Efficient and Accurate Modeling of Conformational Transitions in Proteins: The Case of c-Src Kinase

Edoardo Milanetti,†,‡ Alexandra G. Trandafir,§ Josephine Alba,§ Domenico Raimondo,∥ and Marco D’Abramo*§,∥

†Department of Physics and §Department of Chemistry, Sapienza University of Rome, P. le A. Moro, 5, 00185 Rome, Italy
‡Center for Life Nano Science@Sapienza, Istituto Italiano di Tecnologia, Viale Regina Elena 291, 00161 Rome, Italy
∥Department of Molecular Medicine, Sapienza University of Rome, Viale Regina Elena 324, 00161 Rome, Italy

ABSTRACT: The theoretical computational modeling of large conformational transitions occurring in biomolecules still represents a challenge. Here, we present an accurate “in silico” description of the activation and deactivation mechanisms of human c-Src kinases, a fundamental process regulating several crucial cell functions. Our results clearly show that by applying an efficient and automated algorithm able to drive the molecular dynamics (MD) sampling along the pathway between the two c-Src conformational states—the active state and the inactive state—it is possible to accurately describe, at reduced computational costs, the molecular mechanism underlying these large conformational rearrangements. This procedure, combining the MD simulations with the sampling along the well-defined principal motions connecting the two conformational states, allows to provide a description well beyond the present computational limits, and it is easily applicable to different systems where the structures of both the initial and final states are known.

1. INTRODUCTION

The Src family kinases (SFKs) are the main group of nonreceptor tyrosine kinases. They are enzymes involved in different regulation processes with important roles in signal transduction following the engagement of many different classes of cellular receptors. They control cell growth, differentiation, migration, and cell survival, and their activity is regulated by the phosphorylation status of two tyrosine residues in key positions and by ligand interaction through kinase domains. These interactions stabilize the kinase in an active or inactive form. A dysfunction in the SFK kinase control mechanisms can cause a change in their activation, which is directly related to the onset of many pathologies such as neoplastic progression,†,‡ preeclampsia,§ and immune disorders.∥

The SFKs are proteins of 52–62 kDa, sharing a conserved modular domain organization as we reported in Figure 1. The human Src gene encodes 536 amino acids (the UniProtKB/Swiss-Prot annotation) is composed by the N-terminal region (N-lobe) and the C-terminal region (C-lobe). The enzymatic activity is regulated by the conformational variation and three-dimensional rearrangements of these structural elements, which, in turn, are modulated by phosphorylation and interaction with other regulatory proteins (Figure 1). A crucial residue involved in c-Src auto phosphorylation is Tyr165.7 Autophosphorylation of catalytically inactive kinases can occur, demonstrating that phosphorylation must proceed in trans.6 SH2 and SH3 domains (residues 151–248 and 84–145, respectively) are crucial elements substrate recruitment and in the catalytic activity modulation.7 In physiological condition, the c-Src enzyme is almost inactive in cells, while dramatic changes in its activity occur when the fine balance between phosphorylation and dephosphorylation is altered. However, if the fine balance between phosphorylation and dephosphorylation is disrupted, changes can occur in the c-Src activity with drastic results.8,9 c-Src contains a catalytic loop (residues 385–403) formed by

