SUPPORTING INFORMATION

A Computational Study of the "DFG-flip" Conformational Transition in c-Abl and c-Src Tyrosine Kinases

Yilin Meng, Yen-lin Lin, and Benoît Roux*

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

*Corresponding author
e-mail: roux@uchicago.edu
929 E 57th Street, Room W323B
Chicago, Il, 60637, USA
Telephone:1-773-834-3557
Fax: 1-773-702-0439

METHODS

String Method with Swarms-of-Trajectory. String method is a computational technique intended to find out the MFEP connecting two stable conformations as well as its associated PMF in a space defined by a set of "collective variables" ($z=(z_1, z_2, \ldots)$). A pathway (string) is represented by an ordered set of images $z(\alpha)$, parametrized by α, where $\alpha=0$ is the starting image and $\alpha=1$ is the ending image. Essentially, a string can be viewed as a curve in the collective variable space. A recent variant of the string method, which is named "string method with swarms of trajectories, is utilized in our study. By employing the string method with swarms of trajectories, (1) a string is prepared, (2) each image in the string is equilibrated with restrained MD simulation, (3) a swarm of short unbiased MD trajectories are launched for each image, (4) the average displacement from each swarm of trajectories is calculated and utilized to update the image in collective variable space, and (5) the string is smoothed and re-parametrized to ensure that images are equally distant. The above 5-step procedure is iterated until the MFEP is found. A detailed description of the algorithm of the string method with swarms of trajectories is presented by Pan et al. Once the iteration of a string is completed, the PMF (denoted as F here) along that string can be computed using mean force or Markovian milstoning calculations.

Mean Force Calculations. The potential of mean force (PMF) along the MFEP were firstly computed by mean force calculations. Once the iteration of a string is completed, the PMF (denoted as F here) along that string can be computed as follows:

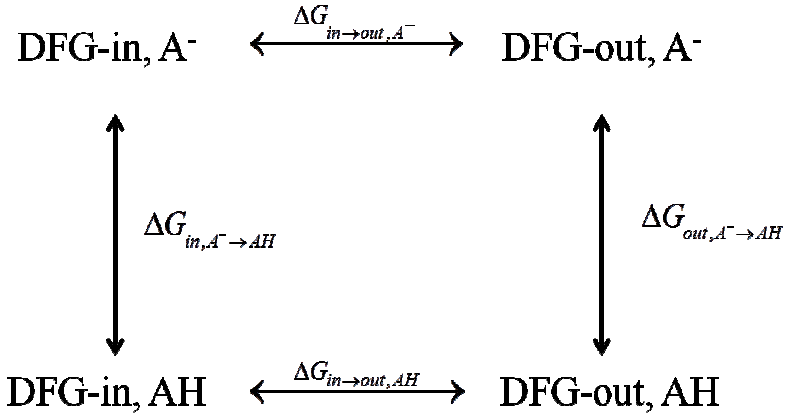
$$F\big(z(\alpha)\big) - F\big(z(0)\big) = \int_0^{\alpha} \frac{dF}{d\alpha'} d\alpha' = \int_0^{\alpha} \sum_{i-1}^{N} \frac{dz_i}{d\alpha'} \frac{\partial F(z(\alpha'))}{\partial z_i} d\alpha'$$

where N is the number of collective variables, z_i denotes the value of the *i*th collective variable, $dz_i/d\alpha'$ represents the curvature at image α', and $dF(z(\alpha'))/dz_i$ is the mean force on collective variable z_i, respectively. As demonstrated by Maragliano et al, the mean force on a collective variable z_i can be estimated from restrained MD simulations. A harmonic biasing potential is used in our study and the mean force is given by:

$$\frac{\partial F(z(\alpha'))}{\partial z_i} = k(z_i - \langle z_i \rangle)$$

where k is the force constant of the harmonic potential and $\langle z_i \rangle$ is the average value of z_i from the simulation. Strings from iteration 49 (iteration index started from 0) of c-Src and pathway #1 of c-Abl were used to represent the MFEPs. A biasing potential with force constant of 500 kcal/(mol*rad^2) and 1 kcal/(mol*Å^2) were applied in the mean force calculations of c-Abl and c-Src, respectively. Each mean force calculation lasted 1 ns in length. The force constant and simulation length were typical for mean force calculations in string method. To examine the robustness of the PMF obtained from mean force calculations, another PMF was computed along iteration 39 of pathway #1 of c-Abl.

Four-state equilibrium and population of the DFG-out conformation in c-Abl. Using two protonation states and two conformational state, an equilibrium among the four species can be described as follows:



