Biophysical Journal, Volume 97

Supporting Material

Atomistic View of the Conformational Activation of Src Kinase Using the String Method with Swarms-of-Trajectories

Wenxun Gan, Sichun Yang, and Benoît Roux

Supplementary Material

1. Preparation of the system

The ANP molecule in the original inactive structure (PDB 2SRC) was mutated into ATP and two magnesium ions were added to coordinate ATP based the structure of ATP-2Mn^{2+} from PKA kinase (PDB 1ATP (1)). Waters from the original crystal structures within 5 Å from the surface of the protein were kept. Hydrogen atoms were built to the initial structures using the HBUILD module of CHARMM (2). The all-atom structures were then solvated in a truncated octahedral solvent box constructed from a 80 × 80 × 80 $Å^3$ cube with TIP3P (3) water molecules such that water extended at least 10 Å 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 approximately 150 mM. Periodic boundary conditions were applied using the CRYSTAL module in CHARMM. Covalent bonds involving hydrogen atoms were constrained with SHAKE and Particle Mesh Ewald (PME) (4) was used to treat long-range electrostatic interactions. After an initial minimization of 200 steps, MD simulations were performed for 2.2 ns at a constant temperature of 300K using the CHARMM program with the CHARMM22 force field. The fully solvated system was first equilibrated with a weak harmonic constraint on the non-hydrogen atoms for 200 ps, and then further equilibrated for 2 ns with no constraint.

2. String method with swarms-of-trajectories

The final snapshots from equilibration were used to generate an initial path using targeted MD (TMD) (5). Instead of applying a harmonic constraint on the RMSD of the inactive structure with reference to the active structure, we applied a weak harmonic constraint on the heavy atoms so that the TMD would converge faster. A total of 51 discretized configurations, including 49 snapshots saved from the TMD trajectory and the two end point structures, form the initial string.

This string method was previously implemented with a set of 553 inter-residue distance as the collective variables (6). To further reduce the large number of degree of freedom in the present study, we use a set of Cartesian coordinates of selected atoms from the kinase A-loop and αC-helix. Specifically, Cα atoms were chosen to characterize protein backbone motion. For amino acids with long charged sidechains, the last carbon atom in sidechain was also included. The entire selection (Table S1) has 50 atoms in total, corresponding to 150 Cartesian position constraints (50 X, Y, Z coordinates).

Images in the string were iterated following the previously published 4-step procedure (6). Briefly, for each of the 51 images, we launched 100 short MD trajectories with different seed to randomize initial velocity, each lasting 1 ps; To remove protein rotation and translation, the 100 configurations at the end of 1 ps trajectories were oriented by best fitting the protein backbone with respect to one reference structure on the basis of the root-mean-square deviation (RMSD). The oriented configurations were then averaged, followed by a linear smoothing and reparametrization according to Eq. (1) and (2):

$$z^{m,*} = (1-s)z^m + \frac{s}{2}(z^{m-1} + z^{m+1}) \quad (1)$$

$$z^m(t+\Delta t) = z^{k-1,*}\frac{L(k)-S(m)}{L(k)-L(k-1)} + z^{k,*}\frac{S(m)-L(k-1)}{L(k)-L(k-1)} \quad (2)$$

where z^m and $z^{m,*}$ are the variables of m^{th} image before and after smoothing, *s* is a parameter that determines to what extent the configurations are smoothed, *L(k)* is the length of the smoothed string up to image *k* and *S(m)* is the target length for image *m* during reparametrization; images in the string were then updated to the reparametrized ones by running 50 ps constrained equilibration with a force constant of 50 kcal/mol·Å^2.

The 4-step procedure was iterated 100 times and the initial TMD pathway relaxed after about 50 to 60 iterations (Figure S1). All the MD simulations of swarm of short trajectories and constrained equilibration were run by NAMD (7) with the CHARMM22 force field. The averaging, smoothing and reparametrization were all done with CHARMM controlled by an external Shell script.

