

ARTICLE

Received 22 Jul 2013 | Accepted 6 Feb 2014 | Published 3 Mar 2014

DOI: 10.1038/ncomms4397

Activation pathway of Src kinase reveals intermediate states as targets for drug design

Diwakar Shukla^{1,2}, Yilin Meng³, Benoît Roux^{3,4} & Vijay S. Pande^{1,2}

Unregulated activation of Src kinases leads to aberrant signalling, uncontrolled growth and differentiation of cancerous cells. Reaching a complete mechanistic understanding of large-scale conformational transformations underlying the activation of kinases could greatly help in the development of therapeutic drugs for the treatment of these pathologies. In principle, the nature of conformational transition could be modelled *in silico* via atomistic molecular dynamics simulations, although this is very challenging because of the long activation timescales. Here we employ a computational paradigm that couples transition pathway techniques and Markov state model-based massively distributed simulations for mapping the conformational landscape of c-src tyrosine kinase. The computations provide the thermodynamics and kinetics of kinase activation for the first time, and help identify key structural intermediates. Furthermore, the presence of a novel allosteric site in an intermediate state of c-src that could be potentially used for drug design is predicted.

¹Department of Chemistry, Stanford University, Stanford, California 94305, USA. ²SIMBIOS NIH Center for biomedical computation, Stanford University, Stanford, California 94305, USA. ³Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Science, The University of Chicago, Chicago, Illinois 60637, USA. ⁴Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, USA. Correspondence and requests for materials should be addressed to V.S.P. (email: pande@stanford.edu).

Protein kinases are a family of enzymes that catalyse the transfer of the γ -phosphate group from ATP to the hydroxyl group of specific serine, threonine or tyrosine residues. Their ability to alternate between catalytically active (upregulated) and inactive (downregulated) states in response to specific signals provides a central switching mechanism in cellular transduction pathways, such as in those regulating cellular growth, proliferation and differentiation¹. Deregulation of kinases can lead to uncontrolled cell proliferation and malignant transformation as observed in numerous cancers¹, which makes them an important target for therapeutic intervention^{2,3}. Selective inhibition of kinases using small molecule inhibitors is gaining importance as an effective therapeutic regimen for treatment of a wide range of cancers with the successful approval of imatinib (Gleevec) for treating chronic myeloid leukaemia^{4,5}. However, despite these encouraging results, this remains very difficult for a number of reasons. Most of the small molecule inhibitors target the ATP-binding pocket in the active/inactive state of the kinase, which is highly conserved owing to the functional similarity between kinases, and therefore provides limited selectivity. Furthermore, targeted protein kinases develop resistance owing to the point mutations in the important structural regions such as the drug-binding site^{6–9}. One of the strategies to overcome molecular resistance and selectivity issues involves targeting kinases by allosteric inhibitors that bind outside the ATP-binding pocket and affect kinase activity by eliciting long-range conformational transformations. While this represents a very promising avenue, the design of allosteric inhibitor is challenging because it not only requires identification of novel allosteric binding sites but also a detailed dynamical picture of allosteric network of residues involved in long-range coupling and activation. Therefore, mechanistic and structural insights into the molecular aspects of kinase conformational change are desired for rational design of selective and robust kinase inhibitors^{6,9}.

The c-src kinase from the Src family of cytoplasmic tyrosine kinases presents an ideal system for mechanistic investigations not only because of its medical and historical relevance¹⁰ but also because of the availability of extensive experimental data related to this kinase including several crystal structures of both inactive and active states¹¹. The key structural differences between active and inactive states are the changes in the orientation and conformation of the C-helix in the amino-terminal lobe and the unfolding of the activation loop (A-loop) in the carboxy-terminal lobe (Fig. 1). In the inactive state, the A-loop with unphosphorylated Tyr416 is closed and folded to occlude substrate entry into the active site. 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 C-helix recruiting the residues needed to form a catalytically competent active site. The function of the catalytic domain of c-src is to phosphorylate the tyrosine side chain in a downstream substrate protein, which can also include the Tyr416 in the A-loop from another kinase. It is the process of transphosphorylation of Tyr416 in the A-loop via a bimolecular encounter with another active Src kinase that 'locks' the domain in its catalytically active state. To become accessible as a substrate, the A-loop of an inactive kinase must transiently adopt conformations that expose Tyr416 to the solvent. These large-scale conformational changes in the structural elements of kinases are also accompanied by subtle changes in the positions and interactions of specific residues. The active and inactive crystal structures of c-src tyrosine kinase show switching of hydrogen bonding partners of Glu310 from Arg409 to Lys295 (Ozkirimli *et al.*^{12,13}), and the alignment of residues comprising R-spine¹⁴ on activation (Fig. 1). These crystal structures reveal differences

between inactive and active conformations but fail to provide insights into the activation mechanism, presence of key intermediates along the activation pathway, the ensemble of pathways connecting the two end states and the timescales of activation/deactivation⁵. It is critical to understand the molecular underpinnings of the kinase activation process to relate these static structures to the kinase function and to provide levers for controlling the activation of aberrant kinase signalling pathways^{2,3,9}.

Atomistic molecular dynamics (MD) simulations have been used to study conformational transitions in proteins such as the flipping of 'DFG' motif^{15,16} in kinases, calculation of the binding free energy of inhibitors¹⁵, elucidating activation pathways^{17–19}. However, owing to the large system size and long timescales associated with kinase conformational dynamics, most of these studies employ simplifications such as coarse-grain representations of enzymes, 'enhanced sampling' techniques, studying mutants with fast conformational transition kinetics^{20–27} and so on. While simulations have been successful at identifying metastable states of proteins, estimates of the kinetics of conformational transitions have remained elusive because of the challenges outlined above.

In the present study, we employ a computational framework that couples transition pathway generation techniques, massively distributed MD simulations (550 μ s) on Folding@home²⁸, Markov state models (MSMs) and novel adaptive sampling algorithms for mapping the conformational landscape of c-src tyrosine kinase. To sample the conformational landscape more efficiently, we start a large number of relatively short simulations from configurations along an optimal pathway connecting the active (pdb id:1Y57 (Cowan-Jacob *et al.*²⁹)) and inactive (pdb id:2SRC³⁰) states determined using the string method with swarm-of-trajectories. A theoretical framework based on MSM in which conformational dynamics is modelled as transitions between kinetically metastable states is then used to investigate the slow kinetics ($\sim 100 \mu$ s) of conformational transitions of kinases by using statistical methods to combine a very large number of short trajectories (~ 20 ns) into a single model^{31,32}. The MSM analysis of c-src conformational dynamics reveals the presence of key intermediates along the ensemble of c-src activation pathways and for the first time, timescales associated with activation and deactivation processes. The results presented in this study would not only considerably increase our knowledge of the structures of possible intermediate conformations sampled during the kinase activation but also provide an extensive database of c-src conformations for design of future novel kinase inhibitors.

Results

Simulations reveal multiple intermediate states of c-src.

Simulations of c-src catalytic domain are performed in this study. The regulatory SH2 and SH3 domains, which are involved in the autoinhibited inactive state, are not included in the simulations performed in this study to mimic the disassembled upregulated form of the enzyme complex^{17,30,33}. ATP molecule is bound in the nucleotide-binding pocket in the study because of the high likelihood of ATP binding at its physiological concentration of 2–8 mM (Frankel *et al.*³⁴). The transphosphorylation of Tyr416 locks the kinase domain in the catalytically active state. Tyr416 is phosphorylated owing to encounter with another kinase only when it is exposed to the solvent in an open active-like state as compared with the inactive state of c-src kinase in which it is buried inside the protein surface. In all our simulations, Tyr416 is in an unphosphorylated state. Therefore, the activation transitions observed in this study represent transition between

