Published in final edited form as:

J Phys Chem B. 2017 April 20; 121(15): 3352–3363. doi:10.1021/acs.jpcb.6b08409.

The Activation of c-Src Tyrosine Kinase: Conformational Transition Pathway and Free Energy Landscape

Mikolai Fajer^{1,‡}, Yilin Meng[‡], and Benoît Roux^{*}

Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois, 60637, USA

Abstract

Tyrosine kinases are important cellular signaling allosteric enzymes that regulate cell growth, proliferation, metabolism, differentiation and migration. Their activity must be tightly regulated, and disruption of this regulation can lead to a variety of diseases, particularly cancer. The non-receptor tyrosine kinase c-Src, a prototypical model system and a representative member of the Src-family, functions as complex multi-domain allosteric molecular switches comprising SH2 and SH3 domains regulating the activity of the catalytic domain. The broad picture of self-inhibition of c-Src via the SH2 and SH3 regulatory domains is well characterized from a structural point of view, but a detailed molecular mechanism understanding is nonetheless still lacking. Here, we use advanced computational methods based on all-atom molecular dynamics simulations with explicit solvent to advance our understanding of kinase activation. To elucidate the mechanism of regulation and self-inhibition, we have computed the pathway and the free energy landscapes for the “inactive-to-active” conformational transition of c-Src for different configurations of the SH2 and SH3 domains. Using the isolated c-Src catalytic domain as a baseline for comparison, it is observed that the SH2 and SH3 domains, depending upon their bound orientation, either promote the inactive or the active state of the catalytic domain. The regulatory structural information from the SH2-SH3 tandem is allosterically transmitted via the N-terminal linker of the catalytic domain.

^{*}Corresponding author: roux@uchicago.edu, Gordon Center for Integrative Science, 929 E 57th Street, Room W323B, Chicago, IL, 60637, USA, Telephone: 1-773-834-3557, Fax: 1-773-702-0439.

[†]Current Address is Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA

[‡]Co-first authors.

ASSOCIATED CONTENT

Supporting Information

Table S1 lists the orientation of W260 in several Src-family kinase crystal structures. The dihedral angles of the N-terminal linker region β -turns in those crystal structures are shown in Table S2. Figure S1 displays the projection of the inactive and active endpoint equilibration simulations onto the CVs used in umbrella sampling simulations. Figure S2 shows the convergence of the swarm-of-trajectories string method calculations. The progress of twelve representative order parameters along the optimized string pathways is shown in Figure S3. The projection of the converged string onto the corresponding PMF is plotted in Figure S4. Figure S5 illustrates the root-mean-squared deviation (RMSD) between any pair of images in the final iteration. The backbone and side-chain dihedral angles for W260 from 100 ns of MD of the wild-type c-Src kinase with the truncated linker in either the inactive or active conformation is shown in Figure S6 (reproduction of Supplemental Figure S9 from Meng *et al.*³¹). Ca-to-Ca distance between the *i*:*i*+3 residues in the β -turns of the N-terminal end of the KD is plotted in Figure S7. Figure S7 plots the compiled histograms of the backbone dihedral angles for W260 in the two SH3/SH2 states along the converged transition ensemble. Figure S9 shows the compiled histograms of the side-chain dihedral angles for W260 in the two SH3/SH2 states along the converged transition ensemble. Histogram of the distances that consist of the electrostatic network during the equilibration runs are displayed in Figure S10. Glu310:CD to Arg409:CZ distance (d_1) of the inactive endpoint is plotted against the index of string method iterations in Figure S11. The coordinates of the converged strings (the last 10 iterations) from the assembled state and the re-assembled state are also given in the supporting information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

The authors declare no conflict of interest.

Analysis of the conformational transition pathways also illustrates the importance of the conserved tryptophan 260 in activating c-Src, and reveals a series of concerted events during the activation process.

Graphical Abstract

INTRODUCTION

Protein kinases are signaling enzymes that serve to regulate cell growth, proliferation, metabolism, differentiation and migration^{1, 2}. The human kinome encodes over 500 protein kinases, representing approximately 2% of the entire genome³. Owing to their critical functional importance, the activity of these protein kinases must be tightly regulated. Malfunction of protein kinases is often associated with a wide range of disease, including inflammation, diabetes, and particularly cancers.

A prototypical system of central interest is provided by the non-receptor tyrosine kinase c-Src, which belongs to the family of nine Src-related non-receptor tyrosine kinases (SFK) sharing a similar regulatory mechanism (Src, Yes, Fyn, Lyn, Lck, Blk, Hck, Fgr, and Yrk)⁴. c-Src was the first proto-oncogenic protein ever described^{5, 6}. It is ubiquitously expressed in all cell types and is discovered to participate in multiple signaling pathways which regulate cell growth, differentiation, etc⁷. All nine SFKs share a common multi-domain topology, which comprises a short membrane-associating SH4 segment and a “unique” (U) segment whose sequence varies widely, followed by two peptide binding modules in the SH2 and SH3 domains, and finally a catalytic tyrosine kinase domain (KD), i.e., SH4-U-SH3-SH2-KD⁷⁻¹⁰. The KD comprises about 300 residues and is highly homologous to those of serine and threonine kinases¹⁰. The SH2 (~100 residues) and SH3 (~60 residues) domains are compact structures that bind to sequences containing a phosphotyrosine⁷ and to a poly-proline (PPII) helix containing the motif PxxP¹¹, respectively¹².

All of the Src-family kinases function as complex allosteric molecular switches whose catalytic activity can be modulated in response to specific cellular signals, such as growth factors or cytokines binding to membrane-bound receptor proteins. Phosphorylation of two distinct tyrosines has opposing effects on catalytic activity¹³. Phosphorylated Tyr527 (chicken c-Src numbering system) in the C-terminus tail of the catalytic domain interacts with the SH2 domain, resulting in an autoinhibition of kinase activity^{14, 15}. Dephosphorylation of Tyr527, or the binding of external ligands to the SH2 or the SH3 domains, results in the activation of the kinase¹⁶⁻¹⁹. Phosphorylation of Tyr416, a residue located in the activation loop (A-loop) near the active site in the catalytic domain, is

necessary for full kinase activity. Up-regulation of the kinase can be induced by auto-phosphorylation of Tyr416^{20, 21}. An active Src kinase implies that the required conformational changes have taken place such that the KD has become catalytically competent. The down- and up-regulation of Src kinase activity are allosterically linked to the assembly/disassembly process of the tertiary structure of the protein.

Crystallographic structures of the auto-inhibited c-Src and Hck show that the SH3 and SH2 domains are assembled on the backside of the catalytic domain, with the SH2 domain binding to phosphorylated Tyr527 and the SH3 domain binding to the PxxP region of the linker connecting the SH2 and the KD^{22–25}. The auto-inhibited conformation of the enzyme is stabilized by the SH2-SH3 domains, which appear to act as a “clamp” that locks the KD in the inactive state. Disruption of either the SH2-pY527 or SH3-linker interaction destabilizes the auto-inhibitory conformation, leading eventually to the activation of the enzyme. Phosphorylation of Tyr416 in the activation (A-) loop locks the KD in a catalytically competent conformation^{26, 27}. Furthermore, a crystal structure with neither Tyr527 nor Tyr416 phosphorylated showed an active-like state with the SH2/S3 domains in a re-assembled conformation on the top of the KD instead of on the backside²⁸. In the aforementioned structure, the KD is in a crucial state for *trans*-autophosphorylation; Tyr416 is not phosphorylated but the KD still adopts a catalytically competent conformation. Although the conformational state of the KD is important for initiating the bimolecular *trans*-autophosphorylation reaction, the importance of the re-assembled SH2/S3 domains conformation is poorly understood.

