tyr416 phosphorylation. Converged free energy profiles were obtained after about 200 ps of sampling in each window amounting to a total amount of sampling for this specific calculation of about 17 ns.

One special note is that, in all figures and for all four systems, the effective free energy surface for the panels corresponding to the restrained TMD simulations (consistently labeled 'R') only represent the structural conformations observed for ease of comparison with the other unrestrained MD simulation panels; that is, unbiasing by the Weighted Histogram Analysis Method (WHAM) algorithm[37,38] is not carried out for the $\Delta D_{\mathrm{RMSD}}$ restraint imposed during the sampling. Residue numbering is always according to the chicken c-Src nomenclature. All molecular pictures were produced using DINO (http://www.dino3d.org). All figures were made using the OPENDX program version 4.2.0 (http://www.opendx.org) or gnuplot version 3.7 (www.gnuplot.info) and the Gnu Image Manipulation Program (GIMP) version 1.2 (http://www.gimp.org).

## RESULTS

### Relative Motion of N- and C-Lobes of the Catalytic Domain

In the conserved architecture of protein kinases, the active site is sandwiched between the small β-sheet containing N-lobe and large, predominantly helical C-lobe of the catalytic domain.[16,12,14] Kinase activation results in a change in the relative orientation of the two lobes, which has been associated with the proper formation of the catalytic site.[39–41] To understand the flexibility and dynamics of the catalytic domain at a coarse level, it is useful to quantify the extent of the relative motion of the two lobes. In Figure 1, two-dimensional plots showing the translation distance and rotation angle of the two lobes are plotted for the four different systems studied.

The procedure used to obtain these values is to initially best-fit orient[42,43] each structure saved during the MD simulations relative to the equilibrated reference inactive structure using C-α atoms of the C-lobe, followed by best-fit orientation of the same structure using the C-α atoms of the N-lobe. The translation distance, the rotation angle, and the vector axis of rotation during the second orientation characterize the relative rigid-body dynamics of the two lobes. To avoid confusion arising from internal dynamics, the C-α atoms chosen exclude residues forming part of the activation loop in the C-lobe (i.e., residues 345–520 chosen excluding residues 400–430), as well as those from the α-C helix in the N-lobe (residues 260–342 chosen excluding residues 300–320). This procedure allows quantification of the relative motion of the N-lobe between the equilibrated reference and instantaneous MD structures. In reality, six variables (three for translation and three for rotation) are required to completely describe the relative orientation of the two domains. However, only two key variables are considered here for simplicity. The three translations are reduced to a single scalar distance of translation, and only the scalar angle of rotation around the rotation axis is plotted to show the extent of the individual rotations. To enhance visualization of stability and variability in population density along the chosen dimensions, the 2D probability distribution $<\rho(d_1,\theta_1)>$ is plotted along the aforementioned distance ($d_1$) and the rotation angle ($\theta_1$) is converted into free energy-like values through the expression:

$$W(d_1, \theta_1) = -k_{\mathrm{B}}T \ln < \rho(d_1, \theta_1) >$$

Each system is unique with respect to the presence of Tyr416 phosphorylation and ATP-2Mg$^{+2}$. In all the analysis carried out in the present study, the dynamic behavior of each system with respect to any chosen order parameter is classified into four sets: (a) the last 2 ns of equilibration of the inactive state structure (labeled 'I'), (b) the restrained TMD run going from the inactive to active state structure (labeled 'R'), (c) the unrestrained relaxation of the TMD run structures (labeled 'UR'), and (d) the last 2 ns of equilibration of the active state structure (labeled 'A'). This classification shows the starting and ending points of the activation pathway for the specific order parameters, the variability of the order parameter in these end-point basins, the pathway followed along these order parameters during the restrained TMD run, and the relaxation around this pathway upon removal of the restraint. This analysis for the two order parameters $d_1$ and $\theta_1$ characterizing the relative dynamic motion of the two lobes is portrayed in Figure 1.

The general trend seen for all four systems is that the active state equilibration structures show larger changes in $d_1$ and $\theta_1$ than the inactive state equilibration structures. The restrained TMD runs and their corresponding unrestrained relaxation structures spread between the two end-point simulations, and appear to even go beyond in some cases. At this coarse level of structural description, the presence or absence of ATP-2Mg$^{+2}$ gives rise to visible differences in the range of conformations visited. This is indicated by a greater separation between the inactive and active state equilibrations, and the appearance of a clear, bistable distribution in the unrestrained relaxation simulations, especially in the systems where ATP-2Mg$^{+2}$ is present. In the presence of both Tyr416 phosphorylation and ATP-2Mg$^{+2}$, the two minima are separate and connected by a broad and shallow barrier, suggesting that lobe motion during Hck activation con-

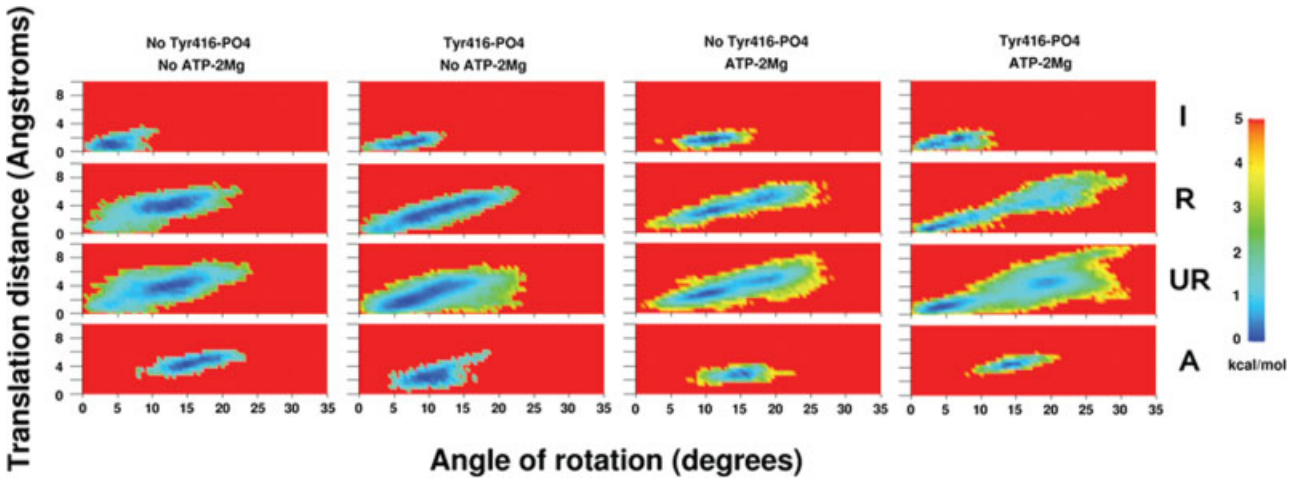


Fig. 1. Variation of relative N-lobe (residues 260–342, excluding residues 300–320) and C-lobe (residues 345–520, excluding residues 400–430) orientation shown in two scalar dimensions of translation distance and rotation angle. Tyr416-PO4 refers to the presence of a phosphate on Tyr416, ATP-2Mg$^{+2}$ refers to presence of ATP with its two associated Mg$^{+2}$ ions in the catalytic site; the equilibration simulation in the inactive state minimum is labeled 'I', the equilibration simulation in the active state minimum is labeled 'A', the restrained TMD simulations between these are labeled R, and the unrestrained simulations starting from the restrained simulation structures are labeled 'UR', this labeling convention is consistent in all figures.

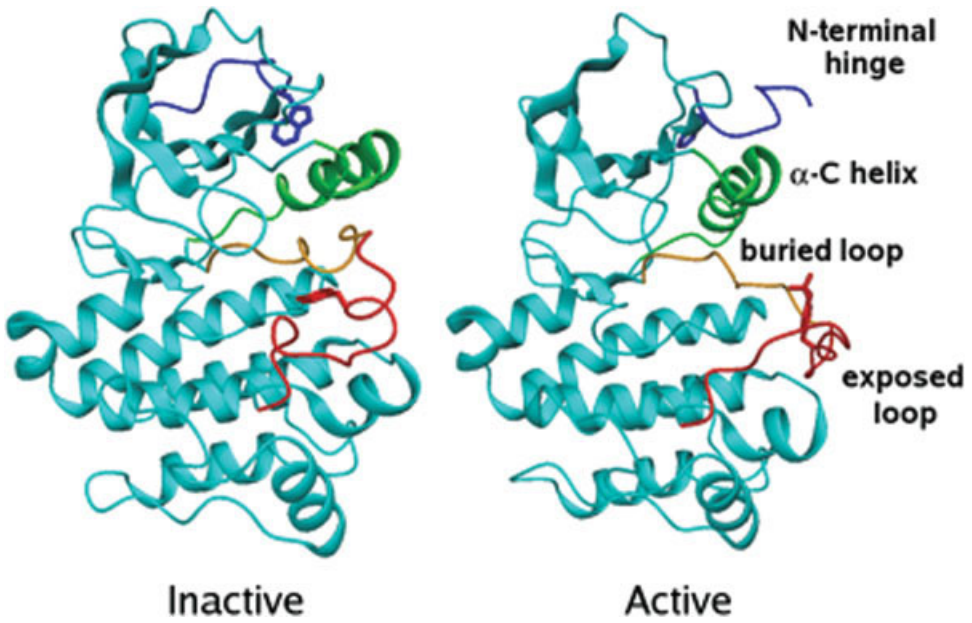


Fig. 2. Subdivision of catalytic domain into four minimal moving parts based on internal conformational change accompanying activation. Solvent-exposed activation loop region (residues 412–424) in red, buried activation loop region (residues 403–411) in orange, α-C helix (residues 299–320) in green, and the N-terminal hinge (residues 253–262) in blue.

sists of diffusive motion between two minimum energy relative orientations.