3. Free energy calculations

The free energy profile of the activation pathway was computed by integrating the mean force along the string (8):

$$F(z(\alpha)) - F(z(0)) = \int_0^\alpha \frac{dF}{d\alpha'}d\alpha' = \int_0^\alpha \sum_{i=1}^{n} \frac{dz_i(\alpha')}{d\alpha'}\frac{\partial F(z(\alpha'))}{\partial z_i}d\alpha' \quad (3)$$

where the mean force for each image was estimated by further running 200 ps constrained MD simulations with a force constant of 5 kcal/mol·Å^2 and calculating the differences of selected collective variables from their constrained reference positions.

$$\frac{\partial F(z)}{\partial z_j} \approx k(z_j - \langle z_j \rangle) \quad (4)$$

and the derivatives of selected collective variables at each image $dz_i(\alpha')/d\alpha'$ were obtained by fitting a cubic spline along the string. Therefore, the path free energy was obtained as a sum over the product of the mean force at each image and the derivative of the path at that point. This free energy calculation method and associated errors are discussed in detail in Ref (8).

To evaluate the quality of the string pathways and make sure the pathways have converged, we also projected the string pathways onto a two-dimensional free energy map of Src family Hck kinase generated by clustering analysis of a large amount of unbiased MD trajectories (9). The difference in the number of native contacts with respect to active and inactive states $\Delta Q = Q^A - Q^I$ in the C-helix and A-loop regions were chosen as the two-dimensional reaction coordinates. Details of the two-dimensional free energy map and clustering analysis are explained in Ref (9).

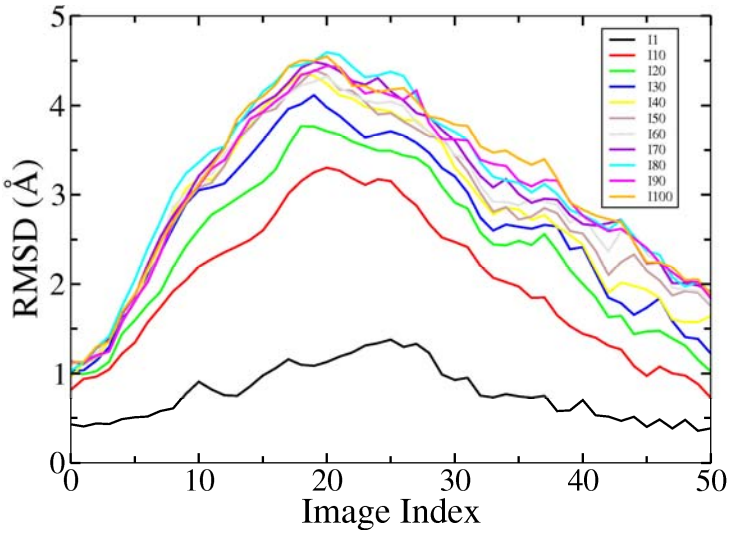
4. Convergence of the string method

To assess the convergence of the pathways obtained from the string method, we examined both the images and free energy profiles along the pathways from different iterations. As shown in Figure S1.A, the RMSD of each image with respect to corresponding image in the initial TMD pathway does not increase after about 50 iterations, indicating that the path has converged afterward. The final converged path was very different from the initial TMD pathway, with an average RMSD of 3 Å and maximum RMSD of 4.5 Å for the 50 selected atoms in table 1 (Figure S1.B). Interestingly, the image with the biggest RMSD corresponds to the metastable intermediate state identified from the one-dimensional free energy profile (image 21). This is understandable because the intermediate state image must have undergone substantial conformational relaxation as it searches for the local minimum. Besides, the paths after 50 iterations projected onto the two-dimensional free energy map of Hck kinase (Figure 3A) look quite similar as they are all contained in a narrow "transition tube". This provide further evidence that the path is indeed well converged after about 50 iterations. One may note that some fluctuations of the path within a "reaction tube" remain even after convergence due to incomplete statistical sampling from the swarms of trajectories. This tube bares some similarity, though is not strictly equivalent, to that obtained via the finite-temperature string method (10-11).

Table S1. Atoms selected to characterize the inactive-to-active transition of Src kinase catalytic domain

CHARMM Atom Type	Residue
CA	Gly300 – Val313, Asp404 – Phe424
CG of Asp, Asn	Asp404, Asp413, Asn414,
CD of Glu, Gln	Glu305, Gln309, Glu310, Gln312, Glu412, Glu415, Gln420
CZ of Arg, Tyr	Arg409, Tyr416, Arg419,
NZ of Lys	Lys295, Lys423

A.



