

The N-Terminal End of the Catalytic Domain of Src Kinase Hck Is a Conformational Switch Implicated in Long-Range Allosteric Regulation

Nilesh K. Banavali and Benoît Roux*
Department of Physiology and Biophysics
Weill Medical College of Cornell University
1300 York Avenue
New York, New York 10021

Summary

Signal transduction in cell growth and proliferation involves regulation of kinases through long-range allostery between remote protein regions. Molecular dynamics free energy calculations are used to clarify the coupling between the catalytic domain of Src kinase Hck and its N-terminal end connecting to the regulatory SH2 and SH3 modules. The N-terminal end is stable in the orientation required for the regulatory modules to remain properly bound only in the inactive catalytic domain. In the active catalytic domain, the N-terminal end prefers a different conformation consistent with dissociation of the regulatory modules. The free energy surface shows that the N-terminal end acts as a reversible two-state conformational switch coupling the catalytic domain to the regulatory modules. Structural analogy with insulin receptor kinase and c-Src suggests that such reversible conformational switching in a critical hinge region could be a common mechanism in long-range allosteric regulation of protein kinase activity.

Introduction

Src kinases are membrane-associated nonreceptor tyrosine kinases consisting of three major components: the catalytic domain sufficient for kinase activity with the conserved architecture first seen in cAMP-dependent protein kinase (PKA) (Taylor and Radzio-Andzelm, 1994); the SH2 regulatory binding module; and SH3 regulatory binding module (Sicheri and Kuriyan, 1997). Phosphorylation or dephosphorylation of two critical tyrosine residues, Tyr416 and Tyr527 (chicken c-Src numbering), by regulatory protein kinases or phosphatases modulates Src kinase activity and coordinates multiple signal transduction pathways in cell differentiation and proliferation (Brown and Cooper, 1996). Crystallographic studies revealed the essential structural features of Src kinases in inactive (Sicheri et al., 1997; Xu et al., 1997, 1999; Schindler et al., 1999) and the active (Yamaguchi and Hendrickson, 1996) states. A fully assembled inactive Src kinase illustrated by the X-ray structure of Hck or c-Src shows the activation loop with an unphosphorylated Tyr416 folded back into the catalytic site, the SH2 domain bound to a phosphorylated Tyr527 residue in the C-terminal tail region, and the SH3 domain bound to a proline-containing linker between the SH2 and catalytic domains (Schindler et al., 1999; Xu et al., 1999). Tyr416 in the activation loop is

separated from Tyr527 in the C-terminal tail by about 40 Å (Schindler et al., 1999; Xu et al., 1999), yet their respective phosphorylations exert opposite regulatory effects. Tyr416 phosphorylation increases enzyme activity, while Tyr527 phosphorylation is inhibitory (Porter et al., 2000). The X-ray structure of the catalytic domain of Lck (Yamaguchi and Hendrickson, 1996) characterizes the activated state in which the activation loop containing phosphorylated Tyr416 moves away from the active site, and the α -C helix rotates inward to properly position key catalytic residues in the kinase active site. In contrast with structural variability observed in kinase-inactivated states, this activated state seems to be structurally conserved amongst several protein kinases (Huse and Kuriyan, 2002).

A number of events, such as dephosphorylation of Tyr527 in the C-terminal tail, binding of a substrate peptide to the SH2 or SH3 domains, or phosphorylation of Tyr416 in the activation loop, can lead to increased activity, accompanied by a substantial change in the structure of the enzyme (Sicheri and Kuriyan, 1997). Biochemical experiments show that Tyr416 phosphorylation increases the availability of the SH2 and SH3 domains to interact with the phosphotyrosine-containing p-YEEI peptide and the polyproline-containing HIV 1-Nef protein, respectively (Porter et al., 2000). The HIV 1-Nef interaction with the SH3 domain can override the inhibitory effect of the SH2 domain interaction with phosphorylated Tyr527 (Moarefi et al., 1997). Disruption of the structural properties of the short linker between the SH2 and SH3 domains by introduction of multiple glycine mutations in c-Src reduces the inhibitory effect of SH2 domain interaction with phosphorylated Tyr527 (Young et al., 2001). Early proteolysis studies with trypsin digestion showed that the susceptibility of the proteolysis site near Trp260 of c-Src increased upon Tyr527 dephosphorylation, indicating that this region represented a bridge between the kinase domain and the regulatory domains (MacAuley and Cooper, 1989). Alanine mutation of the strictly conserved Trp260 residue in the N-terminal end region affects regulation by increasing basal kinase activity, accelerating autophosphorylation, decreasing sensitivity to regulatory domain ligands, and changing accessibility of ligand binding surfaces for the regulatory domains (Lafevre-Bernt et al. 1998), even with inhibitory Tyr527 phosphorylation being present (Gonfloni et al., 1997).

When considered together, these results suggest that the peptide chain linking the N-terminal end of the catalytic domain to the SH2 domain forms part of a long-range allosteric pathway coupling the activation loop of the kinase domain and Tyr527 in its C-terminal tail. The currently available X-ray structures provide essential information about the active (Yamaguchi and Hendrickson, 1996; Cowan-Jacob et al., 2005) and inactive (Schindler et al., 1999; Xu et al., 1999) catalytic domain conformations of the kinases. This crystallographic structural analysis also indicates the conformational change in the N-terminal linker region surrounding Trp260, accompanying the activation of the catalytic

*Correspondence: benoit.roux@med.cornell.edu

domain, thereby defining the “end points” of the conformational change for the N-terminal end. However, the exact mechanism by which information about the conformational state of the activation loop is allosterically transmitted to the distant SH2 domain along the peptide chain of the linker, and whether this process is concerted or not, remains unclear. To address this issue, the free energy surface governing the conformation of the N-terminal end linker (residues 253–272, chicken c-Src numbering) was calculated by using molecular dynamics umbrella sampling simulations with explicit solvent for two different activation states of the catalytic domain.

The calculated free energy profiles, describing a conformational change at the N-terminal end, show that the N-terminal end conformation is clearly dependent on the conformational state of the rest of the catalytic domain. Dissociation of the SH2 module from the C-terminal tail can be linked to the activation of the catalytic domain simply through a change in the free energy landscape affecting the preferred conformation of the N-terminal end linker. Structural analogy with insulin receptor kinase (IRK) and c-Src suggest that the ability of the N-terminal end to act as a reversible two-state conformational switch mediating the communication between the catalytic domain and the regulatory modules may be a mechanism generally utilized for long-range allostery in protein kinases.