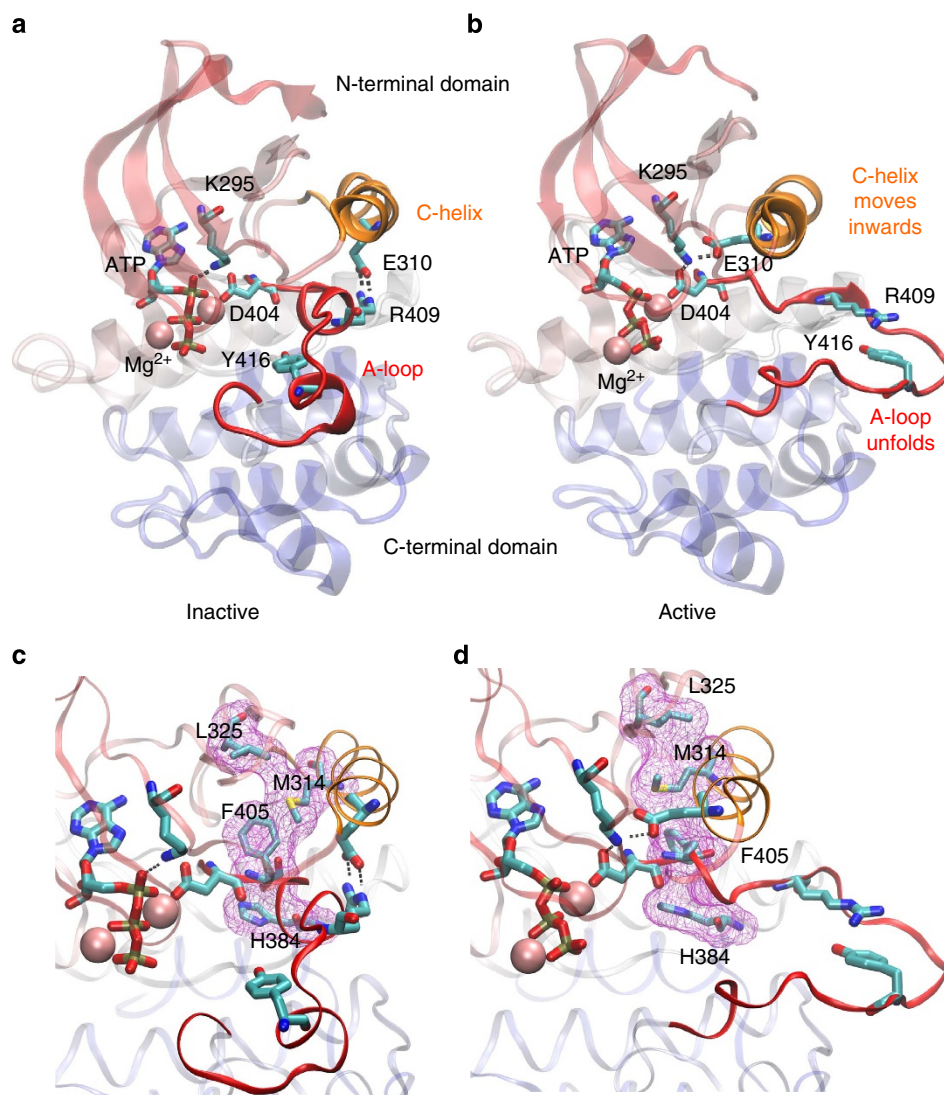


Figure 1 | Conformational changes associated with the c-src activation. The (a) inactive and (b) active crystal structures show structural changes in the activation loop (A-loop; red) and C-helix (orange), switching of the electrostatic network formed between Lys295, Glu310, Arg409 and Tyr416, and alignment of residues L325, M314, F405 and H384 (shown in licorice and surface representation) to form a hydrophobic regulatory spine (R-spine) in active state (d) as compared with the inactive state conformation (c). R-spine forms a continuous hydrophobic region linking the two lobes of the catalytic domain and it is critical for the catalytic activity of the kinase.

inactive state and an open ‘active-like’ conformation that exposes Tyr416 transiently to another kinase¹⁸. The starting structures were taken from a minimum free energy pathway (MFEP) connecting the active and inactive state of c-src kinase generated using the string method with swarms-of-trajectories²⁴. Analysis of the raw trajectories and MSMs built from the simulation data reveal a detailed conformational landscape of the transitions between active and inactive states of c-src kinase including the kinetics of the dynamical processes occurring on the landscape. The overall strategy that consists of combining the information from multiple trajectories through the use of MSM bears some similarities to that used by Yang *et al.*¹⁸, although the massive configurational sampling involved in the present computations far exceeds any previous study by several orders of magnitude.

Analysis of the conformational landscape based on order parameters reveals key intermediates between inactive and active states but fails to provide detailed sequence of events leading to activation, probability of transitions between them, elucidating parallel activation paths and so on. MSMs provide a natural

framework for simplified analysis of large simulation data sets by discarding dynamics faster than a certain timescale called lag time³⁵ instead of projecting the data along a chosen set of order parameters. The Markov state-building process involves separation of the conformational landscape into states with high structural similarity. The transitions between states and the population of each state are obtained by recording the states visited along individual trajectories. A maximum likelihood estimate of the state population and transition probability between states is then obtained, which provides the best model for the kinetics possible, given the data at hand³⁶. To ascertain the robustness of the MSM analysis, MSMs with 1,000, 2,000, 3,000, 4,000, 5,000 and 10,000 microstates were built from the simulation data. The implied timescale of the 2,000 state model were converged indicating that the model is Markovian in nature (Supplementary Fig. 1). Therefore, the 2,000 microstate model was chosen for further analysis.

Two-dimensional conformational landscape of the c-src tyrosine kinase with respect to several order parameter pairs are

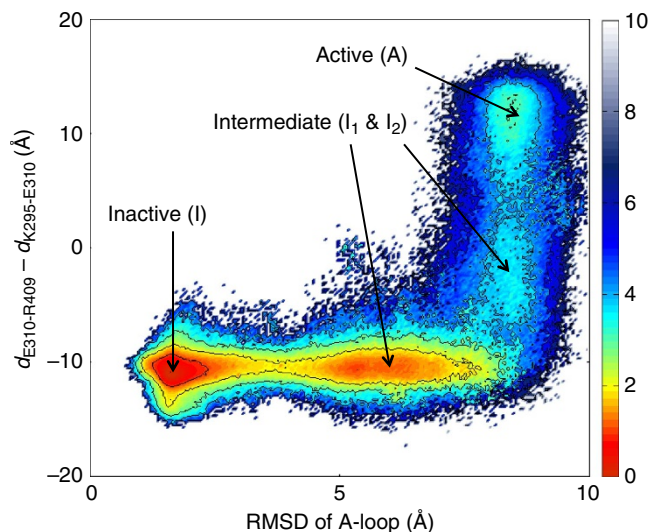


Figure 2 | Conformational landscape of c-src tyrosine kinase. The conformational landscape generated using r.m.s.d. of A-loop residues (404–424) and difference of the distance between Glu310–Arg409 and Lys295–Glu310 residue pairs as the order parameters reveals multiple intermediates along the activation pathway. The free energy values are reported in kcal mol^{-1} . The landscape is obtained by summing over all the microstates of the 2,000 state MSM from $i=1$ to N using the following equation $W(x, y) = -k_B T \ln \left[\sum_i^N \pi_i h_i(x, y) \right]$ where π_i is the probability of state i in the MSM, and $h_i(x, y)$ is the normalized histogram of the variables x and y restricted to the MSM state i .

shown in Fig. 2 and Supplementary Fig. 2. Two intermediate states (I_1 and I_2) are found along the pathway linking the inactive and active states of c-src kinase. In the intermediate state I_1 (Fig. 3c), the A-loop is unfolded but the electrostatic network involving Lys295, Glu310 and Arg409 is in its inactive configuration. In the intermediate state I_2 (Fig. 3d), the Glu310–Arg409 hydrogen bond is broken but C-helix has not rotated and moved inwards to facilitate the formation of Lys295–Glu310 hydrogen bond.

These results suggest that the A-loop has to unfold before the conformational changes in the C-helix. This observation is broadly consistent with previous computational studies related to not only Src family tyrosine kinases^{17,18,23–25,27} but also serine-threonine kinases²². In particular, the Src kinase conformational landscape obtained from independent umbrella sampling simulations by Meng and Roux²⁷ also reveals similar intermediate states. In the current study, extensive sampling of possible catalytic domain conformations is obtained, whereas Meng and Roux²⁷ have reported conformational landscape with the A-loop in only open conformation. To present a direct comparison between these studies, we have now reported a conformational landscape obtained using MSM and umbrella sampling results²⁷ including all conformations of A-loop (Supplementary Fig. 3). The computed population of active (unphosphorylated active-like) conformation from the umbrella sampling calculations reported by Meng and Roux²⁷ was found to be 0.002, which is in agreement with the value of 0.0018 obtained in this study. Similarly, a previous computational study of Hck kinase (a member of the Src family of tyrosine kinases with high structural similarity to c-src) also reported the existence of two intermediate states along the pathway between the inactive and active state kinase¹⁸. However, there are noteworthy differences with the intermediate states reported in the present study that could be potentially exploited for drug design purposes. The

intermediate (I_1) of Hck involves partial unfolding of A-loop and has an intact Glu310–Arg409 ion pair, which is similar to the Intermediate I_1 of c-src (the difference lies in the extent of unfolding of A-loop). However, the second intermediate of Hck is distinct from the intermediate I_2 of c-src. The A-loop in Hck intermediate I_2 is not fully unfolded, whereas it is completely unfolded in c-src. The Lys295–Glu310 ion pair is not formed in c-src, whereas it is formed in the Hck kinase. The binding of 8-anilino-1-naphthalene sulphonate (ANS) to the allosteric site in c-src requires complete unfolding of A-loop. Furthermore, Glu310–Arg409 ion pair stabilizes the partially unfolded states of A-loop, with the DFG-motif end of the A-loop in its inactive conformation. Therefore, ANS can only bind to intermediate I_2 of c-src with broken Glu310–Arg409 ion pair. Therefore, intermediate states of Hck kinase would not be able to bind ANS molecule because both these intermediate states have partially unfolded A-loop. Furthermore, intermediate I_2 of Hck kinase has partially formed Lys295–Glu310 ion pair that blocks the primary ANS-binding site. These observations and the simulation results show that the methods outlined in this study could be used for successfully identifying the subtle structural differences between the kinases with high structural and sequence similarity. Recent computational studies on epidermal growth factor receptor kinase have revealed a partially unfolded conformation of C-helix. Our results indicate that for the Src kinase, the C-helix remains folded during activation as indicated by the high helical content and stable solvent accessible surface area of the C-helix (Supplementary Fig. 4). This observation is consistent with the H–D exchange experiments and bioinformatic analysis of the intrinsic disorder at the C-helix in Src kinase³⁷.

