Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories

Sichun Yang[a], Nilesh K. Banavali[b], and Benoît Roux[a,1]

[a]Department of Biochemistry and Molecular Biology, University of Chicago, 929 East 57th Street, Chicago, IL 60637; and [b]Laboratory of Computational and Structural Biology, Division of Genetics, Wadsworth Center, New York State Department of Health, Empire State Plaza, P.O. Box 509, Albany, NY 12201

Edited by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved January 8, 2009 (received for review August 21, 2008)

The Src-family kinases are allosteric enzymes that play a key role in the regulation of cell growth and proliferation. In response to cellular signals, they undergo large conformational changes to switch between distinct inactive and active states. A computational strategy for characterizing the conformational transition pathway is presented to bridge the inactive and active states of the catalytic domain of Hck. The information from a large number (78) of independent all-atom molecular dynamics trajectories with explicit solvent is combined together to assemble a connectivity map of the conformational transition. Two intermediate states along the activation pathways are identified, and their structural features are characterized. A coarse free-energy landscape is built in terms of the collective motions corresponding to the opening of the activation loop (A-loop) and the rotation of the αC helix. This landscape shows that the protein can adopt a multitude of conformations in which the A-loop is partially open, while the αC helix remains in the orientation characteristic of the inactive conformation. The complete transition leading to the active conformation requires a concerted movement involving further opening of the A-loop, the relative alignment of N-lobe and C-lobe, and the rotation of the αC helix needed to recruit the residues necessary for catalysis in the active site. The analysis leads to a dynamic view of the full-length kinase activation, whereby transitions of the catalytic domain to intermediate configurations with a partially open A-loop are permitted, even while the SH2-SH3 clamp remains fully engaged. These transitions would render Y416 available for the transphosphorylation event that ultimately locks down the active state. The results provide a broad framework for picturing the conformational transitions leading to kinase activation.

conformational landscape | connectivity map | intermediate | kinase inhibitor

Non-receptor Src tyrosine kinases are a family of large allosteric enzymes that are critical for cell growth (1–4). They play a key regulatory role inside the cell by turning downstream target proteins "on" or "off" through phosphorylation, during which the γ-phosphate group of ATP is transferred to a tyrosine residue in the target protein (5). Their own activity is also tightly regulated by phosphatases and upstream kinases in multiple physiological processes. Uncontrolled Src kinase activation is linked to a number of diseases, particularly cancer, making them important targets for therapeutic intervention (6, 7). Discoveries of kinase inhibitors targeting the Src tyrosine kinase fold, including the Src family members and relatives, have been one of the most exciting advances in molecular design (8–13).

Members of the Src kinase family share a common structural organization, which consists of 2 regulatory SH3 and SH2 binding domains, followed by a catalytic domain. The catalytic kinase domain is highly conserved among many protein kinases, and its overall architecture closely resembles that of other protein kinases such as protein kinase A (14–16), Csk (17, 18), and Abl (7, 8, 19). It comprises an N-terminal lobe (N-lobe) and a C-terminal lobe (C-lobe) (see Fig. S1), between which the active site is located.

Structures of inactive and active conformations have been characterized by X-ray crystallography (20–25). The primary structural distinctions between the 2 conformations concerns the phosphorylation of Tyr-416 located in the central activation-loop (A-loop) and the orientation of helix αC. In the inactive state, the A-loop with unphosphorylated Tyr-416 is closed and folded to occlude substrate entry into the active site, while the orientation of the αC helix is stabilized by hydrophobic contacts and salt bridges with the loop and neighboring structural motifs to keep important catalytic residues away from the active site (22, 24, 26). In the active state, an outward movement of the A-loop opening up the active site to substrate binding is accompanied by an inward rotation of the αC helix recruiting the residues needed to form a catalytically competent active site. The fully active state is ultimately stabilized by the phosphorylation of Tyr-416, although the kinase domain can also adopt an open active-like conformation with unphosphorylated Tyr-416 (25). It is the process of *trans*-phosphorylation of Tyr-416 (via a bimolecular encounter with another active Src kinase) that ultimately "locks" the domain in its catalytically active state (27). For this final process to occur, the A-loop of the kinase domain must adopt an open active-like conformation, at least transiently, to expose Tyr-416 to another kinase.

In the full-length inactive state, the SH2 and SH3 regulatory domains serve to down-regulate kinase activity by acting as a "clamp" pushing on the back side of the enzyme (23), presumably by interfering with the ability of the kinase domain to undergo a transition toward its open active-like conformation. In the absence of the SH2 and SH3 regulatory domains, the isolated catalytic domain is constitutively active (20), which suggests that the transition toward the open active-like conformation is allowed under these conditions. In this context, characterizing the conformational transition in the isolated catalytic domain is an important first step to understand the inhibitory action of the regulatory domains. A more complete knowledge of the conformational transition within the kinase domain is also expected to broaden the configurational space to search for possible kinase inhibitors, which could target any number of intermediate states (6).

A detailed mapping of the conformational transition pathway leading to Src activation is very difficult to achieve with experiments because all of the intermediate states are short lived. Molecular dynamics (MD) simulations based on detailed atomic

Author contributions: S.Y., N.K.B., and B.R. designed research; S.Y. performed research; S.Y. and N.K.B. analyzed data; and S.Y., N.K.B., and B.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: roux@uchicago.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0808261106/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.0808261106

models and explicit solvent provide a powerful and alternative approach to investigate such processes. However, the scope of simple brute-force simulations, which are limited to relatively short time scales, must be expanded to yield useful information. In this article, the accumulated information from multiple all-atom MD trajectories with explicit solvent is used to map the conformational landscape from the inactive to active state of the catalytic domain of Hck. The initial structures of these unbiased MD simulations were taken from various points along a targeted MD (TMD) simulation trajectory (28).