Suppose the population of all species is normalized to unity, the following equations can be estabilished:

$$p(DFG-in,A^-)+p(DFG-in,AH)+p(DFG-out,A^-)+p(DFG-out,AH)=1$$

$$\Delta G_{in\to out,A^-}=k_BT\ln(p(DFG-in,A^-)/p(DFG-out,A^-))$$

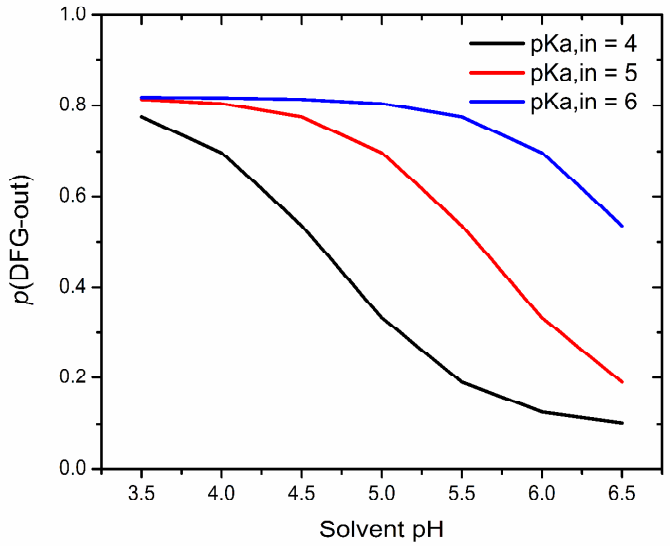
$$\Delta G_{in\to out,AH}=k_BT\ln(p(DFG-in,AH)/p(DFG-out,AH))$$

$$\Delta G_{in,A^-\to AH}=k_BT\ln(p(DFG-in,A^-)/p(DFG-in,AH))=2.303k_BT(pH-pK_{a,in})$$

Therefore, using $\Delta G_{in\to out,A^-}$ = 1.4 kcal/mol and $\Delta G_{in\to out,AH}$ = -0.9 kcal/mol, p(DFG-out) which is the sum of p(DFG-out,A$^-$) and p(DFG-out,AH) can be written as:

$$p(DFG-out)=(\exp[2.303\cdot(pH-pK_{a,in})]+45)/(11\cdot\exp[2.303\cdot(pH-pK_{a,in})]+55)$$

Assuming p$K_{a,in}$ is an adjustable parameter, the population of DFG-out conformation as a function of pH at selected p$K_{a,in}$ values are given below:



CV index	Definition of CV
CV1	C_β(Ala380)−C_α(Ala380)−C_α(Asp381)−Cγ(Asp381)
CV2	C_β(Ala380)−C_α(Ala380)−C_α(Phe382)−Cγ(Phe382)
CV3	Cγ(Asp381)−C_α(Asp381)−C_α(Phe382)−Cγ(Phe382)

Table S1: Collective variables used in the study of c-Abl by string method with swarms of trajectories. c-Abl numbering is utilized in this table. The three residues are conserved in c-Src, corresponding to Ala403, Asp404, and Phe405.

Pathway Index	Rotating Direction		Method to Generate Initial Path
	Asp381	Phe382	
#1	CW	CCW	TMD
#2	CCW	CCW	Direct pulling in dihedral angle space
#3	CW	CW	Direct pulling in dihedral angle space
#4	CCW	CW	TMD

Table S2: The indices of possible DFG-flip pathways in the study of c-Abl. CW: clockwise; CCW: counter-clockwise.

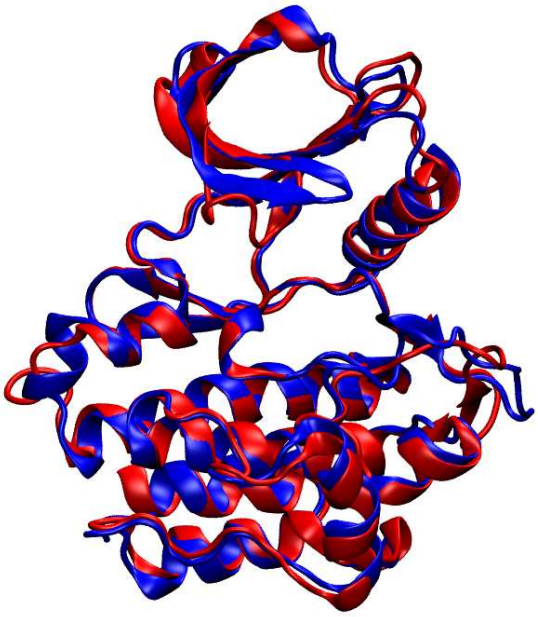


Figure S1. Structural comparison of c-Abl and c-Src kinase domains in their DFG-in (active) conformation. c-Abl (PDB code 2F4J) and c-Src (PDB code 1Y57) is colored in red and blue, respectively. This figure shows that kinase domains of c-Abl and c-Src in their active conformations are very similar, except for the P-loop region. The P-loop in c-Src (residue 273 to 282, c-Src numbering) shows an anti-parallel β-sheet conformation, whereas it (residue 248 to 257, c-Abl numbering) displays a W-shaped conformation in c-Abl.