A reasonable hypothesis suggested by the available data is that the activating conformational transition within the KD is allosterically modulated, somehow, by the configuration of the regulatory SH2 and SH3 domains. Nevertheless, basic questions about the nature of the allosteric coupling between the various structural elements and the molecular mechanisms underlying the regulation of c-Src remained unanswered. For example, how is the SH2-SH3 “clamp” inhibiting the conformational transition of the KD toward a catalytically competent conformation in the assembled state? Does it act mainly by increasing the energy barrier separating the inactive and active states (kinetic control), or by raising the overall free energy of the active conformation to make it less stable (population shift)? And how the stability of the catalytically competent conformation enhanced by the SH2-SH3 tandem in the re-assembled state?

By answering such questions using all-atom molecular dynamics (MD) simulations, our goal is to clarify the mechanism by which the SH2 and SH3 regulatory domains control the activation of the kinase. Several MD simulation studies have studied the regulation of SFKs and focused on SH2 and the C-terminal tail binding²⁹, the role of the linker^{30–32}, allosteric network³³, and the conformational changes and phosphorylations of the isolated KD^{26, 34–45}. Pathway methods, umbrella sampling (US) calculations, and Markov state model (MSM) constructed from aggregate data from unbiased trajectories have proved to be effective in understanding the conformational changes in SFKs. An incomplete, though cogent picture, has emerged from the above computational studies, providing a broad framework for understanding the sequence of events leading to full kinase activation. According to this picture, it is proposed that the catalytic domain with Tyr416 in the A-loop

unphosphorylated makes transient short-lived visits to an active-like catalytically competent conformation. These transient visits increase the availability of Tyr416 to be phosphorylated during a bimolecular encounter, a process that ultimately locks the enzyme in its catalytically competent state^{26, 27}.

In the present effort, a full-length kinase atomic model comprising the SH3, SH2, linker, and catalytic domain was employed to study the role of the SH3-SH2 tandem on the inactive-to-active conformational transition taking place within the KD. We used a combination of targeted molecular dynamics (TMD), swarms-of-trajectories string method^{39, 46}, and umbrella sampling (US) simulations to investigate the inactive-to-active transition pathway within the catalytic domain, and the associated free energy landscape for two different configurations of the regulatory SH3 and SH2 domains. To the best of our knowledge, this is the first computational study that investigates conformational changes in a full-length c-Src kinase. Our hope is that the computations will provide significant new insights to understand the function and regulation of tyrosine kinases.

THEORETICAL METHODS

a) System preparation

The inactive assembled state and the active re-assembled state were modeled from two X-ray crystal structures, with the PDB codes 2src²⁴ and 1y57²⁸, respectively (see Figure 1A and 1B in Ref. 26). These two systems were prepared with the CHARMM-GUI⁴⁷. A truncated octahedron box of TIP3P water molecules⁴⁸ with a buffer of 10 Å from the protein to the edge was used to solvate the systems, using 0.15 M NaCl in addition to neutralizing ions. An ATP molecule and two magnesium ions coordinating the ATP were added based on a high-resolution structure of protein kinase A (PDB id: 1ATP⁴⁹). Y416 in the A-loop is not phosphorylated in any of the construct, while Y527 is phosphorylated only in the assembled construct. The assembled, and re-assembled constructs comprise of approximately 66,000 and 148,000 atoms respectively. The isolated KD, construct is smaller and comprises approximately 38,000 atoms. All additional simulations were performed with the NAMD package⁵⁰ version 2.10, using the all-atom CHARMM27 force field⁵¹. The solvent was minimized while holding the protein fixed for 1000 steepest descent steps, and followed up with another 1000 steps for the entire system. The minimized system was then brought to a temperature of 300K by initializing the velocities at 100K and increasing the temperature by 20K every picosecond (ps). Finally, a short equilibration run of 5ns was performed. All MD simulations, including the swarms-of-trajectories calculations, were performed in the isobaric-isothermal (NPT) ensemble. The pressure and temperature were controlled by Langevin piston method⁵² and Langevin dynamics (using a coupling constant of 1.0 ps⁻¹) and kept at 1 atmosphere pressure and 300 K, respectively. Periodic boundary conditions were applied and the Particle Mesh Ewald (PME)⁵³ was used to treat long-range non-bonded interactions. The short-range non-bonded interactions were truncated at 12 Å, with a switching function turned on from 10 to 12 Å. The non-bonded list updates were truncated at 14 Å. Covalent bonds involving hydrogen atoms were constrained at their equilibrium distances so that a 2 fs time step was used throughout all simulations.

b) String pathway and swarms-of-trajectories simulations

The string method formulated by Vanden-Eijnden represents the pathway between two endpoints as an ordered sequence of M evenly distributed discrete conformations called “images” in a space of collective variables, $[\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \dots, \mathbf{Z}^{(M)}]$ ^{46, 54, 55}. Starting from an initial path, we use the mean drifts of each image $\langle \dot{\mathbf{Z}} \rangle$ calculated from swarms of unbiased MD trajectories to iteratively refine the string^{39, 46}. At each iteration, the string is re-parametrized to impose an equal Euclidean distance between adjacent images in the space of collective variables $\mathbf{Z}$ in order to maintain a well-resolved path over the entire transition⁵⁴. The swarms-of-trajectories string method was recently implemented into NAMD using the multiple copies interface⁵⁶, allowing for extremely scalable computations. Here, a total of 64 images and 32 individual swarms per image were used to calculate the mean drift, for a combined 2048 simultaneous copies. These were distributed over the Blue Gene/P and Blue Gene/Q supercomputers at the Argonne Leadership Computing Facility. The re-parametrization phase used a smoothing constant of 0.1 and a staged harmonic restraint that increased the force constant on the CVs from 0 kcal/(mol·Å²) to 50 kcal/(mol·Å²) in 5 steps. The number of MD steps spent in the re-parameterization phase was set equal to the length of swarms. The progress of the swarms-of-trajectories iterations was monitored with the discrete Fréchet distance⁵⁷ of the CVs between all pairs of strings. The Fréchet distance respects the ordering of the string as opposed to the Hausdorff distance, and emphasizes the maximal difference between two strings rather than an average like in the root mean squared distance (RMSD). By virtue of the thermal noise added from the swarms-of-trajectories, the algorithm samples the ensemble of plausible (high probability) paths between the two endpoints. For this reason, the last 20 iterations for each system were taken as a converged ensemble, so each of the 64 images along the string had 32×20 structures from which projections onto other degrees of freedom could be made. The projections were made using VMD⁵⁸ and plotted using the Python Matplotlib package⁵⁹. The structural images were rendered using the Tachyon ray tracing library or the PyMOL software package⁶⁰.