Diversity within intermediate states of c-src. The conformational landscapes also show a rich diversity within the intermediate states because of the partially unfolded states of A-loop and the multiple conformational states of DFG-motif (Asp404–Gly406) and R-spine (Leu325, Met314, Phe405 and His384; Supplementary Figs 5–8). DFG residues (Asp404, F405 and G406) at the N-terminal end of the A-loop play a critical role in activation of kinases. The crystal structures of active and inactive states of c-src have the DFG motif in the ‘DFG-in’ states^{29,30}. The DFG-out state is not accessible in the conformational landscape of ATP-bound c-src (Supplementary Fig. 8) owing to the interaction between the DFG-aspartate residue and the Mg^{2+} ions bound to the phosphate groups of ATP and steric clash between DFG-phenylalanine and the ATP in the DFG-out state. The DFG motif does not completely flip in the simulations but it still shows considerable fluctuations along the activation pathways. The variation of root mean squared deviation (r.m.s.d.) of the DFG motif (from the active state) with respect to the r.m.s.d. of the A-loop (from the inactive state) shows that DFG motif adopts multiple intermediate conformations because of the movement of the individual loop residues during activation (Supplementary Fig. 6). In the active ‘DFG-in’ state, F405 participates in the formation of the R-spine, which is the hallmark of the active c-src. For the formation of R-spine, F405 has to flip partially to make space for the M314 (in the C-helix) to occupy the gap between L325 and F405 (Fig. 3f). The fluctuations of the positions of these individual residues also show that a rich diversity of conformations exists within individual metastable basins corresponding to active, inactive and intermediate states.

Deactivation timescales are faster than activation timescales. The first eigenvector of the transition probability matrix T_{ij} (includes transition probabilities from state i to state j) provides the estimate of the equilibrium population of all states in MSM.

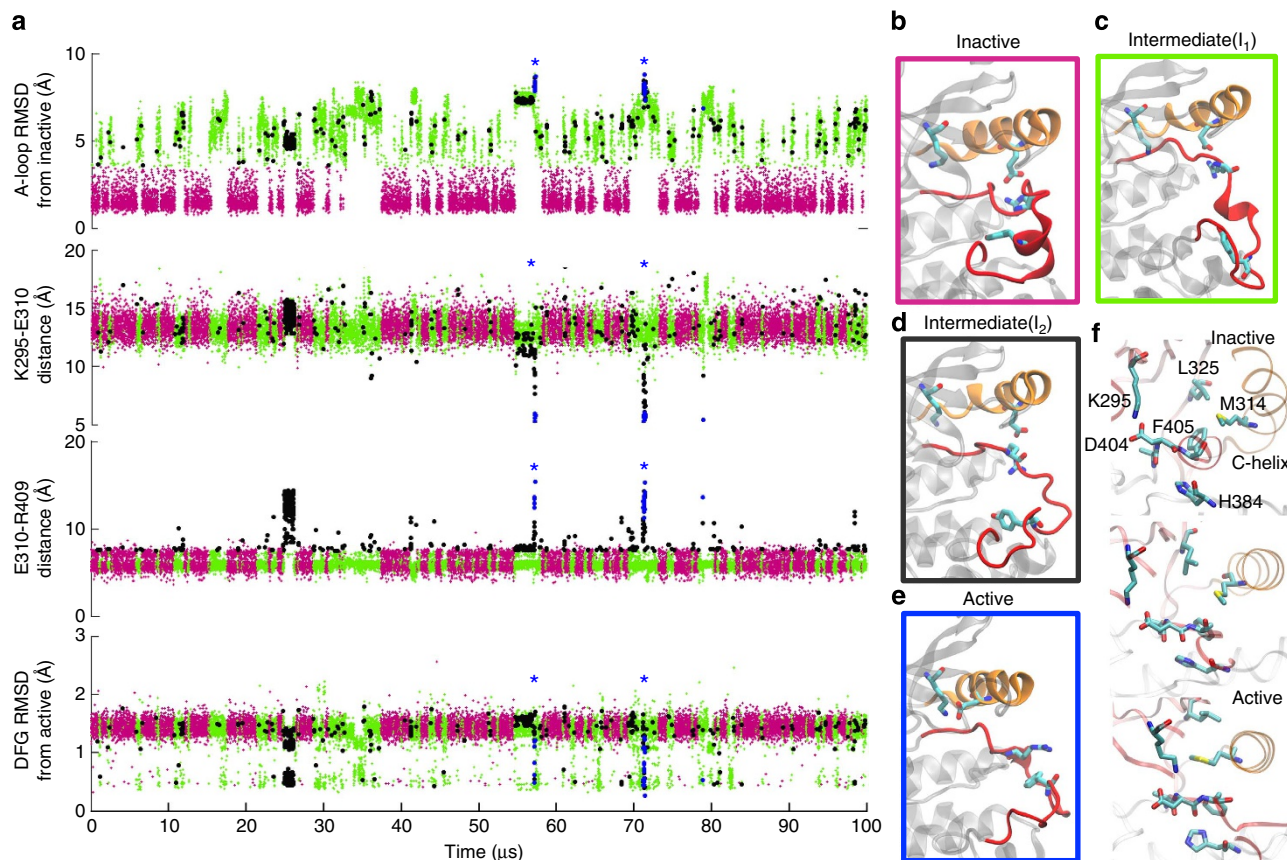


Figure 3 | Kinetics of the c-src kinase activation. MSM of the kinase conformational change reveals novel intermediate states along the activation pathway and provide a measure of the activation/deactivation timescales. **(a)** Variation of key structural metrics as a function of time along the activation trajectory obtained using the MSM. MSM trajectories are calculated using a kinetic Monte Carlo algorithm to generate a trajectory of ($\tau = 5$ ns) microstate jumps, and selecting at random (uniformly) a simulation snapshot to report observables at each time step. The r.m.s.d. of the activation loop is calculated using heavy atoms of residues 404–424. The following atoms were used for the calculations of distances between residues: Lys295 (NZ atom in the NH_3^+ group), Glu310 (CD atom in the COO^- group) and Arg409 (CZ atom in the guanidinium group). Different colours represent the different conformational states of the c-src kinase. Inactive **(b)**, intermediate states I_1 **(c)**, I_2 **(d)** and active **(e)** states are shown in magenta, green, black and blue, respectively, with active state also marked with an asterisk. **(f)** These four conformational states could be further subdivided into states with different conformations of DFG-motif and R-spine.

The equilibrium probabilities estimated from the microstate MSM with respect to the r.m.s.d. of the state from the inactive and active states show that the inactive state is the most populated state ($P_{\text{inactive}} \sim 0.56$) and the active state is one of the least populated states ($P_{\text{active}} \sim 0.001$; Supplementary Fig. 9). These observations are consistent with a free energy barrier of $3.8 \text{ kcal mol}^{-1}$ between the inactive and active states. These values also suggest that the active state is visited transiently in the conformational landscape and requires either binding of a substrate or phosphorylation of Tyr416 to remain in the active state¹¹. The A-loop remains extended in the intermediate and active states, thereby increasing the possibility of stimulatory transphosphorylation of Tyr416. The slowest process occurring on the conformational landscape of c-src involves population flow from inactive state to a partially active state with A-loop fully unfolded and R-spine partially aligned. (Supplementary Fig. 10) Furthermore, the Glu310–Arg409 were hydrogen bonded in the partially active state indicating that the slowest dynamical process involves transition between the inactive state and a microstate with a fully unfolded A-loop.

Figure 3a shows a 100- μs trajectory generated from the MSM using kinetic Monte Carlo scheme starting from the inactive state. Two activation events are observed in this trajectory as denoted

by an asterisk. Colours of the dots indicate the four macrostates (active, inactive, I_1 and I_2) as shown in Fig. 3b–e. The mean first passage time (MFPT) for activation (defined as transition between states with r.m.s.d. closest to the inactive and active state crystal structures) is calculated to be $\sim 106 \mu\text{s}$, whereas the MFPT for deactivation is found to be $\sim 21 \mu\text{s}$. Such long timescales extracted from MSM analysis may appear inconsistent with the relatively small free energy barriers between the inactive, active and intermediate states observed in the potential of mean force (PMF) plot shown in Fig. 2. However, it is worth pointing out that absolute transition rates in complex multidimensional systems depend on a multiple structural and dynamical factors that may remain hidden in a PMF projection onto a subspace of a few order parameters. Other projections for the 100- μs trajectory are shown in Supplementary Fig. 11. The R-spine alignment is coupled to the C-helix rotation and displacement but the couplings are loose. The R-spine fluctuates between active, inactive and intermediate conformations (Fig. 3f) regardless of the conformation of the catalytic domain. R-spine alignment takes place more frequently in the intermediate states indicating that partial unfolding of the A-loop facilitates the rotation of the C-helix. Our results indicate that, although there is one broad pathway connecting the active and inactive states but if the

fluctuations in individual residues comprising 'DFG'-motif and R-spine are considered, it leads to an ensemble of activation pathways with comparable flux, which provides a more fine-grained view of the kinase activation (Supplementary Fig. 12).