Received: July 25, 2018
Revised: August 31, 2018
Published: September 5, 2018
Asp and Asn residues that are highly conserved and very important for the Mg--ATP complex coordination. Residues 404–432 form the activation loop. This region plays a key role in the kinase activity because it is involved in the phosphorylation of Tyr416, a protein modification required for full enzymatic activation. In particular, c-Src lives in "open" active and "closed" inactive conformations that are structurally well understood, a behavior similar to other Src family members. In order to bind Mg--ATP and the protein substrate, the enzyme has to assume the open form. Starting from this open conformation, the protein starts converting to the closed one as the catalysis occurs. As the catalysis ends, the enzyme is reconverted to the open form, the phosphorylated protein and Mg--ADP are released, and the next catalytic cycle can occur. The open form of the active enzyme binds Mg--ATP and the protein substrate; this is accompanied by the conversion to the closed form as catalysis occurs. After catalysis, the phosphorylated protein and then Mg--ADP are released as the enzyme is reconverted to the open form prior to the next catalytic cycle. The two main phosphorylation sites on c-Src are represented by Tyr416 and Tyr527. Phosphorylation of Tyr527 is one of the mechanisms for the regulation of the c-Src activity. When Tyr527 is phosphorylated, the molecule adopts a folded, inaccessible bundle, and the inactive configuration is mediated by an intramolecular interaction between the phosphorylated Tyr527 and the SH2 domain of c-Src. This "inactive conformation" is stabilized by a network of interactions between several residues. In particular, Asp404 forms an electrostatic interaction with Lys295, and in this way, it disrupts the Lys295-Glu310 interaction, which is essential for the maintenance of the active state. As a consequence of the Asp404–Lys295 interaction, Phe424 moves toward the α-C-helix of the c-Src N-lobe, forcing the helix to move away from the catalytic cleft. Moreover, Tyr416 is oriented toward the catalytic center of the kinase domain when residues 413–418 of the activation loop form a short α-helix (A-loop helix). In this way, Tyr416 is not only able to form hydrogen bonds with Arg385 and Asp386 but is also buried into a hydrophobic cavity in order to be protected from autophosphorylation. In addition, the active conformation of the α-C-helix is stabilized by the formation of a salt bridge between Glu310 of the α-C-helix and Arg409. When Tyr527 becomes dephosphorylated, this bond is disrupted, the kinase is opened to an active state, and the autophosphorylation of the residue Tyr416 occurs. Protein interactions also act to regulate c-Src by either directly activating c-Src or moving c-Src to the sites of action.

As such, it results that the complex network of interactions determining the structural behavior of c-Src suggests a conformational plasticity intrinsic to this protein family. On the other hand, the ATP analogs such as tyrophostin and pyrimidine compounds directly inhibit the tyrosine kinase activity of c-Src and/or related kinases. However, the inhibition of c-Src by means of small molecules remains quite difficult because of the very likely to generate mechanisms of resistance such as point mutations in the drug binding site 21. Therefore, alternative routes to develop selective drugs represent an invaluable resource. Among others, the identification of allosteric binding sites able to affect the c-Src conformational behavior represents an effective strategy, in principle, to overcome the drug resistance and selectivity issues. The effectiveness of such a strategy strongly depends on the knowledge of the structural modifications occurring between the inactive and the active states of the kinase. In this way, the complex regulatory network—which dynamically evolves along the transition pathway—could be effectively handled to regulate the protein activity. From a computational point of view, the description of the pathway—although possible in principle—is strongly limited by the process timescale, preventing the effectiveness of computational techniques such as molecular dynamics (MD). To overcome such a limit, several methods able to enhance the sampling have been applied to study the kinase conformational transitions, such as coarse-grained representations and enhanced sampling techniques.

Recently, an unprecedented computational effort based on distributed computing has been used to describe the c-Src activation pathways by means of all-atoms unbiased MD simulations. This work has not only been able to map the c-Src activation conformational transition without the use of external biases but it also provides an invaluable reference to measure the effectiveness and the accuracy of different enhancing sampling techniques.

In this framework, we present here the study of the c-Src conformational transition pathways by means of the essential dynamics sampling (EDS), an enhanced sampling technique able to efficiently sample such slow and wide conformational rearrangements. The EDS rejects movements (in the conformational space) not in the desired direction, driving...
the sampling toward a known structure (target). The advantages of EDS are that (i) no ad hoc terms are introduced in the Hamiltonian and (ii) the system moves along its own essential eigenvectors as provided by unbiased MD. Moreover, the contraction procedure used in this work guarantees—by construction—that the system reaches the final state, being only slightly constrained all along the pathway.29

With respect to other theoretical methods potentially able to model these kinds of processes (i.e., free-energy methods or coarse-grained models), it is worth noting that in the EDS procedure, (i) the model is not biased and thus it is easily transferable to other systems/processes; (ii) the model does not require ad hoc parametrization; and (iii) the description is at the atomistic level. The main limitation of EDS is the difficulty to estimate the relative weight of the intermediate conformations sampled along the path.