Results and Discussion

Free Energy Profile of N-Terminal End Release

The catalytic domain used in the present study (residues 253–523, chicken c-Src numbering) can be conceptually divided into two dynamic regions whose states are expected to be closely coupled: the N-terminal end comprising residues 253–272, and the remaining body of the catalytic domain comprising residues 273–523. The N-terminal end can possibly be either in a so-called “restricted” conformation, as observed in the crystal structure of the fully assembled inactive Hck (Schindler et al., 1999), or in a “released” conformation, as observed in the crystal structure of active Lck (Yamaguchi and Hendrickson, 1996). The main body of the catalytic domain can be either in an inactive or active conformation, also corresponding to the Hck (Schindler et al., 1999) and Lck (Yamaguchi and Hendrickson, 1996) kinase crystal structures, respectively. Four combined end-point states are possible for the two coupled regions: restricted-inactive, restricted-active, released-inactive, and released-active. The model of the restricted-inactive state corresponds to the crystal structure of Hck (Schindler et al., 1999), whereas the model of the released-active state was generated by homology modeling on the basis of the crystal structure of Lck (Yamaguchi and Hendrickson, 1996) followed by equilibration by using molecular dynamics with explicit solvent. The other end point states, i.e., released-inactive and restricted-active, were generated by using restrained MD simulations.

The free energy profile monitoring the progress between the restricted and the released conformations of the N-terminal end is shown in Figure 1A for the active and inactive catalytic domains. When the catalytic do-

main is inactive, the N-terminal end is stabilized in a deep free energy well corresponding to the restricted state. A secondary free energy well corresponding to the released conformation is also observed. When the catalytic domain is in the active state, the free energy well corresponding to the restricted state disappears, and the well corresponding to the released state is considerably broadened and shifted. Structural activation of the catalytic domain is thus reflected as a fundamental change in the character of the free energy surface governing the dynamic conformational freedom of its N-terminal end. Since these features exist even in the absence of the SH2 and SH3 domains, ATP, internal tyrosine phosphorylation, and substrate peptide, they appear to be an intrinsic architectural feature coupling the N-terminal linker and the main body of the catalytic domain. The recent crystal structure of unphosphorylated c-Src also shows N-terminal linker structural changes in the same region in conjunction with the catalytic domain adopting an active conformation (Cowan-Jacob et al., 2005). Figure 1B shows the two-dimensional free energy surface along the rmsd for the N-terminal end residues from their restricted and released end point states on the x and y axes, respectively. This two-dimensional map as well as other local structural parameters are estimated through the sampling obtained during determination of the original one-dimensional free energy profile, and not through separate individual free energy estimations. Even with this qualification, such a two-dimensional map is more informative because it offers greater structural resolution than the original one-dimensional free energy profile. The area of conformational space, corresponding to an energy minimum for the inactive catalytic domain, changes to a high-energy region for the active catalytic domain, whereas a previous high free energy area changes conversely to a free energy well.

Additional validation of the incompatibility of the restricted conformation of the N-terminal end with the active state of the catalytic domain is provided by unconstrained simulations starting from the restricted-active state generated by targeted MD simulations by using constraints on the main body of the catalytic domain. As shown in Figure 2, the ΔD_{rmsd} parameter for all heavy atoms belonging to residues 253–272 evolves spontaneously from a value closer to the restricted-active state (ΔD_{rmsd} around -3 \AA) to a value closer to the released-active state (ΔD_{rmsd} around 2 \AA) within 1 ns. The spontaneous structural change causing this ΔD_{rmsd} change mostly spans the residues Pro254 (showing a large translational motion, but no internal change) through Trp260 (showing a marginal internal change, and very little translational motion), corresponding exactly to the behavior during restrained umbrella sampling MD simulations by using constraints on the N-terminal end shown in the bottom panel of Figures 1A and 1B.

If the SH2 and SH3 regulatory domains (not included in the present calculations because of the much larger size of the resulting system and the prohibitively large computational cost) were added back with their original orientation with respect to the N-terminal end for the minimum energy released-active state, they would undergo a dramatic displacement. The two assumptions implicit in this reconstruction are that the present results on

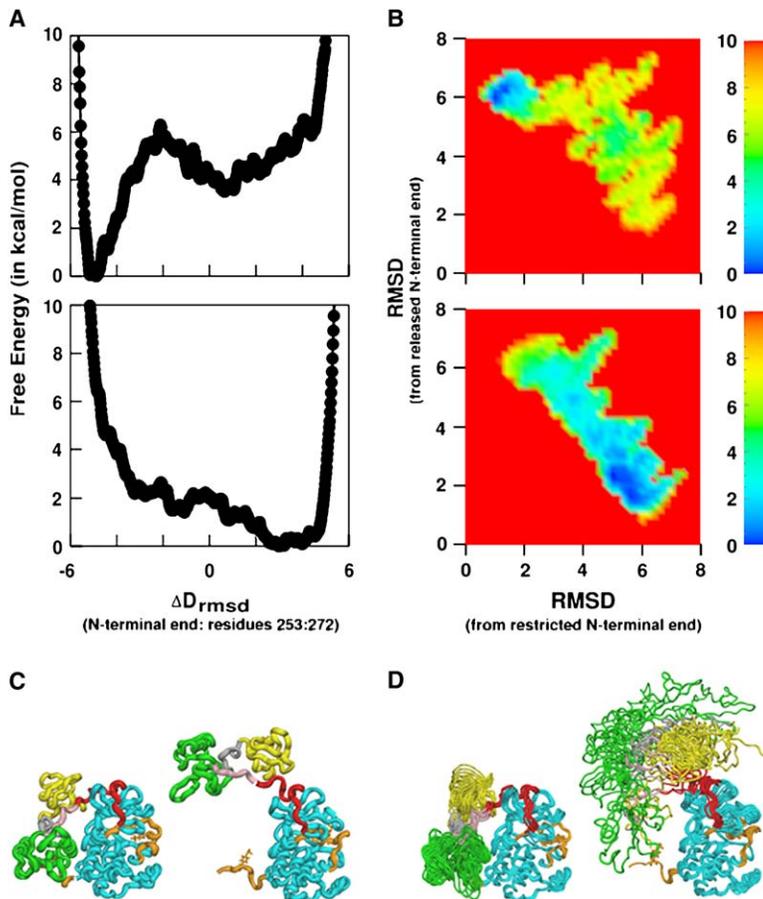


Figure 1. Free Energy Profiles and Structures Indicating the N-Terminal End Conformational Switch