Relating fast activation to slow rate of autophosphorylation.

The experimentally observed timescale for Src-family tyrosine kinase autophosphorylation at residue Tyr416 is on the order of minutes³⁸, whereas the timescale of the conformational transitions associated with kinase activation obtained from the present computational analysis is 110 μ s. To relate these vastly disparate timescales, we constructed a kinetic model describing the time evolution of activated kinase population using the thermodynamic and kinetic data extracted from the atomistic simulations following the reaction scheme included below.



where I is the inactivated state, A is the unphosphorylated active state, A* is phosphorylated (locked) active state, k_{IA} is the activation rate and k_{AI} is the deactivation rate. The rate $k_{trans-p}$ is the rate of kinase phosphorylation, which requires intermolecular encounter with another active kinase (see Methods for model description). A key result from the present analysis is that the catalytic domain undergoes fast exchange between the inactive state and an active-unphosphorylated conformation. Because of the low bimolecular encounter frequency with catalytically active kinases, the fast exchange effectively leads to kinase populations that are in pseudo equilibrium. Time evolution of active phosphorylated kinase population shows a lag or induction time, which is an experimentally observed feature (Supplementary Fig. 13a)^{38,39}. The lag time indicates the slow initial build up of active kinases. As some kinases reach active state by fluctuations, they become available to seed and initiate the autophosphorylation of other kinases, which then leads to a sharp increase in active kinase population. The model also correctly predicts the experimentally observed relationship between time evolution of active kinase population and enzyme concentration, rates of autophosphorylation and activation. Our model results predict that the time evolution of active kinase population is sensitive to changes in the effective equilibrium constant (K_{eq}) between inactive and active-unphosphorylated

states (Supplementary Fig. 13). The typical experimental values of $t_{1/2}$ (time taken for half the kinases to be in active state) is in minutes. Under conditions that are representative of experiments, the limits on K_{eq} values consistent with the range of experimental $t_{1/2}$ values can be estimated using the model. Assuming the rate of autophosphorylation to be $10 \mu\text{M}^{-1} \text{s}^{-1}$, total enzyme concentration $[C_{tot}] = 1 \mu\text{M}$ and the value of $K_{eq} = 740$ calculated from the MSM equilibrium populations, the $t_{1/2}$ values is found to be 8.2 min, which lies in the experimental observed range of $5 < t_{1/2} < 10$ min (Moarefi *et al.*³⁸).

Keeping c-src inactive by stabilizing intermediate states.

Mechanistic understanding of the kinase activation process could be harnessed for the design of novel inhibitors. In this study, we have predicted the existence of metastable intermediate states I₁ and I₂, which could be stabilized to block the activation process. Coincidentally, a recent crystal structure of cyclin-dependent kinase 2 (CDK2; from the CMGC family of the serine/threonine kinases) bound to ANS has structural features similar to the intermediate I₂ of c-src (Supplementary Fig. 14)⁴⁰. Furthermore, ANS molecules were bound to the pocket between the C-helix and β_4 , which is different from typical kinase inhibitors that bind to the ATP-binding pocket or the adjacent hydrophobic pocket (Supplementary Fig. 15)⁴¹. From the activation mechanism of c-src kinase, it is evident that the binding of any drug in the region between C-helix and the β -sheets in the N-terminal lobe would block the displacement and/or rotation of the C-helix, thereby trapping the kinase in a partially active intermediate state.

Simulations of ANS-bound intermediate I₂ from c-src were performed to assess the interactions of the drug molecule with the c-src residues and to ascertain structural changes induced by the drug binding. The structures obtained from the simulations of ANS-bound c-src reveal outward movement of the C-helix as compared with the ATP-bound c-src, which is significantly different from both the active and inactive crystal structures of c-src (Fig. 4). The displacement of the C-helix would not only block the activation of the kinase but also interfere with the binding of the substrate. The R-spine alignment is also disrupted owing to ANS binding, indicating that the ANS also interferes with the rotation of the C-helix, which could potentially suppress the catalytic activity of the kinase (Supplementary Fig. 16). The simulations also show that ANS molecule sits directly

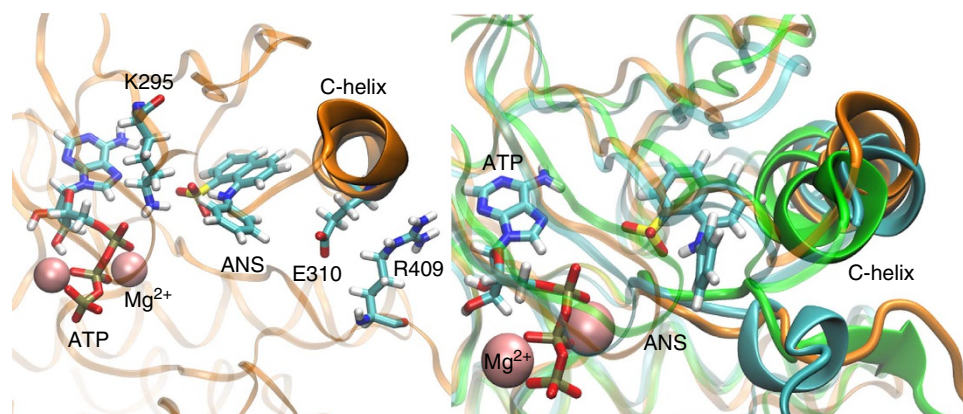


Figure 4 | Mechanism of ANS-induced stabilization of intermediate I₂. (Left) ANS binding to the allosteric site adjacent to C-helix in c-src kinase stabilizes the intermediate conformation by blocking the interactions between Lys295 and Glu310. The hydrogen bond formation between Lys295 and Glu310 is required for the locking of the C-helix in the active conformation. The sulphonate group in the ANS forms a hydrogen bond with the Lys295, thereby locking it in its inactive conformation. (Right) ANS binding also pushes the C-helix away from the ATP-binding pocket. Superimposition of the crystal structures of the inactive (cyan) and active (green) states of ATP-bound c-src kinase with the ANS-bound src-kinase (orange) reveals the distinct conformation of the C-helix in presence of ANS.

between the Lys295 and Glu310 residues, thereby blocking any possibility of hydrogen bond formation between these residues, which is essential for the inward movement of C-helix (Supplementary Fig. 17). Furthermore, strong hydrogen bond formation is observed between the sulphonate group of ANS and Lys295 (Fig. 4). The primary ANS-binding pocket is hydrophobic in nature that interacts favourably with the naphthalene and the phenyl rings in the ANS molecules (Supplementary Fig. 18). The phenyl ring was found to interact with Phe278, Phe307 and Leu297 residues, whereas the naphthalene ring interacts with Phe405, Ala403, Val323, Leu325 and Ile336 residues.

Long range dynamic coupling between residues in *c-src*.

Allosteric communication lies at the heart of the kinase activation process with long-range inter and intradomain communication networks controlling the conformation of the enzyme⁴². Mutual information analysis that provides a measure of long-range coupling between residues by calculating the correlation between the motions of their backbone and side-chain torsion angles is used to identify allosteric couplings in *c-src* kinase⁴³. Variation of mutual information as a function of C_{α} atom distance between residue pairs shows that there is significant correlation between residue pairs even 20 Å apart from each other (Supplementary Fig. 19). The long-range coupling between residues is also evident from the mutual information values between each residue pair in *c-src* kinase (Fig. 5). Furthermore, a visual inspection of the matrix also shows the presence of strong local coupling in different regions of kinase (Supplementary Fig. 20). For example, activation segment (residues 404–432) not only shows long-range coupling, but it also forms a strong local network of coupled residues indicating a cooperative unfolding during activation. The second locally coupled network of residues comprises the loop at the C-terminal end of αE , which forms a β -sheet with the unfolded A-loop in the active state. Mutual information values for the key residues in catalytic domain show that all the residues in electrostatic network, R-spine and DFG motif are strongly

coupled with the A-loop (Supplementary Figs 21,22). Furthermore, residues F405 (DFG motif), Glu310, Arg409 and Tyr416 also show a strong coupling with C-terminal lobe residues. Such strong coupling not only offers an avenue for targeted drug design but also explains the origin of the molecular resistance and the deleterious effects of large number of point mutations in the catalytic domain on the kinase activity.