In this work, the EDS technique has been applied to the both opening and closing processes of c-Src to provide a detailed atomistic description of the corresponding pathways, unrevealing for the first time the differences at the base of such complex conformational transitions. The EDS trajectories have been analyzed in detail and compared with the reference trajectories as obtained by means of massive distributed computing (m.d.c.).27 The results clearly show that these kinds of processes can be described by the proposed procedure with very limited computational efforts and with an accuracy in the pathway description comparable to unbiased MD simulations.

2. COMPUTATIONAL METHODS

MD simulations were performed on c-Src using the Gromacs software package version 5.0.5.30 The initial structures were the c-Src in the closed (inactive state, PDB ID: 2SRC) and open conformations (active state, PDB ID: 1YS7), respectively. Only the kinase domains were considered: residues from Ala175 to Leu449 for 1YS7 and Ala259 to Leu533 for 2SRC. The two conformational states of the kinase domains were (separately) solvated in rectangular boxes large enough to contain the enzyme and 1 nm of the solvent on all sides, resulting in 84 514 atoms and 72 049 atoms for the open and closed forms, respectively. Sodium ions were added to provide a neutral simulation cell. During the simulations, the temperatures were maintained at 300 K using the velocity rescaling algorithm.31 The Amber99sb force field was used.32 For the solvent, the water SPC model33 was used at a standard density (1 kg/dm3). The Verlet cutoff scheme was used;34 long-range electrostatic interactions were treated by means of the particle mesh Ewald method.35,36 For all simulations, the protein secondary structures were monitored to ensure that no unfolding occurred. All the analyses were performed using the Gromacs tools included in the package, unless otherwise specified. The first 20 ns of the two trajectories was considered as equilibration and has been removed from the analysis. In line with our previous work,37 the principal component analysis (PCA) was performed using the covariance matrix of the fluctuations of the C-α motions as obtained by the combined trajectories of both the conformational states. In this way, the eigenvectors describe the motion directions connecting the close and the open forms of c-Src, and they can be used in the EDS procedure to decrease the distance from the reference conformation (i.e., the contraction procedure). In such a procedure, at each time frame, a classic MD step is performed, and the distance between the current and the reference conformation in the subspace is calculated. The step is accepted if the distance decreases. If this does not occur, the atomic coordinates are projected radially on the subspace hypersphere centered on the target conformation. The hypersphere is defined by the set of the essential eigenvectors of the C-α atoms given by the PCA analysis performed on the two concatenated (free) MD trajectories. The displacement in the essential subspace during one time step of MD can be represented by

$$\Delta \xi = \Delta \xi_d + \Delta \xi_c$$

where $\Delta \xi_d$ is the displacement produced by the MD step and $\Delta \xi_c$ is the correction for the application of the constraint. The constraint in the essential subspace can be defined as

$$G(\xi(t); t = 0)$$

To obtain a unique solution to this equation, the requirement is that the total perturbation $|\Delta \xi_d|$ is minimized by adding one Lagrangian multiplier (the mathematical details can be found in ref 29).

Following our previous work,37 we performed a preliminary set of EDS simulations varying the number of eigenvectors used to define the subspace. In this way, we assess the minimum number of eigenvectors able to describe the transitions but still leading final structures very similar to those extracted from the free MD simulations and used as targets. We found that the use of the first 100 eigenvectors leads to structures which are still very similar to the corresponding targets (see Table S1 in the Supporting Information), providing at the same time a very limited bias on the trajectory. To select proper reference structures representative of the two conformational states, a cluster analysis of the unbiased MD simulations has been performed (the Gromacs tool g_cluster with the default cutoff value of 0.1 nm on the root-mean-square deviation (rmsd) was used), providing nine representative structures for both the open and the closed forms. Every possible structure pair has been then used as the starting and arrival states in the EDS procedure, providing 162 (81 × 2) possible paths connecting the two protein conformational states. A scheme of our approach is reported in the Supporting Information (Figure S1). These conformational pathways have been compared to those obtained by Shukla et al.27 In that study, a m.d.c. approach was used to reproduce the transition between the closed and the open forms of c-Src. To this end, the authors employed a very large number of small trajectories which required a huge computational effort. In that work, the authors collected about 500 ms of all-atoms unbiased MD trajectories, and ~20 000 structures have been considered as the representative of the whole set of MD simulations (freely available from https://searchworks.stanford.edu/view/cm993jk8755). This representative trajectory was converted in the “dcd” format to import and analyze it by means of the “bio3d” package of R software. To compare our approach with the m.d.c. results, we calculated for each c-Src structure as sampled by both m.d.c and EDS approaches, two structural observables (defined in Shukla et al.27 and we were able to describe the conformational pathways): (i) the rmsd of the activation loop (A-loop) and (ii) the distance difference between the two pairs of the residues (i.e., Glu310 and Lys295). The rmsd was calculated using the “rmsd” function of the “bio3d” package of R software.

