Supplementary Information

Dynamic, structural and thermodynamic basis of insulin-like growth factor 1 kinase allostery mediated by activation loop phosphorylation

Yaozong Li[a] and Kwangho Nam*[a,b]

[a]*Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden*

[b]*Department of Chemistry and Biochemistry, University of Texas at Arlington, Arlington, TX 76019-0065, USA. E-mail: kwangho.nam@uta.edu; Tel: +1-817-272-1091*

*****Corresponding authors*

CONTENTS (35 pages in total)

1. Simulation and analysis details page S2~S7
2. Supplementary tables page S8~S11
3. Supplementary figures page S12~S35
4. Supplementary references page S36~S37

Details of energy minimization and equilibration MD

To alleviate high energy contacts present in the prepared systems, a series of constrained and restrained energy minimizations were performed before molecular dynamics simulations. First, the model built regions of the kinase, including the missing and restored to the wild-type residues, were minimized with the rest of the structure held fixed to the starting X-ray structure position. Second, all water molecules were minimized with harmonic restraints applied to the rest of the system, followed by minimizations of Na^+ and Cl^- positions. Then, the entire system was minimized with restraints applied to the protein heavy atoms. Finally, the entire system was minimized without any restrain. The entire procedure resulted in 10,000 steps of minimizations. For the holo systems, ATP/ADP, two Mg^{2+} ions and their coordinated water molecules were restrained to their original positions during the entire minimizations with the force constant of 5 kcal/mol-$Å^2$ times the mass of each atom.

Following the minimizations, each system was equilibrated in several steps. First, we heated the system to 300 K over 20 picosecond (ps) and equilibrated for 50 ps. Then, the system was heated again to 600 K over 20 ps and equilibrated for 50 ps. Finally, the system was cooled down back to 300 K over 20 ps and equilibrated for 50 ps. During the heating and cooling cycle, the protein, ATP or ADP, Mg^{2+} and their neighboring residues were harmonically restrained. We then released the harmonic restraints gradually over 100 ps, followed by 100 ps equilibration at 300 K without any restrain. Once the completion of the NVT equilibration, the constant temperature and pressure simulations (i.e, NPT simulations) were performed for 400 ps at 300 K and 1atm.

Details of the accelerated molecular dynamics (aMD) simulations

All accelerated MD (aMD) simulations were performed with the dual boost strategy.[1,2] The boost potential used was

$$\Delta V(\mathbf{r}) = \frac{(E - V(\boldsymbol{r}))^2}{\alpha + E - V(\boldsymbol{r})}, \quad \text{(S1)}$$

where ΔV is the boost potential that was added to the original potential V, E the reference potential energy, and α the acceleration factor. In the dual boost strategy, the reference potential E and the acceleration factor α were determined individually for the dihedral angle energy and the rest of the system, according to the following empirical formulae[3]

$$E_{dihed} = V_{dihed\_avg} + 0.3 \times V_{dihed\_avg};\ \alpha_{dihed} = 0.3 \times \frac{1}{5} V_{dihed\_avg}, \quad \text{(S2)}$$

$$E_{total} = V_{total\_avg} + 0.2 \times N_{atoms};\ \alpha_{total} = 0.2 \times N_{atom}, \quad \text{(S3)}$$

where E_{dihed} and E_{total} are the reference energies applied to all dihedral angles and all atoms (N_{atoms}) of the kinase, and α_{dihed} and α_{total} are their corresponding acceleration factors. The V_{dihed_avg} and V_{total_avg} values were determined using the average dihedral and total potentials obtained from the 300 ns conventional MD (cMD). Starting from the 50 ns time frame of the corresponding cMD simulations, each aMD simulation was performed for 200 ns at 300 K and 1 atm. All other simulation options were the same to the cMD simulations.

Principal component analysis

The principal component analysis (PCA), also known as the quasi-harmonic analysis,[4] was performed using the coordinates saved during the production MD simulations (between 50 ns and 300 ns time frames). The analysis was carried out first by superimposing each saved structure against the corresponding energy minimized

structure and determining the average structure for each system. Each structure was superimposed again but this time, against the average structure to remove the center of mass translational and rotational motions, followed by determination of the covariance matrix of atomic fluctuations. Then, the determined covariance matrix was diagonalized to determine eigenvectors and eigenvalues, which correspond to the protein's vibrational modes (i.e., PCA modes) and frequencies, respectively. Among them, the three lowest frequency motions are presented in **Fig. S4**. In all superposition and construction of covariance matrix, we considered only the backbone of the rigid part of the kinase and ignored flexible regions of the protein to avoid them affecting the determined PCA modes. Specifically, the residues between 958 and 966 (the N-terminus region), 1064 and 1079 (the insert region) and 1240 and 1255 (the C-terminus region) were excluded.

Thus obtained PCA modes were analyzed in two different approaches. In the first approach, we have projected each saved coordinates onto the two lowest PCA modes (i.e., PC1 and PC2) of corresponding system (**Fig. S4**), i.e.,

$$p_i(t) = \frac{\boldsymbol{v}_i \bullet (\boldsymbol{r}_t - \langle \boldsymbol{r} \rangle)}{\|\boldsymbol{v}_i\| \bullet \|(\boldsymbol{r}_t - \langle \boldsymbol{r} \rangle)\|} \quad \text{(S4)},$$

where $p_i(t)$ is the normalized projection of the coordinates saved at time *t*, $\boldsymbol{v}_i$ either PC1 or PC2 and $(\boldsymbol{r}_t - \langle \boldsymbol{r} \rangle)$ the displacement vector of each saved coordinate from the average structure, respectively. $\|\boldsymbol{v}_i\|$ and $\|(\boldsymbol{r}_t - \langle \boldsymbol{r} \rangle)\|$ are norm of corresponding vectors. In the second appraoch, the contribution of the lowest PCA mode (PC1) to RMSF of each residue was measured by the so-called R-component analysis,[5]

$$\text{R-component}_j = \sqrt[2]{\lambda_{PC1}^2 \, \boldsymbol{v}_j^{PC1} \bullet \boldsymbol{v}_j^{PC1}} \quad \text{(S5)},$$

where λ_{PC1} is the eigenvalue of PC1 and $\boldsymbol{v}_j^{PC1}$ is the eigen vector of PC1 for residue j. The results are presented in **Fig. S2** together with the overall RMSF determined from the saved coordinates for each system.

Community network analysis

To perform community network analysis,[6] a spatially contacted network was constructed first. The network was comprised of a set of isolated nodes and edges connecting them. In the present work, each node was defined by a single residue of a protein, the site of which is represented by the Cα atom of each residue. An edge was defined between any residue pair, if any heavy atom of one residue was within 4.5 Å from any heavy atom of the other residue more than 75% of the time during the last 250 ns cMD simulations (i.e., between 50 ns ~ 300 ns),[6] and if the two residues were separated more than 2 residues on the sequence. The edges were then weighted by the probability of information transfer between them,[6] i.e., the weight factor, $w_{ji} = -log|C_{ji}|$, where C_{ji} is the pair correlation value between residues j and i.[7]

To partition the network into different dynamically connected regions (i.e., "communities"), several indicators were considered. First, the length of a path connecting given two nodes i and j (D_{ij}) was defined as the sum of the weights along the path connecting the two nodes. The shortest path is then the path with the smallest D_{ij} value, determined by the Floyd-Warshall algorithm.[8,9] Second, the "betweenness" value was assigned to each edge as the number of shortest paths that crossed the edge.[10] Then, the Girvan-Newman algorithm was applied to partition the entire network into different communities.[11] Namely, the edge with the highest betweenness was removed, and the betweenness values were re-evaluated based on the remaining edges. These two steps

were iterated until the optimized community partition was found by maximizing the indicator "modularity",[12] which is a measurement of difference in probability of intra- and inter-community edges. In the present study, the optimal modularity values were around 0.65, which were within the reasonable range of 0.4~0.7 for biological systems.[13] In the analysis, the correlation map was generated by Carma,[14] network building by NetworkView plugin[6] in VMD 1.9.1,[15] and community partition by the "gncommunities" routine,[6] respectively. The presentation of the communities was prepared by an in-house PyMOL script.

Root-mean-square deviation (RMSD) analysis of MD simulation trajectories

During the 300 ns production MD simulations, none of the systems underwent large-scale structural reorganization. Instead, they fluctuated around the corresponding reference structures (i.e., energy-minimized X-ray structures) with 1.5~3.0 Å backbone RMSD values (**Fig. S1**). Nevertheless, each system displayed distinctive dynamics, especially, in the αC-helix and A-loop regions, which varied depending on the kinase's conformation (inactive versus active), A-loop phosphorylation (0P versus 3P) and bound states (apo versus holo).

The most pronounced difference was observed in the inactive conformation systems. Their backbone RMSDs fluctuated by around 2.5 Å or larger irrespective of A-loop phosphorylation level (**Fig. S1**a). In contrast, in the active conformation systems, the backbone RMSDs stayed at around 2 Å and fluctuated only slightly (**Fig. S1**c). The αC-helix and A-loop of the two inactive conformation systems also exhibited larger RMSDs and fluctuations more than those of their corresponding active conformation systems.

These differences suggest that the inactive conformation has intrinsically larger flexibility than the active conformation.

RMSD comparison of the three A-loop tyrosines suggested that the inactive conformation A-loop was destabilized by phosphorylation, whereas that of the active conformation was stabilized (**Fig. S3**). In particular, Tyr1131 and Tyr1136 of the inactive conformation kinase exhibited substantially enhanced RMSDs after their phosphorylation (**Fig. S3**a and c), suggesting local structural rearrangement. In the active conformation systems, the opposite trend was observed. Namely, phosphorylation suppressed the deviation and fluctuation of the three A-loop tyrosine orientations (**Fig. S3**d-f).

Table S1. Summary of simulated systems, their states and notations.