c) String pathway initialization

During the activation process, the KD also undergoes important conformational changes involving primarily the α C-helix and the A-loop (where Y416 is located). This inactive-to-active conformational transition in the KD enables the catalytic function and has been the subject of our previous studies. To study the role of the regulatory tandem (i.e. SH3-SH2) on the inactive-to-active conformational transition in the KD, we constructed three strings in this investigation, with one in the absence of the regulatory tandem and with the tandem in the assembled and re-assembled states respectively. The initial guess for the inactive-to-active transition pathway for c-Src kinase was taken from our previous work on the KD.³⁸ The same 50 Cartesian collective variables (CVs) from Gan *et al.*³⁹ were used to describe the string, and represent the C α position and a single atom in the amino acid side-chain (details in the Supplementary Material of Ref 39). The assembled inactive system from PDB 2src was then pulled along the string from the inactive to the active state by TMD, using a pulling rate of 10 Å/ns and a harmonic restraint of 10 kcal/mol on the CVs. A total of 64 equidistant structures were saved along the pathway and represent the initial assembled state string. The re-assembled state string was constructed in a similar manner using the string

from our previous work on the KD, except the TMD was carried out in the reverse direction, going from the active (X-ray structure) to the inactive conformation.

d) Umbrella Sampling calculations of the full-length constructs

Two sets of two-dimensional (2D-) US calculations were performed to characterize the free energy associated with the inactive-to-active conformational change of the KD when the SH3-SH2 regulatory tandem is in the assembled and the re-assembled state respectively. The order parameters utilized to monitor the inactive-to-active conformational transition were the same ones previously used to study the role of Y416 phosphorylation: the average distance of the three $i:i+4$ hydrogen bonds in the short α -helix of the A-loop (D413:O to T417:N, N414:O to A418:N, and E415:O to R419:N; d_3) is used to reflect the opening of the A-loop. The difference between the distances of the E310 to R409 salt-bridge (d_1) and E310 to K295 salt-bridge (d_2) is used to characterize the rotation of the α C-helix, either facing the A-loop at negative values or the N-lobe for positive values. Graphical representations of the two order parameters and the conformational change subjected to our umbrella sampling calculations can be found in Figure 2 of Meng *et al.*²⁶

The self-learning adaptive strategy⁶¹ was also employed in both sets of US calculations. To further reduce the computational cost and focus on the regions surrounding the minimum free energy pathway, the refined transition pathways for the assembled and re-assembled systems were used to seed the self-learning adaptive US calculations. The starting structures in the first cycle of the adaptive strategy were all provided by the converged strings, i.e. the last 20 iterations of the swarms-of-trajectories string method calculation. These converged strings were first projected onto the 2D- order parameter space. Umbrella sampling windows whose centers were within 1.5 Å to the projected strings are used in the first cycle, and the nearest image to an umbrella center was served as the starting structure of that particular window. The free energy threshold was adjusted on-the-fly, and a free energy threshold of 9.0 kcal/mol was finally utilized in the self-learning adaptive procedure.

A harmonic restraining potential with a force constant of 10 kcal/(mol·Å²) and a uniform spacing of 0.5 Å were applied on each order parameter. A total of 555 and 338 umbrella windows were used for the re-assembled and the assembled states, respectively. A total of 40 and 30 ns of MD propagation were carried out per US window for the re-assembled and the assembled states, respectively. The shorter simulation time per umbrella window for the assembled state mainly results from a faster convergence rate, compared with the umbrella sampling calculations of the re-assembled state (i.e., the SH2-SH3-linker tandem fluctuates in a greater extent in the re-assembled construct as shown in Fig. 3B, necessitating a longer simulation time to improve the sampling). In total, approximately 32 μ s of aggregated sampling time was spent in our US calculations. The weighted histogram analysis method (WHAM)⁶² was used to combine the time-series of the order parameters from each set of 2D-US calculations and to yield the potential of the mean force (PMF).

e) Umbrella sampling calculations of the linker-KD construct

An additional set of 2D-US calculations were carried out to construct the free energy landscape associated with the inactive-to-active transition of the KD in a construct consisting

of the linker and the KD. The linker is restrained to adopt the same conformation as that in the assembled state using a harmonic potential with a force constant of 10 kcal/(mol·Å²) applied to its non-hydrogen atoms. Essentially, the purpose of this set of US calculations is to evaluate the role of the linker in the auto-inhibition of the KD. If the free energy landscape of activation from this set of US calculation is the same as that from the SH3-SH2-KD construct when the SH3-SH2 adopts an assembled conformation, the SH2-pY527 docking and the SH3-linker binding do not contribute directly to the auto-inhibition of the KD. All the 338 umbrella windows from the assembled construct were utilized. The SH3 and SH2 domains were removed. The same harmonic potentials were applied to the two order parameters as described above. Each umbrella window was propagated for 11 ns (3.7 μs of aggregated sampling time), while the last 10 ns were utilized for WHAM post-processing.

RESULTS AND DISCUSSION

a) Stability of the string endpoint states

The activation pathway in the KD was determined the isolated KD, included here for the sake of comparison, and for two configurations of the SH2 and SH3 regulatory domains in the full-length kinase. Those configurations are the assembled state, which is illustrated by the X-ray structure with PDB code 2src²⁴, and the so-called re-assembled state, which is illustrated by the X-ray structure with PDB code 1y57²⁸. The assembled state (2src) is believed to represent an autoinhibited inactive form of the kinase. In contrast, the status of the re-assembled state (1y57) is less clear. While Tyr416 in the A-loop is unphosphorylated, as would be expected for a kinase locked in a constitutively active state, the key residues in the active site are properly configured to allow catalysis. Our view, based on previous studies, is that the configuration of the KD in the re-assembled state represents a catalytically competent active-like conformation that can exist transiently prior being locked via the phosphorylation of Tyr416 in the A-loop.

As a first step, the four endpoint conformational states in the two strings were simulated without any bias and restraints for 50 ns to ascertain their stability before being used in the pathway refinement. Each string has one endpoint state that corresponds to a conformation observed by X-ray crystallography, namely, the assembled inactive kinase, and the re-assembled active-like kinase. In addition, each string has an artificial or virtual endpoint that has not been observed by crystallography, namely the re-assembled inactive kinase and the assembled active kinase. To validate the stability of the virtual re-assembled inactive and the assembled active endpoint states more extensively, they were further simulated for 507 and 514 ns, respectively. The stability of the conformations was monitored by time-series of the two CVs used in the umbrella sampling, shown in the supplementary information Fig. S1. For comparison, the 50 ns MD trajectories for the (crystallographically-observed) re-assembled active state and the assembled inactive state are also shown in Fig. S1. These trajectories remain centered on the X-ray crystal structure values and show a moderate amount of fluctuation. In contrast, the artificially created endpoints display larger fluctuations and adopt slightly different active and inactive states. They do remain in either the re-assembled inactive or the assembled active state, though, suggesting that at least on

the 500 ns timescale these kinase conformations are stable, indicating that the re-assembled inactive state and the assembled active state can be used as endpoints in their corresponding strings.