The native contact analysis has been performed considering the distance between all the C-α atoms of the protein. For any frame obtained by the EDS, the C-α–C-α distance is

DOI: 10.1021/acs.jpcb.8b07155


C
measured, and the residues pair is considered in contact if their distance is less than 6 Å.

Two residues are not considered in contact if they have a sequence distance of less than 4, even if their distance in structure is less than 6 Å. The first frame of each EDS trajectory (which represents the protein in the starting conformation) is considered as the reference structure for the analysis. For each frame, we defined the fraction of native contacts (Q) as the ratio between the number of contacts in common with the reference structure and the number of contacts of such a frame.

3. RESULTS AND DISCUSSION

To obtain an accurate description of the c-Src kinase domain in the two reference states (i.e., open and closed), we first performed a 200 ns long classical MD simulation for both the active and the inactive forms of c-Src. As expected, the secondary structure elements as well the overall fold do not change along the trajectories, and both the proteins remain in the conformational basin common to the corresponding starting structure. Following our previous work, we apply the EDS technique using the set of the first 100 eigenvectors and eigenvalues obtained by the diagonalization of the atomic fluctuation covariance matrix of the trajectory obtained by merging the active and the inactive MD simulations (hereafter called the concatenated trajectory). The resulting first two eigenvectors obtained by the ED analysis on the concatenated trajectory represent the 92% of the variance of the whole protein motion, which includes the open and the closed conformational basins. As expected, the first eigenvector—characterizing the motion direction linking the two conformational states—describes almost the total motion of the system (≈87%). By means of a clustering analysis, 18 representative structures (9 for “active” form and 9 for “inactive” form) have been selected from the two unbiased MD and used as the starting and target structures for the EDS procedure. That is, the EDS runs were performed to connect each of the nine representative structures of the open state with each one of the nine structures of the closed states (and vice versa). In this work, the EDS algorithm has been used in the contraction scheme, that is, to decrease—along the simulation—the distance in the essential subspace between the selected reference structure and the target structure. By such an approach, 81 trajectories from close to open conformations and 81 trajectories from open to close conformations have been obtained.

3.1. Activation Pathway Identified through the EDS Technique. The analysis of the EDS trajectories performed along the activation pathway reveals that the main changes are carried—as expected—by the A-loop and the C-helix (see Figure 2). In particular, in the opening process, the rearrangements of such regions occur in a concerted fashion: the A-loop runs outward, whereas the C-helix moves inward, in completely agreement with previous computational studies.