(A) The one-dimensional free energy profile along the ΔD_{rmsd} order parameter.
 (B) The two-dimensional free energy surface along the rmsd from the restricted state (x axis) and the rmsd from the released state (y axis) (in both [A] and [B]: top panel, inactive catalytic domain; bottom panel, active catalytic domain).
 (C) The most probable conformations with SH2 and SH3 domains added back assuming their original orientation with respect to the N-terminal end.
 (D) Nine conformations overlaid with respect to the catalytic domain equally spaced in relative free energy value from 0 kcal/mol to 2 kcal/mol (in both [C] and [D]: left, inactive catalytic domain; right, active catalytic domain); the inactive catalytic domain restricts the N-terminal end motion much more than the active catalytic domain. The constrained N-terminal residues 253–273 with Trp260 are shown in red, the SH2 domain is shown in green, the SH3 domain is shown in yellow, the catalytic domain is shown in cyan, the activation loop and C-terminal tail are shown in orange, and ΔD_{rmsd} is given in Å. Free energy axes values are in kcal/mol.

N-terminal end dynamics are not affected greatly by the addition of the regulatory modules, and that the SH2-kinase polyproline linker together with the SH3 and SH2 modules move as a relatively rigid unit in response to motion of the N-terminal end. While corroboration of the first assumption awaits its computational feasibility in the future, the recent unphosphorylated c-Src structure does support the second assumption by showing relatively rigid motion of the SH2 and SH3 domains together in response to a localized structural change at the N-terminal end (Cowan-Jacob et al., 2005).

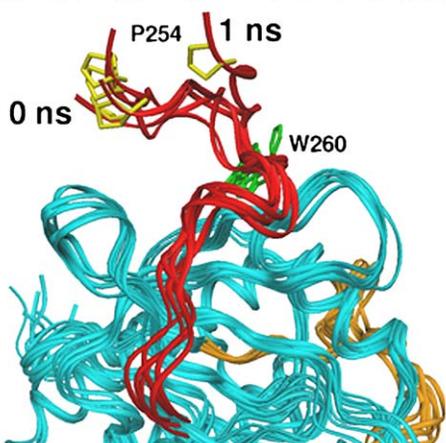
As seen in Figure 1C, the most probable conformation of the N-terminal linker in the inactive catalytic domain positions the SH3 and SH2 domains as in the assembled, inactive state crystal structure (Schindler et al., 1999). In contrast, the most probable conformation of the N-terminal end, when the catalytic domain is active, is not consistent with the required orientation of SH3 and SH2 domains for the assembled state to form. With the inactive catalytic domain, the N-terminal end cannot escape its restricted state conformation without crossing a barrier greater than 5 kcal/mol, effectively locking it in this conformation. With the active catalytic domain, a wide range of conformations around the most preferred released state of the N-terminal end seem to be thermally accessible. In Figure 1D, nine overlaid conformations chosen from regions of the two-dimensional rmsd map corresponding to relative free energy values ranging equally between 0 kcal/mol and 2 kcal/mol for the inactive and active catalytic domains

are shown with regulatory domains added back. This flexibility of the N-terminal end can be hypothesized to be important in reversing the activation of the catalytic domain by allowing the SH2 domain to return in proximity of Tyr527 in the C-terminal tail, when the catalytic domain is in its active conformation. A stable interaction could then form between the phosphorylated Tyr527 in the C-terminal tail and the wandering SH2 domain. The N-terminal end would consequently be pushed toward its restricted conformation, thereby shifting the equilibrium of the catalytic domain toward its inactive state. Such bidirectional communication may result in reversible, long-range allosteric regulation of the kinase.

Localized Mechanics of the Hinge Region

Modest changes in the N-terminal end of the catalytic domain can cause significant reorientation of the regulatory modules; therefore, rather than simply acting as a passive flexible linker undergoing an order-to-disorder transition, the N-terminal end possibly acts as an active hinge that can be switched from one conformation to another depending on the state of the catalytic domain. The residues that critically control this hinge should be important in regulation of kinase activity. To investigate the behavior of individual residues contributing to the N-terminal end hinge motion, free energy maps of backbone (ϕ, ψ) torsions and side chain (χ_1, χ_2) torsions were obtained by using Equations 3–6 (shown in Figure 3). The behavior of specific residues for the same N-terminal end motion is noticeably different for the

Unconstrained N-terminal end release



(N-terminal end: residues 253:272, shown in red)

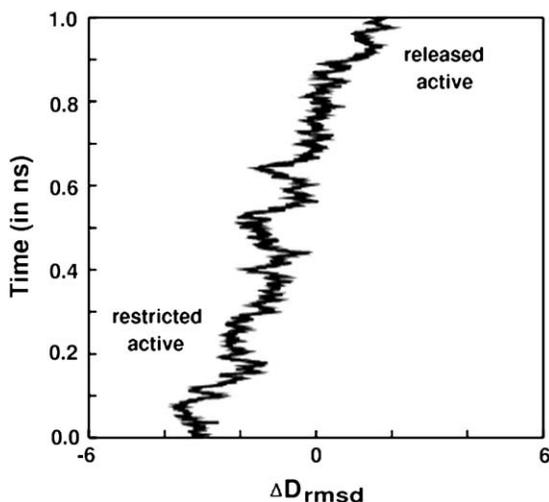


Figure 2. Spontaneous Relaxation of the N-Terminal End from the Restricted-Active State to the Released-Active State

The change in the ΔD_{rmsd} parameter occurring spontaneously during a 1 ns unconstrained MD simulation starting from the restricted-active state (generated by targeted MD through constraints on the main body of the catalytic domain) indicates that the restricted state of the N-terminal end is unstable in the active catalytic domain structure and drifts toward the released state. Six overlaid structural snapshots from 0.0 ns, 0.2 ns, 0.4 ns, 0.6 ns, 0.8 ns, and 1.0 ns shown on top indicate the structural change accompanying N-terminal end release. The N-terminal end (residues 253–272) is shown in red, Trp260 is shown in green, Pro254 is shown in yellow, the catalytic domain is shown in cyan, the activation loop is shown in orange, and ΔD_{rmsd} is given in Å.