b) Convergence of the swarms-of-trajectories

It is known that the length of the short MD trajectories in the string method with swarms-of-trajectories can strongly affect the outcome, causing the pathway to either become stuck in local minima when the trajectories are too short, or depart considerably from the minimum free energy path (MFEP) when the trajectories are too long⁶³. These observations were made for small systems such as the 2D Mueller potential or the blocked alanine dipeptide, and tests on larger systems have not yet been performed. The KD-only swarms-of-trajectories iterations performed by Gan *et al.* had utilized a swarm trajectory length of 1 ps³⁹. To ascertain the sensitivity of the results to the length of the trajectories, the smaller, assembled full-length construct was selected and three independent iterations were run with swarm trajectory lengths of 1, 5 and 10 ps. The discrete Fréchet distance⁵⁷ of each iteration to the initial string was used to monitor whether the string was becoming stuck in a local minima at the lower trajectory lengths. The result of this analysis is shown in Fig. 1a. The 1 ps swarm trajectory length started to reach a plateau well before the 5 and 10 ps swarm trajectory lengths, and we chose to stop the iterations early. In the case of the re-assembled state, the 5 ps trajectory length reached a plateau before the 10 ps iterations. The final strings from the 10 ps trajectory length iterations were taken to represent the most accurate transition pathway for each system. In light of this decision, the string for the isolated KD was re-optimized with a 10 ps trajectory length for the sake of consistency. The convergence of the string was monitored with the discrete Fréchet distance of each iteration to the final string, shown in Fig. 1b. The last third of both the 5 and 10 ps swarm trajectories for the assembled full-length construct shows a plateau at low Fréchet distances, indicating that the string has stopped evolving. In the case of the 10 ps re-assembled full-length and isolated KDs, there remains a gradual downward slope near the end of the iterations. The slope was deemed small enough to consider the strings converged. One can also study the Fréchet distance between every pair of string iterations, although more complicated than the general divergence from a single structure, and this is shown in the Fig. S2. Contiguous regions of low Fréchet distance show local minima of the strings and each of the three states show an extended region of low Fréchet distances for at least the last 20 iterations. Thus, the strings for the final 20 iterations of each state were treated as an ensemble describing the local minimum free energy path of the inactive to active transition.

The converged inactive-to-active transition string pathways for the assembled and re-assembled states and the isolated KD are broadly very similar. This is supported by monitoring the evolution of a set of representative order parameters following the inactive-to-active transition shown in Fig. S3, and the 2D the projection shown in Fig. S4. Generally, the differences in the paths are related to the differences at the endpoints. For example, the re-assembled active A-loop has a different W260 conformation, and the α -helix turn in the inactive A-loop between residues 413 to 419 (d_3 in Fig. 2) is a little more structured. According to Fig. S3, there are also some differences regarding the timing of specific events along the string, although the overall ordering is generally preserved. For example, the

distance d_1 breaks around image index 44 for the assembled state, while it breaks around image index 50 for the re-assembled state, or the distance d_2 forms around image index 48 for the assembled state while it forms around image index 58 for the re-assembled state. In both paths, the order of the microscopic event remains the same (d_1 breaks and then d_2 forms), although there is a slight shift of the image indices between the assembled and re-assembled state. The slight shift in image index is not directly prevented by the reparametrization procedure, which aims at keeping adjacent images equidistant. This is the main reason why we used the Fréchet distance in assessing the convergence of the string to try and avoid the shifting of image indices. Some difference involving the charged residues forming a network of electrostatic interactions are noted, and will be analyzed below.

c) Free energy landscape of the full-length construct

Free energy landscapes determined from (US) calculations^{26, 31} and from Markov state models (MSM)³⁸ can provide extremely valuable information to understand the conformational transitions underlying the activation of protein kinases. To investigate the allosteric role of the SH3-SH2 tandem on the auto-inhibition/activation of the KD, we calculated the potential of mean force (PMF) along two order parameters associated with the motion of the A-loop and the α C helix using US simulations for the assembled and the re-assembled full-length constructs. The results are displayed in Fig. 2a, 2b, and 2d. For comparison, the PMF of the isolated KD previously reported in Meng *et al.*³¹ is also reproduced in Fig. 2b. In all the 2D-PMFs shown in Fig. 2, the inactive KD at the bottom left and the active-like at the top right of the states consistently correspond to local minima, indicating that the endpoints represent meaningful and stable conformations. Nonetheless, the relative population of these endpoint states depends strongly on the context presented by the different configurations of the SH3-SH2 tandem. One can also observe that an intermediate state with the A-loop open and the α C-helix oriented outward (bottom right of the 2D-PMF), is only observed in the 2D-PMF of the isolated KD (Fig. 2b). The results, obtained from an isolated KD, can serve a model for the free energy landscape in a full-length kinase in which the SH3-SH2 tandem would be disengaged and decoupled completely from the KD.

The calculations demonstrate that the presence of the SH3-SH2 tandem in the assembled conformation imposes a significant thermodynamic penalty on the active conformation of the KD, as revealed by a comparison of Fig. 2a and 2b. After integrating the PMF of the isolated KD (Fig. 2b), the free energy of activation ($\Delta G(\text{activation}) \equiv G(\text{active-like}) - G(\text{inactive})$) is 3.6 kcal/mol. However, the same quantity is ~8 kcal/mol in the presence of the SH3-SH2 tandem in the assembled conformation. The transition from the inactive conformation to the active-like one is essentially prohibited when the SH2-SH3 clamp is bound to the backside of the KD. This result reveals why the assembled configuration is truly autoinhibited. According to the crystallographic structures of the assembled c-Src^{23, 24, 64}, the tandem does not block access to the catalytic site of the KD to neither ATP + Mg²⁺ nor the substrate, i.e. the KD can catalyze the transfer of the γ -phosphate from ATP to a substrate when it adopts an active conformation even though the regulatory tandem is assembled. The autoinhibition is thus believed to be caused by the inability of the KD to adopt the catalytically competent active conformation. The calculated free energy landscape

for the assembled full-length kinase is in accord with this explanation. The molecular determinants of the mechanism of auto-inhibition are investigated further in the following section.

The most prominent result of moving the SH2-SH3 tandem to the re-assembled, configuration is a shift of the free energy minimum from the inactive toward the active-like conformational state. The calculated free energy $\Delta G(\text{activation})$ is -1.0 kcal/mol, even though Tyr416 is unphosphorylated. The re-assembled full-length kinase is, thus, expected to be intrinsically very active. However, the calculated free energy landscape indicates that the inactive conformation remains thermodynamically accessible, with an estimated population of $\sim 16\%$. This means that a significant fraction of the structural ensemble is still in an inactive conformation, which explains why the phosphorylation A-loop is required to lock the kinase in the active state to achieve maximum activity²⁶. In addition, we also observe a shift of the position of the free energy minimum corresponding to the active state to higher values of d_3 relative to the assembled full-length kinase construct, consistent with the observations from the unbiased MD simulations of the endpoints as shown in Fig. S1.