Other conformational changes involve residues buried within the protein domain, Leu325, Met314, Phe405, and His384. They form a skeleton of four nonconsecutive hydrophobic residues that constitute a regulatory region, called R-spine, linking the two lobes of the catalytic domain. It is critical for the catalytic activity of the kinase because it is usually broken in the deactivation process. Our analysis confirms that this interaction pathway is rapidly broken during the transition from the active to the inactive form. Our results also show that His384, Arg385, and Asp386 (called HRD motif) maintain the same orientation during the opening process. In the open conformation, the HRD motif does not interact with the A-loop nor with the C-helix. Only after that the C-helix has reached the open conformation and the A-loop moved outward, Arg385 starts to interact with Tyr416 belonging to the A-loop. The electrostatic interactions between Arg385, Tyr416, Arg409, Glu310, Lys295, and Asp404 are known to be involved in the conformational changes occurring during the activation and the deactivation processes. Previous studies have shown that in the closed conformation, there is a peculiar electrostatic interaction network formed by the following couples: Glu310 with Arg409, Arg385 with Tyr416, and Lys295 with Asp404. In the open conformation, these interactions do not take place. Our simulations, confirming such a behavior (Figure 3A,B), also outline that in the activation process, the first interaction to be broken is between Arg385 with Tyr416, anchoring the A-loop to the HRD motif (Tyr416 moves away so that the A-loop starts to move outward and, almost simultaneously, the C-helix starts moving inward). As a result of these motions, the residues Arg409 (belonging to the A-loop) and Glu310 (in the C-helix) no longer interact (Figure 3A).

3.2. Deactivation Pathway Identified through the EDS Technique. In the deactivation process, we found that the HRD motif interacts with Tyr416. As the A-loop moves inward, Tyr416 (belonging to the A-loop) starts to interact with R385 as observed in a previous study of Src kinase conformational activation.

During the deactivation process, our simulations show that the interaction between Tyr416 and Arg409 disappears early; the A-loop starts to fold, and the C-helix is almost in the closed conformation. However, the interaction between Lys295 and Glu310 is still in place. At the same time, the interaction between Lys295 and Glu310 vanishes. The C-helix is in the closed conformation and starts to twist and tilt, driving Glu310 outward, where it finally interacts with Arg409. At the end of the deactivation process, the interactions between Arg385 and Tyr416 take place, stabilizing the structure in the closed conformation (Figure 3B).

The closed conformation is stabilized by the interaction between the A-loop with the HRD motif, the C-helix interacting with the N-lobe. After the Lys295 and Glu310

Figure 2. Superimposition of active (in blue) and inactive (in gold) c-Src conformations along the activation pathway obtained by using the EDS approach. The conformational changes of the αC helix and A-loop (residues 402–424) reveal outward movements of the A-loop while the αC-helix moves inward.
interaction is broken, the A-loop and C-helix show a concerted movement until the interactions stabilizing the closed conformation take place.

3.3. Activation Pathway: Comparison between the EDS and m.d.c Approaches. As highlighted in the Introduction section, the main goal of the present work has been the validation of the sampling as obtained by the EDS technique. To this end, the EDS trajectories were compared to the recent work of Shukla et al.,27 where the conformational transition pathways of the c-Src activation process were described by m.d.c.. In that work, making use of huge computational resources, the authors performed ~24 000 MD simulations totally lasting 50 ms. The results were summarized in a single trajectory (using an adaptive sampling algorithm based on the Markov state model approach), which we used to compare our EDS data.

In the work of Shukla et al.,27 the authors characterized two intermediate states of the c-Src kinase domain by means of two simple structural observables: (i) the rmsd of the A-loop with respect to the inactive state and (ii) the distance between the two residue pairs (Glu310−Arg409 and Lys295−Glu310).

The comparison between the behaviors of these two observables in the two different approaches is reported in Figure 4. We can clearly notice that EDS essentially provides the same intermediate structures of the unbiased simulations with respect to these observables (rmsd of the A-loop and the difference between the Glu310−Arg409 and Lys295−Glu310 distances). In fact, starting from similar (i.e., sampled within the same conformational basin) closed structures, both the methods describe the activation process by an initial change in the A-loop conformation followed by the formation of the Lys295−Glu310 H-bond and the rupture of the Glu310−Arg409 one. The differences in the point densities are due to the fact that the EDS algorithm does not provide the correct weight of the sampling as the m.d.c. does. Very importantly, such results strongly suggest that—at the orders of magnitude faster—the EDS algorithm actually provides the same intermediates as those sampled by an unbiased m.d.c. approach (at least in the subspace described by the A-loop rmsd and by the difference between the Glu310−Arg409 and Lys295−Glu310 distances).