inactive and active catalytic domains. These internal effects are focused between the mostly unaffected Pro254 residue and the partially affected Glu261 residue. The rotamers describing the exact conformations of the minima for each affected N-terminal end residue in the inactive and active catalytic domains are shown in Table S1 (see the Supplemental Data available with this article online). Considering both sequence conservation and conformational changes, the most interesting residues are Trp255, Asp258, and Trp260. By looking at these residues individually, it is possible to clarify their respective

roles in the N-terminal end switch. Hydrophobic residues occupying the sequence position of Trp255 are known to be important in the regulation of Src kinases (Gonfloni et al., 1999). In the active catalytic domain, Trp255 dissociates from a hydrophobic pocket where it is bound in the inactive catalytic domain. Upon activation, its preferred backbone conformation changes from (ϕ, ψ) coordinates of $(-75, 150)$ to $(-120, -30)$, which is a β sheet-to- α helix transition in the (ϕ, ψ) Ramachandran map. Its preferred side chain conformation also shows a shift in (χ_1, χ_2) coordinates of $(-165, 90)$ to $(60, 105)$. This residue seems to be involved in hydrophobic stabilization of the N-terminal end in its restricted state when the main body of the catalytic domain is in its inactive state. In the active state of the main body of the catalytic domain, an internal conformational change accompanies its dissociation from its hydrophobic pocket that is distorted due to rotation of the α -C helix during activation.

Asp258 is conserved among all nine members of the Src kinase family (Sicheri and Kuriyan, 1997). The unusual left-handed α helix (ϕ, ψ) backbone coordinates $(75, 0)$ are more stable in the inactive catalytic domain, whereas the more typical (ϕ, ψ) coordinates $(-105, -150)$ are more stable in the active catalytic domain. Its side chain conformational preference also changes correspondingly, with the (χ_1, χ_2) coordinate minimum shifting from $(-60, 120)$ to $(-165, -105)$. Dissociation of the Lys257-Asp258 side chain H bond at the center of the hinge region accompanies these internal changes. The distance between the two side chains goes from 2–5 Å (hydrogen bonded) to 10–12 Å (dissociated). The free energy profile of this hydrogen bond indicates that it is stable in the inactive catalytic domain (black line) and unstable in the active catalytic domain (red line) (Figure 3C). Instead, three consecutive peptide moieties (residues 259–261) reorient their backbone amide groups to coordinate with the free Asp258 carboxylate. The free energy profile as a function of the distance between an Asp258 carboxylate oxygen and the center of mass of the three backbone amide N-H groups shows that this interaction is stable only in the active catalytic domain (red line) (Figure 3D). The choice for Asp258 seems to be between an internally strained hydrogen bond with adjacent Lys257 or an unstrained electrostatic interaction with reoriented backbone amide groups of three consecutive adjacent residues. These two conformations occur in response to Trp260 side chain motion caused by α -C helix rotation during structural activation of the catalytic domain. The inability of other amino acid side chains to participate in this intricate balancing act linked to the released conformation of the N-terminal end may explain why aspartate is completely conserved at this position in Src kinases.

The highly conserved Trp260 is well known by site-specific mutagenesis to have an important impact on regulation (Gonfloni et al., 1997; Lafevre-Bernt et al., 1998). The dissociation of the Trp260 side chain occurs in conjunction with α -C helix motion in the active catalytic domain that changes its hydrophobic binding pocket (Sicheri et al., 1997). The Trp260 side chain makes a (χ_1, χ_2) coordinate transition from $(-60, 120)$ to $(-165, -105)$, but the backbone only makes a small adjustment from (ϕ, ψ) coordinates of $(-105, -60)$ to $(-105, -30)$. The lack of drastic change in the backbone