A clustering scheme is then used to regroup similar conformations together from independent trajectories. This procedure allows us to construct a connectivity map and characterize the collective dynamics of the conformational transitions linking the inactive to the active state. From the connectivity map analysis, we identify 2 intermediate states and characterize their structural features. The results complement previous simulation studies on Src activation, based on coarse-grained (29) and all-atom models (7, 28, 30–34). The computational strategy is general and should lead to important advances in our ability to analyze protein allosteric transitions.

Results

Conformational Landscape: A Connectivity Map. We consider the transition pathway between the inactive conformation and an active-like conformation with Tyr-416 unphosphorylated (hereafter referred to as active for the sake of simplicity). Systems relaxed and equilibrated at 78 conformations between the inactive and active states were released (28) and simulated by free all-atom MD simulations with explicit solvent, for a total time of ≈1 μs (see *Methods* for details). From the simulations, the entire configuration space is then partitioned into 25 clusters, and a connecting diagram is constructed. This connectivity map is displayed in Fig. 1, which covers the conformational landscape ranging from the inactive to the active state. On the landscape, similar structures are partitioned into the same cluster by using a K-means clustering algorithm (35). The neighboring clusters are connected by filtered transitions as collected from MD trajectories. A few examples of trajectories undergoing a transition from one cluster to another are indicated by colored connecting lines on the map (Fig. 1). The distances between clusters are plotted to be inversely proportional to the intercluster transition rates. The interconversion rates were observed using a time interval (lag time) of 100 ps (see *SI Text* and Fig. S2 for details).

The procedure for constructing the connectivity map can potentially lead to very noisy statistics because of brief "spurious" recrossing events of the MD trajectories between 2 clusters (see a diagram in Fig. S3). To filter out those recrossings at the boundary between clusters, we used a distance cutoff to assign the transition between clusters i and j. The crossing from i to j is considered as a true transition only when a system has really escaped the state i. By definition, a transition occurred if the point-to-centroid distances satisfy $d_i\,(t + \delta t) > \gamma\, d_i(t)$, where $d_i(t)$ and $d_i\,(t + \delta t)$ are the distances of an individual structure to its cluster center at time t and $t + \delta t$, respectively. Accordingly, if the configuration is crossing the cluster-cluster boundary $d(t)$ at time t, the crossing is considered as a true transition only if a distance of $d(t + \delta t) > \gamma d(t)$ is reached within the time interval δt (Fig. S3). This procedure filters out the "false-alarm" recrossings that occur at the cluster boundaries without leading to a definite transition. For the Src map construction, $\delta t = 10$ ps was used and $\gamma = 1.1, 1.2, 1.3$ were applied. Additional check on intercluster transitions shows that different simulations overlap with each other in the cluster space, ensuring that the clusters are well connected (Fig. 1 *Lower*). Thus, we achieved a complete ensemble of conformations spanning the transition from the inactive to the active state.

The connectivity map serves to chart the territory visited by the macromolecule during the state-to-state conformational transitions. The concept of the connectivity map bears some similarities with the protein folding networks (36–40), which have been used to describe the folding landscape from unfolded states to a folded state.

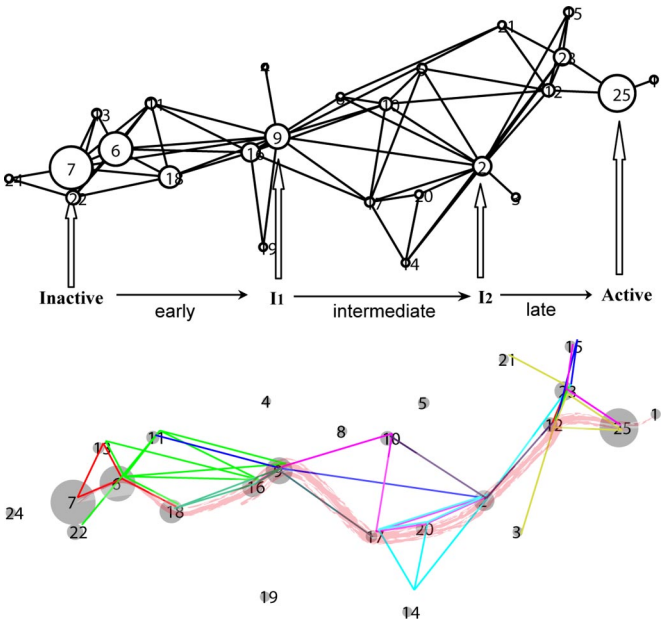

Fig. 1. Connectivity map for conformational transitions during Src activation. (*Upper*) Connectivity diagram linking the inactive to the active catalytic domain of Hck generated by performing a K-means clustering on 78 unbiased MD simulations. The structural network maps the configurational space of 25 clusters that are kinetically connected by MD traces. Each of the 25 vertex corresponds to a cluster of configurations extracted from all-atom MD. Four hubs of the structural network are revealed to represent the transition: Inactive, I_1, I_2, and Active, which divide the conformation space into 3 transition phases: Inactive → I_1, I_1 → I_2, and I_2 → Active. (*Lower*) Overlap of independent simulations by several representative color-coded MD traces. The conformational clusters are kinetically connected and an ensemble of transition pathways is achieved from the accumulation of independent simulations. A representative transition collectively taken from multiple simulations is shown as the brushed line and uploaded online (see Movie S1 for animation).