Notation	Kinase conformation	A-loop phosphorylation[a]	Ligand occupancy	State
I_{Apo}^{0P}	Inactive	0P	None	apo
I_{Apo}^{3P}	Inactive	3P	None	apo
A_{Apo}^{0P}	Active	0P	None	apo
A_{Apo}^{3P}	Active	3P	None	apo
A_{RS}^{0P}	Active	0P	ATP, peptide, and 2 Mg^{2+}	holo reactant
A_{RS}^{3P}	Active	3P	ATP, peptide, and 2 Mg^{2+}	holo reactant
A_{PS}^{0P}	Active	0P	ADP, phospho-peptide,[b] and 2 Mg^{2+}	holo product
A_{PS}^{3P}	Active	3P	ADP, phospho-peptide,[b] and 2 Mg^{2+}	holo product

[a]0P represent the unphosphorylated A-loop and 3P the fully phosphorylated A-loop, respectively.
[b]Phospho-peptide indicates the peptide, whose tyrosine residue is phosphorylated.

Table S2. Negatively charged Asp/Glu residues that interact with each of the inactive conformation A-loop tyrosine in different kinases presented in **Fig. S16**.

Name	A-loop tyrosine	Negatively charged residue	PDB ID
IGF1R	Tyr1131, Tyr1135, Tyr1136	Asp1056, Asp1105, Glu1189	1P4O[16]
IRK	Tyr1158, Tyr1162, Tyr1163	Asp1083, Asp1132, Glu1216	1IRK[17]
TrkA	Tyr676, Tyr680, Tyr681	Asp596, Asp650	4GT5[18]
TrkB	Tyr718, Tyr722, Tyr723	Asp640, Asp692, Glu780	4AT4[19]
Ror2	Tyr641, Tyr645, Tyr646	Asp560, Glu563, Asp615, Glu703	4GT4[18]
MuSK	Tyr750, Tyr754, Tyr755	Asp661, Asp724, Glu808	1LUF[20]
DDR1	Tyr792, Tyr796, Tyr797	Asp708, Asp766, Glu851	3ZOS[21]
KIT	Tyr823	Asp792	1T45[22]
Abl	Tyr421	Asp382	1OPJ[23]
Src	Tyr461	Asp386	2SRC[24]

Table S3. Simulated systems and their corresponding simulation type and length.

System	Simulation type[a]	Length (ns)
I_{Apo}^{0P}	cMD	300
I_{Apo}^{3P}	cMD	300
A_{Apo}^{0P}	cMD	300
A_{Apo}^{3P}	cMD	300
A_{RS}^{0P}	cMD	300
A_{RS}^{3P}	cMD	300
A_{PS}^{0P}	cMD	300
A_{PS}^{3P}	cMD	300
I_{Apo}^{0P}	aMD	200
I_{Apo}^{3P}	aMD	200
A_{Apo}^{0P}	aMD	200
A_{Apo}^{3P}	aMD	200
A_{RS}^{0P}	aMD	200
A_{RS}^{3P}	aMD	200
A_{PS}^{0P}	aMD	200
A_{PS}^{3P}	aMD	200
$I_{Apo,\ mutant}^{0P}$	cMD	200
$A_{RS,mutant}^{3P}$	cMD	200
$A_{PS,mutant}^{3P}$	cMD	200
$I_{Apo}^{0P\leftrightarrow3P}$	alchemical	66
$A_{Apo}^{0P\leftrightarrow3P}$	alchemical	66
$A_{RS}^{0P\leftrightarrow3P}$	alchemical	66
$A_{PS}^{0P\leftrightarrow3P}$	alchemical	66

[a]"cMD" refers the conventional MD, "aMD" the accelerated MD, "alchemical" the alchemical transformation simulation, respectively.

Table S4. Free energy difference before and after the A-loop phosphorylation from alchemical transformation simulations.

Free energy / Transformation	$\Delta F_{unchanging}$[a]	ΔF_{vdW}[b]	$\Delta F_{charging}$[c]	ΔF_{tol}[d]	$\Delta\Delta F$[e]
$\mathbf{I_{Apo}^{0P\rightarrow 3P}}$	66.9	12.5	-693.0	-613.6 $\pm$ 0.6 (-613.3 $\pm$ 0.6)	12.7 $\pm$ 0.6
$\mathbf{A_{Apo}^{0P\rightarrow 3P}}$	66.6	12.9	-706.4	-626.9 $\pm$ 0.6 (-626.1 $\pm$ 0.6)	-0.5 $\pm$ 0.6
$\mathbf{A_{RS}^{0P\rightarrow 3P}}$	66.8	13.8	-708.6	-628.0 $\pm$ 0.6 (-627.5 $\pm$ 0.6)	-1.6 $\pm$ 0.6
$\mathbf{A_{PS}^{0P\rightarrow 3P}}$	67.2	13.3	-705.1	-624.6 ± 0.6 (-625.1 $\pm$ 0.6)	1.7 $\pm$ 0.6
Tyr → pTyr[f]	22.5	5.2	-236.5	-208.8 ± 0.3	--

[a]The free energy change for removing the partial atomic charges of each A-loop tyrosine.
[b]The free energy change for transforming the van der Waals (vdW) interactions of the uncharged tyrosines to those of the uncharged phosphotyrosines (pTyrs).
[c]The free energy charge for charging each pTyr.
[d]The total free energy change, $\Delta F_{tot} = \Delta F_{uncharging,} + \Delta F_{vdW} + \Delta F_{charging}$. The numbers in parenthesis were the free energy values determined by extending each λ simulation to 2 ns to check the convergence of the simulated alchemical transformation. All values showed that they are within the error ranges for each transformation.
[e]The free energy change at each conformational state relative to the free energy change for phosphorylating 3 free tyrosines in water. Each value was determined as a difference relative to the 3 times of the Tyr→pTyr transformation free energy.
[f]The free energy change for transforming a tyrosine to phosphotyrosine in water.

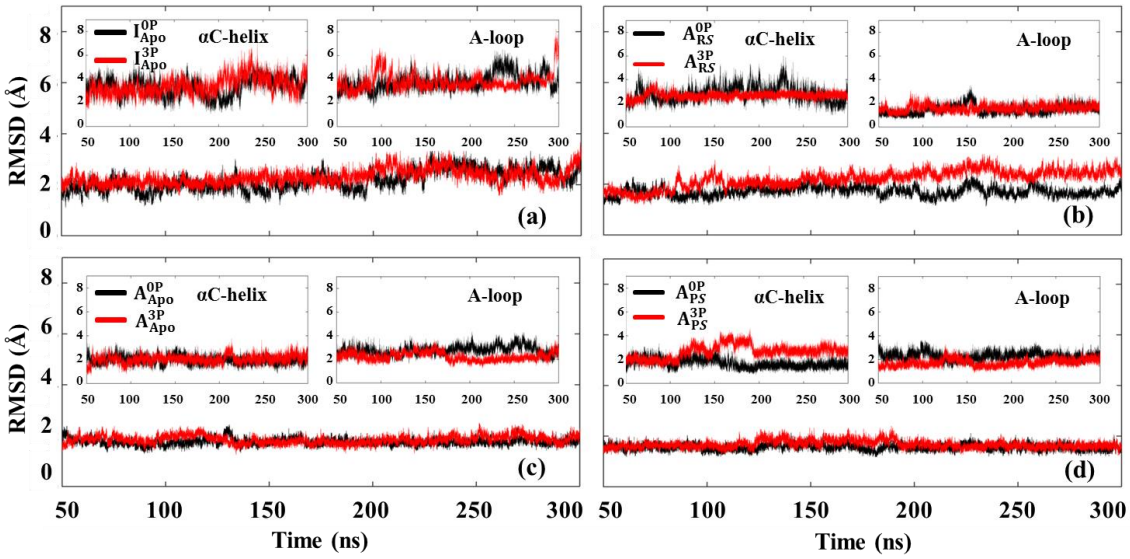


Fig. S1 Structural deviation and fluctuation comparisons between the phosphorylated (red) and unphosphorylated kinases (black): a, b, c and d are the RMSD plots for apo inactive conformation systems, reactant state systems, apo active conformation systems and product state systems, respectively. In each figure, the RMSDs of the entire kinase backbone are plotted at the bottom. The RMSDs of the αC-helix and A-loop backbones are plotted at the upper left and upper right of the plots, respectively. The RMSD values were calculated by taking the energy minimized crystal structures as a reference with flexible protein regions (i.e., the N-terminal and C-terminal tails and insert part) excluded.