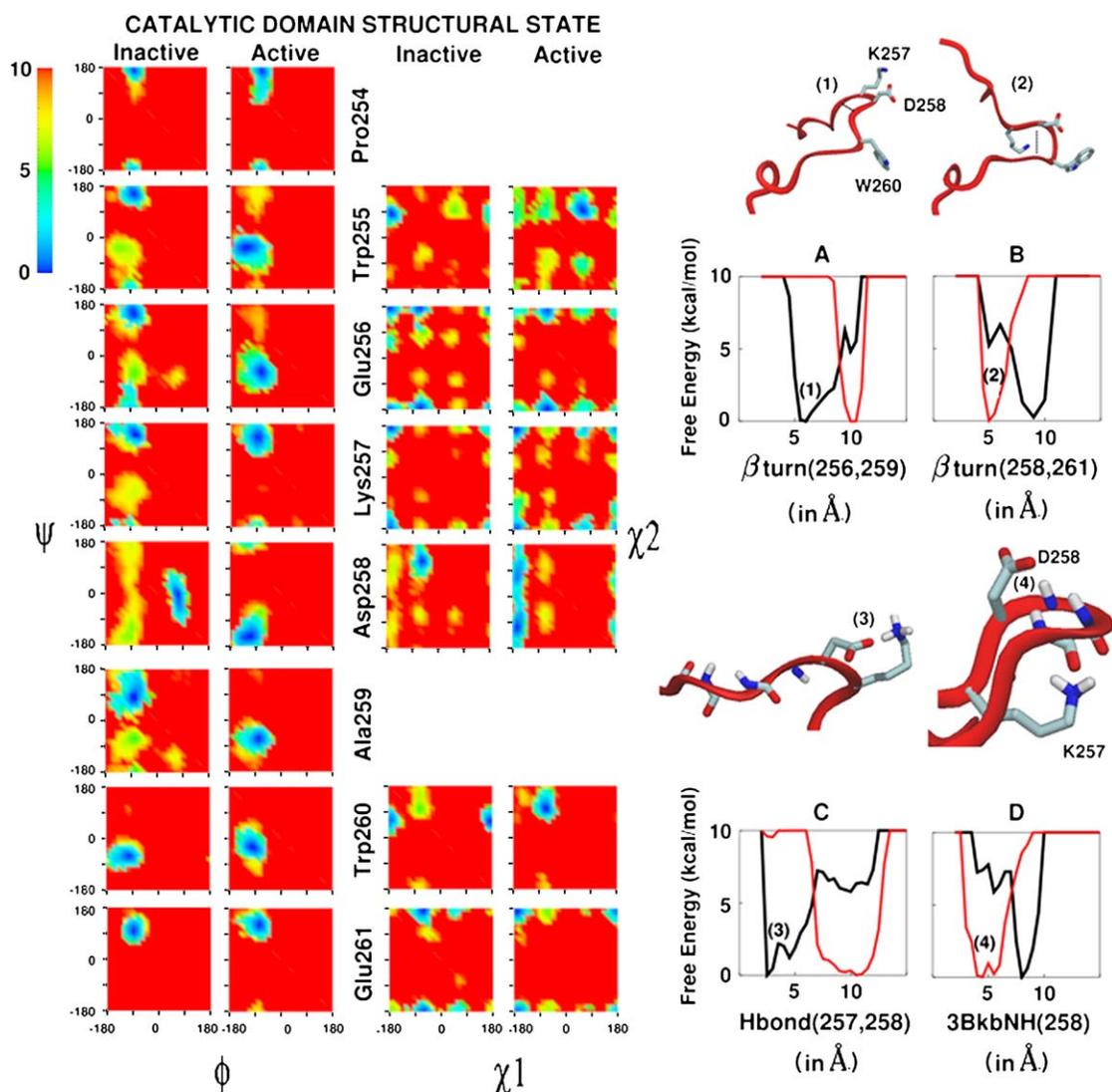


Figure 3. The Local Conformational Changes Involved the N-Terminal End Conformational Switch

Left half: the Ramachandran (ϕ, ψ) and (χ_1, χ_2) conformational free energy maps for residues around Trp260, color coding is in kcal/mol, and all torsions are given in degrees. Right half, clockwise: (A) The free energy profile along the $C\alpha-C\alpha$ distance for the Glu256-Ala259 β turn labeled as β turn(256,259). (B) The free energy profile along the $C\alpha-C\alpha$ distance for the Asp258-Glu261 β turn labeled as β turn(258,261). (C) The free energy profile along the distance between the Asp258 carboxylate oxygen and the center of mass of three consecutive backbone NH groups (residues 259–261) labeled as 3BkbNH(258). (D) The free energy profile along the Lys257-Asp258 side chain hydrogen bond distance labeled as Hbond(257,258). Conformations labeled (1)–(4) are located where indicated in the respective free energy profiles. The dotted line in structures A and B connects the β turn end residue $C\alpha$ atoms. The inactive catalytic domain is shown with a black line; the active catalytic domain is shown with a red line.

dynamics of this critical residue suggests that it may modulate N-terminal end release in Hck primarily through its side chain interactions.

The concerted backbone conformational change during N-terminal end release can also be characterized as a change in the conformational motifs called β turns (Chou, 2000). In the complete, assembled crystal structure of Hck (Schindler et al., 1999), the residues Glu256 and Ala259 form a β turn that is replaced in the modeled active catalytic domain by a β turn between residues Asp258 and Glu261. The two β turns are anticorrelated in their appearance and disappearance as a function of N-terminal end conformation. The free energy as a function of the distance between Glu256 $C\alpha$ and

Ala259 $C\alpha$ atoms shows that the β turn between residues Glu256 and Ala259 is stable only in the inactive catalytic domain (black line, Figure 3A). The free energy as a function of the distance between Asp258 $C\alpha$ and Glu261 $C\alpha$ shows that the β turn between residues Asp258 and Glu261 is stable only in the active catalytic domain (red line, Figure 3B). The same behavior is apparent in the inactive (Xu et al., 1997, 1999) and active (Cowan-Jacob et al., 2005) c-Src crystal structures in which, upon activation, the distance between Asp258 $C\alpha$ and Glu261 $C\alpha$ changes from 9.2 Å to 5.9 Å, and the distance between Ala256 $C\alpha$ and Ala259 $C\alpha$ changes from 5.5 Å to 10.7 Å. The shift in β turns (from a β turn between residues 256–259 to a β turn between residues

258–261) therefore seems to be an integral part of N-terminal end release in both Hck and c-Src.

Crystallographic Evidence Supporting N-Terminal End Release

Aspartate side chains are distinguished from other amino acid side chains by their ability to form stable interactions with backbone amide N-H groups. A distance-based survey of protein crystal structures in the PDB database was conducted to identify any reoccurrences of the Asp258 interaction with three consecutive backbone amide moieties in other known protein crystal structures. All interactions with a distance between the aspartate side chain oxygen and the backbone amide nitrogen less than 3.5 Å were included. Structures deposited in the PDB database as of August, 2004, with resolution better than 2.0 Å were included in the survey. Out of all of the aspartate side chains that showed distances less than 3.5 Å with backbone amide NH moieties of other residues (a total of 126,759 aspartate side chains), 11.6% (14,641 aspartate side chains) showed such interactions with three consecutive backbone amide moieties immediately C-terminal to the specific aspartate (i.e., backbone amide NH groups belonging to $n + 1$, $n + 2$, and $n + 3$ residues in the protein sequence, n being the sequence position of the aspartate). As illustrated in Figure 4, a significant percentage of aspartate moieties participate in the specific conformational motif in which their side chain is interacting with dipoles provided by three consecutive backbone amide NH groups.

There are some striking similarities between the present calculations and the structural changes observed in a recent crystal structure of c-Src (Cowan-Jacob et al., 2005), in which the SH2 domain is dissociated from the unphosphorylated C-terminal tail and the catalytic domain is in the active state. As mentioned above, a β turn shift from residues 256–259 to residues 258–261 also occurs in the restricted-to-released transition of the N-terminal end in c-Src. The N-terminal end hinge region is also localized within the same residue boundaries (residues 254–262). The exact rotameric states differ between c-Src and Hck (Tables S1 and S2), which is also apparent in the different orientation of the released N-terminal end in the c-Src structure and the released Hck N-terminal end-equilibrated reference structure in the present study. These differences could be due to the sequence variation in the residues 253–256 between c-Src and Hck. One feature conserved even in the variable sequence region is a 180° shift in backbone torsion ψ conformation for residue 255 (Leu in c-Src, Trp in Hck), which occurs for both c-Src and Hck, in opposite fashion. Thus, some of the structural properties of the N-terminal hinge region anticipated by our computations in the case of Hck appear to also be observed in c-Src activation.