The variability in the relative orientations of the two lobes during the structural activation of the isolated Hck catalytic domain is substantial, but the maximum extent of the rotation of the N-lobe does not exceed 30°–35° while its translation distance does not exceed 4–6 Å. This range encompasses the rotation (16°) and the translation (1.9 Å) originally associated with the opening and closing of the active site cleft in cAMP-dependent protein kinase,[39] as well as the rotation (18°) and translation (3.4 Å) relating the inactive Hck[13] and active Lck[20] crystal structures. The presence of ATP-2Mg$^{+2}$ seems to subtly affect the conformational range of relative lobe motion in Hck.

## Minimal Moving Components of the Catalytic Domain

All residues in the catalytic domain do not change conformation to the same extent during activation.[2] Differ-

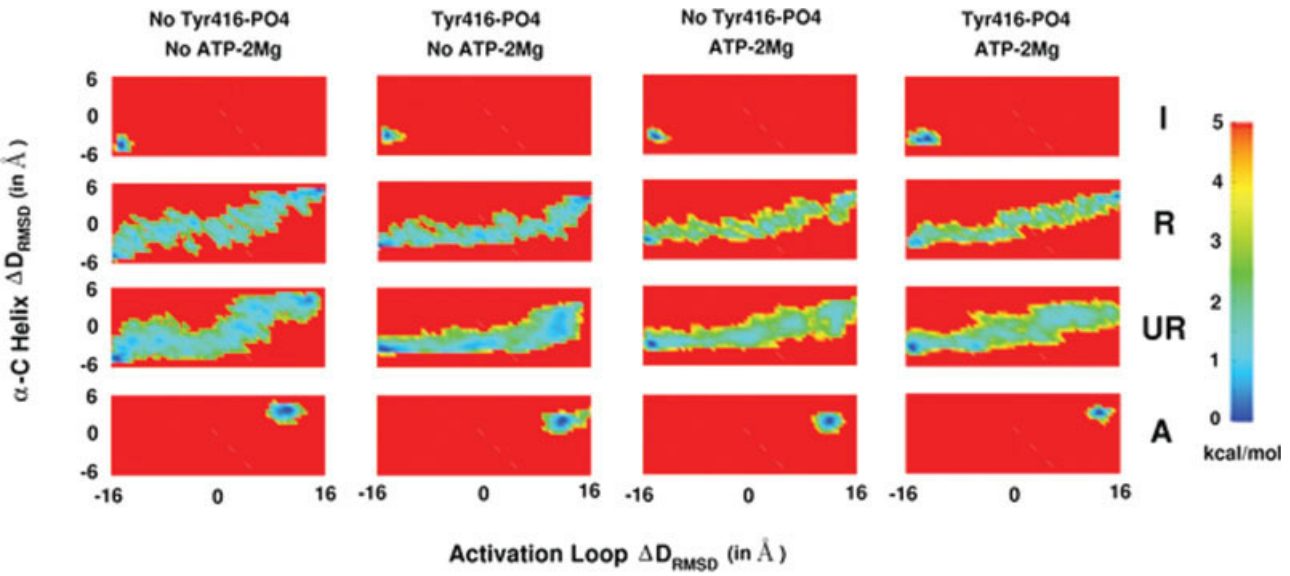


Fig. 3. Free energy distributions along difference RMSD ($\otimes D_{RMSD}$) coordinates describing conformational change in the mobile segment of the activation loop (C-α atoms of residues 412–425), and the core of α-C helix (C-α atoms of residues 300–320). The free energy surface for the second row panel for each system is only a representation of the sampling observed, no unbiasing is carried out for the $\otimes D_{RMSD}$ restraint imposed during the sampling. The distributions indicate the coarse correlation between activation loop opening and α-C helix rotation in the four systems.

ent regions display different degrees of conformational plasticity.[19] Identification of regions showing relatively correlated motions, and the intricate residue-level conformational changes involved, has not yet been achieved. A tentative subdivision of the catalytic domain can be obtained by focusing on secondary structure elements undergoing large conformational change between the inactive and active state structures.[2] With this approach, the catalytic domain can be subdivided into four minimal "moving parts": the buried segment of the activation loop (residues 403–411), the solvent-exposed segment of the activation loop (residues 412–424), the α-C helix (residues 299–320), and the N-terminal end (residues 253–262) (Fig. 2). In Figure 3, the correlated changes in the conformation of the solvent-exposed part of the activation loop and the α-C helix are indicated.

Each individual MD structure provides a $\Delta D_{RMSD}$ order parameter value for each moving part through the following procedure: the MD structure is initially best-fit oriented with respect to the equilibrated inactive and active state reference structures using all heavy atoms in residues except those from residues that are part of the moving regions themselves. The difference in RMSD from the inactive and active state structures for each specific moving region is calculated for these specific orientations. The probability density of the sampling in the 2D space along $\Delta D_{RMSD}$ for the activation loop and the α-C helix is then plotted as a free energy-like distribution (as described in previous section). This information can provide a robust coarse-grained picture of the dynamics of the moving components of the Src kinase catalytic machinery. It is clear from Figure 3 that the start (inactive state) and end (active state) are in completely distinct regions that do not substantially change their relative locations for any of the four systems in this 2D difference RMSD space. The spread of configurations in the inactive and active states is quite small in this 2D $\Delta D_{RMSD}$ space. Unlike the N-lobe rotation in the previous section, there is absolutely no overlap between the inactive and active states. This suggests that unbiased brute force trajectories would need to be prohibitively long to simulate a spontaneous transition from the inactive to the active state.

The restrained and unrestrained TMD simulations visit conformations in the intermediate regions, with some system-dependent variations. For example, in the absence of Tyr416 phosphorylation and ATP-$2Mg^{+2}$, the activation pathway assumes a step-wise character. The restrained MD simulations show partial α-C helix motion, followed by partial loop opening, followed by correlated α-C helix motion and loop opening, followed finally by the remaining loop opening motion. The unrestrained MD simulations show a spread of configurations mostly around those visited during the restrained MD simulation. Such a step-wise behavior is not seen in the system with Tyr416 phosphorylated and no ATP-$2Mg^{+2}$. Here loop opening is about 75% complete before it becomes correlated with α-C helix rotation to reach the final activated state. Similarly, in the system with no Tyr416 phosphorylation and ATP-$2Mg^{+2}$, loop opening also precedes significant α-C helix rotation followed by correlated changes to reach the activated state. In the system where Tyr416 phosphorylation and ATP-$2Mg^{+2}$ are both present, similar behavior is observed, except that the correlated change of α-C helix rotation and loop opening starts earlier than the half-way point in loop opening. This seems to suggest that among all four systems studied, this system displays the most