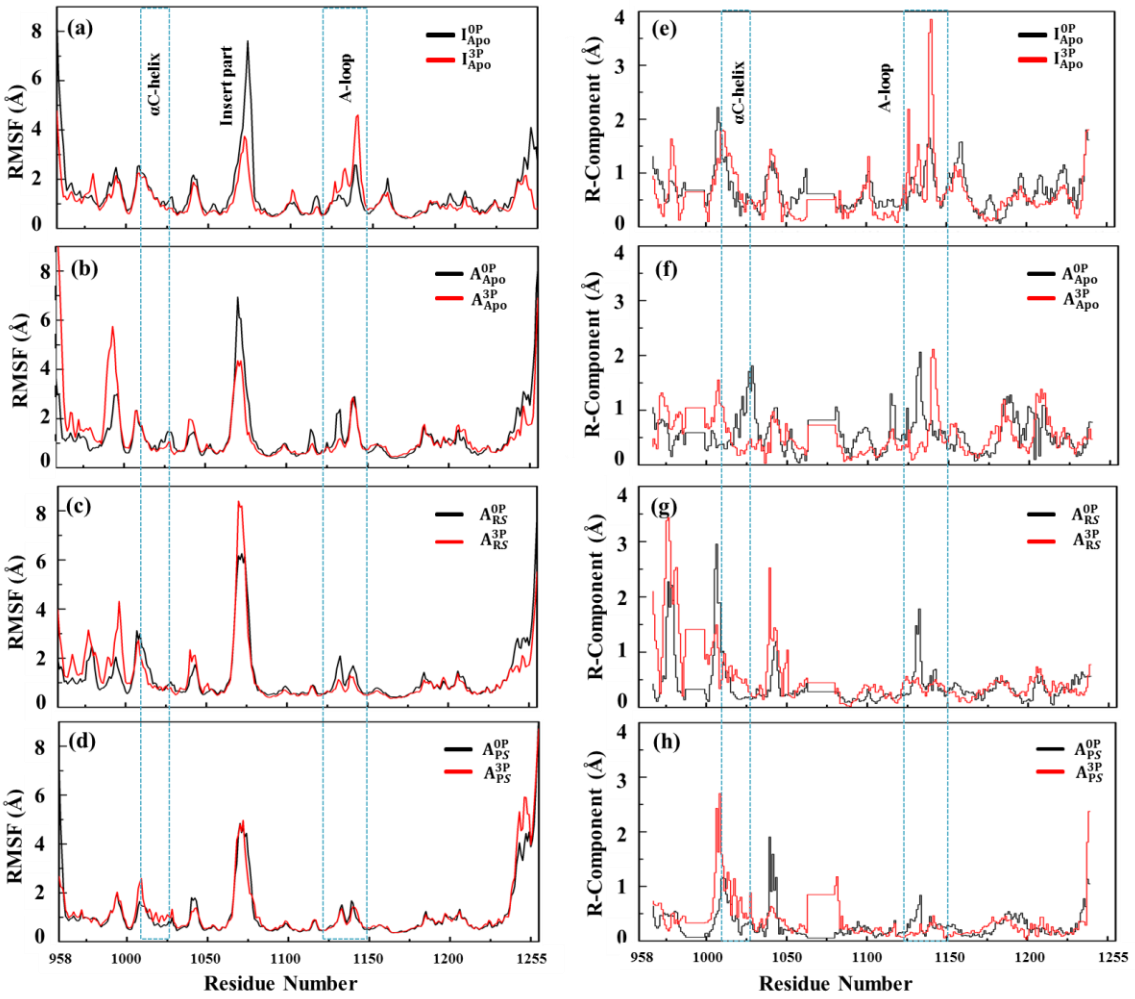


Fig. S2 RMSF comparison between the phosphorylated (red) and unphosphorylated kinases (black): (a) ~ (d) are the RMSF plots for the apo inactive conformation, apo active conformation, reactant state and product state systems, respectively. Figure (e) ~ (h) are the contribution of the lowest frequency PCA vector to the overall RMSF (i.e., R-component) shown in (a) ~ (d), respectively.

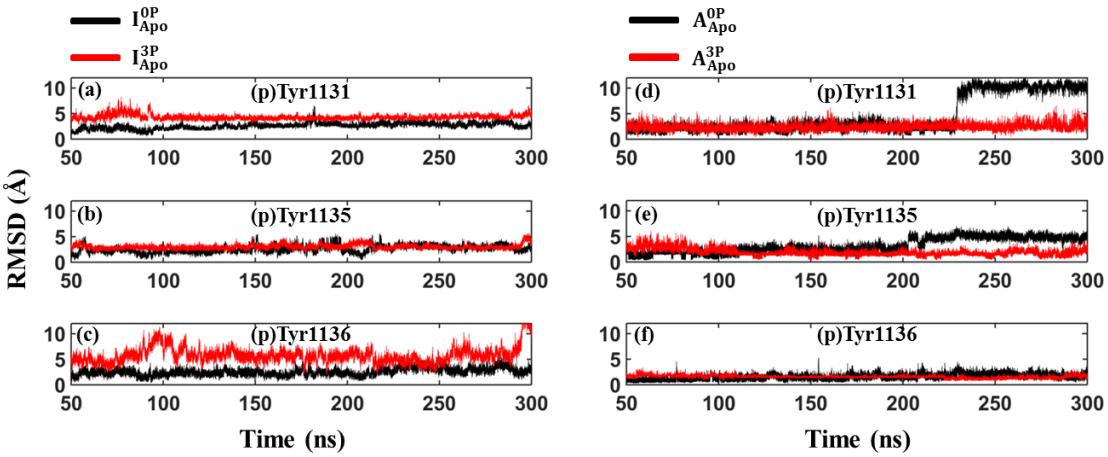


Fig. S3 RMSDs of the A-loop tyrosine residues. The inactive conformation systems are shown in (a) to (c), and the apo active conformation systems in (d) to (f), respectively. The RMSD values of the unphosphorylated and phosphorylated systems are shown by the black and red lines, respectively. The reference structures for calculating RMSD were the corresponding minimized crystal structures.

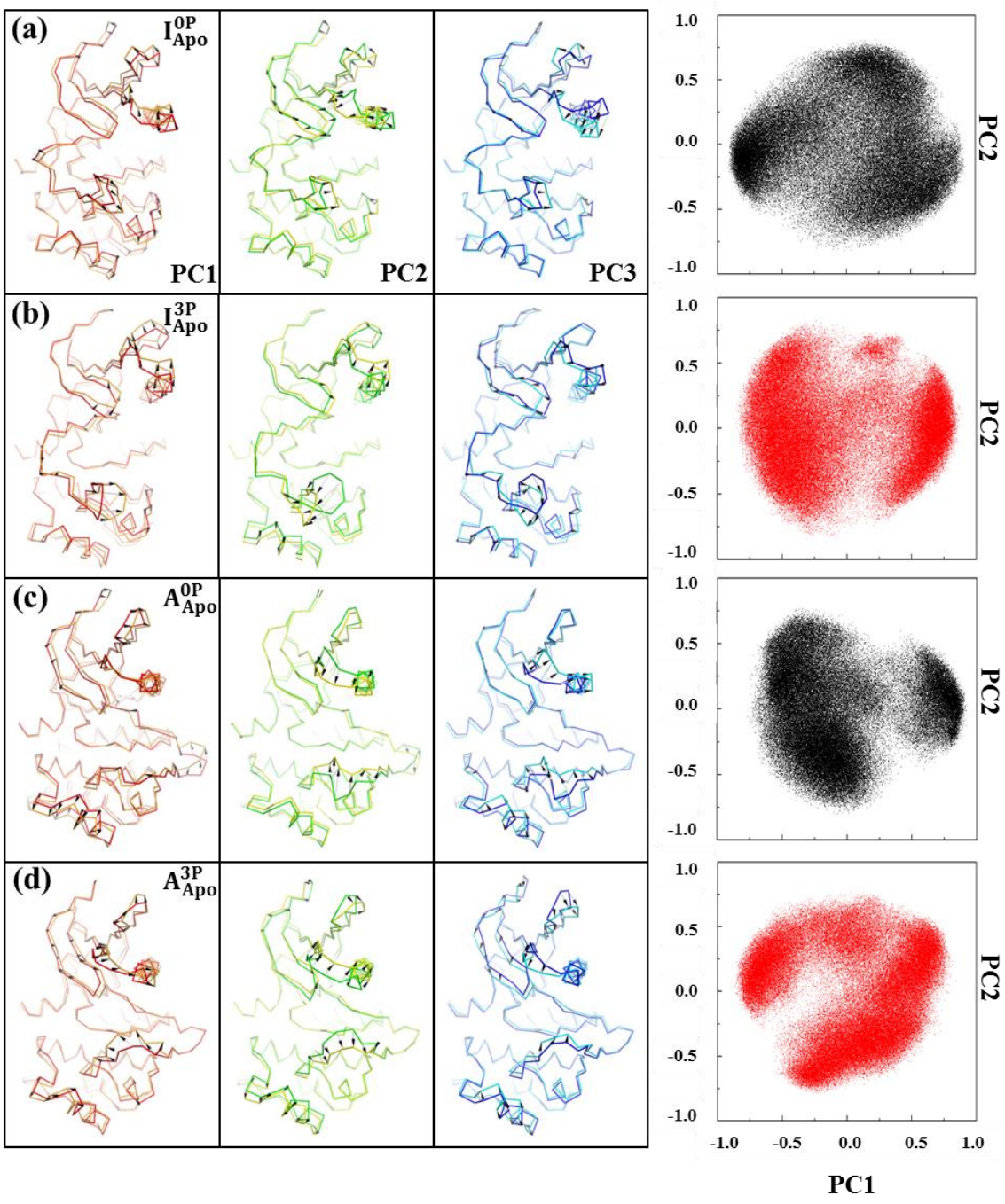
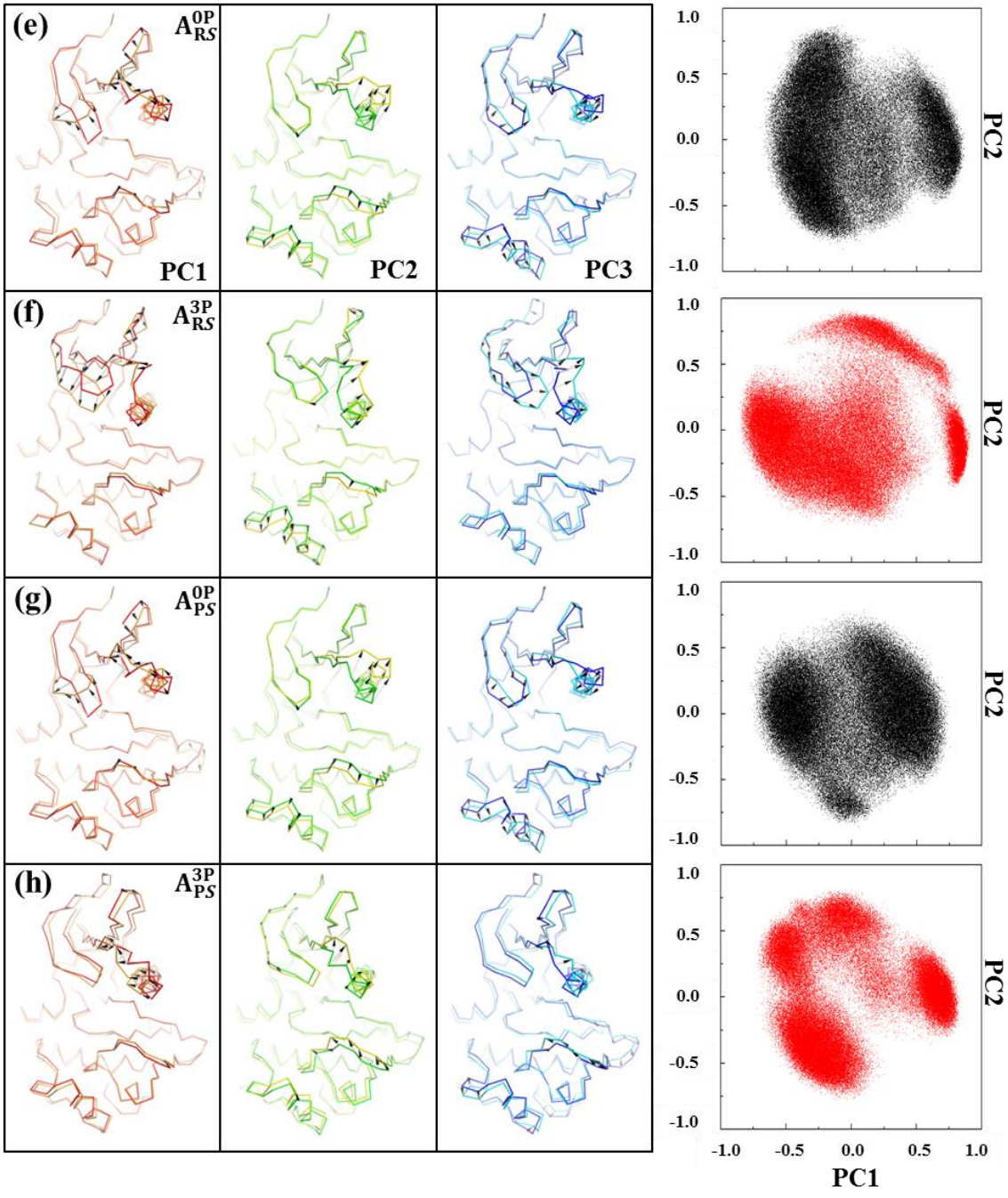
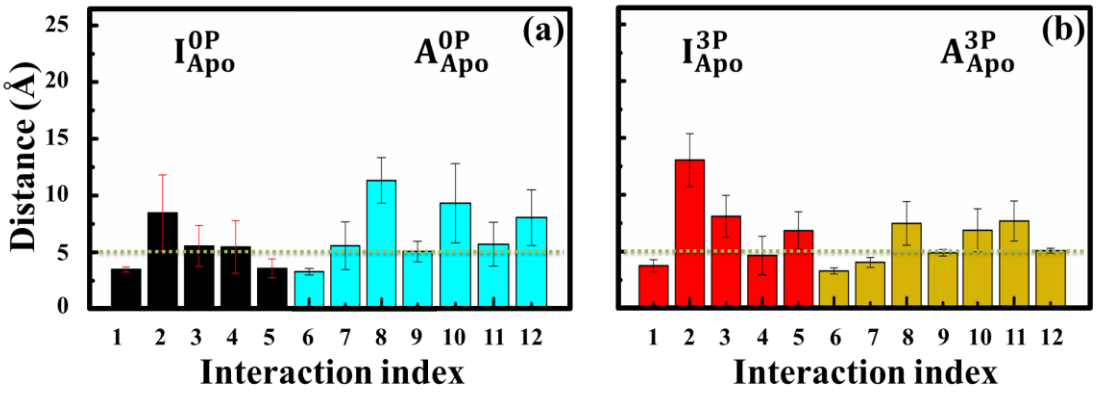