Crystal structures of inactive (Hubbard et al., 1994) and active (Hubbard, 1997) catalytic domains of the insulin receptor kinase (IRK) also show features similar to those identified in this study for Hck N-terminal end release. By sequence alignment (Sicheri and Kuriyan, 1997), it is known that IRK possesses an N-terminal end (987–991) DEWEV sequence analogous to the N-terminal end (258–262) DAWEI sequence in Hck. The altered orientation of the N-terminal end of the activated

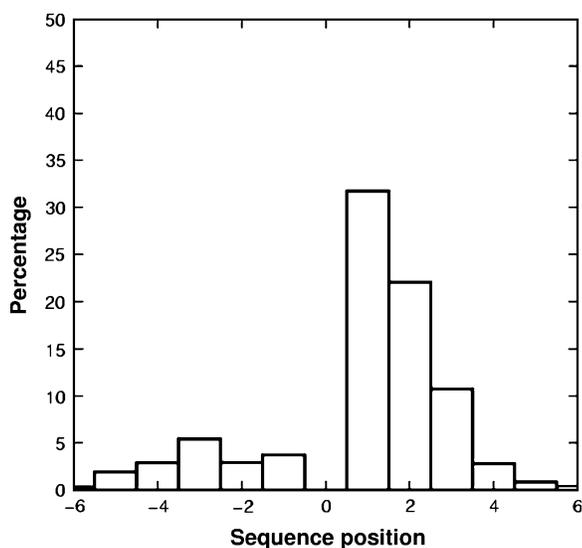


Figure 4. Percentage Breakdown of Aspartate Interaction with Backbone Amide NH Moieties of Other Residues According to Relative Sequence Position

–1 on the x axis indicates the previous residue; +1 indicates the next residue in the protein sequence. Within the known protein crystal structures, a substantial fraction (11%) of aspartate side chains that interact with backbone atoms do so with three adjacent consecutive backbone amide moieties of residues in the C-terminal direction.

IRK as compared to the inactive IRK suggests the presence of an N-terminal end conformational switch analogous to Hck and c-Src. The localized N-terminal end “released” state in IRK accompanying catalytic domain activation also shows side chain motion of Trp989, a shift in backbone turns from residues Pro986:Trp989 to residues Asp987:Glu990 and a possible stabilization of Asp987 by three consecutive backbone NH moieties (Figure S1). These parallels suggest that the hypothesized N-terminal end conformational switch may be a common feature in long-range allosteric regulation of different classes of protein kinases.

Implications for Allosteric Mechanism

Allostery in multimeric proteins has classically been described through multistate diagrammatic schemes describing the progressive changes of individual structural units between end point functional states (Monod et al., 1965; Koshland et al., 1966; Weber, 1972). Recent efforts aim to develop an atomistic perspective of the nature of the coupling between the various structural units, and how the progressive conformational changes occur. Specifically, allosteric conformational changes have been envisioned as concerted quasiharmonic, low-frequency normal mode motions that occur within the individual energy basins of the end point states (Miyashita et al., 2003). In a different approach, evolutionary conservation and correlated mutations have been used to identify networks of interacting residues that play a role in the transmission of allosteric information between remote protein sites (Lockless and Ranganathan, 1999; Suel et al., 2003). Allostery has also been described as a shift between preexisting equilibria of conformational states in activation of the signaling protein

NtrC by a phosphorylation event (Volkman et al., 2001; Kern and Zuiderweg, 2003).

The current study sheds some light on the underlying atomistic details of the allosteric mechanism in the specific case of Src kinases. The computations suggest a mechanism by which long-range allosteric coupling can arise from changes in the free energy conformational landscape of a relatively small and localized dynamic component of a larger system. Monitoring the system in terms of the conformational states adopted by individual coupled units is broadly consistent with the classical description of allostery (Monod et al., 1965; Koshland et al., 1966; Weber, 1972). The bistable equilibria displayed by the calculated PMFs for the conformation of the N-terminal end of the catalytic domain of Hck also agree with the previously proposed allosteric mechanism for NtrC (Volkman et al., 2001). While the PMFs clearly show anharmonic characteristics, the identified low-frequency concerted motions governed by a shallow free energy landscape may be possible to approximate by using quasiharmonic approaches (Miyashita et al., 2003). Finally, the present study suggests a possible interpretation of the existence of correlated residues identified from statistical evolutionary considerations involved in the transmission of allosteric information (Lockless and Ranganathan, 1999; Suel et al., 2003). Since we postulate that allosteric function is controlled by the effective free energy landscape, experimentally observed statistical correlations should mirror the joint effects of these mutations on the free energy landscape, which may not be simply due to a direct interaction between coupled residues. The established method of identifying a network of coupled residues through their functional properties, that can be directly compared to the calculated changes in the conformational free energy landscape, is therefore an ideal experimental test for the present results.

Experimental Procedures

Equilibration and Simulation Protocol

The CHARMM program (c30a1 version) (Brooks et al., 1983) with the all22 CHARMM force field for proteins (MacKerell et al., 1998) was used in all calculations. The 2.0 Å resolution assembled Hck crystal structure (PDB ID: 1QCF) provided the initial inactive catalytic domain conformation; the 1.7 Å resolution Lck active form structure (PDB ID: 3LCK) (Yamaguchi and Hendrickson, 1996) aligned by sequence homology to Hck by using CLUSTALW (Higgins et al., 1994) was used in subsequent homology modeling by using the program MODELLER (Marti-Renom et al., 2000) to get the initial active catalytic domain conformation. These initial structures (comprised of Hck residues 253–523) were solvated in a 150 mM KCl solvent box with dimensions 77.6 × 62.1 × 52.4 Å³ and were minimized with 5 kcal/mol harmonic constraints on nonhydrogen atoms for 500 Steepest Descent steps. Periodic boundary conditions were used with the images generated by using the CRYSTAL module in CHARMM (Field and Karplus, 1992). Covalent bonds involving hydrogen were constrained with SHAKE (Ryckaert et al., 1977) to enable use of an integration time step of 0.002 ps. Long-range electrostatic interactions were treated by using the Particle Mesh Ewald (PME) approach (Darden et al., 1993) 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 cut-offs of 10 Å were used with non-bond 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 constraints by using the MMFP module of CHARMM (Beglov and Roux, 1994).