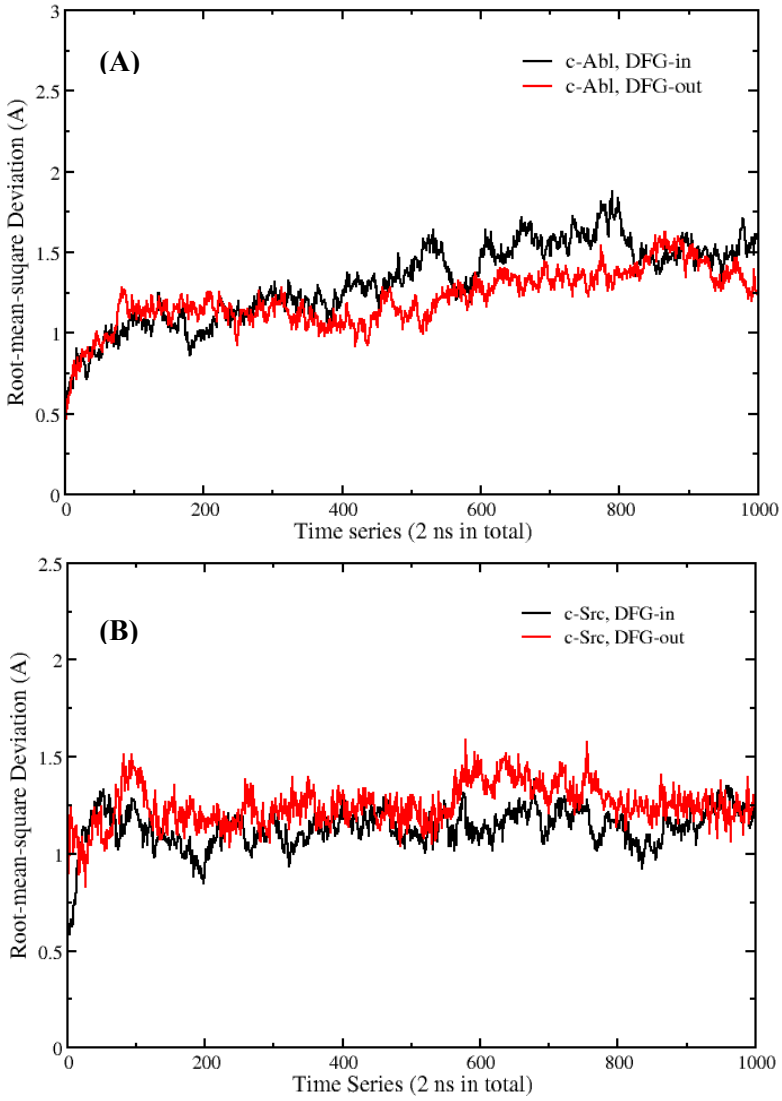


Figure S2. Root-mean-square deviation (RMSD) of C_α atoms vs crystal structure in the unrestrained 2 ns equilibration stage. (A) c-Abl. (B) c-Src.