Fig. S4 PCA analysis. (a) ~ (h) present the three lowest-frequency PCA modes of IGF-1RK in different functional and conformational states. For each state, modes 1 to 3 are arranged in ascending order by mode frequency. For each mode, two structures were generated from the maximum and minimum amplitude positions, and the displacement vectors between the two structures were determined for selected residues to be shown by arrows in each figure. In the right panel, the projection of each saved coordinates to the first two PCA modes are shown for each step.



Continued Fig. S4



(c)

Interaction index	Residue 1	Residue 2
1	Lys1003 (NZ)	Asp1123 (CG)
2	Arg1104 (CZ)	Glu1020 (CD)
3	Arg1128 (CZ)	Asp1056 (CG)
4	Arg1012 (CZ)	Glu1016 (CD)
5	Lys1058 (NZ)	Asp1134 (CG)

Interaction index	Residue 1	Residue 2
6	Lys1003 (NZ)	Glu1020 (CD)
7	Arg1128 (CZ)	Glu1016 (CD)
8	Arg1012 (CZ)	(p)Tyr1136 (OH)
9	Arg1104 (CZ)	(p)Tyr1136 (OH)
10	Lys1100 (NZ)	(p)Tyr1131 (OH)
11	Lys1100 (NZ)	Asp1134 (CG)
12	Arg1137 (CZ)	(p)Tyr1135 (OH)

Fig. S5 Average interaction distances between the selected residue pairs (a) for the unphosphorylated inactive and active conformation kinases and (b) for the fully phosphorylated inactive and active conformation kinases. In each (distance) bar, the error bar indicates the standard deviation. The dashed green line shows a 5 Å distance. (c) Details of the interaction index and residue pairs.

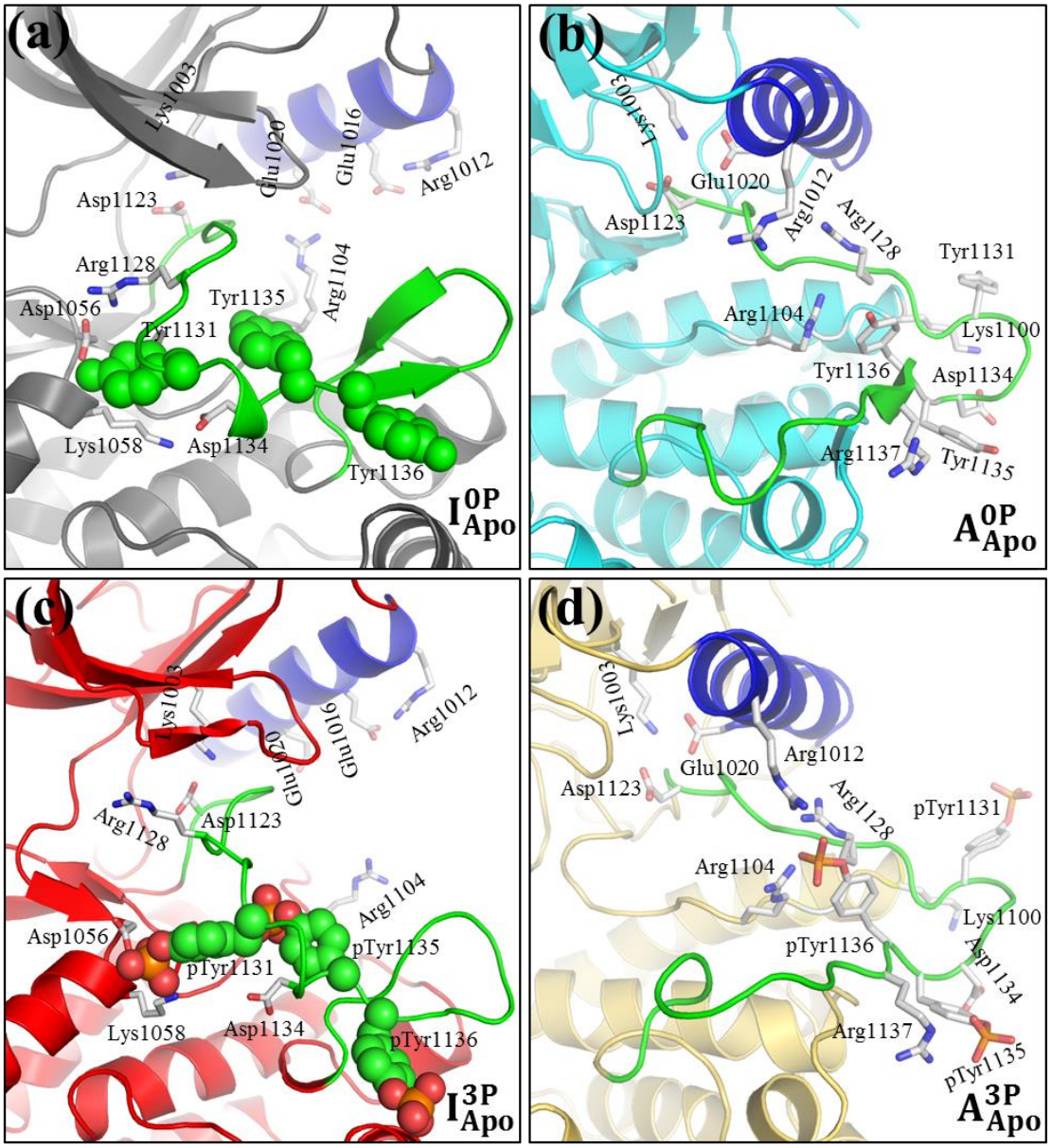


Fig. S6 Side chain interactions of the systems (a) I^{0P}_{Apo}, (b) A^{0P}_{Apo}, (c) I^{3P}_{Apo}, and (d) A^{0P}_{Apo}. Each figure was generated using the last saved coordinates from the 300 ns MD simulations. The αC-helix and A-loop are colored in blue and green, respectively. The rest of the kinase is colored differently for different systems: dark gray in I^{0P}_{Apo}, cyan in A^{0P}_{Apo}, red in I^{3P}_{Apo}, gold in A^{0P}_{Apo}. In (a) and (c), the A-loop tyrosines are presented by green sphere.

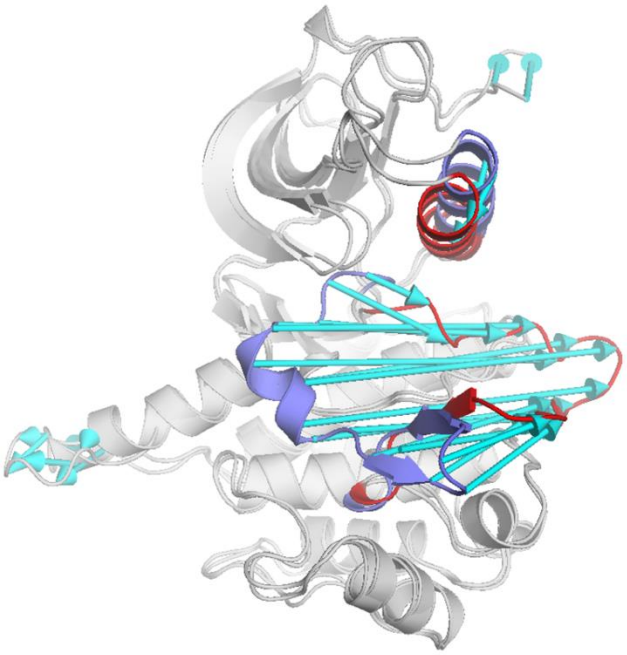


Fig. S7 The displacement vector (cyan) representing the conformational change of the kinase from the inactive to the active conformation. The inactive and active conformations of αC-helix and A-loop are colored in blue and red, respectively.