It is of interest to examine the converged string pathways, optimized in the multi-dimensional space of collective variables, projected onto the reduced 2D subspace of the PMF. The projection, displayed in Fig. S4, provides a combined view of the transition pathway and the underlying free energy landscape. From Fig. S4, it appears that all three strings roughly follow the minimum free energy path on this 2D free energy surface. It should be noted that the images of the string are no longer equidistant when projected onto a space of lower dimensionality. For instance, the clustering of images near the inactive, intermediate, and active states suggests that the pathway in this region is oversimplified by the 2D projection. The images are re-parameterized during the swarms-of-trajectories iterations to remain equidistant along the entire length of the string in the space of collective variables (Fig. S5), therefore, this clustering in 2D is due to movement along collective variables that are orthogonal to the two order parameters used to calculate the 2D-PMF. Furthermore, the 10 ps string of the isolated KD demonstrates that a small shift in $d_1 - d_2$ also occurs (Fig. S4a), and is investigated further below in the section entitled “Electrostatic Network”.

d) N-terminal end of the kinase domain

The linker region connecting the SH2 through the N-terminal end of the KD has long been identified as one of the important factors controlling the auto-inhibition of c-Src kinases. The impact of the linker region has been studied by site-specific mutagenesis experiments^{65–68}, e.g. the L255V⁶⁵ and W260A⁶⁸ mutations in c-Src resulted in a highly elevated kinase activity. To identify the molecular determinants that inhibit the opening of the A-loop and rotation of the α C-helix, two events associated with the inactive-to-active conformational transition, we initially searched for residues within 5–6 Å from the α C-helix or the A-loop from the SH2 and SH3 domains. However, this search uncovered only D117 in the SH3 domain within a distance of 6 Å at which non-bonded interactions are negligible. The search was then extended to include the linker region, leading to the identification of Q253, L255, and W260. Interestingly, both L255 and W260 have been recognized to be

crucial for auto-inhibition of the kinase. This simple analysis suggests that whereas the SH2 and SH3 domains, themselves, do not couple strongly to the KD, the linker region clearly plays a more critical role. This observation leads us to hypothesize that the role of the SH2 and SH3 domains is essentially to stabilize the N-terminal linker region in an “inhibiting conformation”, but it is the linker region itself that is actually responsible for the direct inhibition of the catalytic domain.

To test this hypothesis, we carried out an additional set of 2D-US calculations in which the SH3/SH2 tandem was removed while the conformation of the linker was restrained to that in the auto-inhibitory conformation. The resulting 2D-PMF, which is displayed in Fig. 2c, bears a striking resemblance (almost quantitative) to that of the assembled full-length construct (Fig. 2a). This similarity reveals that auto-inhibition of the KD is indeed completely attributed to the conformation of the linker N-terminal region. Consequently, we posit that the importance of the interactions between the SH2 domain and the pY527 in the C-terminal tail, as well as between the SH3 domain and the PxxP motif in the linker in the assembled state is essentially to stabilize the N-terminal of the KD in its “inhibiting conformation”. To further validate this hypothesis, we also examined the flexibility of the N-terminal linker region in the 2D-US calculations of both full-length constructs. The 2D-PMFs of α C-helix rotation and the RMSD of the linker with respect to the corresponding X-ray structure are constructed and are displayed in Fig. 3 (note that this result was obtained with no additional US simulations). In the assembled full-length construct (Fig. 3a), it is observed that the deviations of the linker from the X-ray conformation are quite small in the assembled full-length construct, which is reflected by small values and the narrow range of the RMSD. In contrast, the linker is much more flexible in the re-assembled full-length construct (Fig. 3b), regardless of the conformation of the KD. This behavior confirms that SH2-pY527 docking and SH3-linker binding restrain the motion of the linker to assure the auto-inhibition is effective. Additionally, the linker demonstrates a greater flexibility when the KD adopts an active conformation in the re-assembled full-length construct than that when the KD adopts an inactive conformation in the same construct. This behavior further shows that the inactive conformation of the KD has an intrinsic propensity to restrain the motion of the linker.

A conserved tryptophan residue (W260 in c-Src) at the interface between the linker and the KD has long been identified to be critical in regulating the kinase activity of tyrosine kinases^{68, 69}. For example, in Src-family tyrosine kinases such as c-Src and Hck, mutating this tryptophan to alanine elevated kinase activity, whereas the corresponding mutation in Tec-family tyrosine kinases such as Btk and Itk show an opposite effect^{70, 71}. A comparison of the X-ray structures for Src kinases (see Fig. 4 and Table S1) shows that in the inactive state of both Hck and c-Src the W260 residue is buried in a hydrophobic cavity between the α C-helix and N-lobe. The active state of both c-Src and Lck shows the W260 rotated out and the cavity collapsed as the α C-helix moves closer into the N-lobe. The linker also changes its conformation, moving from a type-II β -turn between residues Ala256-Ala259 in the inactive state to a type-I β -turn between residues Asp258-Glu261 in the active state, which has drastically changed where the SH2/SH3 domains are with respect to the KD (Table S2).

A truncated linker and KD (residues 253–523) has been studied with unbiased molecular dynamics and umbrella sampling^{30, 31}. The umbrella sampling of the truncated construct by Meng *et al.*³¹ confirms that the Ala256-Ala259 β -turn is stable when the kinase is in the active state, and adopts the Asp258-Glu261 β -turn when the kinase is in the inactive state. Furthermore, the free MD of the truncated constructs show that the backbone (ϕ , ψ) distributions for W260 are not affected by changing the linker orientation from inactive to active (reproduced as Fig. S6). The inactive state has a single, well defined (χ_1 , χ_2) for W260 seen in both Hck and c-Src crystal structures^{24, 25}, as well as the MD simulation. The active state has several possible conformations, one of which is (–75, 90).

The inactive-to-active conformational activation pathway from the string method allows us to examine the structural behavior of the N-terminal linker region in the context of the assembled or re-assembled state. Throughout the transition, the linker remains in either the Ala256-Ala259 β -turn, in the case of the assembled state, or the Asp258-Glu261 β -turn, in the case of the re-assembled state (Fig. S7). The backbone (ϕ , ψ) also remain constant throughout the transition, closely resembling the results of the unrestrained MD by Meng *et al.* (Fig. S8). Strikingly, it is only the conformational activation pathway for the re-assembled state that exhibits a rotation of the W260 side-chain (Fig. S9). A comparison of the snapshots along the conformational activation pathway for the assembled state in Fig. 4 shows that the W260 has been shifted out of the hydrophobic pocket to allow the α C-helix to rotate and move closer to the N-lobe. This is expected in the active state, but here this was achieved by pushing the β_1 strand out of position. This is presumably caused by the competing tension of the N-terminal linker being locked into the assembled state and the α C-helix rotation. The re-assembled orientation of the N-terminal linker is more flexible, though, and a rotation of the W260 side-chain is sufficient to move from the inactive hydrophobic pocket (red in Fig. 4c–d) to the active state (magenta in Fig. 4c–d).

e) Electrostatic network

A network of electrostatic interactions between charged residues that are altered upon activation of the kinase has previously been identified^{44, 45}. These include the salt bridge between E310 and R409 or K295 (d_1 and d_2) interactions used in this study to monitor the rotation of the α C-helix. The progress of these interactions along the three different conformational activation pathways from the string method is shown in Fig. 5. The salt bridge between E310 and R409 occasionally broke during the equilibration of the inactive endpoint (Fig. S10), although this conformation is stable for tens of nanoseconds. The kinase-only inactive endpoint equilibration finished with the broken E310 to R409 interaction, and it did not reform for the endpoint during the swarms-of-trajectories iterations (Fig. S11). The re-assembled state inactive endpoint began with the E310-R409 interaction, but this broke after approximately 75 iterations without reforming. Regardless of whether the E310-R409 salt-bridge was formed at the inactive endpoint, in both the assembled and re-assembled states the salt-bridge was present during the early parts of the inactive-to-active transition, indicating that it is important during the opening of the A-loop.