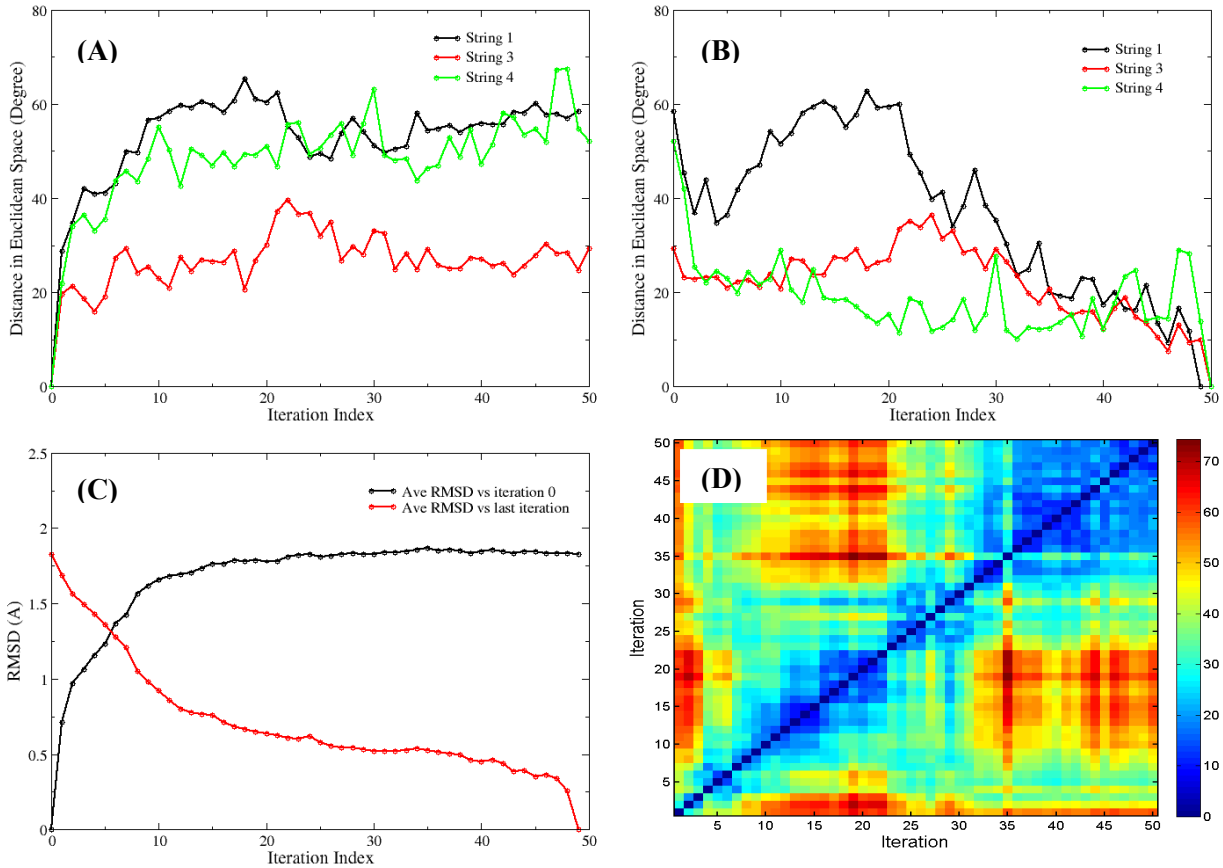


Figure S3. Convergence of string method iterations. The convergence was monitored by examining the distance of an image relative to the corresponding image in the initial/last string, as a function of iteration index. Therefore, 31 distances were generated at each ineration and the average distance as a function of iteration index was reported here. (A) c-Abl, relative to initial pathway. (B) c-Abl, relative to the last pathway. For pathway #2 of c-Abl, the unbiased MD simulation in the 5-step procedures converted the non-endpoint images to either the "in" or the "out" conformation in the first iteration of string method. The same behavior persisted, even though the unbiased MD simulation was reduced to 1 ps of computational time. Later umbrella sampling calculations suggested that pathway #2 was in a very unstable region, no local MFEP could be found. Therefore, pathway #2 was discarded from the beginning. (C) c-Src, relative to initial and final pathway. (D) Euclidean distance between a pair of strings, as a function of iteration index. The distance between two strings is calculated in the same manner as described above. Small Euclidean distances between any two strings can be found among the last 15 iterations. This behavior suggests that images of a string only drift slightly during the late stage of iterations, which further validates the convergence is achieved. One should also note that a minimum free energy pathway is not a static object. A minimum free energy pathway can still fluctuate within the valley of a free energy landscape, as shown in Figure S5. The Euclidean distance is color-coded and is in the unit of degree because dihedral angles are used as CVs. The units of the y-axis are different for c-Abl and c-Src because the c-Abl pathway used dihedral angles as CVs, while the c-Src pathways used Cartesian coordinates. We chose to change the type of CV from Cartesian coordinates to dihedral angles because using dihedral angles as CV is consistent with the umbrella sampling coordinates.

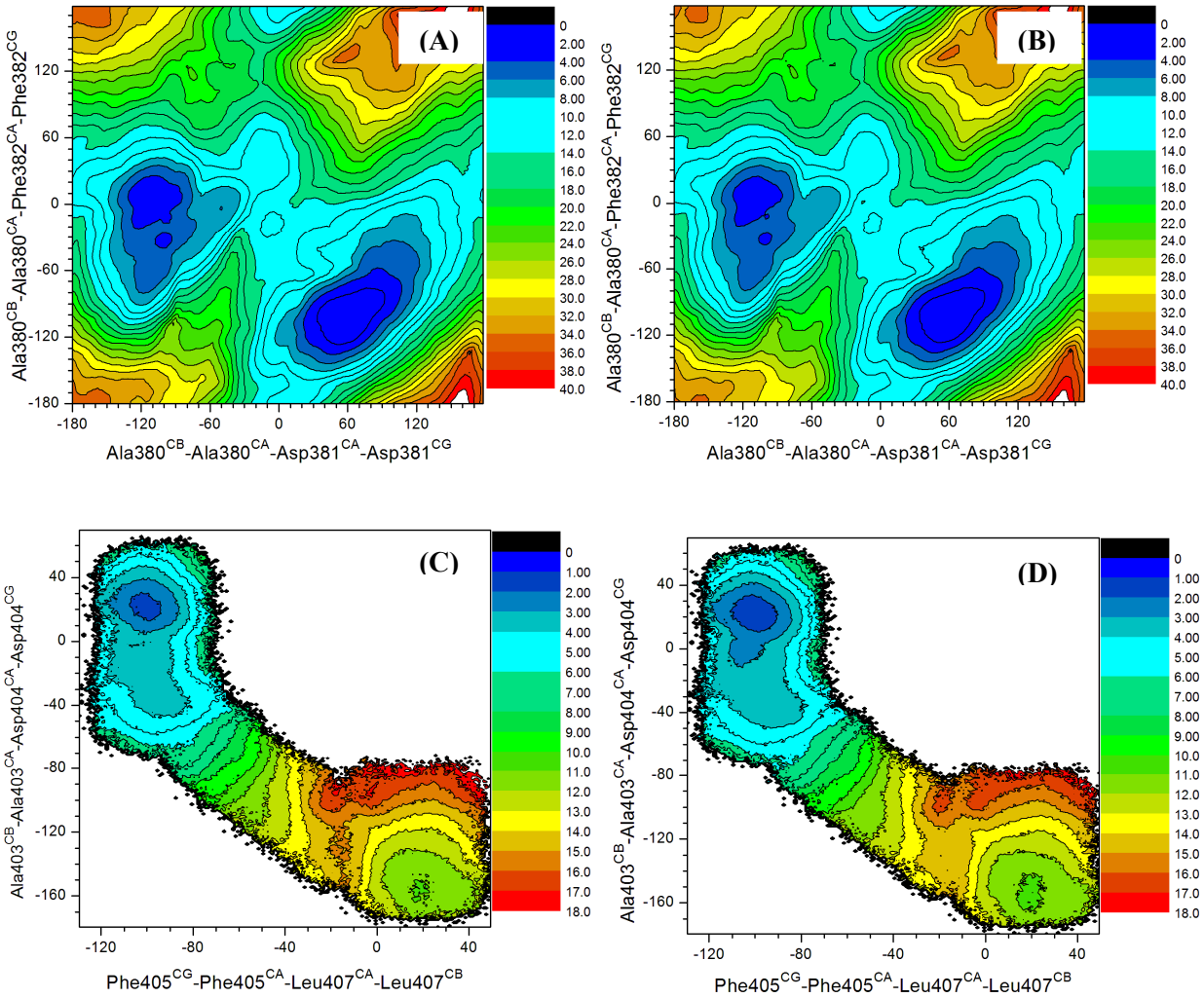


Figure S4. 2D-PMFs of DFG-flip in c-Abl and c-Src. All units of the 2D-PMFs are in k_BT, where T= 300 K. Asp381(Asp404) is deprotonated. (A) c-Abl, obtained from 3 ns of umbrella sampling simulation per window. (B) c-Abl, obtained from 4 ns of umbrella sampling per window. (C) c-Src, obtained from 3 ns of umbrella sampling simulation per window. (D) c-Src, obtained from 4 ns of umbrella sampling per window.