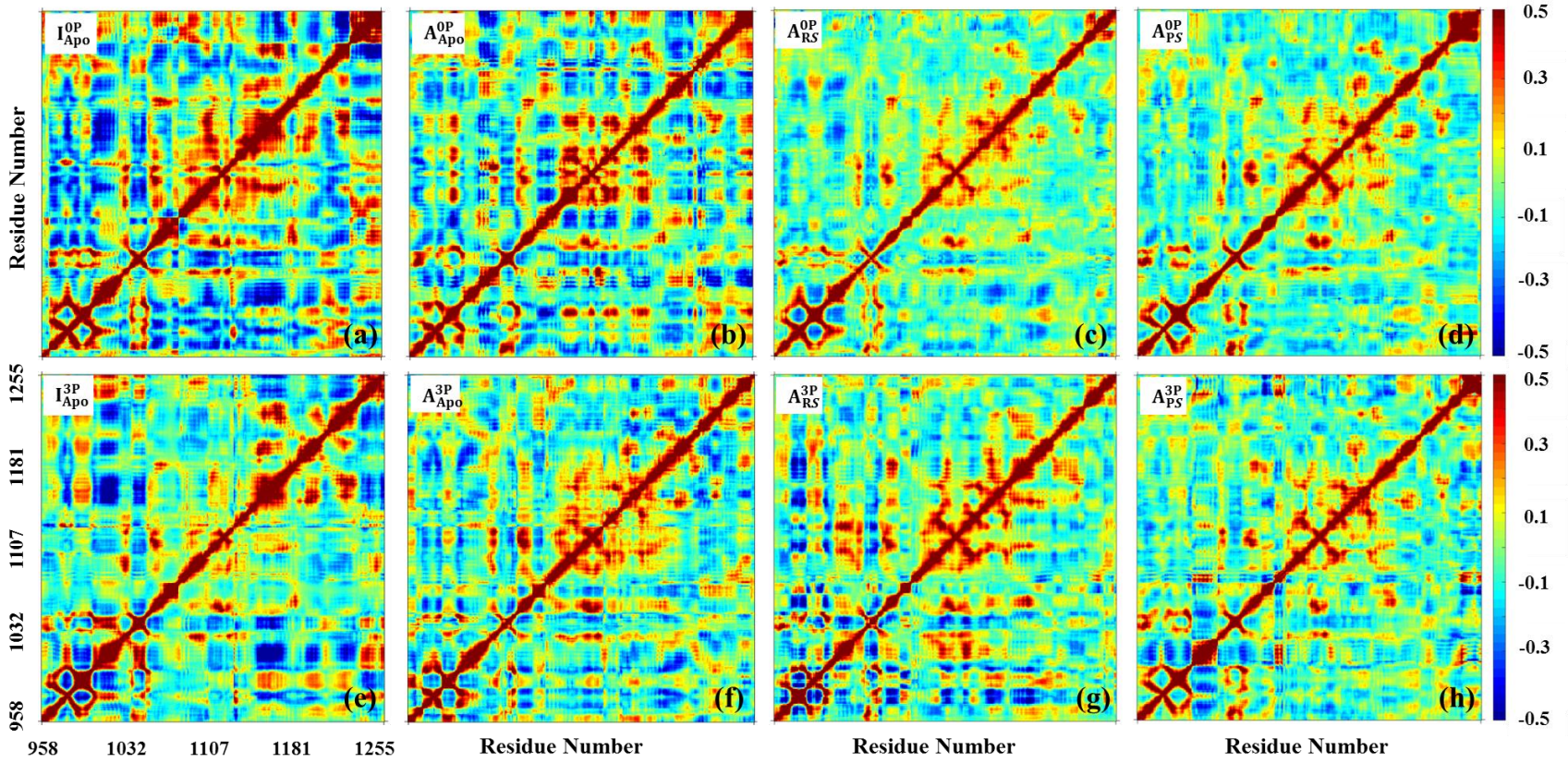


Fig. S8 Covariance matrix. The color scale changes from blue (<-0.5 and smaller), to green (0.0), and to red (>0.5) for the correlation between a pair of residues.

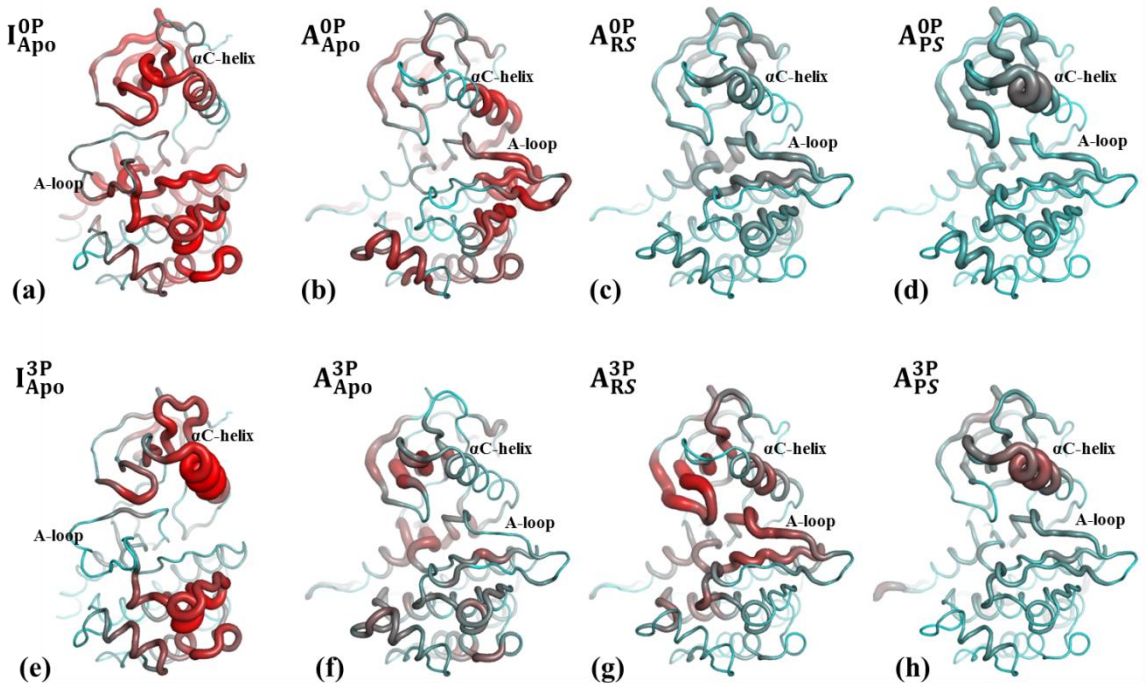


Fig. S9 Structural representation of each residue's correlation with the rest of the kinase. The average correlation value was calculated along each column (or row) of the covariance matrix by the formula, $\langle C_i \rangle = \frac{1}{N_T} \sum_{j=1}^{N_T} |C_{ji}|$, where $\langle \dots \rangle$ is the average over the total number of residues (N_T), and C_{ji} is the pair correlation value of the j and i residue pair. Each figure was generated by mapping the determined average covariance values onto each corresponding structure, i.e., the MD-averaged structure of I^{3P}_{Apo} in (a) and (e) and the MD-averaged structure of A^{0P}_{Apo} in (b) ~ (d) and (f) ~ (h). The thickness and color of backbone vary smoothly between the thin cyan-colored tube for the average correlation value ≤ 0.07 and the thick red-colored tube for the average correlation value ≥ 0.27.

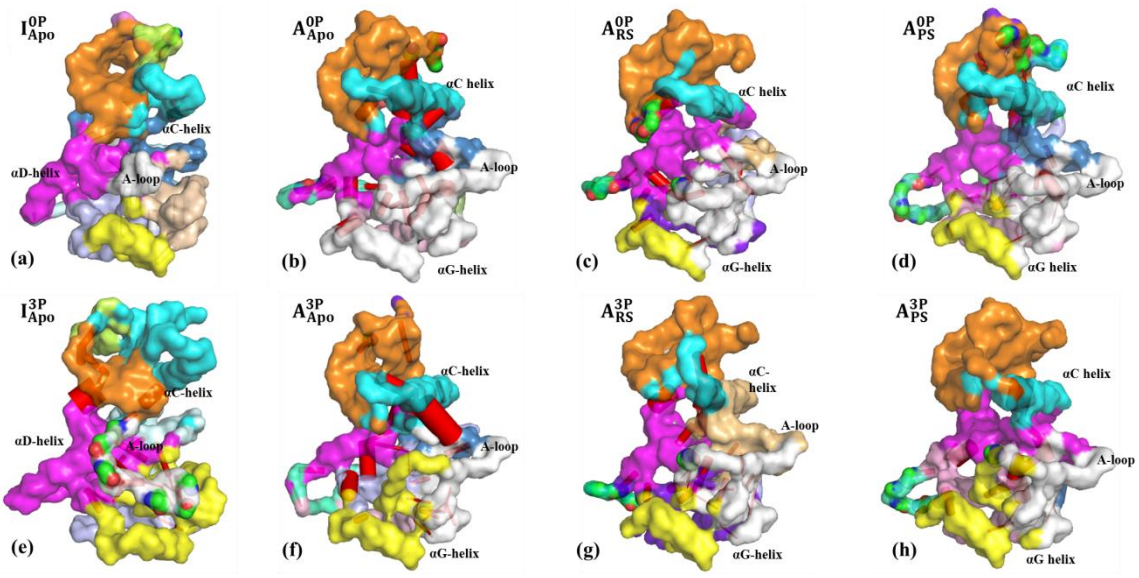


Fig. S10 Community networks. In each figure, different colors indicate different communities and are assigned to be as consistent as possible between the different systems. The red stick indicates the indicator 'betweenness' and the strength of 'betweenness' is shown by the thickness of the stick. In (e), to highlight the isolation of A-loop from the rest of the kinase in I^{3P}_{Apo}, the A-loop is shown with atom-based colors, i.e., green for Carbon, red for Oxygen, and blue for Nitrogen atoms, respectively.