The A-loop region is also coupled to myristate-binding pocket region in *c-src* at the base of the C-terminal lobe. Myristoylation has been shown to regulate *c-src* activity and stability⁴⁴ but there are no crystal structures of the myristoylated *c-src*. It has been shown via NMR experiments that myristate can bind to a pocket in *c-src*, which is similar to the myristate-binding pocket in *c-Abl* (a tyrosine kinase with similar fold as *c-src*)²⁹. Drugs targeting the myristate-binding pocket have been shown to allosterically modulate the conformation of A-loop and ATP-binding pocket in *c-Abl* kinase³. Our results indicate that drugs targeting the myristate-binding pocket in *c-src* could also serve as allosteric modulators of *c-src* kinase activity.

Discussion

In this study, we have used a combination of pathway generation technique (string method in collective variables^{45,46}) and massive distributed computing coupled with MSMs for studying long timescale phenomena⁴⁷. We showed that this method can efficiently capture kinetics with short independent trajectories several orders of magnitude shorter than the timescale of the process under investigation. This method also represents a much more efficient use of simulation time than simply running a long brute force simulation of the same length. For example, to observe 10 activation events of *c-src* kinase in a single trajectory, a total simulation time of $\sim 1,100 \mu\text{s}$ would be required. Here we obtained converged timescales and conformational landscapes in $\sim 100 \mu\text{s}$ starting from a conformational transition pathway determined via the string method.

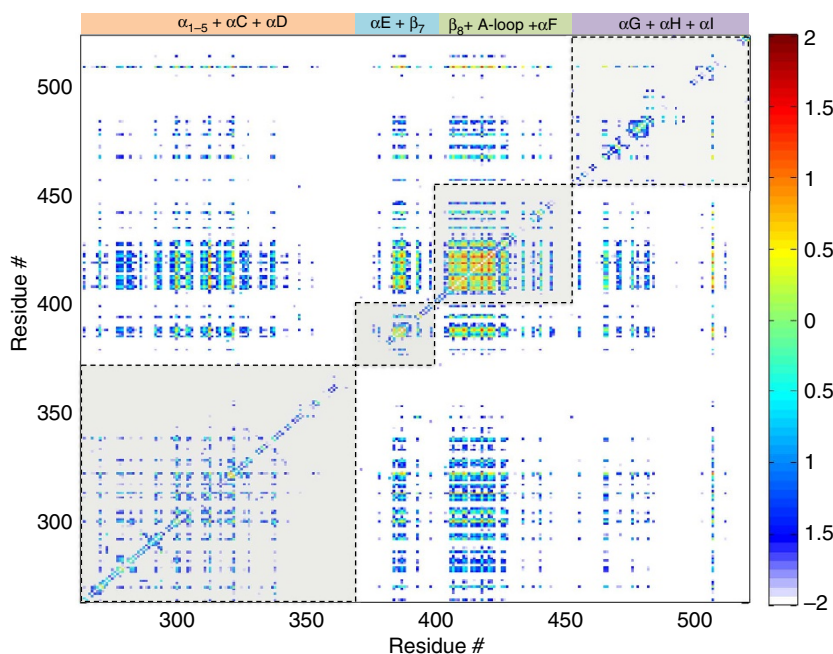


Figure 5 | Mutual information between residues in *c-src* kinase. Hierarchical clustering of significant mutual information values identifies four regions (shown as squares) within the catalytic domain, which have significant dynamical correlation between them. The most strongly coupled cluster involves the residues in the A-loop region that cooperatively unfold during the activation process. Colours indicate the log of the mutual information value.

Our results also show that network models of conformational change present a more natural framework for describing enzyme dynamics on a complex conformational landscape as compared with a linear reaction coordinate analysis or few state models. MSMs are well developed for studying global conformational rearrangements such as protein folding³⁵, but these models have not been used extensively for studying conformational changes in enzymes. This study represents one of the first application of novel MSM approaches for studying complex conformational changes in enzymes and provides pointers for further methodological advances in this direction.

Long timescale simulation of c-src kinase reported in this study has provided novel insights into the mechanism of kinase activation. However, there are several caveats worth noting. First, we have not performed the simulations of the catalytic domain along with its two regulatory domains, which must presumably influence the kinetics of activation. Second, it is likely that the activation mechanism is also dependent on the phosphorylation state of Tyr416. This study provides an extensive database of possible conformations of the catalytic domain, which can be used as a set of templates for obtaining starting configurations for simulations of catalytic domain with phosphorylated Tyr416 and catalytic domain with bound regulatory domains. Owing to these extensive set of distinct starting configurations, their conformational landscapes could be quickly sampled in parallel.

Simple model of kinase autophosphorylation predicts the presence of induction time in the time evolution of active kinase population. According to the activation timescales reported in this study, the rate of autophosphorylation is the limiting step in the kinase phosphorylation process. However, the time evolution of active kinase population was found to be sensitive to the equilibrium constant (K_{eq}) relating inactive to active (unphosphorylated) state equilibrium populations. For Src family kinases in the downregulated state (both regulatory domains bound to the catalytic domain and C-terminal Tyr527 phosphorylated), it has been observed that induction time increases significantly as compared with the kinase with unphosphorylated Tyr527 (Moarefi *et al.*³⁸). Our results suggest that shift in the equilibrium population of active (unphosphorylated) kinases from 10^{-3} to 10^{-4} would change the induction time from several minutes to hours. These results support the view that fast kinase conformational fluctuations at the 10–100- μ s timescale control the observed macro timescale behaviour.

We also predict the existence of binding pocket for novel allosteric inhibitors, the ANS-binding pocket adjacent to C-helix in c-src. On the basis of the binding interactions of ANS with c-src, novel allosteric inhibitors that span the ATP-binding pocket and the ANS-binding site or inhibitors that span both the ANS sites could be designed⁴¹. This fragment-based inhibitor design approach has been used to design CDK2 inhibitor AT7519, which consists of two covalently linked fragments that bind to different regions of the targeted kinase^{48–50}. Allosteric inhibitors targeting this site would not only help in selective inhibition of kinases but would also be useful for understanding the function of a single kinase in a complex cellular signalling milieu⁵¹.

The present results provide a description of the thermodynamic and kinetic factors underlying the activation of c-src kinase with unprecedented atomistic detail. Src-family kinases share a conserved protein fold and are highly homologous in sequence. This raises a host of fascinating questions about the similarities of their activation mechanisms. If there are subtle differences in their activation mechanisms (which could be studied using the methodologies used in this study), then these differences could be harnessed for future design of selective inhibitors for individual members of Src family.

Methods

Generation of initial pathway using string method. The inactive (PDB ID: 2SRC)³⁰ and active (PDB ID: 1Y57)²⁹ X-ray structures of human c-src kinase were used as the two end point conformations for the generation of the MFEP. Only the catalytic domain residues (W260–T521) are included in the simulation set-up. The initial pathway used in adaptive sampling simulation described below was generated by a two-steps procedure to filter out the noise in pathways computed using the string method in high-dimensional spaces in the space of collective variables^{45,46}. The methodology combines principal component analysis to identify a number of relevant modes that can effectively describe the transition with a smoothing procedure of a representative path from the ensemble of projected ones. Principal component analysis was performed on an ensemble of 30 paths that corresponded to iteration 71–100 of the inactive-to-active human c-src tyrosine kinase conformational transition that was obtained in Gan *et al.*²⁴ Fifteen principal components were identified such that the difference with the original ones is less than the difference between the projected pairs of paths. The average of the 30 projected paths was taken and a smoothing algorithm was applied. The strategy we employ to smoothen a string defined in a high-dimensional space is to minimize the following objective function:

$$c = \frac{1}{2} k_a \sum_{i=1}^{M-1} (t_i \cdot t_{i+1})^2 + k_d \sum_{i=1}^M \frac{1}{(R_i + \delta)} \quad (2)$$

where t_i is the unit tangent vector between a pair of images along the string and R_i is the distance of image i from all the replicas of the same image in the ensemble of paths. k_a and k_d are scalar parameters to be chosen by the user that describe the weight of different terms and are equal to 1.0 and 0.03, respectively, while δ is a threshold value. One thousand steps of steepest decent minimization were performed. The final path was then transformed back to original Cartesian coordinate system and was used to initiate the adaptive sampling process⁵² on Folding@home. Starting conformations for the simulations were taken from the 51 snapshots along the smoothed MFEP.