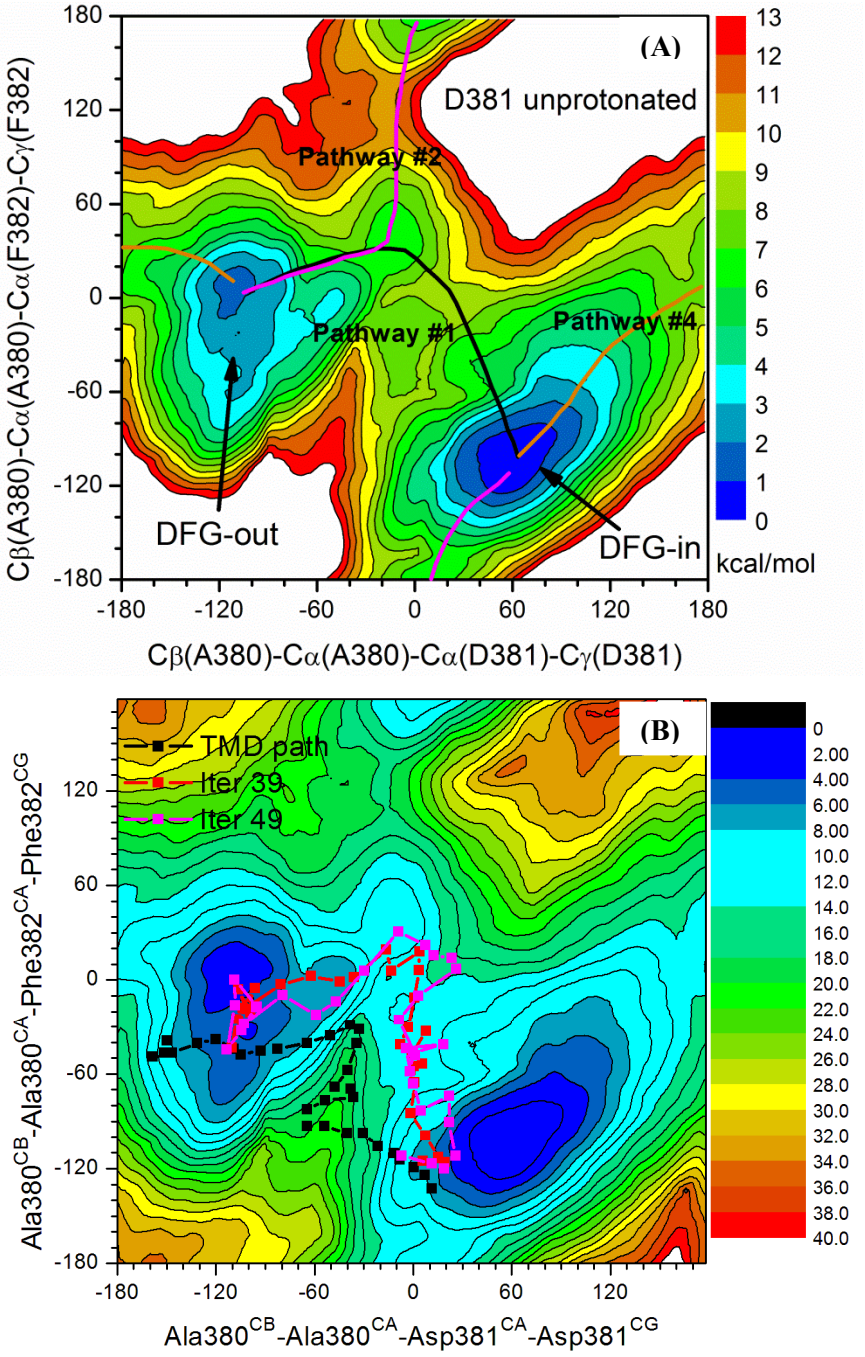


Figure S5. (A) Projected pathways on the free energy landscape of c-Abl. The unit of free energy is in kcal/mol. (B) Projection of strings at three iterations onto the free energy landscape of c-Abl. The unit of free energy is in in k_BT, where T= 300 K. Asp381 is deprotonated in both cases.