To further compare the two sampling methodologies, the EDS and the m.d.c., we projected the two sets (i.e., EDS and m.d.c.) of structures sampled along the activation pathway on the subspace spanned by the two principal eigenvectors (Figure 5). As the eigenvalues correspond to the mean square eigenvector coordinate fluctuation and, therefore, contain the contribution of each principal component to the total fluctuation, the projections have been scaled according to the ratio between the two eigenvalues (0.057). The projections along the first principal component as obtained by the EDS technique reported in Figure 5 closely resemble those provided by the m.d.c. approach. The discrepancy along the second principal component, between the EDS and m.d.c. inactive state (still very small), is due to the different conditions used to define the productive phase of the corresponding simulations.

In fact, Shukla et al. have reconstructed a unique trajectory to describe the conformational transitions of the Src kinase combining several short trajectories—lasting 20 ns plus 1 ns considered as the equilibration time—into a single model.27 On the other hand, the nine reference structures used as the starting point and the nine structures representing the target (arrival) points of the EDS trajectories were extracted from the corresponding conformational basins as provided by two single and independent “free” 200 ns MD runs (note that these 200 ns long MD simulations were only used to describe the two forms, whereas the paths connecting the two forms have been modeled by means of EDS). In such two MD simulations, the first 20 ns was considered as an equilibration phase and hence discarded from the analysis. Including such 20 ns of sampling (Figure 5), it is clear that the differences between our approach and the m.d.c. one is due to such a difference in the starting points of the pathway. In summary, the presented data clearly provide evidence that the EDS trajectories describe the
conformational changes on the activation pathway very similar to those obtained by means of the m.d.c. approach. These results strongly suggest that our methodology can provide a reliable sampling along the opening and closing processes occurring in the c-Src kinase.

### 3.4. Activation and Deactivation Pathways

Because of the EDS computational efficiency, the EDS procedure has also been applied to model the deactivation process. The major structural observables changing along the process have been highlighted in the previous section. Nevertheless, additional information arises when comparing the opening and the closing pathways. First, the projection of the closing pathways on the essential subspace indicates no relevant differences with respect to the c-Src activation within such a subspace (Figure 6). On the contrary, the plot of the distances between Glu310, Arg409, and Lys295 versus the rmsd of the A-loop as occurring along the open-to-close pathway (Figure 7) shows some interesting differences with respect to the activation process (Figure 4). In fact, although in the opening process the H-bond network changes follow the A-loop “unfolding” process (Figure 5), in the closing process, the H-bond network starts to change quite quickly (Figure 7).

In other words, our data suggest that, differently from the activation pathway, the open-to-close mechanism might involve the rupture of the H-bond between Lys295 and Glu310 in the very first part of the process. The differences between the two pathways are also confirmed by the C-helix tilt angle. Considering that the C-helix movement is one of the major conformational changes involved in the processes, in Figure 8, we report the probability to have a tilt angle variation greater than 60° along the paths. That is, we consider that the A-helix conformational change occurred when the tilt angle (variation) reaches this value, taking as zero its initial value (for the sake of clarity, the graphical representations of c-Src structures at various tilt angle values are given in Figure S3).

---

**Figure 5.** Projection along the first and second principal components of nine EDS opening trajectories (blue points) connecting the projections given by the MD of the open and closed states (black points in left and right regions, respectively). The projections of the closed and intermediate structures as provided by the m.d.c approach are represented as gray and red points.

**Figure 6.** Projection along the first and second principal components of nine EDS deactivation trajectories (green points) connecting the projections given by the MD of the open and closed states (black points in left and right regions, respectively). Projections of the closed and intermediate structures as provided by the m.d.c approach are represented as gray and red points.

**Figure 7.** Changes in the structural observables as provided by the EDS during the deactivation process. In x-axis is reported the rmsd of A-loop, whereas in the y-axis is reported the difference between the two distances: the first and the second distances are related to the pairs Glu310–Arg409 and Lys295–Glu310, respectively.

**Figure 8.** The probability to observe the C-helix tilt transition, as measured by a corresponding tilt angle greater than 60°, vs the number of the EDS frames for the opening (the blue line and points) and for the closing (the green line and points) processes.
shown in Figure S3, the C-helix transition occurs in a wide interval, indicating that different EDS trajectories employ a different number of EDS steps to move the C-helix in the final inactive state.