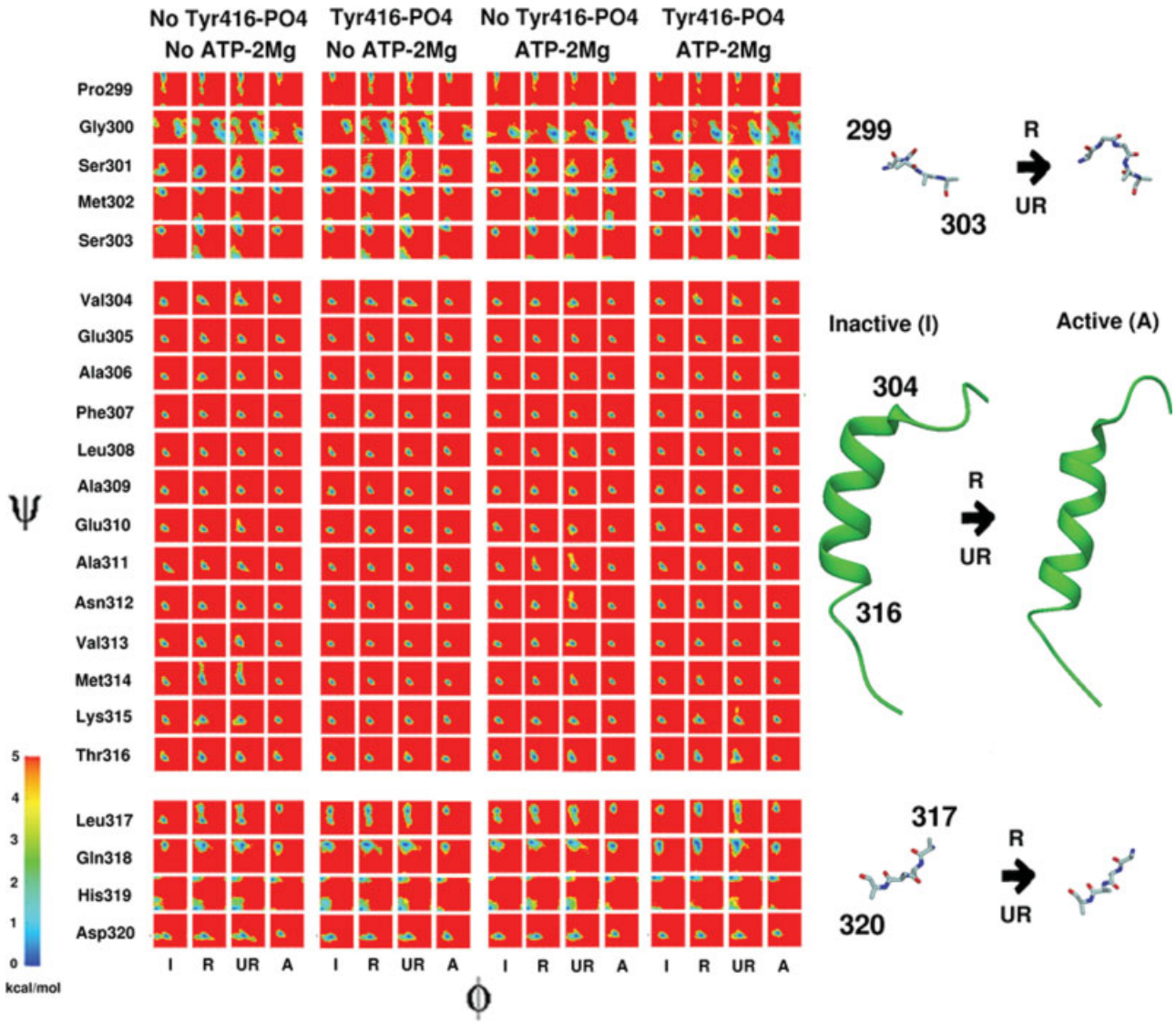


Fig. 4. Ramachandran free energy maps for the α-C helix and its neighboring loops (residues 299–320), for four systems considered. All torsions are in degrees, each 2D torsional surface panel ranges from −180° to 180° in both axes.

direct activation trajectory from the inactive to the active state.

The fourth moving part of the catalytic domain, which corresponds to the N-terminal end (residues 253–262), was not restrained in the activation trajectories. The behavior of this region has been characterized in a previous study[44] using umbrella sampling MD simulations in detail. Its spontaneous behavior in response to enforced change in the activation loop and α-C helix regions is consistent with previous observations[44] and is described in more detail in the supplementary material.

## Rotation of the α-C Helix

The coarse description of the four moving regions of the catalytic domain based on $\Delta D_{RMSD}$ illustrates the broad aspects of the conformational transition. The overall structural change, however, can be traced back to a combination of conformational transitions in backbone torsions in individual residues. To understand the features of α-C helix rotation at this localized level, the 2D ϕ-Ψ free energy surfaces for residues 299–320 are plotted in Figure 4. The residues that form the α-C helix itself (residues 304–316) displays no significant internal change between the inactive state and active states. They reside in a narrow conformational minimum in the standard right-handed α helix region of the Ramachandran plot. Distortions from this α helix region to the β-sheet region of the Ramachandran plot occur in the restrained and unrestrained MD simulations only for the residues 311, 312, and 314, and even these distortions are not consistently seen in all four systems.

The lack of conformational variability for the residues 304–316 indicates that this region essentially moves as a

rigid body during the activation of the catalytic domain. Similar observations were made regarding rigid body motions in PKA in an earlier molecular dynamics study, where highly flexible regions were mapped onto the loop regions connecting secondary structure elements.[45] Several residues on both sides of the stable helical stretch show much greater variability in their backbone conformations. This is true for residues 299–303 and residues 317–320, but is especially significant for residues Gly300 and Leu317. In the systems without ATP-2$Mg^{+2}$, Gly300 has a bistable distribution in the integrated restrained and unrestrained MD simulations, with one minimum each corresponding to the inactive and active state equilibrium structures. In the systems with ATP-2$Mg^{+2}$, Gly300 seems to be pre-organized in a conformation corresponding to the active state, even in the inactive state. A similar behavior is also shown by Leu317, though a transition from the right- handed α helix region to the β-sheet region of the Ramachandran plot correlated with activation is shown by this residue only in the system with no Tyr416 phosphorylation and no ATP-2$Mg^{+2}$. While the active state equilibration conformation of this residue lies firmly in the β-sheet region in all four systems, it tends to be bistable between the right-handed α helix and the β-sheet in unrestrained simulations in all four systems.

These results indicate that α-C helix rotation occurs mechanically due to local backbone conformation change in the end-loop regions bordering the helix. It has been shown by experimental studies that certain mutations in the linkers connecting the β-3 and α-C regions and α-C and β-4 regions can affect regulation of Src kinases[23]. The α-C and β-4 linker is especially interesting because it interacts with the SH2-kinase linker as well as serves as a hinge for α-C helix movement. This does not necessarily attribute a causative role to these end-loop residues, i.e. the mean force causing in α-C helix rotation may be exerted primarily by other residues. In particular, it is likely that the formation of a hydrogen bond between Glu310 and Lys2952, and rearrangement of intervening hydrophobic residues such as Leu407 (discussed in a subsequent section) plays an important role in the α-C helix rotation.

### The α-C Helix-Associated Part of the Activation Loop

The activation loop located between the N- and C-lobes of the Hck catalytic domain can be conceptually divided into two regions: residues 403–411 that are relatively buried and intimately associated with the active site and the α-C helix, and residues 412–425 (including Tyr416) that are relatively solvent exposed, especially in the active state conformation. Many of the residues of the buried segment (403–411) have hydrophobic characteristics. In Figure 5, the 2D ϕ-Ψ free energy surfaces for residues 403–412 are plotted to understand the conformational characteristics of the buried part of the activation loop associated with the α-C helix.

The first residue in the buried segment, Ala403, shows stability in a right-handed α helix-like conformation except in the inactive state equilibration of the system with Tyr416 phosphorylation and ATP-2$Mg^{+2}$. The next three residues: Asp404, Phe405, and Gly406 form the highly conserved DFG motif[46] at the base of the activation loop. Asp404 is involved in crucial contacts with an ATP-coordinated $Mg^{+2}$ ion and with Lys295 in facilitating formation of a catalytically competent active site.[19] This residue prefers a β-sheet-like conformation except in the inactive state equilibration of the system with Tyr416 phosphorylation and ATP-2$Mg^{+2}$ where it assumes a left-handed α helical state.

There is some variability in the preferred backbone conformation of Phe405 in the equilibration MD simulations of the four systems. In the inactive state equilibrations, an α-helical state is preferred for the system with no Tyr416 phosphorylation and no ATP-2$Mg^{+2}$, whereas a β-sheet-like conformation is preferred for all other systems. In the active state simulations, an α-helical state is preferred for all systems except when Tyr416 is phosphorylated and ATP-2$Mg^{+2}$ is present, where a β-sheet-like conformation occurs with some appearance of an alternate α-helical state. In all restrained and unrestrained simulations of the structural activation pathway, however, a bistable equilibrium between α-helical and β-sheet conformations is observed for Phe405. The behavior of Gly406 is also sensitive to Tyr416 phosphorylation and presence of ATP-2$Mg^{+2}$ as reflected in the different behavior of equilibration states in the four systems. This residue also shows a very broad distribution of its ϕ-Ψ conformational landscape in the restrained and unrestrained MD simulations. The significant flexibility of the backbone at this position in the active site may be important in facilitating the structural intermediates involved in activation. This position is conserved and glycine is the only amino acid that is amenable to such backbone variation, suggesting that it may play a dynamic role in the structural activation.

In addition to their backbone conformational behavior, the aspartic and phenylalanine side chains in the DFG motif display some interesting characteristics (data shown in Figures in the supplementary material). In the unrestrained trajectories for the transition from the inactive to the active state (with Tyr416 phosphorylation and ATP-2$Mg^{+2}$ present), the side chain carboxyl oxygens of Asp404 interact directly with the backbone amide moieties of both Gly406 and Leu407. A similar electrostatic interaction with the backbone amides is observed for the conserved aspartate in the N-terminal hinge of Hck.[44] In going from the inactive to the active state, the interaction of Asp404 with Asn391 is generally improved while the interaction with Lys295 is reduced, though there are specific differences depending on Tyr416 phosphorylation and presence of ATP-2$Mg^{+2}$. Because of the rotation of the α-C helix and rearrangement of the hydrophobic pocket during activation, the hydrophobic contacts made by Phe405 also change. Specifically, contacts with Val313 are increased, while contact with Leu322 and Tyr382 are reduced. The role played by these specific conformational and interaction