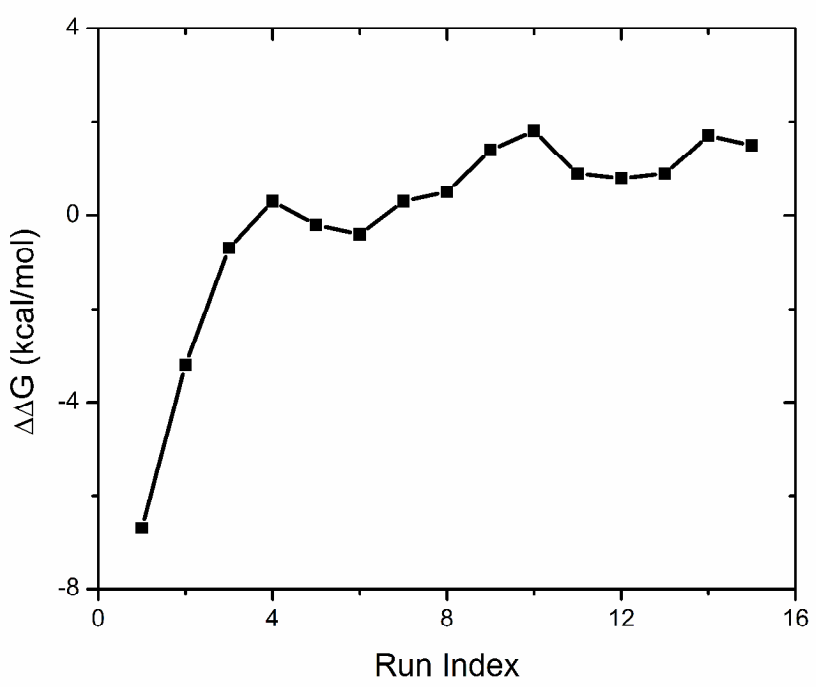


Figure S6. $\Delta\Delta G_{binding} \equiv \Delta G_{binding}(AspH) - \Delta G_{binding}(Asp)$ as a function of alchemical free energy calculation cycles. $\Delta G_{binding}$ is the binding free energy of Gleevec to the DFG-out conformation of c-Abl. "AspH" represents that Gleevec binds to c-Abl with Asp381 protonated, whereas "Asp" stands for Gleevec binding to c-Abl with Asp381 deprotonated.

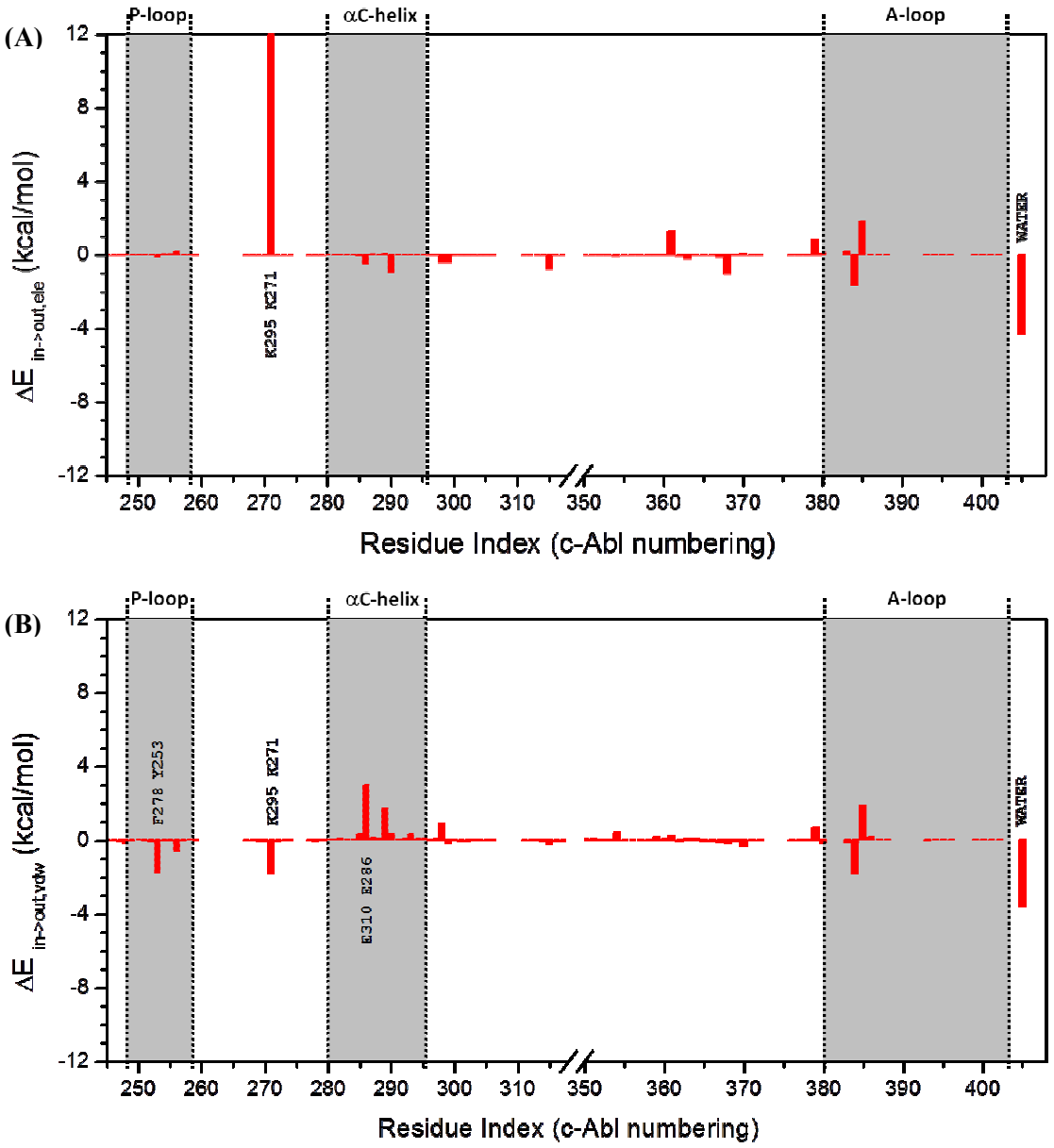


Figure S7. (A) Electrostatic component of $\Delta E_{in \rightarrow out}$ of c-Abl. (B) van der Waals component of $\Delta E_{in \rightarrow out}$ of c-Abl. Asp381 is deprotonated in the calculation of interaction energies.