Two Intermediates States Are Identified. A careful examination of the connectivity map reveals the existence of 2 intermediate states along the pathway linking the inactive and active conformation of the catalytic domain. These 2 intermediates, henceforth called I_1 and I_2, form the most connected "hub"-like clusters on the connectivity map of Src activation (Fig. 1) and are the most robustly connected structural states for a range of analysis parameters. This robustness is demonstrated by varying 2 criteria used to construct the map: the lag times and the boundary cutoffs. With different lag times (50 ps, 100 ps, and 200 ps), the overall connectivity among clusters remains intact (Fig. S2). The topology of the map is also conserved with various boundary cutoffs ($\gamma = 1.1$, 1.2 and 1.3), which filter out spurious recrossing events at the boundary between 2 clusters (Fig. S3). I_1 and I_2 are also likely to be metastable intermediates along the activation path because of the estimated longer average lifetime of each cluster to which they belong (Fig. S4). The lifetime is defined as the time period for which the conformation remains in its own cluster before making a transition to another cluster. It should be noted that another cluster (no. 21) has a long lifetime but is not classified as an intermediate because it is close

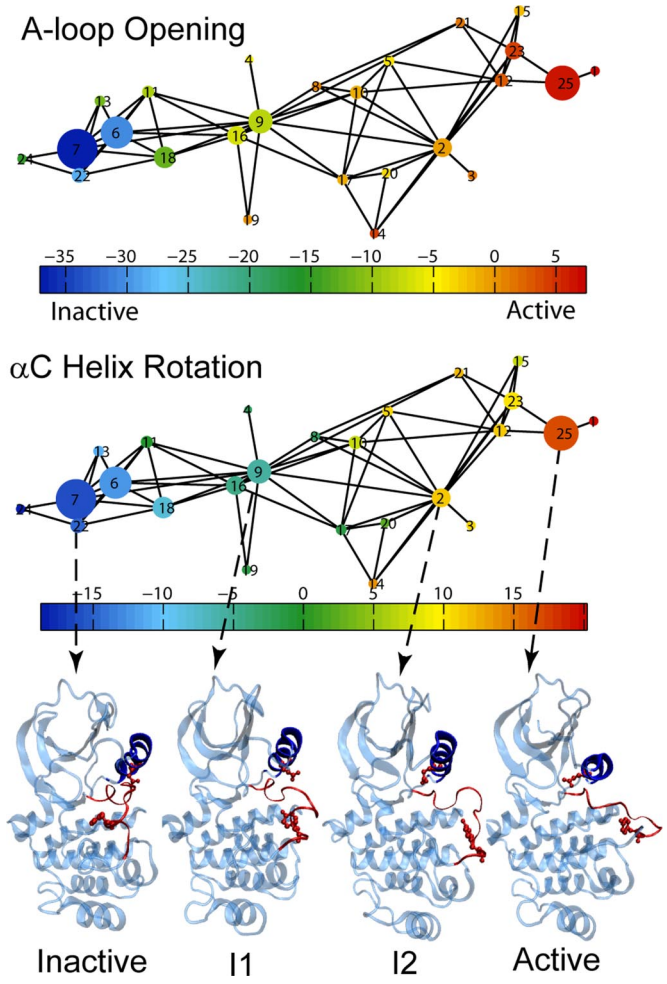

Fig. 2. Mechanisms of A-loop and αC helix and structural details of I_1 and I_2. The color-coded maps for the A-loop (by $\Delta Q_{\text{A-loop}}$) and the αC helix (by $\Delta Q_{\alpha C}$) indicate that the loop opens earlier than the αC helix starts to rotate. For example, the αC helix is in its inactive conformation at I_1, while the A-loop adopts a partially extended form. The most probable structures are shown in the *Lower* to illustrate their structural features. Residues I411 to P425 are used to define the flexible region of the A-loop. Residues V304 to K315 are used to define the αC helix. $\Delta Q_{\text{A-loop}}$ and $\Delta Q_{\alpha C}$ are the difference of the number of contacts. For an easier comparison with the results from a previous study based on a coarse-grained model (29), an identical list of native contacts was used in the definition of ΔQ. A list of 55 (48) and 30 (59) distances was included to describe the helix rotation (the A-loop opening) for the inactive and active state, respectively, defined on the basis of the crystal structure of Hck (PDB entry 1QCF) and the homology model of c-Src (PDB entry 1Y57). A contact between 2 residues (*i* and *j*) is considered formed if $|r_{ij} - \sigma_{ij}| < (\gamma - 1)\sigma_{ij}$ where $\gamma = 1.2$. r_{ij} and σ_{ij} are the instantaneous and native distances between 2 residues (*i* and *j*), respectively.

to the edge of the connectivity map and gets disconnected as the cutoff γ increases (Fig. S3).

Collective Dynamics During Src Activation. The crystallographic structures display a relative shift of the N-lobe and the C-lobe between the inactive and active state (24). To investigate the large collective domain-domain motion, we first align all configurations generated from simulations according to 3 main α-helices in C-lobe as shown in red in Fig. S5 (αE: L360-Q379, αF: I441-T457, and αH: E489-W499 in c-Src numbering). We then determine 3 principal axes of 3 main β-strands in N-lobe in blue (β2: G279-T285, β3: K291-M297, and β5: I336-E339). The rotation angles around 3 principal axes are finally computed and projected on the connectivity map (Fig. S5). The color-codes are based on the average rotation angles. From the map, a major rotation around the third axis is observed, which separates I_1 from I_2. Similarly, the translational motion between 2 lobes is also projected on the map (Fig. S5), according to the center-of-mass distances of Cα atoms between the red and blue regions in 2 lobes. I_1 and I_2 are ≈0.23 Å away from the inactive and active clusters, respectively, with a translation of 0.68 ± 0.67 Å between I_1 and I_2. Overall, the relative lobe-lobe orientation around the third axis is one of the main differences between I_1 and I_2.