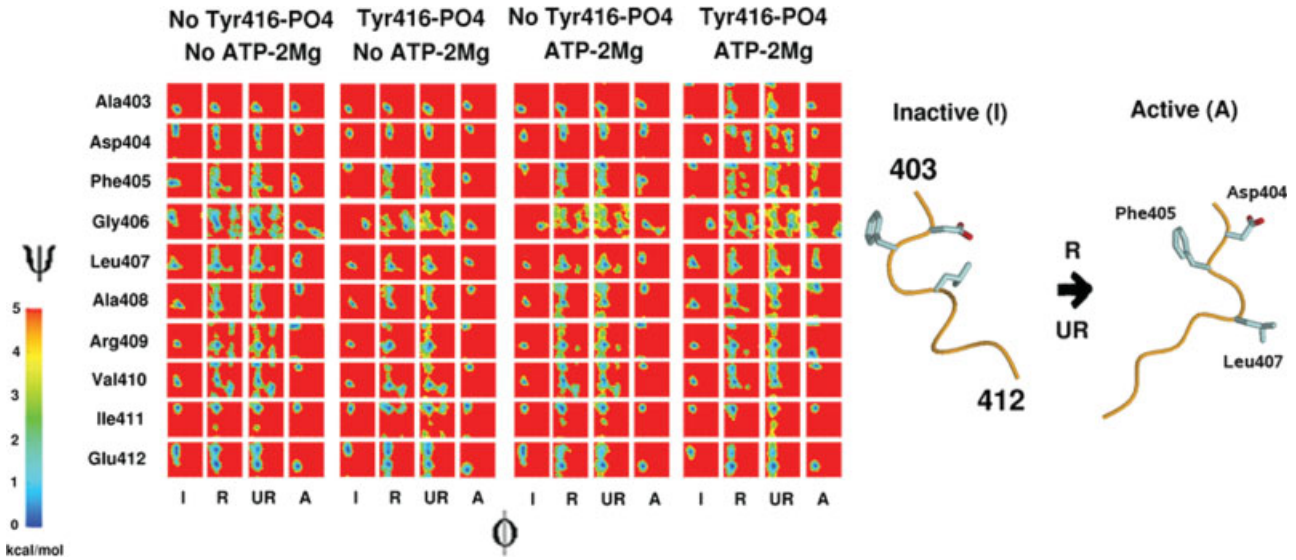


Fig. 5. Ramachandran free energy maps for the base of the activation loop closely associated with the α-C helix (residues 403–411), for four systems varying in presence of Tyr416 phosphorylation and ATP-2Mg$^{+2}$. All torsions in degrees, each 2D torsional surface panel ranges from −180° to 180° in both axes.

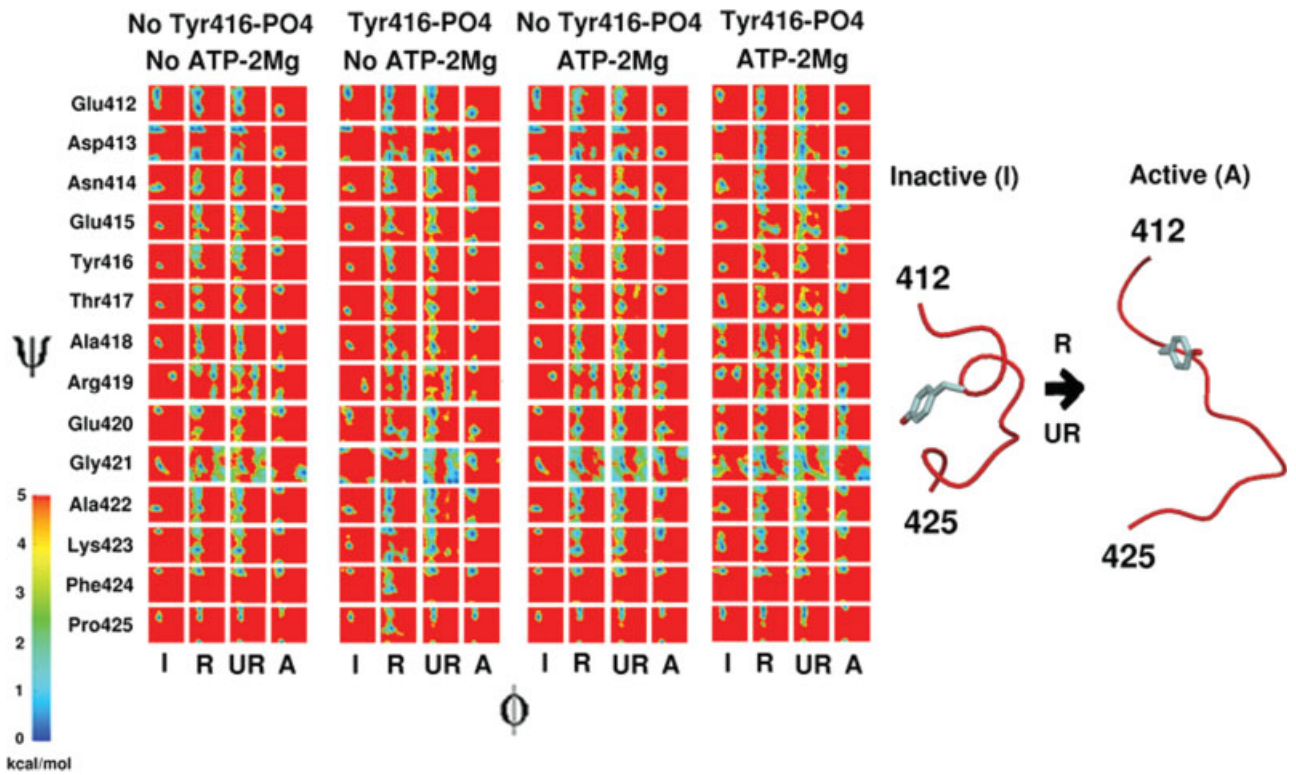


Fig. 6. Ramachandran free energy maps for the flexible part of the activation (residues 412–425), for four systems varying in presence of Tyr416 phosphorylation and ATP-2Mg$^{+2}$. All torsions in degrees, each 2D torsional surface panel ranges from −180° to 180° in both axes.

changes for the conserved DFG motif in the chemical catalysis of phosphoryl transfer is difficult to assess without the knowledge of the transition state conformation with the peptide substrate.

In the inactive structure, Leu407 forms hydrophobic contacts with other residues in the N-lobe and is located directly in between the Lys295 and Glu310 residues.[15,13] The Lys295-Glu310 hydrogen bond, critical for catalysis,[19] thus cannot form without the Leu407 sidechain moving out of the way. This residue shows a marginal shift of its α-helical preference to a conformation intermediate between the α-helical and β-sheet conformations

in the Ramachandran plot. The next three residues, Ala408, Arg409, and Val410 show consistent behavior in all four systems, with transitions from an α-helical state to a β-sheet state accompanying the structural activation. The last residue in this region of the activation loop, Ile411, mostly remains anchored in a β-sheet conformation throughout the structural activation in all four systems, and is the residue separating the relatively hydrophobic and hydrophilic parts of the activation loop.

To summarize, the structural transition occurring in the buried part of the activation loop thus involves a complex mixture of hydrophobic rearrangements (Phe405, Leu407, Val410, Ile411), and exchange of hydrogen bonds between electrostatic residues (Glu310-Lys295). The significant flexibility of Gly406 in the conserved DFG motif may be important in facilitating the activation transition.

### The Solvent-Exposed Part of Activation Loop

The residues 412–425 (containing the critical Tyr416) form the most flexible region of the activation loop. A part of this region was not resolved in an earlier crystal structure of Hck, underlining its conformational variability.[12] In Figure 6, the 2D ϕ-Ψ free energy surfaces for residues 412–425 are plotted to assess the flexibility of this loop region. This region shows the greatest conformational change during activation in the entire catalytic domain. Some residues can be grouped together depending on the types of conformational transitions during activation. Glu412 and Asp413 show a β-sheet to α-helical transition in all four systems. Asn414, Phe424, and Pro425 show no substantial differences between the inactive and active state equilibrations, but show a conformational variability during the structural activation process itself, signifying a possible role in stabilizing one or more metastable intermediates. Glu415, Tyr416, Thr417, Ala418, Ala422, and Lys423 all undergo a transition from α-helical to β-sheet conformation during activation.

Arg419 is unique in that it populates a left-handed α-helical conformation in the inactive state, which converts into a β-sheet conformation during activation. Glu420 shows a dynamic equilibrium between α-helical and β-sheet conformations, independent of the activation process, when Tyr416 is phosphorylated and ATP-2$Mg^{+2}$ is present. In the system with no Tyr416 phosphorylation and no ATP-2$Mg^{+2}$, Glu420 prefers a β-sheet conformation, with some transitions to an α-helical conformation during the activation process. In the presence of Tyr416 phosphorylation alone or ATP-2$Mg^{+2}$ alone, this residue shows a clear β-sheet to α-helical transition during activation. The backbone behavior of Gly421 is highly dependent on Tyr416 phosphorylation and ATP-2$Mg^{+2}$ presence, however, it consistently shows a large conformational variability during the structural activation.

To summarize, the solvent-exposed segment of the activation loop displays a wide conformational heterogeneity in the activation transition of the catalytic domain. This variability is probably related to the electrostatic perturbation created by the phosphorylation of Tyr416.

### Electrostatic and Hydrophobic Rearrangements

Although a richly, detailed mechanical description of the global structural change can be obtained through local backbone torsion analysis, the forces exerted by hydrophobic and electrostatic interactions are probably the underlying determinants of the conformational change. A reasonably complete picture of large protein structural changes cannot be obtained without some understanding of these sidechain-dependent interactions and their variability. Figure 7 monitors the movement of some key residues that interact through electrostatic or hydrophobic interactions. The residues chosen lie in the close vicinity of the active site and play a role in the formation of the critical electrostatic interaction between Glu310 and Lys295. Glu310, Lys295, Arg409, and Tyr416 form an electrostatic network that is reorganized when the kinase goes from the inactive state to the active state. Such a switching in the electrostatic network during activation was also recently observed during MD simulations of the Lyn kinase.[47] Phe307, Met314, and Leu407 form a hydrophobic cluster lying squarely between Glu310 and Lys295 in the inactive state structure. This cluster has to break in order to allow complete rotation of the α-C helix and approach of Glu310 in proximity to Lys295.