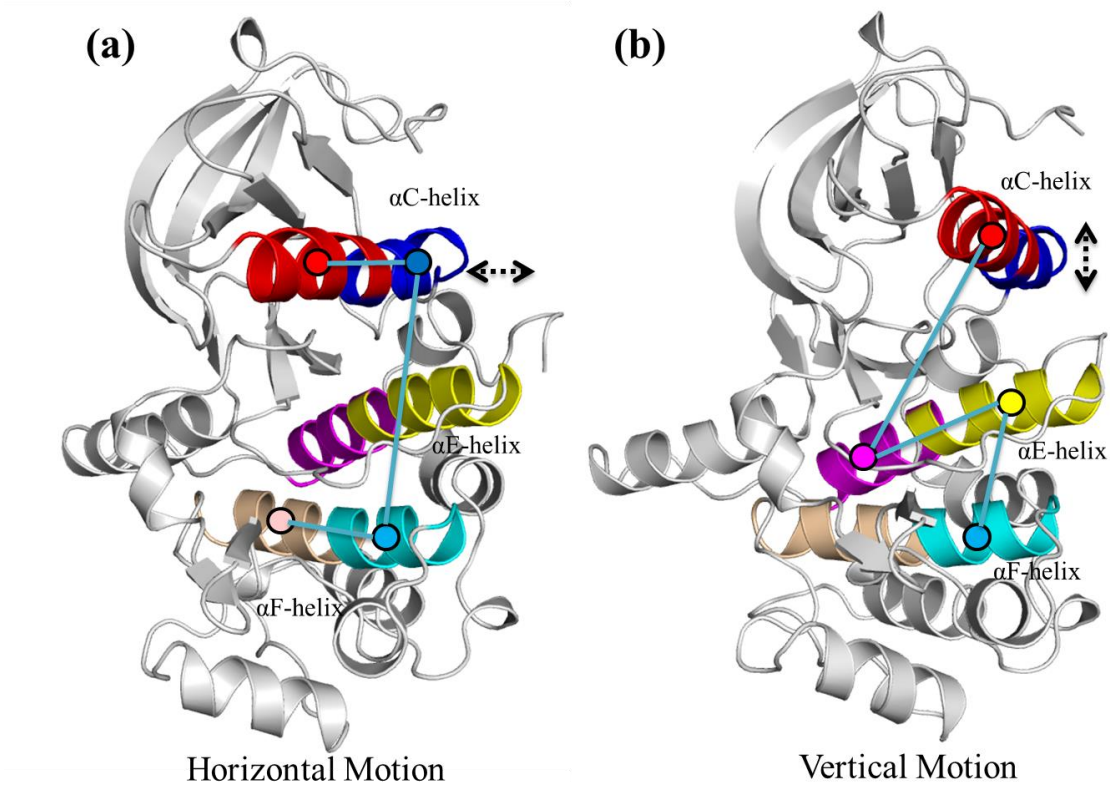


Fig. S11 Definition of the two pseudo-dihedrals, i.e., (a) horizontal and (b) vertical motions, used to describe the rotation motion of αC-helix. Each pseudo-dihedral is defined by 4 centers of mass indicated by the black edged cycles. The protein backbone atoms used in the definition of each center of mass are indicated with different colors.

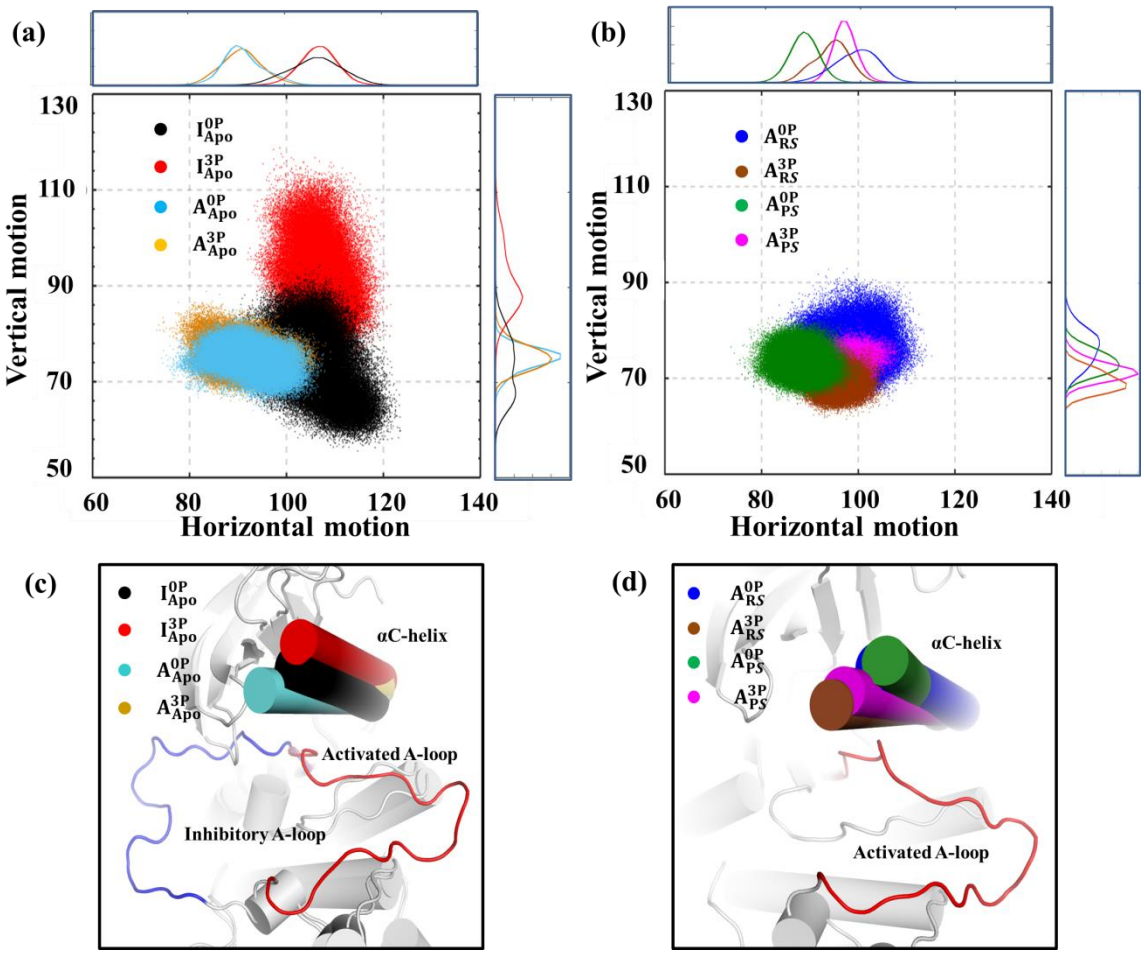


Fig. S12 αC-helix orientation distributions along the horizontal and vertical motions for (a) the apo and (b) holo systems. In (a) and (b), the distribution is also presented in the form of the density distribution along each motion (in the upper panel for the horizontal motion and in the right panel for the vertical motion). The definition of each αC-helix motion is shown in **Fig. S11**. The data points in figures (a) and (b) were generated from coordinates saved from the 300 ns conventional MD simulations. Comparison of the αC-helix orientations (c) between the apo systems and (d) between the holo systems determined from the 200 ns aMD simulations. The orientation of each αC-helix was determined by the same manner described in **Fig. 3**. The color scheme is also the same as **Fig. 3**.

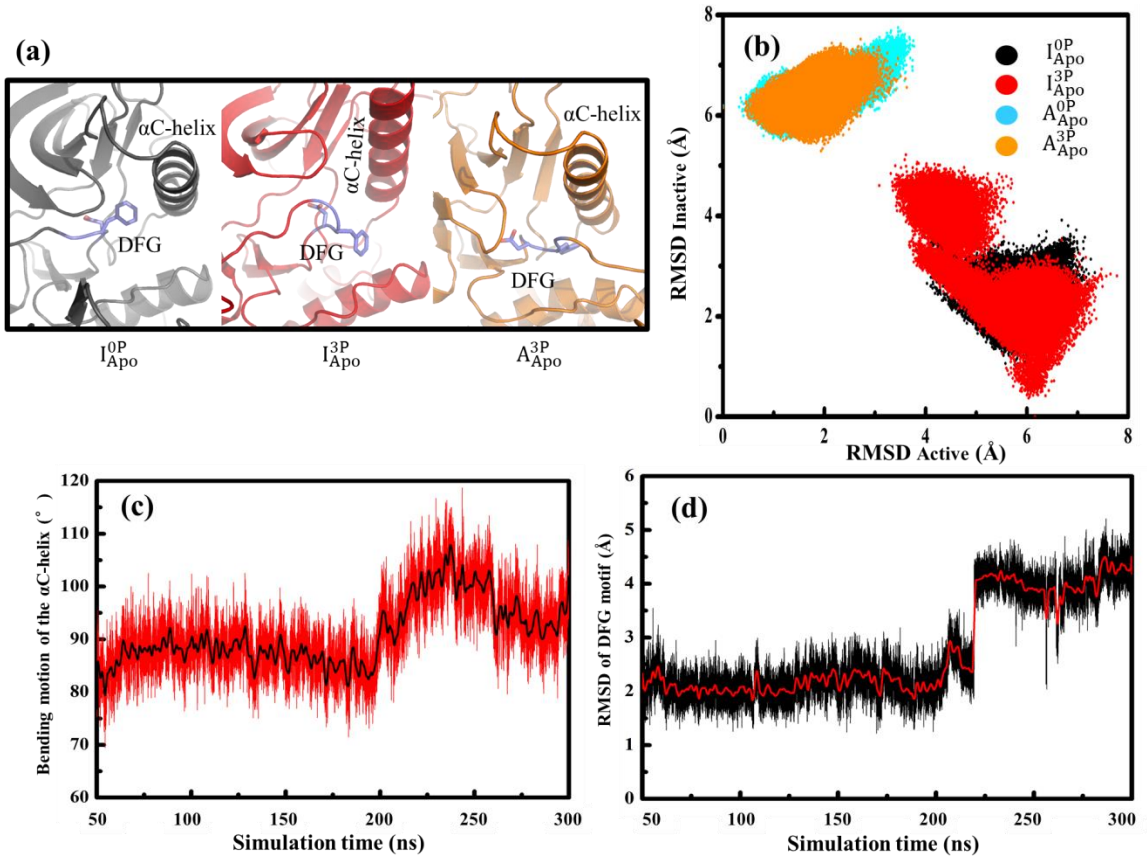


Fig. S13 (a) Variation of the DFG (blue stick) motif orientation between the I^{0P}_{Apo}, I^{3P}_{Apo} and A^{3P}_{Apo} systems. Each structure was prepared using the last set of coordinates saved during the 300 ns MD simulation. (b) RMSD distribution of the DFG motif with respect to its position in the active (X-axis) and inactive conformation X-ray structures (Y-axis), showing the partial flipping of the DFG motif in I^{3P}_{Apo} (red dots) toward that of the active conformation. The time courses of (c) the outward rotation of αC-helix and (d) flipping of DFG motif as a function of the MD simulation time, showing the correlated behavior between the two motions in the I^{3P}_{Apo} system.