Fig. 2 shows the connectivity map for the A-loop opening and the αC helix rotation colored by structural similarity defined by $\Delta Q_{\alpha C}$ and $\Delta Q_{\text{A-loop}}$. $\Delta Q_{\alpha C} = Q^{\text{A}}_{\alpha C} - Q^{\text{I}}_{\alpha C}$ was used to characterize the conformational state of the αC helix from V304 to K315, where $Q^{\text{I}}_{\alpha C}$ and $Q^{\text{A}}_{\alpha C}$ are the number of contacts made between any residue in the αC helix and any other residues for the inactive and active state, respectively. A similar definition for $\Delta Q_{\text{A-loop}}$ was used for the A-loop represented by the segment from I411 to P425. The conformation is close to the inactive cluster if ΔQ is negative (blue) and close to the active cluster if ΔQ is positive (red). Representative structures of I_1 and I_2 are shown in Fig. 2. At I_1, the αC helix is relatively close to the inactive state, while the A-loop has progressed to the middle of its transition toward the active state (green). At I_2, the αC helix starts to rotate toward the active state (green), while the A-loop is opening further (yellow). The maps clearly show the motion of 2 key motifs. Initially, the A-loop opens up without much rotation of the helix (Inactive → I_1), and then a concerted movement of both of them is required to make the complete transition ($I_1 \rightarrow I_2 \rightarrow$ Active). Such sequential movements were also observed in the activation of cyclin-dependent kinase cdk2 upon binding to cyclin (41, 42).

The structural difference between 2 intermediates suggests that the αC helix plays a more important role than the A-loop in terms of stabilizing the inactive conformation. This is because the A-loop is dynamic and the αC helix is relatively rigid. For transphosphorylation to occur, the A-loop must be open to expose Tyr-416 to the active site of the other kinase, at least transiently. Transient dynamic openings of the A-loop are consistent with the existence of significant structural fluctuations of Src (22, 21). In the initial X-ray structure of Hck (22), the A-loop could not be resolved, indicating that it was either very mobile or adopted a number of conformations. However, the inward rotation of αC helix is rather late within a rather narrow region, indicating that it is relatively rigid. This notion has been observed in both coarse-grained models (29) and atomistic models (34). From a practical standpoint, this structural difference suggests that potential kinase inhibitors that target stabilization of the αC helix might be able to better trap the Src kinase in inactive or I_1 states and block the transition to I_2 or the active state.

Switching of Ion-Pairs and Hydrophobic Packing. From the simulations, we observe that a set of key residues involved in ion-pair formation proximal to the A-loop and the αC helix are switching their interaction partners during activation. Among these residues are Glu310 in the αC helix, Lys295 in the β3 strand, Thr338 in the β5 strand, Arg385 from the "HRD" motif (21, 22), Asp404 from the DFG motif (7), and Arg409 in the A-loop (Fig. 3). During activation, Glu310 from the αC helix switches its interacting partner from Arg409 to Lys295, while Lys295 appears to move away from Asp404 to Glu310. The mass-weighted center-of-mass distance changes between the side-chains of these residues as shown in Fig. 3 and Fig. S6. These observations are consistent with previous results obtained with nonequilibrium MD simulations (32). Notably, the ion-pair formed by Glu310 and Arg409 is maintained in I_1, but the separation starts during the transition of I_2 → active. We also find that the distance between Glu310 and Arg385 gradually increases by ≈2 Å, correlated with the αC helix rotation (Fig. S6). We note that a similar

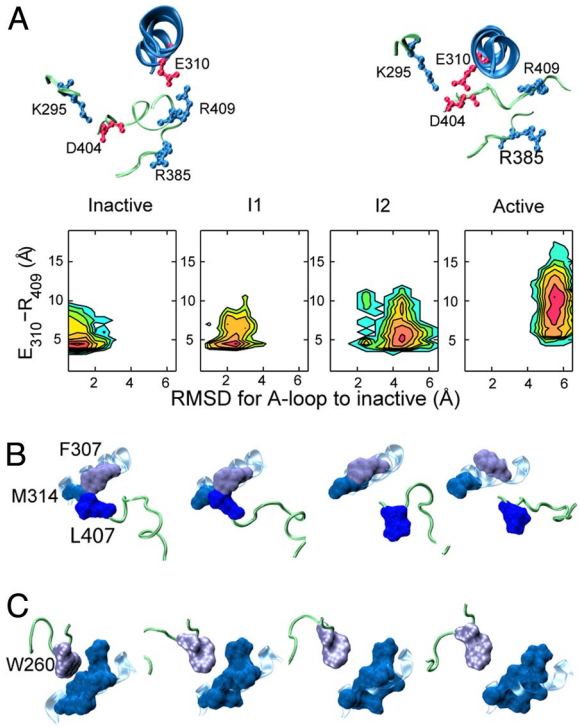

Fig. 3. Switching of ion-pairs and hydrophobic packing. (*A*) Switching of a network of charged residues of K295, E310, R385, D404, and R409. In the inactive state, E310 points toward R409. In the active state, E310 switches its partner from R409 to K295 when the αC helix is rotated. This switching mechanism is clearly shown from contour plots of mass-weighted center-of-mass distances between the side-chains of 2 residues and the RMSD of the A-loop (relative to its inactive conformation). Additional ion-pair formations such as E310-R385 and K295-D404 are also examined along the pathways (Fig. S6). (*B*) Hydrophobic rearrangement around the central activation-loop and around the αC helix among F307-M314-L407. (*C*) The hydrophobic packing between Trp260 and hydrophobic residues in the αC helix.

concerted motion was inferred from the interactions of Lys295, Glu310, and Arg 409 in recent restrained MD simulations (49).