The Constraint and Analyzing Unconstrained Order Parameters

The initial structures for the transitions between restricted and released conformations of the N-terminal region (residues 253–272) were generated by using a one-dimensional rmsd constraint pulling all non-hydrogen atoms in the N-terminal end gradually toward the opposite conformation. Each intermediate was generated at 0.1 Å rmsd intervals with 10 ps of constant volume and temperature (NVT) sampling before pulling it to the next intermediate value –0.1 Å rmsd away. The final structure was constrained to an rmsd value of 0.0 Å from the opposite reference structure. Thus, 61 intermediate windows were generated spanning the range of 6 Å rmsd separating the 2 reference structures. These initial windows were then used as starting points for umbrella sampling MD simulations with the constraint w_j on the ΔD_{rmsd} order parameter implemented in CHARMM (Banavali and Roux, 2005).

$$w_j = K_{\text{rmsd}} (\Delta D_{\text{rmsd}} - \Delta D_{\text{min}})^2, \quad (1)$$

where ΔD_{min} specifies the minimum value for the harmonic potential. The ΔD_{rmsd} order parameter is the difference between the rmsd values of each structure from the two reference states:

$$\Delta D_{\text{rmsd}} = \text{rmsd}(X_t, X_{\text{restricted}}) - \text{rmsd}(X_t, X_{\text{released}}), \quad (2)$$

where X_t is the instantaneous structure and $X_{\text{restricted}}$ and X_{released} are the restricted and released reference structures, respectively. In this constraint, the intermediate structures can sample conformational regions away from both end points without any energetic penalty as long as ΔD_{rmsd} remains the same, thereby allowing appropriate relaxation of the intermediates. The 0.1 Å intervals in the constraint used for pulling to generate initial structures in each window correspond to 0.2 Å intervals for the same structures in the ΔD_{rmsd} order parameter. The force constant for this harmonic constraint was gradually reduced from 500 kcal/mol/Å² to 20 kcal/mol/Å² 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 one-dimensional ΔD_{rmsd} constraint of 20 kcal/mol/Å² by using an NVT MD simulation for an additional 140 ps. To remove any long-lasting correlation in the solvent surrounding the solute protein, the final structure in each umbrella sampling window was extracted, resolvated, and subjected to 210 ps of NVT MD simulation with the exact same protocol of slowly relaxing constraints as before. By excluding the first 70 ps of each separate 210 ps MD simulation as equilibration, a total of 280 ps of sampling per window was used in the subsequent analysis. The Weighted Histogram Analysis Method (WHAM) algorithm (Kumar et al., 1992; Souaille and Roux, 2001) was used to get the Potential of Mean Force (PMF) along the one-dimensional difference rmsd reaction coordinate from the time series of this variable saved at every step of the 280 ps production dynamics. This ΔD_{rmsd} order parameter used for the umbrella sampling calculations incorporates many other order parameters contributing to the detailed dynamics of the overall motion. If the convergence of the ΔD_{rmsd} free energy profile is assumed to indicate adequate sampling of these other order parameters, then their free energy profile can be obtained (Banavali and Roux, 2005). First, the biased probability distribution, $\langle \rho_{\text{bias}}(\xi_1, \xi_2) \rangle$, for the two variables, $\xi_1 = \Delta D_{\text{rmsd}}(t)$, $\xi_2 =$ any other order parameter that can be extracted through structural analysis, is obtained from the MD simulation trajectory. The unbiased probability distribution $\langle \rho(\xi_1, \xi_2) \rangle$ can then be obtained from $\langle \rho_{\text{bias}}(\xi_1, \xi_2) \rangle$ by using the WHAM equations:

$$\langle \rho(\xi_1, \xi_2) \rangle = \frac{\sum_i N_i \langle \rho_{\text{bias}}(\xi_1, \xi_2) \rangle}{\sum_j N_j \exp[F_j - w_j(\xi_1)/k_B T]}, \quad (3)$$

where N_i is the number of histogram bins, N_j is the number of umbrella sampling windows, n_i and n_j are the number of time points in each bin or window, respectively, F_j are the free energy constants for each window, and $w_j(\xi_1)$ is the biasing harmonic potential imposed along ΔD_{rmsd} indicated by Equation 1. The values of F_j can be obtained from the equation,

$$\exp(F_j/k_B T) = \int \langle \rho(\xi_1, \xi_2) \rangle \exp(-w_j(\xi_1)/k_B T) d\xi_1 d\xi_2. \quad (4)$$

Since there are two unknowns, F_j and $\rho(\xi_1, \xi_2)$, to get optimal values of $\rho(\xi_1, \xi_2)$, Equations 3 and 4 have to be solved by self-iteration by using appropriate convergence criteria (change of less than 0.0001 in successive F_j values is used as a convergence criteria here). Finally, the unbiased density $\rho(\xi_2)$, for the order parameter ξ_2 , can be obtained from:

$$\langle \rho(\xi_2) \rangle = \int \langle \rho(\xi_1, \xi_2) \rangle d\xi_1, \quad (5)$$

and the free energy surface represented by $W(\xi_2)$ can be obtained from:

$$W(\xi_2) = -k_B T \ln(\rho(\xi_2)) + C, \quad (6)$$

where C is an arbitrary constant. $W(\xi_2)$ is obtained without explicitly sampling the entire range of ξ_2 , but the assumption here is that the sampled range will be the one most relevant to the overall motion studied in the original one-dimensional constrained sampling along ΔD_{rmsd} . The extension of this method to two or more order parameters is straight forward, and it only requires the saved MD trajectory snapshots sufficient to generate the initial multidimensional biased histogram. This method cannot only elucidate the free energy behavior of motions that remain obscured in the original one-dimensional free energy profile, but it can also serve as an excellent analysis tool by to gain a qualitative localized understanding of the overall process. This method is used in the present study to calculate the two-dimensional free energy map in Figure 1B as well as all of the local torsional profiles reported.

The accuracy and convergence of the free energy profiles was judged by the extent of change in the free energy profiles upon further sampling (not shown): i.e., sampling was extended until no further change occurred. While this test is adequate for judging the convergence along the path obtained through the targeted MD simulations, it does not preclude the existence of additional paths that may show different converged free energy profiles. Since the one-dimensional ΔD_{rmsd} reaction coordinate allows intermediate structures to simultaneously move further away from both end point states, such additional paths should be sampled during the present simulations, assuming that they are separated from the original path by thermally accessible kinetic barriers. It is prohibitively expensive, however, to explicitly test convergence along many different targeted MD trajectories to rule out the presence of kinetically inaccessible additional paths. In the near future, when computational resources become an order of magnitude greater than those presently available, it may be possible to test convergence along many different paths as well as extend the present study to the different Src kinase family members. All molecular pictures were produced by using DINO (<http://www.dino3d.org>), all graphs were made by using either gnuplot (<http://www.gnuplot.info>) or OPENDX (<http://www.opendx.org>) software, and all figures were composed by using the Gnu Image Manipulation Program (GIMP) (<http://www.gimp.org>).