Simulation details. Distributed MD simulations were performed using GRO-MACS⁵³ on the Folding@home²⁸ computing platform. The CHARMM22 (Brooks *et al.*⁵⁴) force field was used for protein and ligands along with TIP3P⁵⁵ water model. The all-atom structures were then solvated in a truncated octahedral solvent box constructed from a $80 \times 80 \times 80 \text{ \AA}^3$ cube with TIP3P water molecules such that water extended at least 10 \AA away from the surface of the protein; 22 Na^+ ions and 19 Cl^- ions were added to the system to neutralize the charge, corresponding to a salt concentration of $\sim 150 \text{ mM}$. Covalent bonds involving hydrogen atoms were constrained with LINCS⁵⁶ and particle mesh Ewald⁵⁷ was used to treat long-range electrostatic interactions. The structures obtained after an initial equilibration for 1 ns at constant temperature and pressure and with constraints on the heavy atom positions were used as the starting conformation for the distributed MD simulations. Production MD simulations were carried out at constant temperature and pressure of 300 K and 1 atm, respectively, with a time step of 2 fs. Two sets of simulations were started using the initial structures generated using the protocol described above. The first set of $\sim 12,000$ simulations were started from the 51 snapshots along the string method pathway for an aggregate simulation time of 150 μ s on the Folding@home platform. To ensure that the simulations started from the string method do not give a biased conformational landscape, another set of 12,000 simulations were performed starting from only the active and inactive states of c-src kinase. The aggregate simulation time of 350 μ s was obtained for this set using the Folding@home platform. The conformational landscapes obtained using the second data set do not reveal new regions of landscape as compared with the conformational landscape obtained using the 150- μ s simulations performed using the snapshots from the string method pathway. Simulations of c-src kinase catalytic domain were also performed in the AMBER99sb-ildn force field (total duration $\sim 50 \mu$ s) to test the dependence of the simulation results on the choice of force field. These simulations were started from the MSM states obtained from the simulations of kinase catalytic domain in CHARMM22 force field. The simulations were found to be independent of the choice of force field.

Adaptive sampling algorithms based on MSMs^{52,58} were used to efficiently sample the conformational landscape of c-src tyrosine kinase. The initial structures for adaptive sampling rounds were chosen from the MSM states with the minimum population. In all, we performed two rounds of adaptive sampling and a total of 24,000 simulations with a total duration of 500 μ s. Trajectory snapshots were recorded every 100 ps. The distribution of total number of simulations versus the simulation length is shown in Supplementary Fig. 23.

Simulations of c-src intermediate with ANS molecules bound to the allosteric site adjacent to the C-helix were performed to assess the ability of ANS to block activation of c-src tyrosine kinase. Two sets of simulations were performed: first, with one molecule of ANS bound to the protein and, the second, with two molecules of ANS bound to the protein. The crystal structure of CDK2 in complex with two molecules of ANS (PDB ID: 3PXF)⁴⁰ was used as a template to align the residues of the c-src intermediate within 5.0 \AA of the ANS-binding pocket. The force field parameters for ANS were obtained using GAFF⁵⁹. The aligned kinase structure along with the ANS molecules was then solvated in a truncated octahedral solvent box with TIP3P water molecules and ions added to neutralize the charge, corresponding to a salt concentration of $\sim 150 \text{ mM}$. The corresponding

systems were minimized using GROMACS for 100,000 steps and then equilibrated for 2 ns at constant temperature and pressure of 300 K and 1 atm, respectively. A total of 2000 distributed MD simulations were started from the two structures for an aggregate simulation time of 50 microseconds using the simulation protocols outlined above.

Markov state models. Our methods draw on kinetic network models called Markov state models (MSMs) that are built from extensive molecular dynamics simulations using Folding@home²⁸ distributed computing architecture to describe a protein-intrinsic dynamics. Similar to a map of a molecule's free energy landscape, an MSM provides a reduced view of the ensemble of spontaneous fluctuations the molecule undergoes at equilibrium. These models provide the population and transition rates between key conformational states of protein^{31,32,36}. The MSMBuilder⁶⁰ software was used to build MSMs of the conformational landscape of c-src tyrosine kinase. Conformations were clustered using a subset of heavy atoms identified using the root mean square fluctuations of atoms after subtracting the inactive-state root mean square fluctuation values. The subset of atoms include all heavy atoms from the N-terminal lobe residues 270–340, and A-loop residues 404–424 were used for the clustering and construction of the MSM (Supplementary Fig. 24).

MSMs describe the conformational dynamics of proteins in terms of jumps between the microstates obtained from the clustering of the simulation data set into discrete states. The transition probability matrix $T(\tau)$ includes the probability of transitions from state i to state j in a certain time interval τ and are estimated by counting the number of transitions n_{ij} observed between time t and $t + \tau$, and normalizing n_{ij} with the sum of all transitions from state i . To enforce detailed balance (which ensures that population of states is conserved), a maximum likelihood estimate of the transition probability matrix that follows the detailed balance is obtained. The transition probability matrix can be used to obtain the population of the system at any time using the following relationship:

$$p(t + \tau) = p(t)T(\tau) \quad (3)$$

where $p(t)$ is the vector of state population at time t . The eigenvalue/eigenvectors spectrum of the transition matrix gives information about the aggregate transitions between subsets of states in the model and what timescales these transitions occur on⁶¹. The equilibrium population of the individual states are estimated from the first eigenvector of the transition probability matrix. The timescales of the dynamical processes occurring on the conformational landscape of c-src can be obtained by estimating the eigenvalues of this matrix. The eigenvalues μ of the transition probability matrix are related to the implied timescales $1/k$ of transitions via the following expression:

$$k = \frac{-\tau}{\ln(\mu)}. \quad (4)$$

The 2000-state model decomposition yielded a good balance between state connectivity and adequate transition sampling. The model also provided converged implied timescales as a function of lag time, which is used as a measure of the Markov nature of the model (Supplementary Fig. 1). A lag time of $\tau = 5$ ns was determined to be suitable by building a series of MSMs at different lag times to find a region where the spectrum of implied timescales are relatively insensitive to lag time. Transition Path theory was used to analyse the distribution of activation pathway fluxes from inactive to active states. Commitor values and MFPTs were computed for each state using methods described in the literature^{62–64}.

Similar simulation and analysis protocol has been successfully used earlier for the study of Hck kinase¹⁸, where initial structures of the catalytic domain were taken from the pathway generated using targeted MD simulations. Short simulations (total simulation time of 1 μ s) from these initial structures were performed and subsequently analysed using kinetic network models for studying the conformational transition between active and inactive state.

Mutual information. The excess mutual information was computed for all protein torsion angles (backbone dihedrals ϕ , ψ and side chain χ angles (only the first χ angle for proline)) throughout the simulations to capture nonlinear correlated motions of residues in an unbiased, statistically robust manner. The following formula was used for the calculation of mutual information between residue pairs⁴³:

$$I_{ij}^{i \neq j} = \sum_{\theta_i} \sum_{\theta_j} \int_0^{2\pi} \int_0^{2\pi} p(\theta_i, \theta_j) \ln \frac{p(\theta_i, \theta_j)}{p(\theta_i)p(\theta_j)} d\theta_i d\theta_j \quad (5)$$

The average of the mutual information computed from 10 iterations of scrambled data was subtracted from the mutual information values computed from the simulation data to filter out correlations that are not statistically significant.

Model for transautophosphorylation of kinases. Residue Tyr416 in c-src kinase is phosphorylated to lock the kinase in the active conformation, thereby enhancing its ability to catalyse the phosphate transfer from ATP to the substrate. The thermodynamic and kinetic data reported in this study are used to build a kinetic model of kinase autophosphorylation following the reaction scheme included

below.



where I is the inactivated state, A is the unphosphorylated active state, A^* is phosphorylated (locked) active state, k_{IA} is the activation rate and k_{AI} is the deactivation rate. The rate $k_{trans-p}$ is the rate of kinase phosphorylation, which requires intermolecular encounter with another active kinase. The time evolution of different state of kinase in the solution can be described using the following ordinary differential equations.

$$\frac{d}{dt}[C_I] = -[C_I]k_{IA} + [C_A]k_{AI} \quad (7)$$

$$\frac{d}{dt}[C_A] = [C_I]k_{IA} - [C_A]k_{AI} - k_{trans-p}[C_A]([C_{A^*}] + [C_A]) \quad (8)$$

$$\frac{d}{dt}[C_{A^*}] = k_{trans-p}[C_A]([C_{A^*}] + [C_A]) \quad (9)$$

The rate of phosphorylation is much slower than the activation/deactivation rates because it involved a diffusion-limited bimolecular encounter between two active kinases. The kinase activation timescales measured in this study are of the order of 100 μ s. The set of equations (7)–(9) could be simplified by assuming a pseudo equilibrium between inactive and unphosphorylated active state.

$$\frac{d}{dt}[C_I] = 0 = -[C_I]k_{IA} + [C_A]k_{AI} \quad (10)$$

$$\frac{[C_I]}{[C_A]} = \frac{k_{AI}}{k_{IA}} = \frac{P_I}{P_A} = K_{eq} \quad (11)$$

The inactive (P_I) and unphosphorylated active state (P_A) equilibrium populations obtained from the MSMs results reported in this study are used for estimating K_{eq} . The pseudo equilibrium assumption is valid under the following condition:

$$k_{trans-p}([C_{A^*}] + [C_A]) \ll k_{IA} \quad (12)$$

$$k_{trans-p}([C_{tot}] - [C_I]) \ll k_{IA} \quad (13)$$

$$r = \frac{k_{trans-p}[C_{tot}]}{k_{IA}} \ll 1 \quad (14)$$

The total concentration of protein $[C_I] + [C_A] + [C_{A^*}] = [C_{tot}]$ is fixed, which provides the following relationship between $[C_A]$ and $[C_{A^*}]$.