The simulations show that a hydrophobic cluster (formed by Phe307, Met314, and Leu407) remains packed for the inactive state and I_1, but starts to break and disperse gradually during the transition from I_1 to I_2 (Fig. 3). The "melting" of the hydrophobic cluster at the intermediate stage leaves room for Glu310 to form an ion-pair with Lys295 (33), and the αC helix to rotate. The hydrophobic rearrangement is near Trp260 in the N-terminal of the catalytic domain, which has been described as a key component for down-regulation acting as a gate-keeper residue (23). Fig. 3 also shows the gradual separation of Trp-260 from the hydrophobic residues in the αC helix.

It is worth noting that the shape of the substrate binding pocket is affected by the conformational transition, especially in the region proximal to residues Ile-336, Thr-338, and Ile-341. Structural plasticity in this region is important because various kinase inhibitors can target different conformations, including intermediate states (6). Analysis of 7 known inhibitors and their steric clashes with protein atoms indicates that the binding pockets of I_1 and I_2 would be able to accommodate a wider range of inhibitors than the inactive conformation (Fig. S7). Thus, I_1 and I_2 could serve as additional target structures for the discovery of new kinase inhibitors.

Coarse-Grained Free Energy Landscape and Kinase Activation. To assess the relative stability of the 4 states along the activation pathway, a coarse-grained free energy landscape associated with the key motions of the A-loop and the αC helix was constructed from the connectivity map and the population of each cluster. The coarse free energy landscape was obtained by summing over Gaussian functions weighted by the population of each cluster p_i according to Eq. **1**. The result represents an equilibrium sum over each individual cluster, "smeared" by the natural fluctuations within each cluster. The procedure bears some similarities with metadynamics (43) and the single sweep method (44). The population of each cluster was estimated as the equilibrium eigenvector (with unit eigenvalue) of the transition probability matrix T built with a time interval of 100 ps from all-atom MD trajectories with explicit solvent (Fig. S8).

Fig. 4*A* shows the reconstructed coarse free energy surface. The free energy landscape supports the notion that the motions of the A-loop and the αC helix are decoupled in the early stage of the activating transition, and that they become increasingly coupled only toward the active state. Four well separated stable states, including I_1 and I_2 (marked by arrows), are clearly identified. In the intermediate state I_1, the A-loop is already half-open even though the helix αC has not started to undergo any rotation. The population of the 4 stable states is estimated to be: 57% (inactive), 20% (l_1), 14% (l_2), and 9% (active) (Fig. S8). Previous umbrella sampling simulations also indicated that transitions from the inactive state to intermediate configurations similar to l_1, in which the A-loop is open but the helix αC has not rotated, were energetically feasible (34). Additional analysis based on the transition probability matrix T provides an estimate of the mean first passage time or the inactive–active transition around 0.1–0.2 μsec (Eq. **2** in *SI Text* and Fig. S8). Although this value is quite uncertain, it could be improved with additional sampling and by refining the Markovian character of the transitions (38). According to Fig. 4*A*, the inactive state has the highest probability and is still dominant for the isolated catalytic domain. Nevertheless, the A-loop displays a sufficiently high propensity of opening up, which would permit the transphosphorylation of Y416 (20).

The free energy landscape extracted from the current all-atom simulations is qualitatively consistent with the results from a previous study based on a simplified 2-state coarse-grained model (29) (see Fig. S9). In both studies, the A-loop can progress without rotation of the αC helix near the inactive state until the 2 structural elements become more strongly coupled as the system approaches the active state. The main difference concerns the 2 intermediates I_1 and I_2, which are observed in the atomistic MD simulations. Those intermediates could not exist previously because only the 2 end-points were stable configurations by construction of the 2-state coarse-grained model. The qualitative accord between the 2 models suggests that the dominant features of the transition were captured in the current study.

Although the present computations did not include the SH2 and SH3 regulatory domains, it is possible to form a hypothesis about their influence on the activation process in the full-length kinase. Fig. 4*C* shows the SH2 and SH3 regulatory domains reconstructed onto the isolated kinase domain for 10 snapshots extracted randomly from each of the 4 stable states identified in Fig. 4*A*. The 2 domains were reconstructed relative to the catalytic domain N-terminal linker (Lys-252 to Trp-260) at their most likely position using the simple assumption of no internal structural change. These putative 3D models of the full-length Src kinase suggest that complete release of the SH2-SH3 clamp is not necessary in the early stages of the conformational transition. Specifically, the majority of the reconstructed configurations for I_1 are compatible with the down-regulated assembled form of the SH2-SH3 clamp. In contrast, the number of reconstructed configurations compatible with the down-regulated assembled form is reduced for I_2 and the active state (although the down-regulated form never appears to be strictly forbidden). This suggests that the catalytic domain should be able