Correlated change in the Glu310-Arg409 and Glu310-Lys295 seems to be the fundamental electrostatic rearrangement involved in formation of the catalytically competent active site.[2,19] While Glu310 exchanges between Arg409 and Lys295, Arg409 itself exchanges between Glu310 and Tyr416. In the inactive state of all four systems, Glu310 is associated with Arg409 and completely separated from Lys295. In the active state, Glu310 and Lys295 are associated while Arg409 may remain partly associated with Glu310. In the system with no Tyr416 phosphorylation, the exchange of Glu310 between Arg409 and Lys295 occurs only at an intermediate state where all three sidechains are in close association followed by separation of Arg409. In the other three systems, a variety of intermediate states are possible where Arg409 can dissociate from Glu310 before the Glu310-Lys295 interactions forms. The system with both Tyr416 phosphorylated and ATP-2$Mg^{+2}$ present offers the greatest range of such intermediate states.

In all systems going from the inactive state to the active state, the Arg409-Glu310 interaction is replaced by the Arg409-Tyr416 interaction. Not surprisingly, the Arg409-Tyr416 interaction resides in a narrower well in the free energy surface in the systems with Tyr416 phosphorylated because of the stronger and more stable interaction between the negatively charged phosphate and the positively charged Arg409 sidechain. The lack of population of the bottom left corner of the panels for the restrained TMD and the unrestrained MD simulations of the systems with Tyr416 phosphorylated and no ATP-2$Mg^{+2}$ (see Fig. 7) indicates that dissociation and association of the Arg409-Glu310 and the Arg409-Tyr416 pairs need not always require an intermediate where all three are in

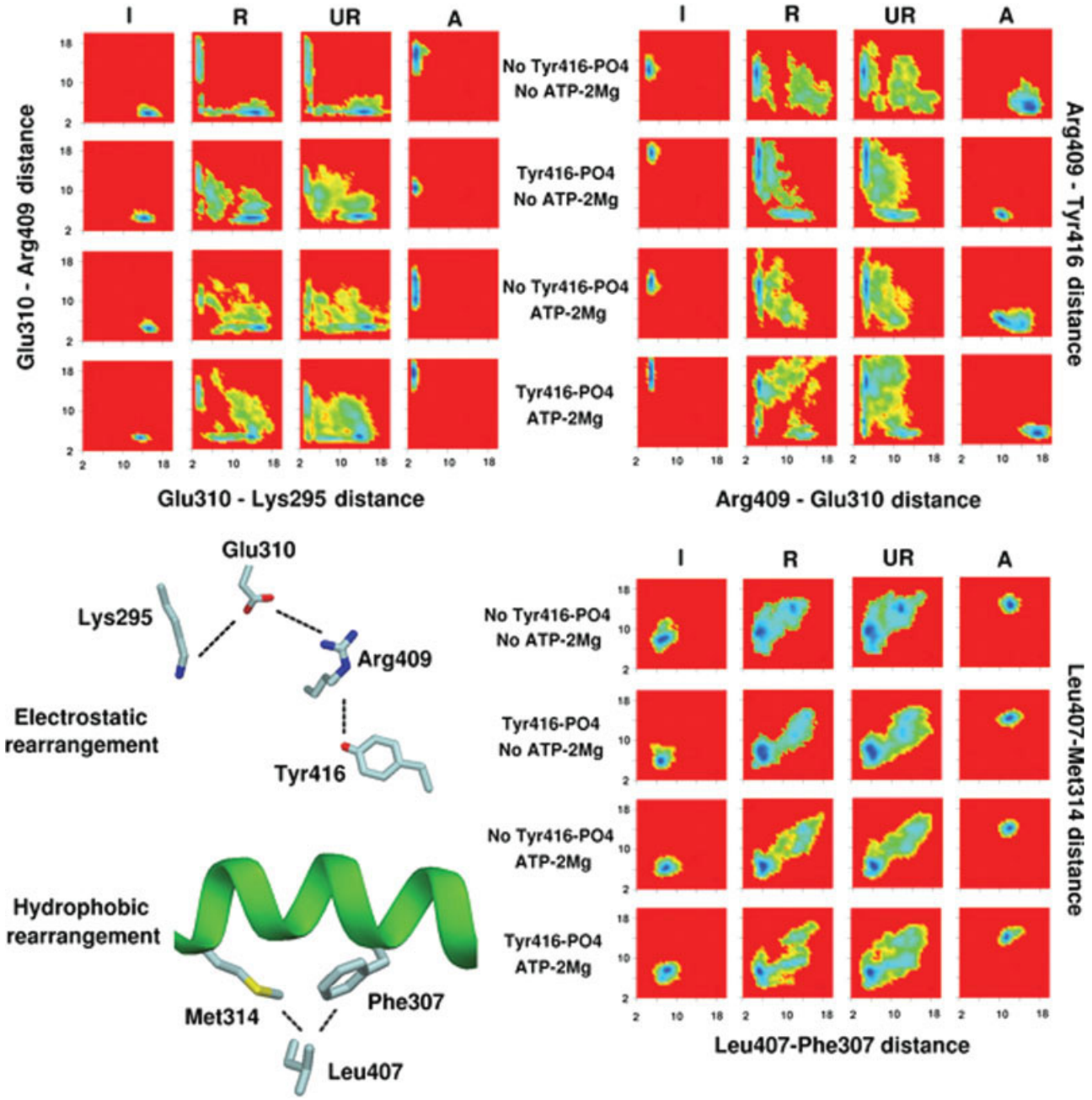


Fig. 7. Free energy surfaces along distance coordinates quantifying various electrostatic or hydrophobic pair interactions crucial to the rotation of the α-C helix and catalytic activation of Hck. The specific distances chosen are shown pictorially on the left. The specific atoms chosen for distance calculations are C-δ of Glu310, C-ξ of Arg409, hydroxyl oxygen of Tyr416, and N-ξ of Lys295. For hydrophobic interactions, distances were calculated between the center of mass of all sidechain non-hydrogen atoms of Phe307, Met314, and Leu407, all distances shown are in Å.

close association. In the presence of ATP-2$Mg^{+2}$, however, there are more configurations in this region indicating the increased likelihood of a state where all three charged residues are in close proximity. Amongst all four systems, it is again the unrestrained MD simulations during activation of the system with both Tyr416 phosphorylated and ATP-2$Mg^{+2}$ present that shows the greatest amount of permissibility in allowing multiple conformational combinations.

The hydrophobic triad formed by Phe307 and Met314 (in the α-C helix) and Leu407 (in the activation loop) is stable in the inactive state equilibration simulations of all four systems. The distances between these associated pairs of hydrophobic residues are typically larger than the effective range of electrostatic interactions (more so because of the present center-of-mass description of the sidechain location of each residue). It is also interesting to note that the spread of the distances around the minimum is quite large

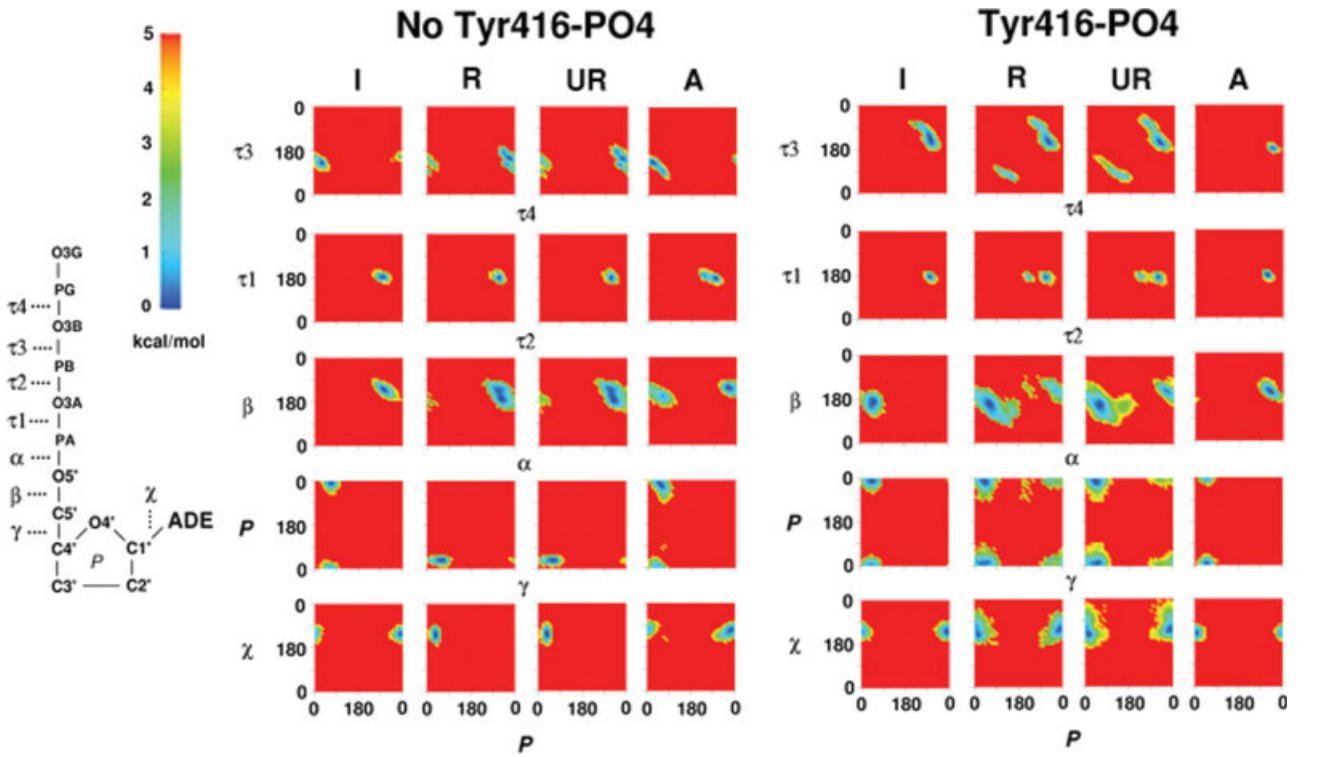


Fig. 8. A condensed view of the conformational dynamics of ATP bound to the catalytic site of Hck, with Tyr416 phosphorylated or unphosphorylated. Pairwise torsional free energy maps (sugar pseudorotation angle *P* vs. χ, γ vs. *P*, α vs. β, I 2 vs. τ1, and τ4 vs. τ3) provide a clear picture of the preferred minima and the torsional landscapes surrounding them that govern ATP flexibility.