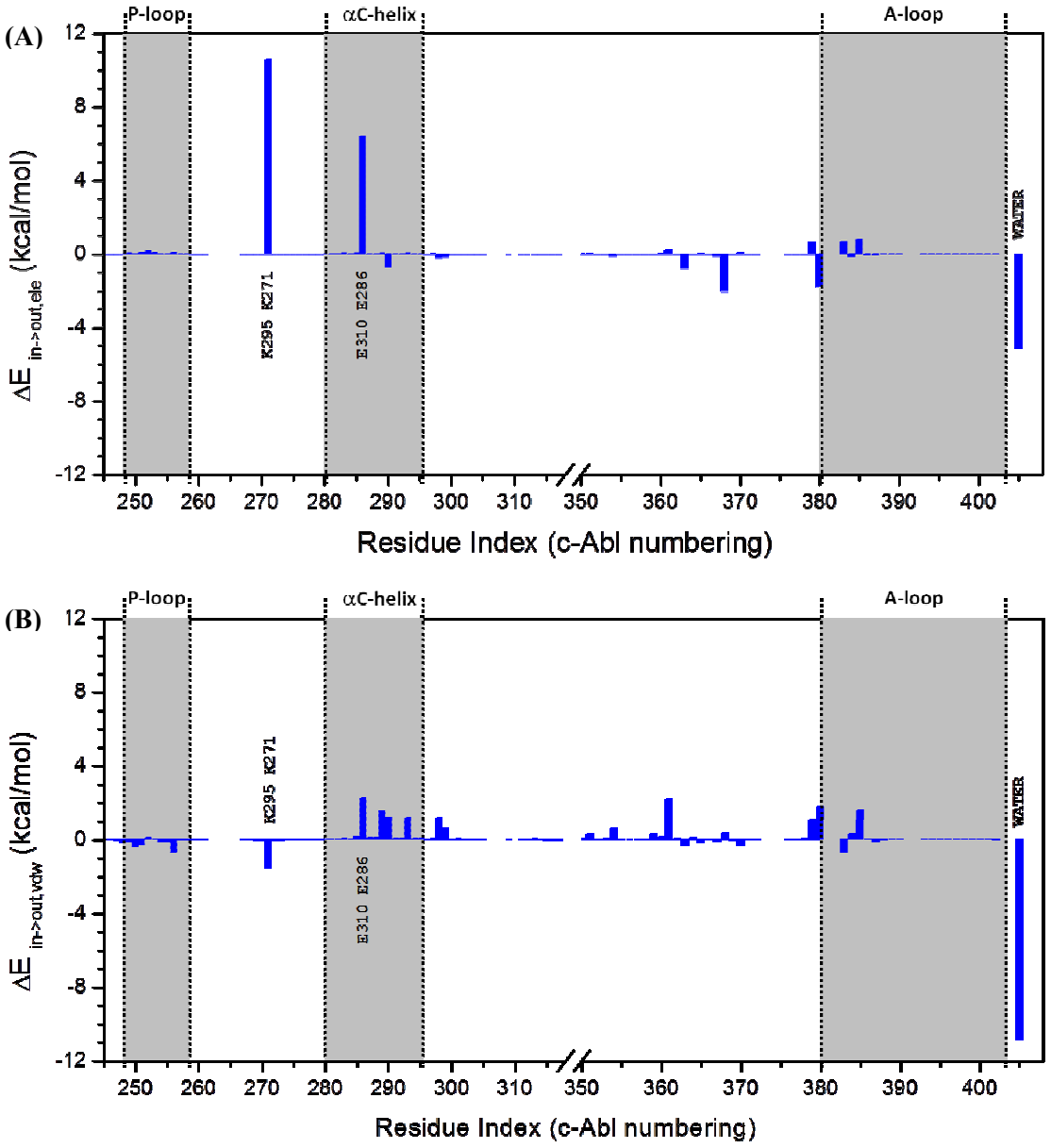


Figure S8. (A) Electrostatic component of $\Delta E_{in \rightarrow out}$ of c-Src. (B) van der Waals component of $\Delta E_{in \rightarrow out}$ of c-Src. Asp404 is deprotonated in the calculation of interaction energies.