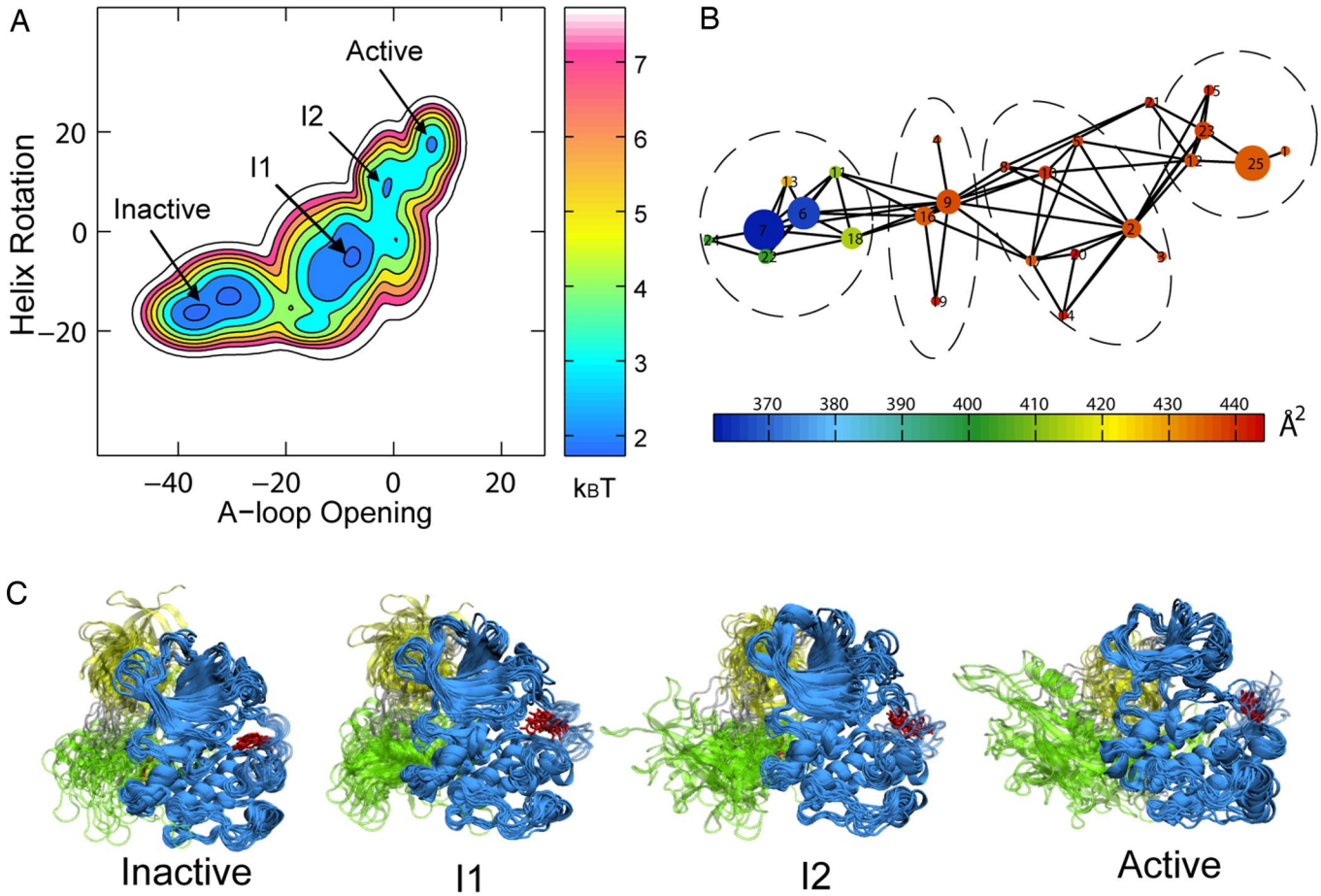

Fig. 4. Coarse-grained free energy landscape for A-loop and αC helix. (*A*) Two-dimensional free energy surface reconstructed from the color-code connectivity maps for the A-loop (by $\Delta Q_{\text{A-loop}}$) and the αC helix (by $\Delta Q_{\alpha C}$) as used in Fig. 2. On the coarse free energy surface, the 2 intermediates I_1 and I_2 are clearly identified. Each color level corresponds to 1 k_BT. (*B*) The solvent exposure of Tyr-416 in the middle of the A-loop on the connectivity map. The solvent-accessible surface areas (SASA) for Tyr-416 are projected and indicated by the color code on the map (in unit of Å^2). The SASA calculations were carried out by using a probe size of 1.4 Å around residue Tyr-416. (*C*) A total of 10 putative conformations for each state of the catalytic domain (blue) with the SH2 (green) and SH3 (yellow) binding domains of Inactive, I_1, I_2, and Active. The orientations of the SH2 and SH3 domains were modeled based on the overlay of backbone atoms of the catalytic domain N-terminal linker region (Lys-252 to Trp-260). The procedure for reconstructing the SH2 and SH3 domains is illustrated in Fig. S10. An ensemble of Try416 is highlighted in red.

to make transient visits to the state I_1, even while the SH2 domain remains fully "engaged" to the C-tail. While Y416 in the middle of the A-loop is buried in the activate site pocket in the inactive state, the A-loop opens up during the transitions to I_1, through which Y416 becomes solvent-accessible (Fig. 4*B*). The functional implication is that a complete transition to the fully active conformation is not required to render Y416 accessible for transphosphorylation. Transient visits of the catalytic domain to intermediate configurations are permitted, when the SH2-SH3 clamp remains fully engaged to the C-tail. Those transitions would periodically open up the A-loop, making Y416 available for the transphosphorylation event that ultimately locks the fully active state.