(about 4 Å), suggesting that such hydrophobic clusters are dynamical at room temperature. In the active state equilibrations, the Leu407-Phe307 and Leu407-Met314 pairs are both disrupted, creating a passageway used by Glu310 to move towards Lys295. There seem to be two coarse pathways for this disruption, one in which the Leu407-Phe307 pair is disrupted to a greater extent than the Leu407-Met314 pair, and the other in which the opposite is true. The system with both Tyr416 phosphorylation and ATP-$2Mg^{+2}$ seems to prefer the former, whereas the system with no Tyr416 phosphorylation and no ATP-$2Mg^{+2}$ seems to prefer the latter. The other two systems display correlated disruption of both pairs. In all cases, the effective free energy surfaces obtained for the restrained and unrestrained simulations during activation show at least one broad and continuous pathway between the inactive and active state equilibration state extremes.

## ATP Dynamics and Coordination With Phosphorylated Tyr416

A number of studies point to the possibility that changes in the ligand at the ATP-binding site affect the conformational properties of regions involved in activation. Kinetic studies of another protein tyrosine kinase, Csk, have indicated that nucleotide release could be the rate-limiting step in the phosphoryl-transfer reaction.[48] Deuterium-protection experiments indicate that there might be conformational changes in Csk that vary depending on whether the bound nucleotide is ADP or ATP.[49] Time-resolved fluorescence anisotropy decay methods show that binding of substrates have a different effect on the dynamic properties of PKA than binding of products.[50] In Hck, a part of the activation loop is in a disordered state when quercetin is bound in the ATP-binding site[12]; it is ordered when the bound inhibitor is PP1.[13] These results suggest that the presence of ATP with its associated $Mg^{+2}$ ions might therefore be an important variable when considering the activation of the catalytic domain.

The flexibility and preferred conformational state of ATP can be assessed from its dynamics in the present simulations. The starting conformation of ATP-$2Mg^{+2}$ assumed here is the same as the conformation of ATP bound to PKA, which is generally considered to be most representative of kinase catalytic function.[30] The absence of the substrate peptide means that no link can be drawn between the conformation of the ATP observed here and its catalytically competent conformation. The assumption of the presence of 2 $Mg^{+2}$ ions also arises from the same structure but it should be noted that the behavior in the presence of $Mn^{+2}$ or a single $Mg^{+2}$ ion cannot be assumed to be the same as the behavior observed here. Since there is no conformational restraint imposed on ATP itself, any change in its structure is in

response to its environment during the structural activation process. In Figure 8, the torsional energy landscape of ATP is mapped through the 2D free energy surfaces along adjacent torsion pairs. The conformational state of ATP can be described quite accurately by knowing the values of 9 torsional variables: $\chi$, pseudorotation angle *P*, $\gamma$, $\beta$, $\alpha$, $\tau 1$, $\tau 2$, $\tau 3$, and $\tau 4$. In the absence of Tyr416 phosphorylation, ATP shows conformational variability mostly in the $\alpha$, $\tau 3$, and $\tau 4$ torsions associated with the phosphate moieties. In the presence of Tyr416 phosphorylation, all torsional degrees of freedom show much greater variability due to the formation of an explicit interaction between the $\gamma$-phosphate of ATP and the phosphate on Tyr416 that is bridged by a $Mg^{+2}$ ion (see below). This interaction must be broken for the loop to assume its fully open state. The presence of single minima in the torsional degrees of freedom describing the conformational behavior of sugar and base components of ATP indicate that most of the conformational variability in ATP can be attributed to its triphosphate region.

In the inactive state equilibration of the system containing both Tyr416 phosphorylation and ATP-2$Mg^{+2}$, an interesting structural transition was observed to occur spontaneously. The canonical coordination of the two $Mg^{+2}$ ions associated with ATP ($\beta$-$\gamma$ phosphate, and $\alpha$-$\gamma$ phosphate, which is stable in the present MD simulations with no Tyr416 phosphorylation) was disrupted due to the proximity of the phosphorylated Tyr416. One $Mg^{+2}$ ion shifted to coordinate the $\gamma$-phosphate of ATP to the phosphate on Tyr416. The strength of this coordination was sufficient to maintain Tyr416 interaction with ATP until the very final stages of the restrained MD simulations enforcing structural activation. Such a coordination state was not observed to occur spontaneously in the active state equilibration in which the phosphorylated Tyr416 was separated by a large distance from the phosphate of ATP. Such a coordination state could conceivably occur if the activation loop continued to fluctuate after phosphorylation of Tyr416, rather than be locked into an open conformation such as seen in the x-ray structure of Lck.[20]

The stability of the observed coordination state cannot be immediately assessed from the simulations. To address this issue, restrained TMD runs enforcing opening of the activation loop alone (without enforcing $\alpha$-C helix rotation) were carried out with the same protocol used for the original restrained TMD simulations for the system with no Tyr416 phosphorylation and ATP-2$Mg^{+2}$ present. Since no reorganization of the $Mg^{+2}$ ions occurs when Tyr416 is not phosphorylated, this procedure allowed generation of intermediate states of the activation loop opening that were not biased by the observed coordination state. Specific intermediate structures spaced at equal intervals of distance between the hydroxyl oxygen on Tyr416 and the phosphorus atom in the phosphate are chosen randomly from the TMD runs. Umbrella sampling MD simulations were then carried out using these as starting structures for windows along the phosphorus–oxygen distance. The umbrella sampling MD simulations were carried out both without and

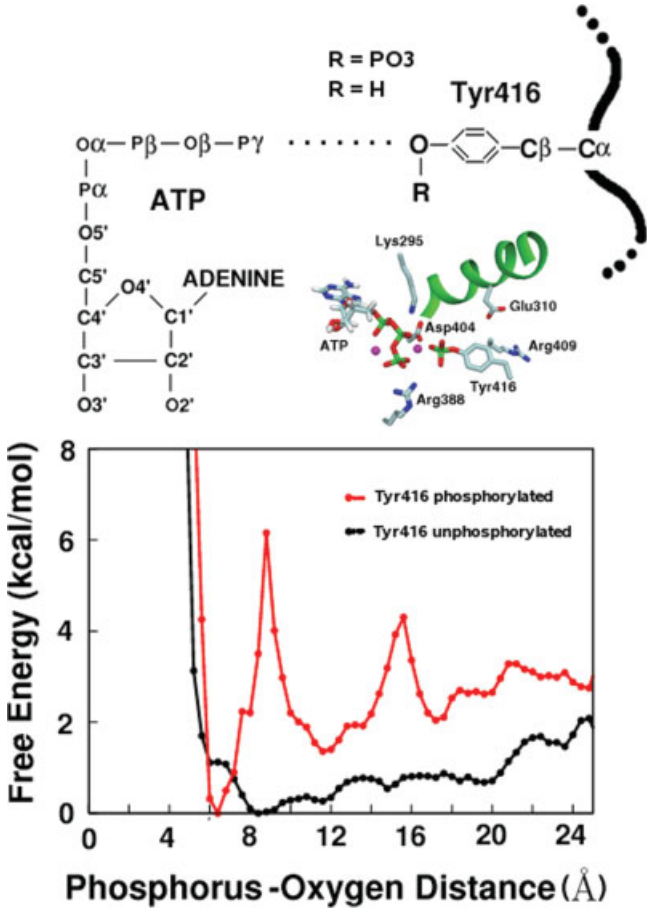


Fig. 9. Free energy profile as a function of the distance between the phosphate of ATP and the hydroxyl oxygen of the Tyr416 sidechain along with a schematic and pictoral representation of the coordination state being tested. This free energy profile is strictly a comparison between Tyr416 having or not having a phosphate group for the same starting structures.

with Tyr416 phosphorylation. Since Tyr416 phosphorylation was added to the starting structures of each umbrella sampling window, the resulting differences can be attributed to the presence of this phosphorylation alone, and not the starting structures themselves. It is important to stress that the absolute height of the free energy barriers in Figure 9 is only suggestive since multiple degrees of freedom other than the distance coordinate were, most likely, not completely sampled in this calculation. Figure 9 shows the free energy profiles along the phosphorus–oxygen distance calculated using the above protocol along with a schematic and pictoral description of the observed coordination state.