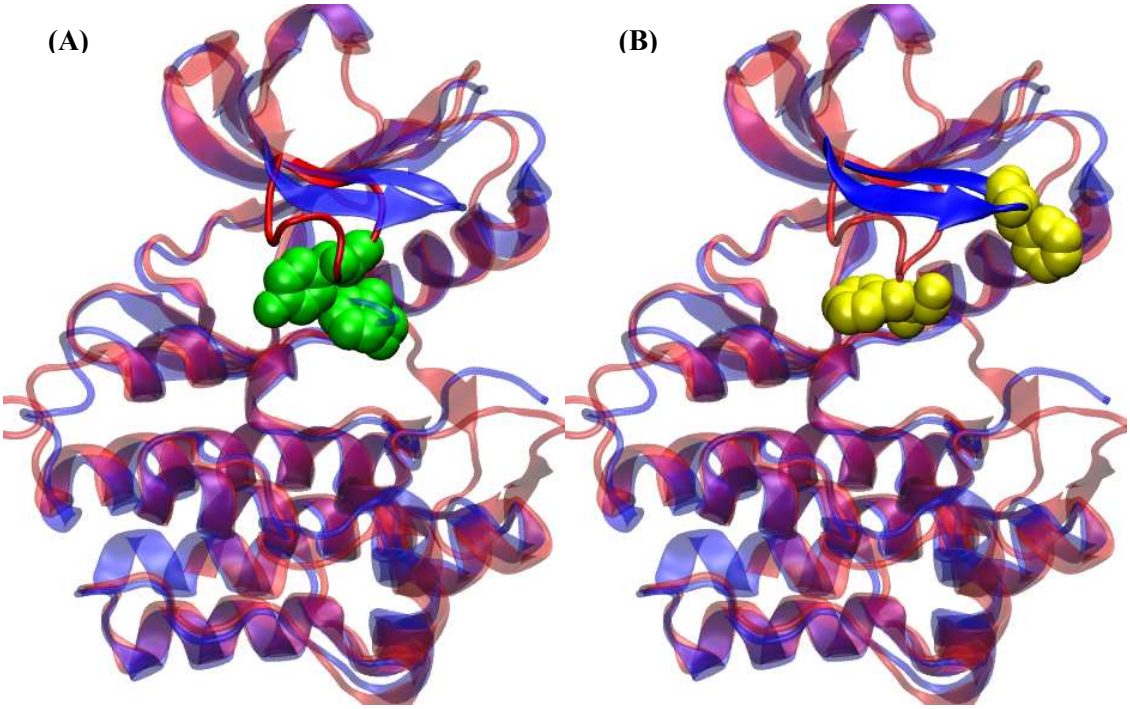


Figure S9. Structural comparison of P-loop conformations and key residue interactions. c-Abl and c-Src is colored in red and blue, respectively. (A) Tyr253 and Phe382 in c-Abl. (B) Phe278 and Phe405 in c-Src which correspond to Tyr253 and Phe382 in c-Abl.

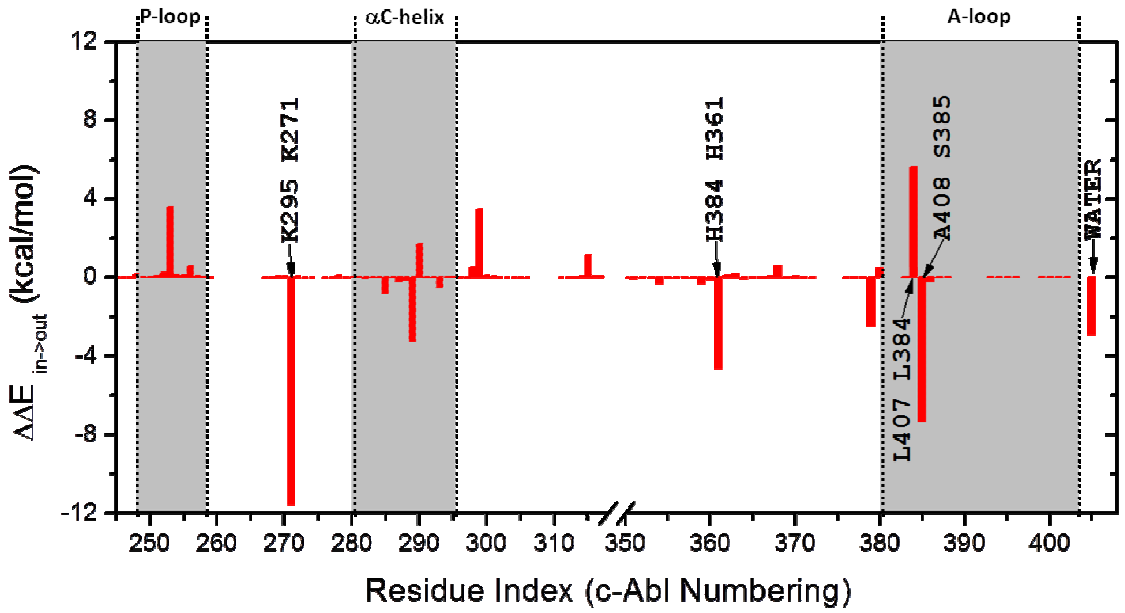


Figure S10. Difference in $\Delta E_{in \rightarrow out}$ ($\Delta\Delta E_{in \rightarrow out}$) as a function of residue index. $\Delta\Delta E_{in \rightarrow out}$ is calculated as $\Delta E_{in \rightarrow out}$(AspH) - $\Delta E_{in \rightarrow out}$(Asp), where "AspH" and "Asp" represents c-Abl with Asp381 protonated and deprotonated, respectively. $\Delta E_{in \rightarrow out}$(Asp) and $\Delta E_{in \rightarrow out}$(AspH) as a function of residue index are displayed in Figure 7 in the main text.