Conclusions

The activation of a multidomain allosteric signaling enzyme such as Src kinases involves complex large-scale conformational changes. We presented a general computational strategy to characterize such a conformational change in the Src catalytic domain by bridging multiple MD trajectories. Using all of the information gathered from relatively short all-atom molecular dynamics simulations, we were able to assemble a connectivity map allowing us to navigate the conformational landscape for the conformational transition of Src. Two intermediate states were identified along the conformational transition pathway and their key structural features were characterized. These intermediates appear as shallow basins in a coarse-grained free energy surface. The connectivity map and the coarse-grained free energy surface help provide a simplified structural description of the concerted motions of Src activation. The broad framework emerging from the present analysis sets the stage for a quantitative characterization of the effect of the SH2 and SH3 regulatory domains on kinase activation.

Methods

Simulation Details. Models of inactive and active states of Src kinases were generated from the crystal structures of Hck (PDB entry 1QCF) and Lck (PDB entry 3LCK), respectively (23, 20). In both cases, 2 associated Mg^{2+} ions and ATP were present in the active sites [modeled based on the ATP-Mn^{2+} conformation in PKA, PDB entry 1ATP (45)] and Tyr416 in the activation loop was not phosphorylated. The 2 structures with the Hck amino acid sequence were equilibrated by using 2 ns MD simulations with heavy-atom restraints to reach their local structural minima (33). All structures were solvated with a 150 mM KCl aqueous solution. Simulations were performed with the MD packages of CHARMM (46) and/or NAMD (47). Additional details can be found in *SI Text*.

Clustering and Mapping Method. To extend the short-time limit of all-atom simulations, we used the following 5-step computational strategy. Step 1: A targeted MD simulation (TMD) pulling the protein from the initial to the final state using a with root-mean-square-deviation (RMSD) restraints (e.g., ref. 48) was performed to generate an initial path from the inactive to active state

(33). Step 2: A large number of representative configurations spanning the TMD trajectories between the active and inactive conformations were selected and further relaxed in the presence of the RMSD restraint (33). Step 3: The restraining potential was released and free unbiased all-atom simulations with explicit solvent were generated. Step 4: The configurations from the unbiased trajectories were collected and clustered into N classes based on structural similarity; we chose $N = 25$, an upper bound for dividing the configurational space of the catalytic domain based on previous coarse-grained simulations (29). Step 5: The conformational landscape was finally built based on the transitions inferred from the unbiased MD trajectories between the various clusters. In steps 1 and 2, the RMSD restraint was applied to the αC helix, the Activation-loop, and 2 β-strands in the N-lobe (H289 to V329 and S397 to G437). For steps 4 and 5, the structural similarity was based on a list of 895 distances between pairwise Cα atoms forming interactions in either the inactive or the active. If the clusters were not all connected, either more configurations (step 2) or longer MD simulations (step 3) or both, were required. The procedure can be repeated until all clusters are connected and a fully connected map is achieved. In the case of the Src kinase domain, we selected 78 configurations (step 2), for which an unbiased MD trajectory of 12 ns was generated. The total resulting simulation time is $\approx$1 μs.

Reconstruction of a Coarse-Grained Free Energy Landscape. A coarse-grained free energy landscape can be reconstructed by computing the relative free energy

$$F(X, Y) = -k_B T \log \sum_{i=1}^{N} p_i e^{-\left(\frac{(X-X_i)^2}{2(\mu\sigma_{x_i})^2}+\frac{(Y-Y_i)^2}{2(\mu\sigma_{y_i})^2}\right)} \quad [1]$$

where X and Y are any 2 variables. k_B is the Boltzmann factor and T is the temperature of the system. In the Src activation, X and Y represent the A-loop opening and the αC helix rotation, respectively. X_i (and Y_i) and σ_{xi} (and σ_{yi}) represent the average and standard deviation for cluster i on the colored connectivity maps shown in Fig. 2. An empirical scaling factor $\mu = 0.45$ was used.

ACKNOWLEDGMENTS. We thank Luca Maragliano and D. Eric Walters for critical reading of this manuscript. This work was supported by the National Cancer Institute (CA-093577) (to B.R.) and Wadsworth Center new investigator funds (N.K.B.). Computational support was provided in part by the San Diego Supercomputer Center and the Teragrid project.