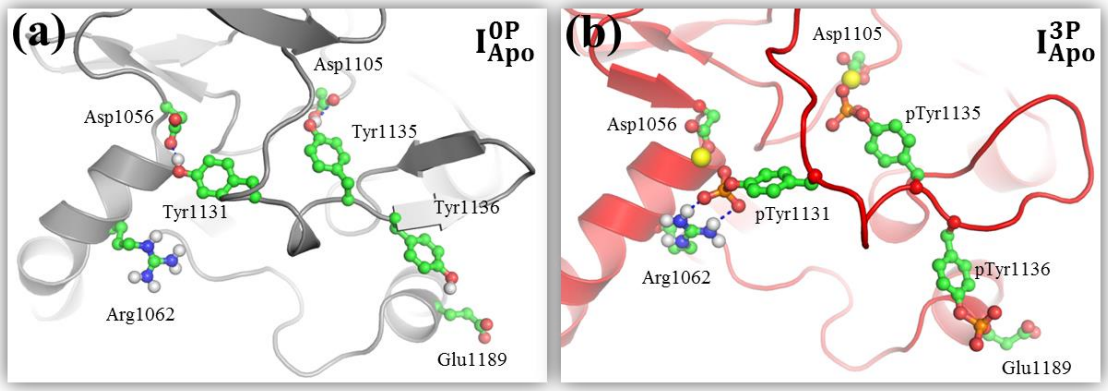


Fig. S14 Interactions of the A-loop tyrosines with nearby residues in the (a) unphosphorylated and (b) fully phosphorylated inactive conformation kinases. In the phosphorylated kinase, two Na^+ ions (yellow sphere) were diffused toward the phosphorylated tyrosines during the MD simulation to from bridged interactions between the Asp and phosphotyrosine residues.

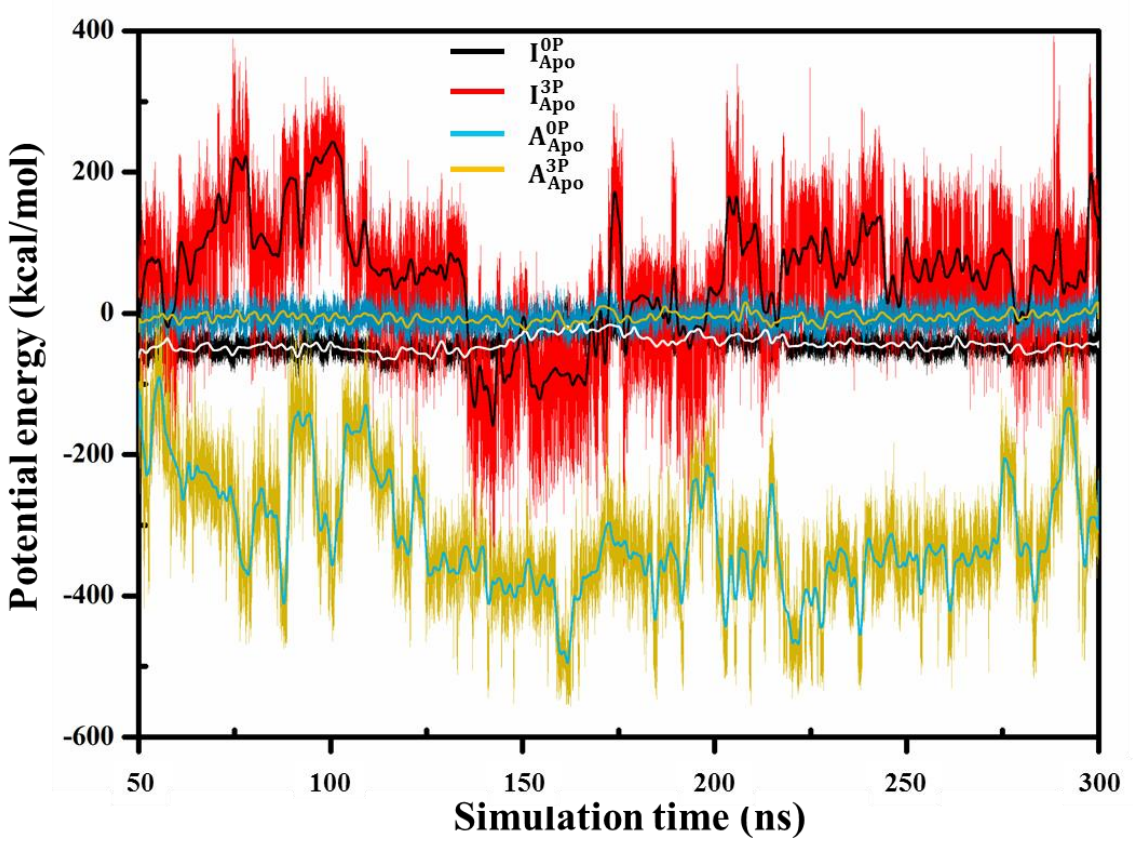


Fig. S15 Interaction energies between the A-loop tyrosines and kinases residues that were within 4 Å from each tyrosine as a function of the simulation time. The thick solid line in each plot represents the smoothed line by averaging the interaction energies from nearby time points.

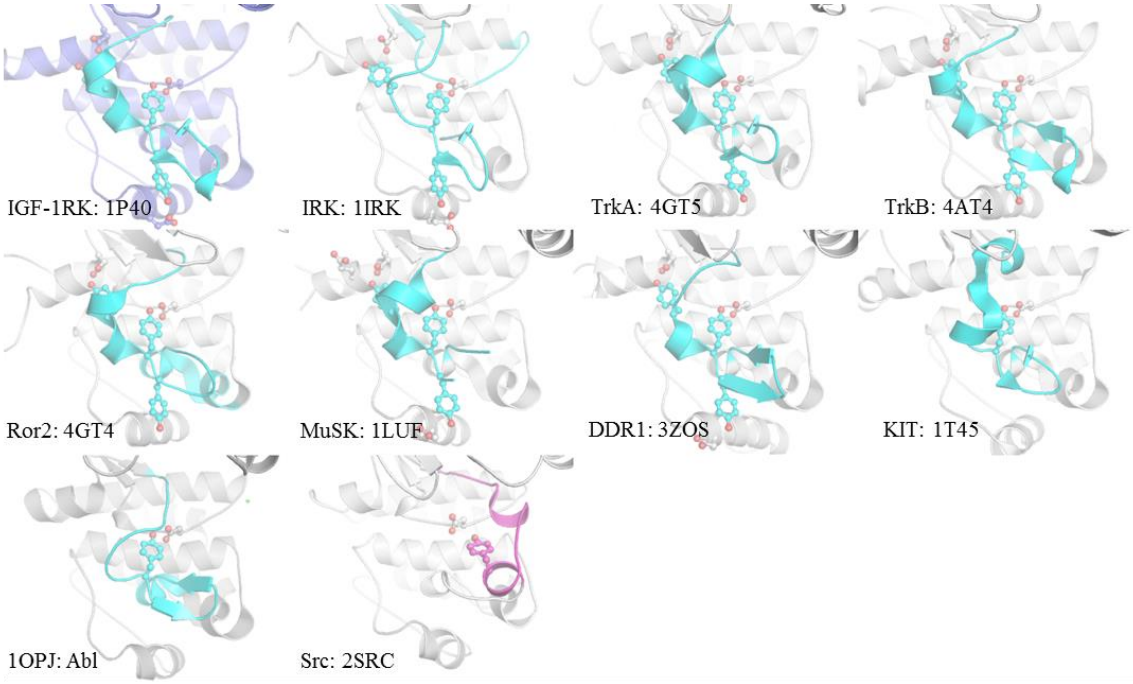


Fig. S16 Interactions of the inactive conformation A-loop tyrosines with nearby Asp/Glu residues in different kinases. In each structure, A-loop is shown in cyan, except the A-loop of Src kinase. The Src kinase A-loop is shown in different color because its orientation is qualitatively different from other kinases presented in the figure. In all structures, the A-loop tyrosines and neighboring negatively charged residues are presented by the ball-and-stick model. The reference for each structure is listed in **Table S2**.

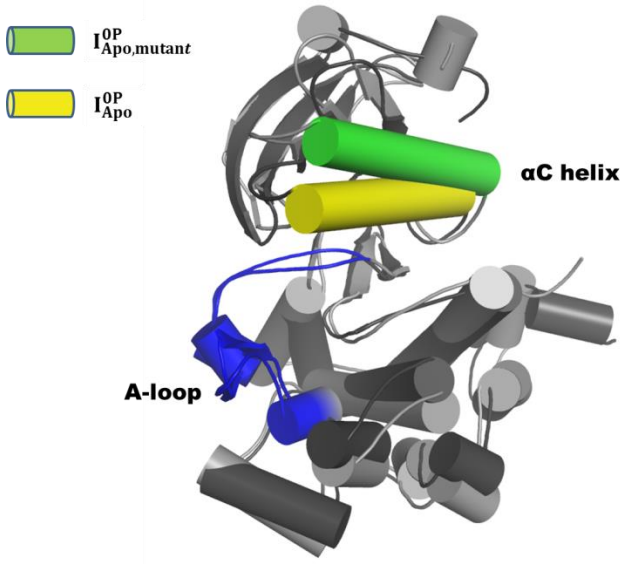


Fig. S17 Comparison of the αC-helix orientation between the wild-type and R1104A mutant of I^{0P}_{Apo}. The αC-helixes of the wild-type and mutant are shown with yellow and green cylinders, respectively, and the A-loops are shown in blue for both systems. The remainder of the kinase is shown in dark and light gray for the wild-type and mutant, respectively.