The energetic cost of moving the unphosphorylated Tyr416 away from the active site is less than 2 kcal/mol, suggesting that this contact may form and break spontaneously quite easily in solution. When Tyr416 phosphorylation is present, however, there is a stable $Mg^{+2}$-coordinated state of the phosphate of ATP and the phosphate on Tyr416 corresponding to a deep

energy well at a phosphorus–oxygen distance of 6 Å extending up to a distance of more than 8 Å. Three additional wells stabilizing non-ATP-coordinated conformations of phosphorylated Tyr416 are also observed at distances of about 12 Å, 17 Å, and 24 Å, respectively. Although the formation of a strong interaction between phosphorylated Tyr416 and ATP seems counterintuitive, the present results suggest that bridging by a divalent cation could counterbalance the electrostatic repulsion between the two negatively charged groups. According to the free energy profile in presence of Tyr416 phosphorylation, spontaneous trajectories starting from the fully activated open loop conformation would have to traverse at least three minima before the phosphorylated Tyr416 could coordinate to the phosphate of ATP. Though the process would be slow, transient visits to such intermediate states during the μs to ms timescale of protein kinase catalytic turnover[51] cannot be ruled out. Although a reorganization of divalent cations around ATP has not been observed in protein kinase crystal structures, some structural similarities do exist in crystal structures of the small molecule kinases phosphoglycerate kinase, phosphofructokinase, and phosphoenolpyruvate carboxykinase (see Fig. 3 in the supplementary material).

Interestingly, it is known that protein kinase activity decreases upon increase in $Mg^{+2}$ ion concentration beyond a stoichiometry of 1 $Mg^{+2}$ per ATP.[51] However, the mechanism behind this decrease in catalytic activity is not well-understood.[51] A catalytically unproductive $Mg^{+2}$-mediated interaction between phosphorylated Tyr416 and ATP could provide a reasonable structural explanation for the inhibitory effect of an additional stoichiometric $Mg^{+2}$ ion. In addition to its stability, the formation of this state also enables proximity of the residues that require to undergo pair wise electrostatic exchanges (Glu310, Lys295, Arg409, and Tyr416). Although the present results are only suggestive, such metastable intermediate states may need to be considered when formulating the complete catalytic cycle of protein tyrosine kinases.

## Concluding Discussion

The present study uses all-atom MD simulations with explicit solvent to shed light on the conformational transition between the inactive and active states of Hck. A detailed characterization of the behavior before, during (with and without restraints), and after the transformation was carried out on three different length scales. In the coarsest description of the dynamics of the Hck catalytic domain, the relative orientation of the N-lobe with respect to the C-lobe displays a significant amount of conformational variability, presumably due to the flexibility of the hinge region between the two lobes. The presence of ATP-2$Mg^{+2}$ has a discernable impact in making the relative lobe motion bistable.

The second scale of description uses a difference in RMSD to describe correlations in rotation of the α-C helix and opening of the activation loop in going from the inactive state to the active state. At this level, the impact of ATP-2$Mg^{+2}$ and Tyr416 phosphorylation starts to become clearer, especially in terms of the order in which the different motions can occur. The third and highest resolution scale describes the dynamic backbone behavior of individual residues, which brings into stark focus the specific conformational transformations that occur for each amino acid position involved in the structural change accompanying activation. This description clarifies the atomistic details of the specific structural activation processes, such as the rigidity of the α-C helix with respect to the hinge flexibility displayed by specific residues bordering the helical region. It also provides information about the allowed backbone conformational pathways for these local equilibria for each individual residue.

These results highlight the multiscale character and complexity of the structural activation process of tyrosine kinases. The existence of multiple local conformational changes unavoidably raises questions about the order in which they occur and the possibility of variations in this order, which will be the focus of future studies. Strict traditional views of structural properties of proteins implicitly assume an architectural stability, which presume narrow global free energy basins corresponding to crystal structures. While this may be true of very stable domains, the present simulations show how much variation can occur within the inactive state and active state minima for specific residues in Hck, without essentially leaving the basin of a given conformational state.

A plausible scenario explaining bidirectional allostery in Src kinase regulation can be proposed based on the present results. Four distinct conformational motifs (the two activation loop segments, the α-C helix, and the N-terminal end) coupled to each other through their specific nearest-neighbor contacts were chosen to monitor the allosteric transition. Phosphorylation of Tyr416 in the flexible, solvent-exposed region of the activation loop causes a shift in the conformational equilibrium for this loop region. The adjacent region (including the buried segment of the activation loop and the α-C helix) responds to this activation loop opening through changes in specific interactions, which can be remodeled to stabilize the inactive or active states. An example of such specific changes is the exchange of Arg409 between Glu310 and phosphorylated Tyr416 during activation. These changed interactions result in a coupled shift of the conformational equilibrium for both regions, due to a change in their underlying conformational free energy landscape. Activation loop opening can thus be coupled to rotation of the α-C helix. Rotation of the α-C helix in turn affects the conformational equilibrium describing the motion of the N-terminal end hinge[44] also through specific nearest-neighbor interactions between the α-C helix and the N-terminal end, thereby transmitting the effect of Tyr416 phosphorylation to the regulatory SH2 and SH3 domains.

There are two noteworthy features in this model. The first is that since the coupling is only necessary between adjacent conformational motifs, there is no need to invoke long range interactions to explain allostery. For example, the N-terminal end need not have any direct interactions with the activation loop to eventually change its conformation in response to Tyr416 phosphorylation, a shift in the conformation of the buried segment of the loop and the α-C helix mediates the response. The second feature is that since the coupling between adjacent secondary structure motifs is postulated to occur through changes in their underlying conformational free energy landscape, the coupling need not have any inherent directionality. Hinge motion in the N-terminal end can therefore cause rotation of the α-C helix, just the same as the rotation of the α-C helix can result in hinge motion in the N-terminal end. This lack of directionality is compatible with the bidirectional nature of the allostery that regulates Src kinase function.[23]

The existence of a metastable $Mg^{+2}$-coordination state between phosphorylated Tyr416 and the γ-phosphate of ATP is suggested by the present calculations. Though counterintuitive, such a state is chemically reasonable and was shown to be stable using umbrella sampling calculations. It is presently difficult to predict the significance of such intermediate states in the catalytic cycle of Hck and more work will be needed to further characterize it.

Inhibitors of protein tyrosine kinases[52,53] mostly bind through hydrophobic interactions in the ATP-binding site and have considerable therapeutic and therefore commercial value. The structural activation process involves hydrophobic rearrangements, which can probably affect the binding of the small molecule inhibitors. The robustness of the binding during such dynamic structural changes, and indeed the ability to hamper such changes, is probably important to the inhibition capacity of these small molecule inhibitors. The present study of conformational variability in kinase activation is therefore likely to suggest alternate conformations against which the binding affinity of these inhibitors could be tested.

The SH2 and SH3 domains were not included in the present simulations. This choice is motivated by the need to first characterize the specific moving parts within the catalytic domain. The SH2 and SH3 domains constitute additional degrees of freedom in the Src kinase machinery that can be coupled to the various moving parts in the catalytic domain. The importance of such coupling is not completely clear and remains to be assessed. For example, the isolated Lck catalytic domain has been demonstrated to be active without the SH2 and SH3 domains[20] and SH2 or SH3 deletion mutants show increased kinase activity.[54] However, a recent X-ray structure of a c-Src complex in an active conformation suggests possible interactions of the SH2 and SH3 with the catalytic domain,[55] and the SH2 and SH3 domains interact directly with the catalytic domain to stabilize the active state structure of Csk, a closely related kinase (35% homology to Src).[56] The SH2 domain also interacts with the C-lobe of the catalytic domain in the inactive assembled structure of Hck and c-Src.[12–14] Clearly, further studies will be needed to better understand the influence of direct interactions of the catalytic domain with the regulatory domains.

The present results, together with those from a previous study focused on the N-terminal end linker region of the catalytic domain[44] suggest a dynamic view of allosteric regulation of the various functional states of the enzyme. A number of structural elements maintain their internal conformation and behave almost as rigid modules that are re-organized relative to one another (e.g., the SH2 and SH3 domains, the α-C helix in the N-lobe). In contrast, other elements can transform their internal conformation significantly (e.g., the N-terminal end of catalytic domain, the activation loop, the linker between the SH2 and SH3 domains). Such "transformative" structural elements (typically corresponding to relatively short peptide regions) display a bi-stable (or multi-stable) character[44]—the conformational free energy surface is a useful concept to display this feature.[44] The ability of the transformative elements to switch between different stable conformational states can lead to long-range spatial rearrangement of the (relatively stable) "modular" elements with respect to one another. This view is consistent with the accepted understanding of allostery as a shift in equilibrium between two overall minima[21] and provides a useful physical framework to seek the causative factors in the long-range allosteric response of a complicated macromolecular assembly.

## ACKNOWLEDGMENTS

We would like to thank Carol Post, John Kuriyan, Susan Taylor, Guillaume Lamoureux, Simon Bernèche, Toby Allen, Hyung-June Woo, Yuqing Deng, Sergei Noskov, José Faraldo-Gomez, and Deniz Sezer for helpful discussions. This work was supported financially by NIH grant CA093577 and by Keck Postdoctoral Fellowship for NKB. Computational support from the Pittsburgh Supercomputing Center (PSC) obtained through the National Resource Allocation Committee (NRAC) was used for the calculations.