While the pathways for the assembled and re-assembled states are broadly very similar (see Figs. S3 and S4), one notable difference concerns the E310-K295 interaction. The unbiased

equilibration MD for the inactive endpoint shows little difference with both the assembled and re-assembled states having these residues separated by approximately 12 Å. However, the E310-K295 distance starts to decrease much earlier in the case of the re-assembled state than the assembled state, passing through a few intermediate states as opposed to the single switch seen in both the kinase-only and assembled states (Fig. 5b). There are small variations of the HRD-motif along the pathway for the assembled and re-assembled states (Fig. 5c–d). Of particular interest is the R385-Y416 interaction, which appears to be strengthened at the re-assembled active endpoint relative to the assembled and kinase-only systems. This is seen by the shift of the active endpoint equilibrium distribution (Fig. S10d), and the contact being fully formed at the activated KD endpoint for the string in the re-assembled state (Fig. 5d). After the phosphorylation of Y416, the R385-pY416 interaction is part of a triad of arginine residues that stabilize the active-like endpoint²⁶.

f) Ordering of events

In order to ultimately understand the activating conformational transition with the KD in atomic detail, we analyzed the progress over a set of representative order parameters along the string activation pathway. We focus the present analysis on the pathway in the case of the re-assembled state for the sake of simplicity. The pathway for the two other cases are fairly similar (Fig. S3). Each order parameter was normalized to show the end-point closest to the inactive state in blue, and the end-point closest to the active state in red. The resulting “progress plots” were then sorted manually and displayed in Fig. 6, allowing inference of the ordering of microscopic events along the activating conformational transition. As is to be expected from previous work³⁹, the first step in the inactive-to-active transition is the breaking of the two short α -helices in the A-loop as it begins to open. This is followed by the breaking of the Arg419 interaction with the backbone carbonyl of Ala434 in the α EF-helix in the C-lobe, allowing the A-loop to completely unfold and form the first intermediate state. The side-chain of Trp260 then rotates out of the hydrophobic pocket involving the α C-helix, although the latter starts to move only later. The second intermediate state is formed, as Phe424 flips away from the α EF-helix and associates with the hydrophobic elements of the α G-helix. The final changes leading to the formation of the Lys295-Glu310 salt bridge typical of the catalytically competent conformation occur almost simultaneously: Glu310 breaks away from Arg409 and unlocks the interaction that anchors the α C-helix to the A-loop, pulling the α C-helix toward the N-lobe as it associates with Lys295. The released Arg409 flips away from the α C-helix and associates with Tyr416, as one of the triad of arginine residues (Arg385, Arg409, and Arg419) that will eventually stabilize the phosphotyrosine of the fully activated kinase. However only Arg385 and Arg409 are closely associated with the unphosphorylated Tyr416 in the active-like state observed in the re-assembled conformation. The hydrophobic regulatory spine consisting of M314, L325, H384, and F405 also locks into place⁷².

CONCLUSION

To understand the allosteric control of the SH2 and SH3 regulatory domains on the inactivation of the catalytic KD, we used a combination of computational methods to determine the inactive-to-active conformational transition pathway and the underlying free

energy landscape under various assembly states of the multi-domain kinase. The impact of the regulatory tandem on activation is demonstrated by the results from umbrella sampling calculations: compared with the free energy landscape of the isolated KD from our previous study, the free energy landscape of the assembled full-length construct displays that the active conformation of the KD is further disfavored by more than 4 kcal/mol, essentially achieving an auto-inhibition of the KD. In contrast, the active-like state of the KD in the re-assembled full-length construct dominates the structural ensemble. Furthermore, our results also reveal that the linker region connecting SH2 and KD domains plays a critical role in auto-inhibition, and the SH2 and SH3 domains contribute to stabilizing the linker conformation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) through grant R01-CA093577. The computations are supported in part by NIH through resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago and Argonne National Laboratory under grant S10 RR029030-01, by the Extreme Science and Engineering Discovery Environment (XSEDE) grant number OCI-1053575, and by the University of Chicago Research Computing Center. The authors want to thank Dr. Matthew Pond and Dr. Avishek Das for many insightful discussions.

REFERENCES

1. Taylor SS, Ilouz R, Zhang P, Kornev AP. Assembly of allosteric macromolecular switches: lessons from PKA. *Nat. Rev. Mol. Cell Biol.* 2012; 13:646–658. [PubMed: 22992589]
2. Taylor SS, Kornev AP. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* 2011; 36:65–77. [PubMed: 20971646]
3. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science.* 2002; 298:1912–1934. [PubMed: 12471243]
4. Bolen JB. Nonreceptor Tyrosine Protein-Kinases. *Oncogene.* 1993; 8:2025–2031. [PubMed: 8336932]
5. Martin GS. The hunting of the Src. *Nat. Rev. Mol. Cell Biol.* 2001; 2:467–475. [PubMed: 11389470]
6. Vogt PK. Retroviral oncogenes: a historical primer. *Nat. Rev. Cancer.* 2012; 12:639–648. [PubMed: 22898541]
7. Thomas SM, Brugge JS. Cellular functions regulated by Src family kinases. *Annu. Rev. Cell Dev. Biol.* 1997; 13:513–609. [PubMed: 9442882]
8. Abram CL, Courtneidge SA. Src family tyrosine kinases and growth factor signaling. *Exp. Cell Res.* 2000; 254:1–13. [PubMed: 10623460]
9. Brown MT, Cooper JA. Regulation, substrates and functions of src. *BBA-Rev. Cancer.* 1996; 1287:121–149.
10. Sicheri F, Kuriyan J. Structures of Src-family tyrosine kinases. *Curr. Opin. Struct. Biol.* 1997; 7:777–785. [PubMed: 9434895]
11. Ren RB, Mayer BJ, Cicchetti P, Baltimore D. Identification of a 10-Amino Acid Proline-Rich Sh3 Binding-Site. *Science.* 1993; 259:1157–1161. [PubMed: 8438166]
12. Kuriyan J, Moarefi I, Sicheri F, Lee CH, Huse M. Structure and regulation of Src-family tyrosine kinases. *Faseb J.* 1997; 11:A876–A876.