$$[C_A] = \frac{[C_{tot}] - [C_{A^*}]}{1 + K_{eq}} \quad (15)$$

The equation (9) is simplified to the following ordinary differential equation by substituting $[C_A]$ expression from the equation (15).

$$\frac{d}{dt}[C_{A^*}] = k_{trans-p} \frac{[C_{tot}] - [C_{A^*}]}{1 + K_{eq}} \left(\frac{[C_{tot}] - [C_{A^*}]}{1 + K_{eq}} + [C_{A^*}] \right) \quad (16)$$

$$\frac{d}{dt} \left(\frac{[C_{A^*}]}{[C_{tot}]} \right) = \frac{k_{trans-p}[C_{tot}]}{1 + K_{eq}} \left(1 - \frac{[C_{A^*}]}{[C_{tot}]} \right) \left(\frac{1 + K_{eq}[C_{A^*}]/[C_{tot}]}{1 + K_{eq}} \right) \quad (17)$$

$$\frac{dP_{A^*}}{dt} = \frac{k_{trans-p}[C_{tot}]}{1 + K_{eq}} (1 - P_{A^*}) \left(\frac{1 + K_{eq}P_{A^*}}{1 + K_{eq}} \right) \quad (18)$$

This initial value problem can be solved numerically or analytically using $P_{A^*}^0 = 0$ and $[C_{tot}] = 1 \mu$ M as the initial conditions. The analytical solution of the above initial value problem has the following form:

$$-\ln \left(\frac{1 - P_{A^*}}{1 + K_{eq}P_{A^*}} \right) = \frac{k_{trans-p}[C_{tot}]}{1 + K_{eq}} t \quad (19)$$

Substituting the values $t = t_{1/2}$, $P_{A^*} = 0.5$ in the above equation provides the expression for the time required for half the kinases to be phosphorylated.

$$t_{1/2} = \frac{\ln(2 + K_{eq})(1 + K_{eq})}{k_{trans-p}[C_{tot}]} \quad (20)$$

This expression is used to obtain limits on the value of K_{eq} by comparing the experimental $t_{1/2}$ values with the model predictions for different values of K_{eq} . For $k_{trans-p} = 10 \mu$ M⁻¹s⁻¹, $[C_{tot}] = 1 \mu$ M and $1 \text{ min} < t_{1/2} < 10 \text{ min}$, the K_{eq} values lie in the range (130,880). The value of K_{eq} calculated from the MSM equilibrium populations is 740. Similarly, for the predicted value of K_{eq} , the limits on $k_{trans-p}[C_{tot}]$ were found to be (8,80) s⁻¹.

ADP release has been identified as the slowest step in the kinase catalytic cycle for a variety of kinases⁶⁵. In the model above, the effect of slow ADP release process could be captured by decreasing the autocatalytic effect of the active phosphorylated kinase population. The ADP-bound active state cannot phosphorylate other active kinases and has to release ADP and bind to ATP to catalyse the next phosphate transfer reaction. This effect can be captured by lowering the value of the rate of autophosphorylation $k_{trans-p}$. The value of $k_{trans-p}$ was chosen to be 10μ M⁻¹s⁻¹ owing to the rate-limiting effect of ADP release.

The value of $k_{\text{trans-p}} = 10 \mu\text{M}^{-1} \text{s}^{-1}$ at the total kinase concentration of $1 \mu\text{M}$ translates to an effective phosphorylation rate of 10s^{-1} . The experimental ADP release rate lies in the range $(1,100) \text{s}^{-1}$ (Adams *et al.*⁶⁵). Therefore, the results from the simplified model proposed above are not expected to change significantly because of incorporation of the ADP release in the model. However, we have also included details of the kinetic model with ADP release in the Supplementary Methods to substantiate this claim.

Procedure for generating MSM trajectories. Several simulation studies have reported methodologies for obtaining kinetics from network model^{66–68}. The time course of observables were obtained by using a kinetic Monte Carlo scheme on the transition probability matrix T , where T_{ij} represent the probability of transition from state i to j . For a MSM trajectory in state i at time t , the next state visited at time $t + \tau$ is estimated using the following scheme:

1. Generate a random number lying between $[0,1]$.
2. Calculate the probability R_{ik} for all states given by $R_{ik} = \sum_{j=1}^k T_{ij}$.
3. If the generated random number lies between $R_{i,k}$ and $R_{i,k+1}$, then the state k is chosen for transition.
4. Select a random snapshot from the conformations assigned to state k . All conformations assigned to a particular microstate are structurally similar (because of the small size of the microstate) and they interconvert at a timescale smaller than the lag time. Therefore, the representative conformation for state k is chosen randomly (uniformly) from all the conformations assigned to this state.
5. Calculate the observables from the selected conformation.

Procedure for calculating MFPTs between states from MSMs. Singhal *et al.*^{69,70} have reported a detailed procedure for the calculation of MFPT from MSMs. In brief, MFPTs are calculated from a set of linear equations that are based on the MSM transition probability matrix. MFPT is defined as the average time taken to get from a particular state i to a any state j in the final (in this case all active or inactive states) set of states. The MFPT can be defined conditionally based on the following expression:

$$\text{MFPT}(i) = \sum_{\text{transition}(i,k)} T_{ik} \times \text{MFPT}(i | \text{transition}(i,k)) \quad (21)$$

where the sum is over all transitions from i to j . The $\text{MFPT}(i)$ is the sum of the time it takes to transition from state k to i and $\text{MFPT}(k)$, which gives

$$\text{MFPT}(i) = \sum_{\text{edge}(i,k)} T_{ik} \times (\tau + \text{MFPT}(k)) \quad (22)$$

where sum is over all the edges leading from state i . Similar equations can be written for MFPT from all states to state j to generate a system of linear equations, which can be solved for $\text{MFPT}(i)$.