Supplemental Data

Supplemental Data including a table indicating rotameric minima for each residue in the N-terminal end hinge region for Hck from the present umbrella sampling MD simulations, a table indicating rotameric states for the same N-terminal end hinge region in inactive and active c-Src crystal structures, and a figure indicating the presence of an N-terminal end release mechanism in Insulin Receptor Kinase are available at <http://www.structure.org/cgi/content/full/13/11/1715/DC1/>.

Acknowledgments

We would like to thank Guillaume Lamoureux, Toby Allen, Hyung-June Woo, Yuqing Deng, Sergei Noskov, Jose Faraldo-Gomez, Marta Murcia, and Deniz Sezer for helpful discussions. This work was supported financially by grant CA93577 from the National Institutes of Health and a Keck Postdoctoral Fellowship for N.K.B. Computational support from the Pittsburgh Supercomputing Center (PSC) obtained through the National Resource Allocations Committee (NRAC) was used for the calculations.

Received: June 20, 2005

Revised: September 7, 2005

Accepted: September 7, 2005

Published: November 8, 2005

References

- Banavali, N.K., and Roux, B. (2005). The free energy landscape of A-DNA to B-DNA conversion in aqueous solution. *J. Am. Chem. Soc.* **127**, 6866–6876.
- Beglov, D., and Roux, B. (1994). Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050–9063.
- Brooks, B., Brucoleri, R., Olafson, B., States, D., Swaminathan, S., and Karplus, M. (1983). CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
- Brown, M., and Cooper, J. (1996). Regulation, substrates and functions of Src. *Biochim. Biophys. Acta* **1287**, 121–149.
- Chou, K.-C. (2000). Prediction of tight turns and their types in proteins. *Anal. Biochem.* **286**, 1–16.
- Cowan-Jacob, S.W., Fendrich, G., Manley, P.W., Jahnke, W., Fabbro, D., Liebetanz, J., and Meyer, T. (2005). The crystal structure of a c-src complex in an active conformation suggests possible steps in c-Src activation. *Structure* **13**, 861–871.
- Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald – an N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
- Field, M.J., and Karplus, M. (1992). CRYSTAL: a program for crystal calculations in CHARMM. PhD thesis, Harvard University, Cambridge, Massachusetts.
- Gonfloni, S., Williams, J., Hattula, K., Weijland, A., Wierenga, R., and Superti-Furga, G. (1997). The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src. *EMBO J.* **16**, 7261–7271.
- Gonfloni, S., Frischknecht, F., Way, M., and Superti-Furga, G. (1999). Leu255 of src couples intramolecular interactions to inhibition of catalysis. *Nat. Struct. Biol.* **6**, 760–764.
- Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
- Hubbard, S. (1997). Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and atp analog. *EMBO J.* **16**, 5573–5581.
- Hubbard, S.R., Wei, L., Ellis, L., and Hendrickson, W.A. (1994). Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372**, 746–754.
- Huse, M., and Kuriyan, J. (2002). The conformational plasticity of protein kinases. *Cell* **109**, 275–282.
- Kern, D., and Zuiderweg, E.R.P. (2003). The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13**, 748–757.
- Koshland, D.E., Jr., Nemethy, G., and Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365–385.
- Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A., and Rosenberg, J.M. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules.1. The method. *J. Comput. Chem.* **13**, 1011–1021.
- Lafevre-Bernt, M., Siccheri, F., Pico, A., Porter, M., Kuriyan, J., and Miller, W.T. (1998). Intramolecular regulatory interactions in the Src family kinase Hck probed by mutagenesis of a conserved tryptophan residue. *J. Biol. Chem.* **273**, 32129–32134.
- Lockless, S., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299.
- MacAuley, A., and Cooper, J.A. (1989). Structural differences between repressed and derepressed forms of p60 c-Src. *Mol. Cell. Biol.* **9**, 2648–2656.
- MacKerell, A., Jr., Bashford, D., Bellot, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998).

All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616.

Marti-Renom, M.A., Stuart, A., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325.

Miyashita, O., Onuchic, J.N., and Wolynes, P.G. (2003). Non-linear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl. Acad. Sci. USA* **100**, 12570–12575.

Moarefi, I., LaFevre-Bernt, M., Sicheri, F., Huse, M., Lee, C., Kuriyan, J., and Miller, W. (1997). Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. *Nature* **385**, 650–653.

Monod, J., Wyman, J., and Changeux, J.P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118.

Porter, M., Schindler, T., Kuriyan, J., and Miller, W. (2000). 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.* **275**, 2721–2726.

Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327–341.

Schindler, T., Sicheri, F., Pico, A., Gazit, A., Levitzki, A., and Kuriyan, J. (1999). Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor. *Mol. Cell* **3**, 639–648.

Sicheri, F., and Kuriyan, J. (1997). Structures of Src-family tyrosine kinases. *Curr. Opin. Struct. Biol.* **7**, 777–785.

Sicheri, F., Moarefi, I., and Kuriyan, J. (1997). Crystal structure of the Src family tyrosine kinase Hck. *Nature* **385**, 602–609.

Souaille, M., and Roux, B. (2001). Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Comm.* **135**, 40–57.

Suel, G.M., Lockless, S.W., Wall, M.A., and Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **10**, 59–69.

Taylor, S.S., and Radzio-Andzelm, E. (1994). 3 protein kinase structures define a common motif. *Structure* **2**, 345–355.

Volkman, B.F., Lipson, D., Wemmer, D.E., and Kern, D. (2001). Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429–2433.

Weber, G. (1972). Ligand binding and internal equilibria in proteins. *Biochemistry* **11**, 864–878.

Xu, W., Harrison, S.C., and Eck, M.J. (1997). Three-dimensional crystal structure of the tyrosine kinase c-Src. *Nature* **385**, 595–602.

Xu, W., Doshi, A., Lei, M., Eck, M.J., and Harrison, M.J. (1999). Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Mol. Cell* **3**, 629–638.

Yamaguchi, H., and Hendrickson, W. (1996). Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature* **384**, 484–489.

Young, M., Gonfloni, S., Superti-Furga, G., Roux, B., and Kuriyan, J. (2001). Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115–126.