B.

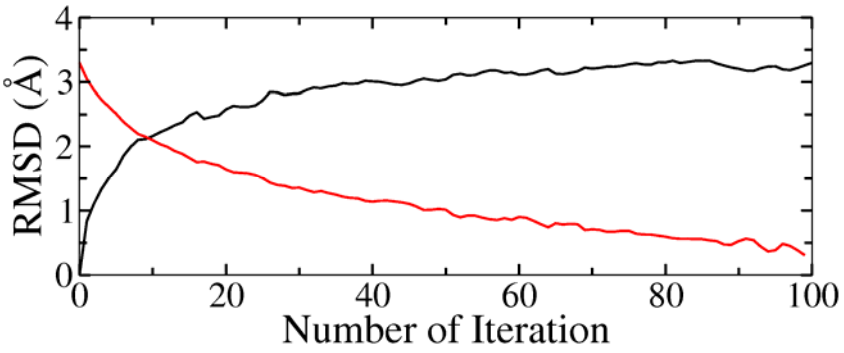


Figure S1. A. The RMSD of each image with reference to the corresponding image in the initial TMD pathway, showing the string gradually converges after about 50 iterations. B. The average RMSD of the total 51 images with reference to the initial TMD pathway (black) and with reference to the final pathway of iteration 100 (red) as a function of iteration number, showing a plateau after about 50 to 60 iterations.

Reference:

1. Zheng, J., E.A. Trafny, D.R. Knighton, N.H. Xuong, S.S. Taylor, L.F. Ten Eyck, and J.M. Sowadski. 1993. 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr D Biol Crystallogr* 49:362-365.
2. Brooks, B.R., R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* 4:187-217.
3. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926-935.
4. Darden, T., D. York, and L. Pedersen. 1993. Particle Mesh Ewald: an N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089-10092.
5. Schlitter, J., M. Engels, and P. Kruger. 1994. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J Mol Graph* 12:84-89.
6. Pan, A.C., D. Sezer, and B. Roux. 2008. Finding transition pathways using the string method with swarms of trajectories. *J Phys Chem B* 112:3432-3440.
7. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781-1802.
8. Maragliano, L., A. Fischer, E. Vanden-Eijnden, and G. Ciccotti. 2006. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* 125:24106.
9. Yang, S., N.K. Banavali, and B. Roux. 2009. Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. *Proc Natl Acad Sci U S A.* 106:3776-3781
10. Ren, W., E. Vanden-Eijnden, P. Maragakis, and W. E. 2005. Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J Chem Phys* 123:134109.
11. E, W., W. Ren, and E. Vanden-Eijnden. 2005. Finite temperature string method for the study of rare events. *J Phys Chem B* 109:6688-6693.

The tarfile “script.tgz” is a compressed tarfile of the directory script, which provides the details of NAMD and VMD/Tcl scripts for implementing the string method with swarm-of-trajectories. This directory contains the main script “string.sh” and the perl script “repara.per” to perform the calculations described in this work. There are 3 directories containing all the files.

./script

Main script files

string.sh	shell script to call scripts in other directory and do the averaging and reparametrization
repara.pl	perl script to reparametrize the string

./script/NAMD directory

general input files

equil.inp	NAMD input file to run constrained equilibration
swarm.inp	NAMD input file to run swarm of trajectories
mini.inp	NAMD input file to run a few steps of minimization

../script/VMD/Tcl directory

trj.tcl	VMD/Tcl script to get the snapshots from DCD files, orient every snapshot to fit a reference PDB
ave.tcl	VMD/Tcl script to average the swarm
rmsd.tcl	VMD/Tcl script to calculate the distance of the string before reparametrization
rmsd2.tcl	VMD/Tcl script to calculate the distance of the string after reparametrization

./script/Data directory

general data files for NAMD and VMD

src-namd.psf	NAMD format psf file of the whole system (protein, ATP, water, ions)
src-p-namd.psf	NAMD format psf file of the protein
src_ref.pdb	reference PDB file used in orientation to remove protein translation/rotation
water.pdb	PDB file used to add ATP, ions and water to protein