References

1. Blume-Jensen, P. *et al.* Oncogenic kinase signalling. *Nature* **411**, 355–365 (2001).
2. Noble, M., Endicott, J. & Johnson, L. Protein kinase inhibitors: insights into drug design from structure. *Sci. Signal.* **303**, 1800 (2004).
3. Zhang, J., Yang, P. & Gray, N. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **9**, 28–39 (2009).
4. Eck, M. *et al.* The interplay of structural information and functional studies in kinase drug design: insights from bcr-abl. *Curr. Opin. Cell Biol.* **21**, 288 (2009).
5. Rabiller, M. *et al.* Proteus in the world of proteins: conformational changes in protein kinases. *Arch. Pharm.* **343**, 193–206 (2010).
6. Daub, H., Specht, K. & Ullrich, A. Strategies to overcome resistance to targeted protein kinase inhibitors. *Nat. Rev. Drug Discov.* **3**, 1001–1010 (2004).
7. Gibbons, D., Priel, S., Kantarjian, H., Cortes, J. & Quintás-Cardama, A. The rise and fall of gatekeeper mutations: the bcr-abl1 t315i paradigm. *Cancer* **118**, 293–299 (2011).
8. Zhang, J. *et al.* Targeting bcr—abl by combining allosteric with atp-binding-site inhibitors. *Nature* **463**, 501–506 (2010).
9. Fang, Z., Grütter, C. & Rauh, D. Strategies for the selective regulation of kinases with allosteric modulators-exploiting exclusive structural features. *ACS Chem. Biol.* **8**, 58–70 (2013).
10. Aleshin, A. & Finn, R. Src: a century of science brought to the clinic. *Neoplasia* **12**, 599 (2010).
11. Roskoski, R. Src protein—tyrosine kinase structure and regulation. *Biochem. Biophys. Res. Commun.* **324**, 1155–1164 (2004).
12. Ozkirimli, E., Yadav, S., Miller, W. & Post, C. An electrostatic network and long-range regulation of src kinases. *Protein Sci.* **17**, 1871–1880 (2009).
13. Ozkirimli, E. & Post, C. Src kinase activation: a switched electrostatic network. *Protein Sci.* **15**, 1051–1062 (2009).
14. Taylor, S. & Kornev, A. Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem. Sci.* **36**, 65–77 (2011).
15. Lin, Y., Meng, Y., Jiang, W. & Roux, B. Explaining why gleevec is a specific and potent inhibitor of abl kinase. *Proc. Natl Acad. Sci. USA* **110**, 1664–1669 (2013).
16. Lovera, S. *et al.* The different flexibility of c-src and c-abl kinases regulates the accessibility of a druggable inactive conformation. *J. Am. Chem. Soc.* **134**, 2496–2499 (2012).
17. Yang, S. & Roux, B. Src kinase conformational activation: thermodynamics, pathways, and mechanisms. *PLoS Comput. Biol.* **4**, e1000047 (2008).
18. Yang, S., Banavali, N. & Roux, B. Mapping the conformational transition in src activation by cumulating the information from multiple molecular dynamics trajectories. *Proc. Natl Acad. Sci. USA* **106**, 3776–3781 (2009).
19. Kohlhoff, K. *et al.* Cloud-based simulations on google exacycle reveal ligand modulation of gpcr activation pathways. *Nat. Chem.* **6**, 15–21 (2014).
20. Masetti, M., Cavalli, A., Recanatini, M. & Gervasio, F. Exploring complex protein-ligand recognition mechanisms with coarse metadynamics. *J. Phys. Chem. B* **113**, 4807–4816 (2009).
21. Saladino, G. & Gervasio, F. New insights in protein kinase conformational dynamics. *Curr. Top. Med. Chem.* **12**, 1889–1895 (2012).
22. Berteotti, A. *et al.* Protein conformational transitions: the closure mechanism of a kinase explored by atomistic simulations. *J. Am. Chem. Soc.* **131**, 244–250 (2008).
23. Shan, Y. *et al.* A conserved protonation-dependent switch controls drug binding in the abl kinase. *Proc. Natl Acad. Sci. USA* **106**, 139–144 (2009).
24. Gan, W., Yang, S. & Roux, B. Atomistic view of the conformational activation of src kinase using the string method with swarms-of-trajectories. *Biophys. J.* **97**, L8 (2009).
25. Banavali, N. & 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* **13**, 1715–1723 (2005).
26. Deng, N.-J., Zheng, W., Gallicchio, E. & Levy, R. M. Insights into the dynamics of hiv-1 protease: a kinetic network model constructed from atomistic simulations. *J. Am. Chem. Soc.* **133**, 9387–9394 (2011).
27. Meng, Y. & Roux, B. Locking the active conformation of c-src kinase through the phosphorylation of the activation loop. *J. Mol. Biol.* **426**, 423–435 (2014).
28. Shirts, M. & Pande, V. Screen savers of the world unite! *Science* **290**, 1903–1904 (2000).
29. Cowan-Jacob, S. *et al.* The crystal structure of a c-src complex in an active conformation suggests possible steps in c-src activation. *Structure* **13**, 861–871 (2005).
30. Xu, W. *et al.* Crystal structures of c-src reveal features of its autoinhibitory mechanism. *Mol. Cell* **3**, 629–638 (1999).
31. Prinz, J.-H. *et al.* Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).
32. Noé, F., Horenko, I., Schütte, C. & Smith, J. C. Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *J. Chem. Phys.* **126**, 155102 (2007).
33. Sicheri, F., Moarefi, I. & Kuriyan, J. Crystal structure of the src family tyrosine kinase hck. *Nature* **385**, 602 (1997).
34. Frankel, M., Bishop, S., Ablooglu, A., Han, Y. & Kohanski, R. Conformational changes in the activation loop of the insulin receptor's kinase domain. *Protein Sci.* **8**, 2158–2165 (2008).
35. Lane, T., Shukla, D., Beauchamp, K. & Pande, V. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **23**, 58–65 (2013).
36. Pande, V., Beauchamp, K. & Bowman, G. Everything you wanted to know about markov state models but were afraid to ask. *Methods* **52**, 99–105 (2010).
37. Shan, Y. *et al.* Oncogenic mutations counteract intrinsic disorder in the egfr kinase and promote receptor dimerization. *Cell* **149**, 860–870 (2012).
38. Moarefi, I. *et al.* Activation of the src-family tyrosine kinase hck by sh3 domain displacement. *Nature* **385**, 650–653 (1997).
39. Cooper, J. A. & MacAuley, A. Potential positive and negative autoregulation of p60-src by intermolecular autophosphorylation. *Proc. Natl Acad. Sci. USA* **85**, 4232–4236 (1988).
40. Betzi, S. *et al.* Discovery of a potential allosteric ligand binding site in cdk2. *ACS Chem. Biol.* **6**, 492 (2011).
41. Martin, M. P. *et al.* A novel approach to the discovery of small-molecule ligands of cdk2. *Chembiochem* **13**, 2128–2136 (2012).
42. Zhang, Y. *et al.* Tyr130 phosphorylation triggers syk release from antigen receptor by long-distance conformational uncoupling. *Proc. Natl Acad. Sci. USA* **105**, 11760–11765 (2008).
43. McClendon, C., Friedland, G., Mobley, D., Amirkhani, H. & Jacobson, M. Quantifying correlations between allosteric sites in thermodynamic ensembles. *J. Chem. Theory Comput.* **5**, 2486–2502 (2009).
44. Patwardhan, P. & Resh, M. D. Myristoylation and membrane binding regulate c-src stability and kinase activity. *Mol. Cell Biol.* **30**, 4094–4107 (2010).
45. Maragliano, L., Fischer, A., Vanden-Eijnden, E. & Ciccotti, G. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **125**, 024106 (2006).

46. Pan, A., Sezer, D. & Roux, B. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* **112**, 3432–3440 (2008).
47. Pan, A. C. & Roux, B. Building markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **129**, 064107 (2008).
48. Wyatt, P. G. *et al.* Identification of n-(4-piperidinyl)-4-(2, 6-dichloro-benzoylamino)-1h-pyrazole-3-carboxamide (at7519), a novel cyclin dependent kinase inhibitor using fragment-based x-ray crystallography and structure based drug design. *J. Med. Chem.* **51**, 4986–4999 (2008).
49. Erlanson, D. A. Fragment-based lead discovery: a chemical update. *Curr. Opin. Biotechnol.* **17**, 643–652 (2006).
50. Gill, A. New lead generation strategies for protein kinase inhibitors-fragment based screening approaches. *Mini Rev. Med. Chem.* **4**, 301–311 (2004).
51. Knight, Z. A., Lin, H. & Shokat, K. M. Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* **10**, 130–137 (2010).
52. Weber, J. & Pande, V. Characterization and rapid sampling of protein folding markov state model topologies. *J. Chem. Theory Comput.* **7**, 3405–3411 (2011).
53. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
54. Brooks, B. *et al.* Charmm: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (2004).
55. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
56. Hess, B. *et al.* Lincs: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
57. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an nlog (n) method for ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993).
58. Bowman, G., Ensing, D. & Pande, V. Enhanced modeling via network theory: adaptive sampling of markov state models. *J. Chem. Theory Comput.* **6**, 787–794 (2010).
59. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
60. Beauchamp, K. *et al.* Msmbuilder2: modeling conformational dynamics on the picosecond to millisecond scale. *J. Chem. Theory Comput.* **7**, 3412–3419 (2011).
61. Bowman, G. R., Beauchamp, K. A., Boxer, G. & Pande, V. S. Progress and challenges in the automated construction of markov state models for full protein systems. *J. Chem. Phys.* **131**, 124101 (2009).
62. Metzner, P., Schütte, C. & Vanden-Eijnden, E. Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.* **125**, 084110 (2006).
63. Metzner, P., Schütte, C. & Vanden-Eijnden, E. Transition path theory for markov jump processes. *Multiscale Model. Simul.* **7**, 1192–1219 (2009).
64. Weinan, E. & Vanden-Eijnden, E. Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* **61**, 391–420 (2010).
65. Adams, J. A. *et al.* Kinetic and catalytic mechanisms of protein kinases. *Chem. Rev.* **101**, 2271–2290 (2001).
66. Marinelli, F., Pietrucci, F., Laio, A. & Piana, S. A kinetic model of trp-cage folding from multiple biased molecular dynamics simulations. *PLoS Comput. Biol.* **5**, e1000452 (2009).
67. Pietrucci, F., Marinelli, F., Carloni, P. & Laio, A. Substrate binding mechanism of hiv-1 protease from explicit-solvent atomistic simulations. *J. Am. Chem. Soc.* **131**, 11811–11818 (2009).
68. Xia, J., Deng, N.-j. & Levy, R. M. Nmr relaxation in proteins with fast internal motions and slow conformational exchange: model free framework and markov state simulations. *J. Phys. Chem. B* **117**, 6625–6634 (2013).
69. Singhal, N., Snow, C. & Pande, V. Using path sampling to build better markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* **121**, 415 (2004).
70. Singhal, N. & Pande, V. S. Error analysis and efficient sampling in markovian state models for molecular dynamics. *J. Chem. Phys.* **123**, 204909 (2005).

Acknowledgements

This work was funded in part by the SIMBIOS NIH National Center for Biomedical Computation through the NIH Roadmap for Medical Research Grant U54 GM07297. We thank the users of the Folding@Home distributed computing project who donated compute time used for the simulations. This work was supported by the NIH Grant CA093577 (Y.M. and B.R.). D.S. thanks Dr Morgan Lawrenz and Mohammad M. Sultan from Department of Chemistry at Stanford University for many insightful discussions.

Author contributions

V.S.P. and B.R. conceived and supervised the project. D.S. performed simulations, analysed data and build MSMs. Y.M. performed the string method calculations. D.S. wrote the manuscript with inputs from Y.M., B.R. and V.S.P.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npq.nature.com/reprintsandpermissions/>

How to cite this article: Shukla, D. *et al.* Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat. Commun.* **5**:3397 doi: 10.1038/ncomms4397 (2014).