On the contrary, in almost all the EDS opening trajectories, the C-helix suddenly reaches its final open conformation. In Figure S4, we also reported the fraction of conserved native contact (Q) trajectories along the conformational transitions for six representative EDS trajectories (three for the activation process and three for the deactivation process). Interestingly, both the processes retain a remarkable fraction (between 70 and 80%) of the native contacts of the corresponding starting forms, indicating that a large part of local contacts is not influenced by such a large conformational rearrangement. As expected, the protein regions mainly determining such a native contact decrease are the C-helix (residues 298–309) and the A-loop (residues 411–419) ones (see Figure S4).

These results clearly point out the different mechanisms between the activation and the deactivation processes: in the former, the H-bond network is maintained (Figure 4), and the C-helix transition can occur on a wide interval of frames (Figure 8, blue line), whereas in the deactivation mechanism, the H-bond (Figure 7) and C-helix changes (Figure 8, green line) are likely to occur in the very first part of the process. The described mechanism of activation/inactivation has the potential to improve the drug selectivity (which is very limited in the case of conserved binding site), for example, designing new molecules targeting a low-conserved protein region, which might affect one specific protein conformational transition pathway. Therefore, the proposed approach, able to describe slow processes occurring in large biomolecular systems with extreme computational efficiencies, seems very well suited for the design of new inhibitors, acting on different protein regions as well as to shed new light on complex conformational equilibria. In addition, it can give insights into the conformational rearrangements occurring in different processes such as small ligand binding or transition channel opening.

4. CONCLUSIONS

In this study, we modeled the conformational transitions occurring along the pathways connecting the active and inactive states of the c-Src kinase. The complex molecular rearrangements occurring along the activation pathways are in line with previous observations based on the unbiased all-atoms simulations obtained by means of m.d.c.

The active to inactive transition is mainly described by the C-helix and A-loop rearrangements as well as the dynamical change of the interactions in important c-Src motifs (R-spine and HRD motif). Our results clearly show that our approach provides a similar accuracy with respect to unbiased MD simulations at the orders of magnitude faster and, very importantly, without arbitrariness on the choice of the parameters needed to describe the process.

Because of the EDS computational efficiency and the agreement of our data with that produced by Shukla et al. for the c-Src activation pathway, the deactivation pathway has also been described. Interestingly, the associated mechanism is slightly different with respect to the activation: we have found that—in contrast to the opening pathway—the C-helix rearrangement occurs in the very first few EDS steps during the deactivation and the characteristic H-bond network of the closed structure involving the residues Lys295, Glu310, and Arg409.

Finally, we think that the possibility to describe the molecular pathways connecting the active and inactive states at reduced computational cost might improve the discovery of new drugs able to interact with different protein regions and traps the protein in a specific conformational state as well as to better understand the effect of point mutations on complex conformational equilibria.

# ASSOCIATED CONTENT

## Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b07155.

RMSD values of EDS; EDS flowchart; projections along the first and second principal components; C-helix tilt during the activation process; and fraction of native contacts and ribbon representation of Src (PDF)

# AUTHOR INFORMATION

## Corresponding Author

*E-mail: marco.dabramo@uniroma1.it

## ORCID

Marco D’Abramo: 0000-0001-6020-8581

## Author Contributions

The manuscript was written through contributions from all authors. All authors have given their approval to the final version of the manuscript.

## Notes

The authors declare no competing financial interest.

# ACKNOWLEDGMENTS

The authors acknowledge the Sapienza University of Rome for the financial support and CINECA and NVIDIA for the computational support.

## REFERENCES


2201. (12) Xu, W.; Doshi, A.; Lei, M.; Eck, M. J.; Harrison, S. C. Crystal
structures of c-Src reveal features of its autoinhibitory mechanism.
(15) Yamaguchi, H.; Hendrickson, W. A. Structural basis for activation of human lymphocyte kinase Lck upon tyrosine
of c-Src with focal adhesions alters their properties and is regulated by
active inhibitor of epidermal growth factor signaling with potential for