13. Cooper JA, Macauley A. Potential Positive and Negative Auto-Regulation of P60c-Src by Intermolecular Autophosphorylation. *Proc. Natl. Acad. Sci. U.S.A.* 1988; 85:4232–4236. [PubMed: 2454466]
14. Cantley LC, Auger KR, Carpenter C, Duckworth B, Graziani A, Kapeller R, Soltoff S. Oncogenes and Signal Transduction. *Cell.* 1991; 64:281–302. [PubMed: 1846320]
15. Macauley A, Cooper JA. Structural Differences between Repressed and Derepressed Forms of P60c-Src. *Mol. Cell. Biol.* 1989; 9:2648–2656. [PubMed: 2474758]
16. Alexandropoulos K, Baltimore D. Coordinate activation of c-Src by SH3- and SH2-binding sites on a novel, p130(Cas)-related protein, Sin. *Gene. Dev.* 1996; 10:1341–1355. [PubMed: 8647432]
17. Briggs SD, Sharkey M, Stevenson M, Smithgall TE. SH3-mediated Hck tyrosine kinase activation and fibroblast transformation by the Nef protein of HIV-1. *J. Biol. Chem.* 1997; 272:17899–17902. [PubMed: 9218412]
18. Liu XQ, Brodeur SR, Gish G, Songyang Z, Cantley LC, Laudano AP, Pawson T. Regulation of C-Src Tyrosine Kinase-Activity by the Src Sh2 Domain. *Oncogene.* 1993; 8:1119–1126. [PubMed: 7683128]
19. Moarefi I, LaFevre-Bernt M, Sicheri F, Huse M, Lee CH, Kuriyan J, Miller WT. Activation of the Src-family tyrosine kinase Hck by SH3 domain displacement. *Nature.* 1997; 385:650–653. [PubMed: 9024665]
20. Hubbard SR, Mohammadi M, Schlessinger J. Autoregulatory mechanisms in protein-tyrosine kinases. *J. Biol. Chem.* 1998; 273:11987–11990. [PubMed: 9575136]
21. Hubbard SR, Till JH. Protein tyrosine kinase structure and function. *Annu. Rev. Biochem.* 2000; 69:373–398. [PubMed: 10966463]
22. Sicheri F, Moarefi I, Kuriyan J. Crystal structure of the Src family tyrosine kinase Hck. *Nature.* 1997; 385:602–609. [PubMed: 9024658]
23. Witucki LA, Huang X, Shah K, Liu Y, Kyin S, Eck MJ, Shokat KM. Mutant tyrosine kinases with unnatural nucleotide specificity retain the structure and phospho-acceptor specificity of the wild-type enzyme. *Chem. Biol.* 2002; 9:25–33. [PubMed: 11841936]
24. Xu WQ, Doshi A, Lei M, Eck MJ, Harrison SC. Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Mol Cell.* 1999; 3:629–638. [PubMed: 10360179]
25. Schindler T, Sicheri F, Pico A, Gazit A, Levitzki A, Kuriyan J. Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor. *Mol Cell.* 1999; 3:639–648. [PubMed: 10360180]
26. Meng YL, Roux B. Locking the Active Conformation of c-Src Kinase through the Phosphorylation of the Activation Loop. *J Mol Biol.* 2014; 426:423–435. [PubMed: 24103328]
27. Yamaguchi H, Hendrickson WA. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. *Nature.* 1996; 384:484–489. [PubMed: 8945479]
28. Cowan-Jacob SW, Fendrich G, Manley PW, Jahnke W, Fabbro D, Liebetanz J, Meyer T. The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. *Structure.* 2005; 13:861–871. [PubMed: 15939018]
29. Gan WX, Roux B. Binding specificity of SH2 domains: Insight from free energy simulations. *Proteins: Struct., Funct., Bioinf.* 2009; 74:996–1007.
30. Banavali NK, Roux B. The N-terminal end of the catalytic domain of Src kinase Hck is a conformational switch implicated in long-range allosteric regulation. *Structure.* 2005; 13:1715–1723. [PubMed: 16271895]
31. Meng YL, Roux B. Computational study of the W260A activating mutant of Src tyrosine kinase. *Protein Sci.* 2016; 25:219–230. [PubMed: 26106037]
32. Young MA, Gonfloni S, Superti-Furga G, Roux B, Kuriyan J. Dynamic coupling between the SH2 and SH3 domains of c-Src and hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell.* 2001; 105:115–126. [PubMed: 11301007]
33. Foda ZH, Shan YB, Kim ET, Shaw DE, Seeliger MA. A dynamically coupled allosteric network underlies binding cooperativity in Src kinase. *Nat. Commun.* 2015; 6:5939. [PubMed: 25600932]
34. Banavali NK, Roux B. Anatomy of a structural pathway for activation of the catalytic domain of Src kinase Hck. *Proteins: Struct., Funct., Bioinf.* 2007; 67:1096–1112.

35. Banavali NK, Roux B. Flexibility and charge asymmetry in the activation loop of Src tyrosine kinases. *Proteins: Struct., Funct., Bioinf.* 2009; 74:378–389.
36. Faraldo-Gomez JD, Roux B. On the importance of a funneled energy landscape for the assembly and regulation of multidomain Src tyrosine kinases. *Proc. Natl. Acad. Sci. U.S.A.* 2007; 104:13643–13648. [PubMed: 17699616]
37. Huang H, Zhao RJ, Dickson BM, Skeel RD, Post CB. alpha C Helix as a Switch in the Conformational Transition of Src/CDK-like Kinase Domains. *J. Phys. Chem. B.* 2012; 116:4465–4475. [PubMed: 22448785]
38. Shukla D, Meng Y, Roux B, Pande VS. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* 2014; 5:3397. [PubMed: 24584478]
39. Gan WX, Yang SC, Roux B. Atomistic View of the Conformational Activation of Src Kinase Using the String Method with Swarms-of-Trajectories. *Biophys. J.* 2009; 97:L8–L10. [PubMed: 19686639]
40. Yang S, Banavali NK, Roux B. Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106:3776–3781. [PubMed: 19225111]
41. Yang S, Roux B. Src kinase conformational activation: Thermodynamics, pathways, and mechanisms. *PLoS Comput. Biol.* 2008; 4:e1000047. [PubMed: 18369437]
42. Lovera S, Sutto L, Boubeva R, Scapozza L, Dolker N, Gervasio FL. The Different Flexibility of c-Src and c-Abl Kinases Regulates the Accessibility of a Druggable Inactive Conformation. *J Am Chem Soc.* 2012; 134:2496–2499. [PubMed: 22280319]
43. Meng YL, Lin YL, Roux B. Computational Study of the "DFG-Flip" Conformational Transition in c-Abl and c-Src Tyrosine Kinases. *J. Phys. Chem. B.* 2015; 119:1443–1456. [PubMed: 25548962]
44. Ozkirimli E, Post CB. Src kinase activation: A switched electrostatic network. *Protein Sci.* 2006; 15:1051–1062. [PubMed: 16597828]
45. Ozkirimli E, Yadav SS, Miller WT, Post CB. An electrostatic network and long-range regulation of Src kinases. *Protein Sci.* 2008; 17:1871–1880. [PubMed: 18687871]
46. Pan AC, Sezer D, Roux B. Finding transition pathways using the string method with swarms of trajectories. *The journal of physical chemistry.* 2008; 112:3432–3440.
47. Jo S, Kim T, Iyer VG, Im W. Software news and updates - CHARNIM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* 2008; 29:1859–1865. [PubMed: 18351591]
48. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 1983; 79:926–935.
49. Zheng JH, Trafny EA, Knighton DR, Xuong NH, Taylor SS, Teneyck LF, Sowadski JM. 2.2-Angstrom Refined Crystal-Structure of the Catalytic Subunit of Camp-Dependent Protein-Kinase Complexed with Mnatp and a Peptide Inhibitor. *Acta Crystallogr. D.* 1993; 49:362–365. [PubMed: 15299527]
50. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 2005; 26:1781–1802. [PubMed: 16222654]
51. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 1998; 102:3586–3616. [PubMed: 24889800]
52. Feller SE, Zhang YH, Pastor RW, Brooks BR. Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J. Chem. Phys.* 1995; 103:4613–4621.
53. Darden T, York D, Pedersen L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* 1993; 98:10089–10092.
54. Weinan E, Ren W, Eijnden E. String method for the study of rare events. *Physical Review B.* 2002; 66:052301.

55. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *The Journal of chemical physics*. 2006; 125:24106. [PubMed: 16848576]
56. Jiang W, Phillips JC, Huang L, Fajer M, Meng YL, Gumbart JC, Luo Y, Schulten K, Roux B. Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD. *Comput. Phys. Commun.* 2014; 185:908–916. [PubMed: 24944348]
57. Aronov B, Har-Peled S, Knauer C, Wang YS, Wenk C. Frechet distance for curves, revisited. *Lect Notes Comput Sc.* 2006; 4168:52–63.
58. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J. Mol. Graph. Model.* 1996; 14:33–38.
59. Hunter JD. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 2007; 9:90–95.
60. Schrodinger, LLC: 2015.
61. Wojtas-Niziurski W, Meng YL, Roux B, Berneche S. Self-Learning Adaptive Umbrella Sampling Method for the Determination of Free Energy Landscapes in Multiple Dimensions. *J. Chem. Theory Comput.* 2013; 9:1885–1895. [PubMed: 23814508]
62. Roux B. The Calculation of the Potential of Mean Force Using Computer-Simulations. *Comput. Phys. Commun.* 1995; 91:275–282.
63. Maragliano L, Roux B, Vanden-Eijnden E. Comparison between Mean Forces and Swarms-of-Trajectories String Methods. *J. Chem. Theory Comput.* 2014; 10:524–533. [PubMed: 26580029]
64. Xu WQ, Harrison SC, Eck MJ. Three-dimensional structure of the tyrosine kinase c-Src. *Nature*. 1997; 385:595–602. [PubMed: 9024657]
65. Gonfloni S, Frischknecht F, Way M, Superti-Furga G. Leucine 255 of Src couples intramolecular interactions to inhibition of catalysis. *Nat. Struct. Biol.* 1999; 6:760–764. [PubMed: 10426955]
66. Gonfloni S, Weijland A, Kretschmar J, Superti-Furga G. Crosstalk between the catalytic and regulatory domains allows bidirectional regulation of Src. *Nat. Struct. Biol.* 2000; 7:281–286. [PubMed: 10742171]
67. Gonfloni S, Williams JC, Hattula K, Weijland A, Wierenga RK, Superti-Furga G. The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src. *Embo J.* 1997; 16:7261–7271. [PubMed: 9405355]
68. LaFevre-Bernt M, Sicheri F, Pico A, Porter M, Kuriyano J, Miller WT. Intramolecular regulatory interactions in the Src family kinase Hck probed by mutagenesis of a conserved tryptophan residue. *J. Biol. Chem.* 1998; 273:32129–32134. [PubMed: 9822689]
69. Gonfloni S, Williams JC, Hattula K, Weijland A, Wierenga RK, Superti-Furga G. The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src. *Embo J.* 1997; 16:7261–7271. [PubMed: 9405355]
70. Chopra N, Wales TE, Joseph RE, Boyken SE, Engen JR, Jernigan RL, Andreotti AH. Dynamic Allostery Mediated by a Conserved Tryptophan in the Tec Family Kinases. *PLoS Comput. Biol.* 2016; 12
71. Guo SL, Wahl MI, Witte ON. Mutational analysis of the SH2-kinase linker region of Bruton's tyrosine kinase defines alternative modes of regulation for cytoplasmic tyrosine kinase families. *Int. Immunol.* 2006; 18:79–87. [PubMed: 16291652]
72. Kornev AP, Taylor SS. Defining the conserved internal architecture of a protein kinase. *BBA-Proteins Proteom.* 2010; 1804:440–444.

Figure 1.

Discrete Fréchet distance of each swarms-of-trajectories iteration to the (a) initial or (b) final string for the various systems and swarm trajectory lengths. The x-axis is scaled by the swarm trajectory length for each series of iterations in order to normalize the computational resources required for producing each curve. Note how the 1 ps plateaued much earlier than the 5 and 10 ps swarm trajectory lengths.

Figure 2.

2D-PMF with respect to A-loop opening and α C-helix rotation. All free energies are in the unit of kcal/mol, and all distances are in the unit of Å. (a) assembled SH2/SH3, (b) catalytic domain only, used as a model for the SH2/SH3 dissociated state, (c) N-linker constrained to the assembled state, and (d) re-assembled SH2/SH3. Each subplot is accompanied by a visual of the characteristic in-active and active state for the specific system. The surfaces of the KD, SH2 and SH3 are shown in blue, green and yellow respectively. The linker, A-loop and phosphorylated C-terminal tail are shown when present.

Figure 3. 2D-PMF with respect to the N-terminal linker flexibility and α C-helix rotation. All free energies are in the unit of kcal/mol, and all distances are in the unit of Å. (a) assembled SH2/SH3, and (b) re-assembled SH2/SH3.

Figure 4.

Comparison of the W260 position in both the X-ray crystal structures (a–b) and converged string endpoints (c–d). The α C-helix and portions of the N-lobe are shown as well. The spheres indicate the direction of the N-terminal linker relative to the KD. The X-ray crystal structures shown are the inactive Hck (red, PDB id: 1QCF), inactive c-Src (green, PDB id: 2SRC), active Lck (magenta, PDB id: 3LCK), and active c-Src (blue, PDB id: 1Y57). The converged string endpoints are the assembled SH2/SH3 inactive (green) and active (blue), followed by the re-assembled SH2/SH3 inactive (red) and active (magenta).

Figure 5.

Progress of four electrostatic interactions during the transition, a subset of the electrostatic network identified by Ozkirimli *et al.* (a) Glu310:CD to Arg409:CZ (d_1), (b) Glu310:CD to Lys295:NZ (d_2), (c) Asp386:CG to Tyr416:OH, and (d) Arg385:CZ to Tyr416:OH.

Figure 6.

The progress of twelve properties along the optimized string pathway were normalized, and then manually sorted for the purpose of showing a few key events in an ordered fashion. For each property, the extrema most closely resembling the inactive and active states were colored in blue and red, respectively, in order to visually display the evolution of the variable. The twelve properties are described in the following: *SASA Y416* refers to the solvent accessible surface area of Y416, its value ranging from 244 (blue) to 266 Å² (red); *A-loop α-turn 1* refers to the average distance of the three $i:i+4$ hydrogen bonds in the short α-helix turn in the inactive A-loop between residues 413 to 419 (called d_3 in Figure 2) used to reflect the opening of the A-loop, its value ranging from 2 (blue) to 10 Å (red); *A-loop α-turn 2* refers to the $i:i+4$ hydrogen bond of the single turn α-helix in the inactive A-loop between residues 407 to 411, its value ranging from 2 (blue) to 10 Å (red); *Arg419 to α-EF* refers to the distance between R419 of the A-loop and the backbone of A434 from the α-EF helix, its value ranging from 2 (blue) to 10 Å (red).; the W260 side chain angles χ_1 and χ_2 , its value ranging from -150° (blue) to -60° (red) and 0° (blue) to 110° (red), respectively; *F424 to α-G* refers to the distance between F424 of the A-loop and N468 of the αG helix, its value ranging from 10 (blue) to 2 Å (red); *A-loop α-turn 3* refers to the $i:i+4$ hydrogen bond between R419 and K423 in the active A-loop, its value ranging from 10 (blue) to 2 Å (red); *α-C to A-loop* refers to the E310 to R409 salt-bridge distance (called d_1 in Figure 2),

its value ranging from 2 (blue) to 10 Å (red).; *α -C to N-lobe* refers to the E310 to K295 salt-bridge distance (called d_2 in Figure 2), its value ranging from 10 (blue) to 2 Å (red).; *R409 to Y416* refers to the distance between R409 and Y416, its value ranging from 10 (blue) to 2 Å (red); *R-spine RMSD* refers to the heavy atom RMSD of the regulatory spine residues (M314, L325, H384, F405, and D444) from the equilibrated re-assembled active structure, its value ranging from 1.8 (blue) to 0.6 Å (red).