1. Martin GS (2001) The hunting of the Src. *Nat Rev Mol Cell Biol* 2:467–475.
2. Thomas SM, Brugge JS (1997) Cellular functions regulated by Src family kinases. *Annu Rev Cell* Dev Biol 13:513–609.
3. Yeatman TJ (2004) A renaissance for Src. *Nat Rev Cancer* 4:470–480.
4. Engen J, *et al.* (2008) Structure and dynamic regulation of src-family kinases. *Cell Mol Life Sci* pp 1–16.
5. Bose R, Holbert MA, Pickin KA, Cole PA (2006) Protein tyrosine kinase-substrate interactions. *Curr Opin Struct Biol* 16:668–675.
6. Noble MEM, Endicott JA, Johnson LN (2004) Protein kinase inhibitors: Insights into drug design from structure. *Science* 303:1800–1805.
7. Levinson NM, *et al.* (2006) Src-like inactive conformation in the Abl tyrosine kinase domain. *PLoS Biol* 4:e144.
8. Schindler T, *et al.* (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* 289:1938–1942.
9. Deininger MWN, Druker BJ (2003) Specific targeted therapy of chronic myelogenous leukemia with imatinib. *Pharmacol Rev* 55:401–423.
10. Baselga J (2006) Targeting tyrosine kinases in cancer: The second wave. *Science* 312:1175–1178.
11. Talpaz M, *et al.* (2006) Dasatinib in imatinib-resistant Philadelphia chromosome-positive Leukemias. *N Engl J Med* 354:2531–2541.
12. Das J, *et al.* (2006) 2-Aminothiazole as a novel kinase inhibitor template. structure-activity relationship studies toward the discovery of *N*-(2-Chloro-6-methylphenyl)-2-[[6-[4-(2-hydroxyethyl)-1- piperazinyl)]-2-methyl-4-pyrimidinyl]amino)]-1,3-thiazole-5-carboxamide (Dasatinib, BMS-354825) as a potent pan-Src kinase inhibitor. *J Med Chem* 49:6819–6832.
13. Kothe M, *et al.* (2007) Structure of the catalytic domain of human polo-like kinase 1. *Biochemistry* 46:5960–5971.
14. Knighton D, *et al.* (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 253:407–414.
15. Taylor S, Radzio-Andzelm E (1994) Three protein kinase structures define a common motif. *Structure* 15:345–355.
16. Gullingsrud J, Kim C, Taylor SS, McCammon JA (2006) Dynamic binding of regulatory subunit RIα. *Structure* 14:141–149.
17. Ogawa A, *et al.* (2002) Structure of the Carboxyl-terminal Src kinase, Csk. *J Biol Chem* 277:14351–14354.
18. Mills JE, *et al.* (2007) Novel disulfide bond in the SH2 domain of the C-terminal Src kinase controls catalytic activity. *J Mol Biol* 365:1460–1468.
19. Hantschel1 O, Superti-Furga G (2004) Regulation of the c-Abl and BcrAbl tyrosine kinases. *Nature Rev Mol Cell Biol* 5:33–44.
20. Yamaguchi H, Hendrickson WA (1996) Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature* 384:484–489.
21. Xu W, Harrison SC, Eck MJ (1997) Three-dimensional structure of the tyrosine kinase c-Src. *Nature* 385:595–602.
22. Sicheri F, Moarefi I, Kuriyan J (1997) Crystal structure of the Src family tyrosine kinase Hck. *Nature* 385:602–609.
23. Schindler T, *et al.* (1999) Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor. *Mol Cell* 3:639–648.
24. Xu W, Doshi A, Lei M, Eck MJ, Harrison SC (1999) Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Mol Cell* 3:629–638.
25. Cowan-Jacob SW, *et al.* (2005) The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. *Structure* 13:861–871.
26. Breitenlechner CB, *et al.* (2005) Crystal structures of active Src kinase domain complexes. *J Mol Biol* 353:222–231.
27. Hubbard SR, Till JH (2000) Protein tyrosine kinase structure and function. *Annu Rev Biochem* 69:373–398.
28. Banavali NK, Roux B (2005) The N-terminal end of the catalytic domain of Src kinase Hck is a conformational switch implicated in long-range allosteric regulation. *Structure* 13:1715–1723.
29. Yang S, Roux B (2008) Src kinase conformational activation: Thermodynamics, pathways, and mechanisms. *PLoS Comput Biol* 4:e1000047.
30. Young MA, Gonfloni S, Superti-Furga G, Roux B, 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.
31. Mendieta J, Gago F (2004) In silico activation of Src tyrosine kinase reveals the molecular basis for intramolecular autophosphorylation. *J Mol Graphics Model* 23:189–198.
32. Ozkirimli E, Post CB (2006) Src kinase activation: A switched electrostatic network. *Protein Sci* 15:1051–1062.
33. Banavali NK, Roux B (2007) Anatomy of a structural pathway for activation of the catalytic domain of Src kinase Hck. *Proteins* 67:1096–1112.
34. Banavali NK, Roux B (2008) Flexibility and charge asymmetry in the activation loop of Src tyrosine kinases. *Proteins* DOI: 10.1002/prot. 22153.
35. Bishop CM (1995) *Neural Networks for Pattern Recognition* (Oxford Unviersity Press).
36. Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels: Kinetic pathways through compact conformational space. *Proc Natl Acad Sci USA* 89:8721–8725.
37. Rao F, Caflisch A (2004) The protein folding network. *J Mol Biol* 342:299–306.
38. Chodera JD, Swope WC, Pitera JW, Dill KA (2006) Obtaining long-time protein folding dynamics from short-time molecular dynamics simulations. *Multiscale Model Simul* 5:1214–1226.
39. Ravasz E, Gnanakaran S, Toroczkai Z (2007) Network structure of protein folding pathways, arXiv:0705.0912v1.
40. Gfeller D, De Los Rios P, Caflisch A, Rao F (2007) Complex network analysis of free-energy landscapes. *Proc Natl Acad Sci USA* 104:1817–1822.
41. Pavletich NP (1999) Mechanisms of cyclin-dependent kinase regulation: Structures of cdks, their cyclin activators, and cip and ink4 inhibitors. *J Mol Biol* 287:821–828.
42. Morris MC, Gondeau C, Tainer JA, Divita G (2002) Kinetic mechanism of activation of the Cdk2/Cyclin A complex. key role of the c-lobe of the Cdk. *J Biol Chem* 277:23847–23853.
43. Laio A, Parrinello M (2002) Escaping free-energy minima. *Proc Natl Acad Sci USA* 99:12562–12566.
44. Maragliano L, Vanden-Eijnden E (2008) Single-sweep methods for free energy calculations. *J Chem Phys* 128:184110.
45. Zheng J, *et al.* (1993) 2.2 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr D* 49:362–365.
46. Brooks B, *et al.* (1983) CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J Comput Chem* 4:187–217.
47. Phillips J, *et al.* (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802.
48. van der Vaart A, Karplus M (2005) Simulation of conformational transitions by the restricted perturbation–targeted molecular dynamics method. *J Chem Phys* 122:114903.
49. Ozkirimli E, Yadav SS, Miller WT, Post CB (2008) An electrostatic network and long-range regulation of src kinases. *Protein Sci* 17:1871–1880.