## REFERENCES

1. Brown M, Cooper J. Regulation, substrates and functions of src. Biochim Biophys Acta 1996;1287;121–149.
2. Sicheri F, Kuriyan J. Structures of Src-family tyrosine kinases. Curr Opin Struct Biol 1997;7:777–785.
3. Waksman G, Shoelson SE, Pant N, Cowburn D, Kuriyan J. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. Cell 1993;72:779–790.
4. Zheng XM, Wang Y, Pallen CJ. Cell transformation and activation of pp60c-src by overexpression of a protein tyrosine phosphatase. Nature 1992;359:336–339.
5. Wu X, Knudsen B, Feller SM, Zheng J, Sali A, Cowburn D, Hanafusa H, Kuriyan J. Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. Structure 1995;3:215–226.

6. Moarefi I, LaFevre-Bernt M, Sicheri F, Huse M, Lee C, Kuriyan J, Miller W. Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. Nature 1997;385:650–653.
7. Azarnia R, Reddy S, Kmiecik TE, Shalloway D, Loewenstein WR. The cellular src gene product regulates junctional cell-to-cell communication. Science 1988;239:398–401.
8. Porter M, Schindler T, Kuriyan J, Miller W. Reciprocal regulation of Hck activity by phosphorylation of Tyr(527) and Tyr(416). Effect of introducing a high affinity\intramolecular SH2 ligand. J Biol Chem 2000;275:2721–2726.
9. Ziegler SF, Marth JD, Lewis DB, Perlmutter RM. Novel protein-tyrosine kinase gene (hck) preferentially expressed in cells of hematopoietic origin. Mol Cell Biol 1987;7:2276–2285.
10. Hu Y, Liu Y, Pelletier S, Buchdunger E, Warmuth M, Fabbro D, Hallek M, Etten RAV, Li S. Requirement of Src kinases Lyn, Hck and Fgr for BCR-ABL1-induced B-lymphoblastic leukemia but not chronic myeloid leukemia. Nat Genet 2004;36:453–461.
11. Podar K, Mostoslavsky G, Sattler M, Tai YT, Hayashi T, Catley LP, Hideshima T, Mulligan RC, Chauhan D, Anderson KC. Critical role for hematopoietic cell kinase (Hck)-mediated phosphorylation of Gab1 and Gab2 docking proteins in interleukin 6-induced proliferation and survival of multiple myeloma cells. Biol J Chem 2004;279:21658–21665.
12. Sicheri F, Moarefi I, Kuriyan J. Crystal structure of the Src family tyrosine kinase Hck. Nature 1997;385:602–609.
13. Schindler T, Sicheri F, Pico A, Gazit A, Levitzki A, Kuriyan J. Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor. Mol Cell 1999;3:639–648.
14. Xu W, Harrison S, Eck M. Three-dimensional structure of the tyrosine kinase c-Src. Nature 1997;385:595–602.
15. Xu W, Doshi A, Lei M, Eck M, Harrison S. Crystal structures of c-Src reveal features of its autoinhibitory mechanism. Mol Cell 1999;3:629–638.
16. Taylor SS, Radzio-Andzelm E. 3 Protein-kinase structures define a common motif. Structure 1994;2:345–355.
17. Hubbard S. Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. EMBO J 1997;16:5573–5581.
18. Hubbard SR, Wei L, Ellis L, Hendrickson WA. Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature 1994;372:746–754.
19. Huse M, Kuriyan J. The conformational plasticity of protein kinases. Cell 2002;109:275–282.
20. Yamaguchi H, Hendrickson W. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. Nature 1996;384:484–489.
21. Kern D, Zuiderweg ER. The role of dynamics in allosteric regulation. Curr Opin Struct Biol 2003;13:748–757.
22. Young M, Gonfloni S, Superti G-Furga, Roux B, Kuriyan J. Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. Cell 2001;105:115–126.
23. Gonfloni S, Williams J, Hattula K, Weijland A, Wierenga R, Superti-Furga G. The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src. EMBO J 1997;16:7261–7271.
24. Mendieta J, Gago F. In Silico activation of Src tyrosine kinase reveals the molecular basis for intromolecular phosphorylation. J Mol Graph Model 2004;23:189–198.
25. Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy minimization and dynamics calculations. J Comput Chem 1983; 4:187–217.
26. MacKerell AD, Jr, Bashford D, Bellot M, Dunbrack R, Evanseck J, Field M, Fischer S, Gao J, Guo H, Ha DJ-MS, Kuchnir L, Kuczera K, Lau F, Mattos C, Michnick S, Ngo T, Nguyen D, Prodhom B, Reiher W III, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102: 3586–3616.
27. Pavelites JJ, Gao JL, Bash PA, Mackerell AD. A molecular mechanics force field for NAD(+), NADH, and the pyrophosphate groups of nucleotides. J Comp Chem 1997;18:221–239.
28. Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
29. Marti-Renom MA, Stuart A, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Ann Rev Biophys Biomol Struct 2000;29:291–325.
30. Zheng JH, Trafny EA, Knighton DR, Xuong NH, Taylor SS, Ten-eyck LF, Sowadski JM. 2.2-Angstrom refined crystal-structure of the catalytic subunit of CAMP-dependent protein-kinase complexed with MnATP and a peptide inhibitor. Acta Crystallographica 1993;49:362–365.
31. Field MJ, Karplus M. CRYSTAL: program for crystal calculations in CHARMM. Cambridge, MA: Harvard University; 1992.
32. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comp Phys 1977; 23:327–341.
33. Darden T, York D, Pedersen L. Particle Mesh Ewald—an N•log(N) method for ewald sums in large systems. J Chem Phys 1993;98:10089–10092.
34. Beglov D, Roux B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. J Chem Phys 1994;100:9050–9063.
35. Feller SE, Zhang YH, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation—the Langevin piston method. J Chem Phys 1995;103:4613–4621.
36. Banavali NK, Roux B. Free energy landscape of A-DNA to B-DNA conversion in aqueous solution J Am Chem Soc 2005;127: 6866–6876.
37. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 1992;13:1011–1021.
38. Souaille M, Roux B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. Comput Phys Commun 2001;135:40–57.
39. Karlsson R, Zheng JH, Xuong NH, Taylor SS, Sowadski JM. Structure of the mammalian catalytic subunit of cAMP-dependent protein-kinase and an inhibitor peptide displays an open conformation. Acta Crystallographica D 1993;49:381–388.
40. Zheng J, Knighton D, Xuong N, Taylor S, Sowadski J, Eyck LT. Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. Protein Sci 1993;2:1559–1573.
41. Johnson DA, Akamine P, Radio E-Andzelm, Madhusudan M, Taylor SS. Dynamics of cAMP-dependent protein kinase. Chem Rev 2001;101:2243–2270.
42. Kabsch W. A solution for the best rotation to relate two sets of vectors. Acta Crystallogr A 1976;32:922–923.
43. Kabsch W. A discussion of the solution for the best rotation to relate two sets of vectors. Acta Crystallogr A 1978;34:827–828.
44. Banavali NK, Roux B. The N-terminal end of the Catalytic Domain of Src kinase Hck is a conformational switch implicated in long-range allosteric regulation. Structure 2005;13:1715–1723.
45. Tsigelny I, Greenberg JP, Cox S, Nichols WL, Taylor SS, Ten Eyck LF. 600 ps molecular dynamics reveals stable substructures and flexible hinge points in cAMP dependent protein kinase. Biopolymers 1999;50:513–524.
46. Johnson L, Lewis RJ. Structural basis for control by phosphorylation. Chem Rev 2001;101:2209–2242.
47. Ozrikirimli E, Post CB. Src kinase activation: a switched electrostatic network. Protein Sci 2006;15:1051–1062.
48. Shaffer J, Sun G, Adams JA. Nucleotide release and associated conformational changes regulate function in the COOH-terminal Src kinase, Csk. Biochemistry 2001;40:11149–11155.
49. Hamuro Y, Wong L, Shaffer J, Kim JS, Stranz DD, Jennings PA, Woods VL, Jr, Adams JA. Phosphorylation driven motions in the COOH-terminal Src kinase, CSK, revealed through enhanced hydrogen-deuterium exchange and mass spectrometry (DXMS). J Mol Biol 2002;323:871–881.
50. Li F, Gangal M, Juliano C, Gorfain E, Taylor SS, Johnson DA. Evidence for an internal entropy contribution to phosphoryl transfer: a study of domain closure, backbone flexibility, and the catalytic cycle of cAMP-dependent protein kinase. J Mol Biol 2002; 315:459–469.

51. Adams JA. Kinetic and catalytic mechanisms of protein kinases. Chem Rev 2001;101:2271–2290.
52. Bridges AJ. Chemical Inhibitors of Protein Kinases. Chem Rev 2001;101:2541–2571.
53. Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, Miller WT, Clarkson B, Kuriyan J. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571)1. Cancer Res 2002;62:4236–4243.
54. Seidel-Dugan C, Meyer BE, Thomas SM, Brugge JS. Effects of SH2 and SH3 deletions on the functional activities of wild-type and transforming variants of c-Src. Mol Cell Biol 1992;12:1835–1845.
55. Cowan-Jacob SW, Fendrich G, Manley PW, Jahnke W, Fabbro D, Liebetanz J, Meyer T. The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. Structure 2005;13:861–871.
56. Ogawa A, Takayama Y, Sakai H, Chong KT, Takeuchi S, Nagakawa A, Nada S, Okada M, Tsukihara T. Structure of the carboxy-terminal Src kinase, Csk. J Biol Chem 2002;277:14351–14354.