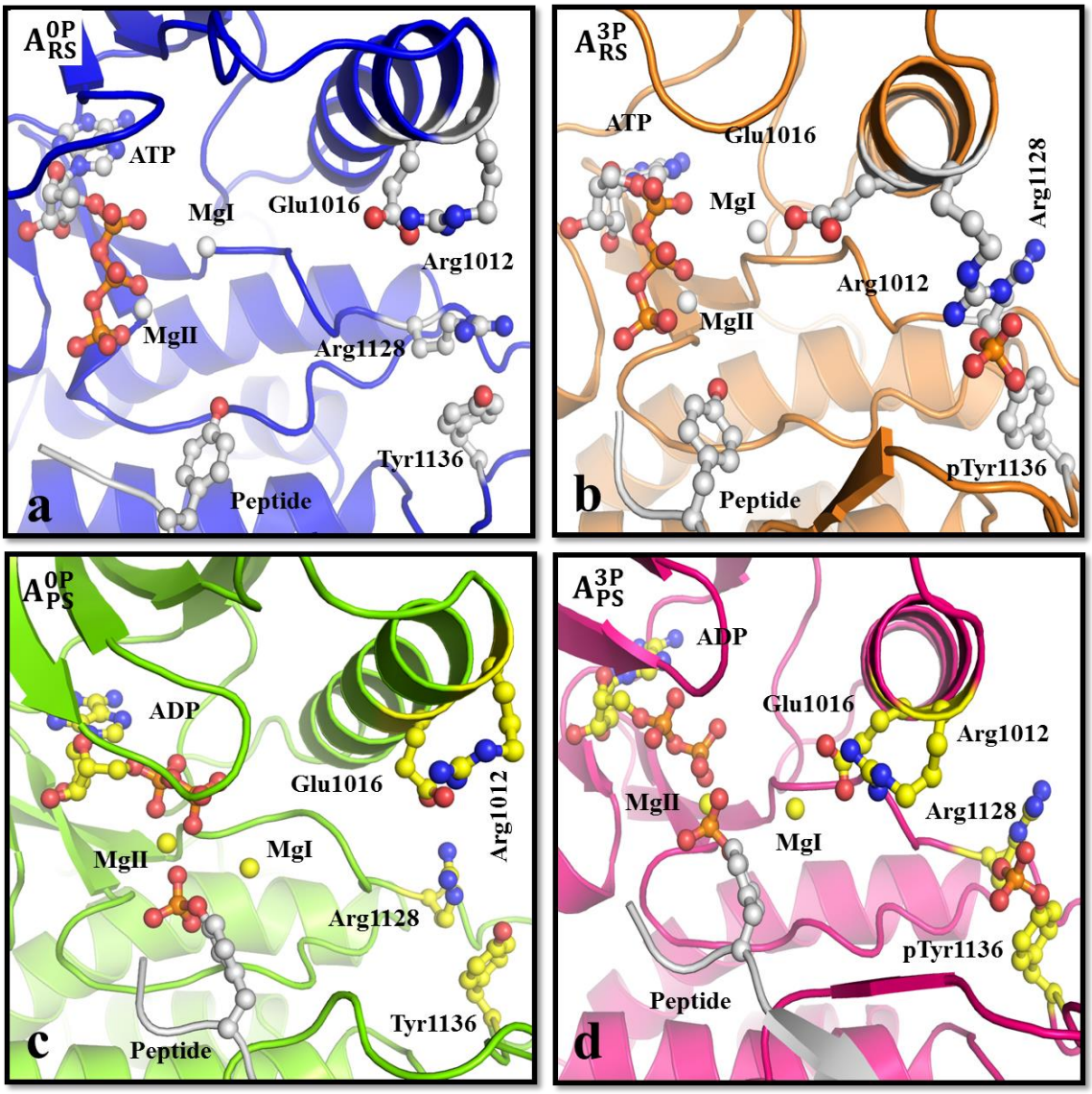


Fig. S18 Changes of salt bridge interaction of Arg1012 before and after the catalytic reaction. In the unphosphorylated systems, i.e., (a) A^{0P}_{RS} and (c) A^{0P}_{PS}, Arg1012 interacts with Glu1016 via the intra-helix salt bridge. In the fully phosphorylated system, Arg1012 interacts with pTyr1136 in (b) A^{3P}_{RS}, and interacts with Glu1016 in (d) A^{3P}_{PS}. In both (b) and (d), Gly1016 also forms the second-solvation shell interaction with MgI. In all figures, each structure was selected to represent closely the MD averaged structure of the corresponding system.

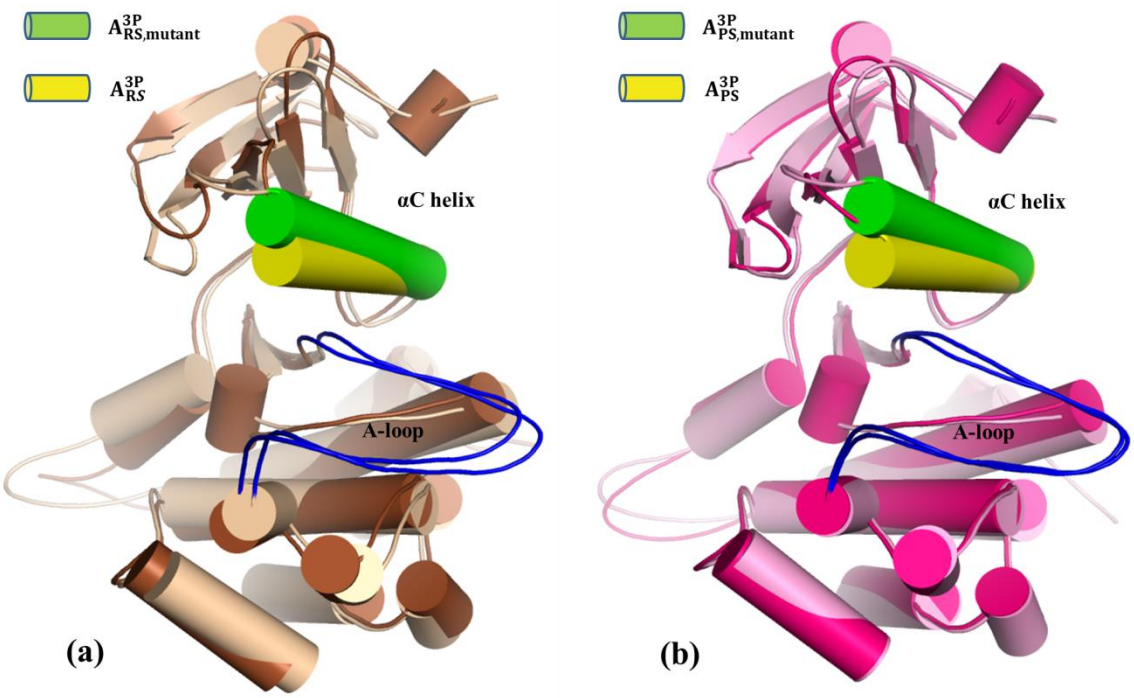


Fig. S19 Comparison of the αC-helix orientation between the wild-type and R1012A-E1016A double mutant of (a) A^{3P}_{RS} and (b) A^{3P}_{PS}. In both figures, the αC-helixes are shown in yellow for the wild-type and green in the mutant, respectively, and the A-loop is shown in blue. In (a), the remainder of the kinase is shown in brown and light brown for the wild-type and mutant systems, and in (b), shown in pink and light pink for the wild-type and mutant systems, respectively.

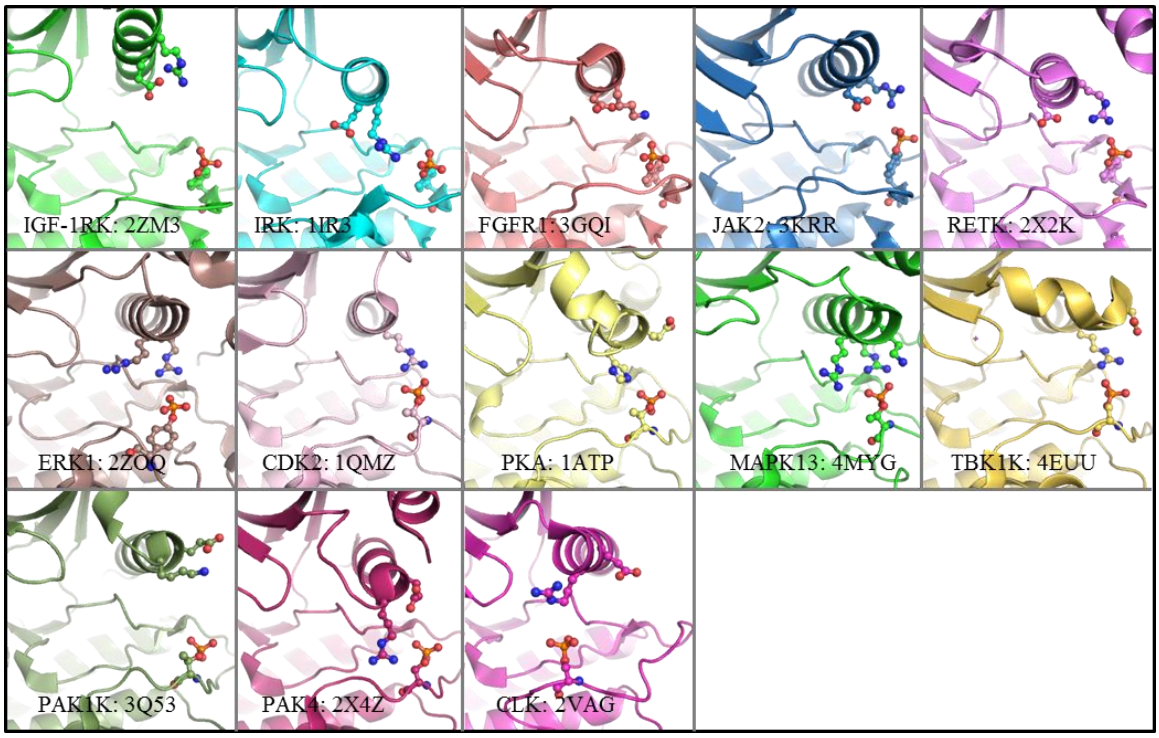


Fig. S20 Interactions between the phosphorylated A-loop tyrosine/serine/threonine and αC-helix arginine in the active conformation of different kinases. In each figure, selected residues are shown by the ball-and-stick model. The name of each kinase corresponding PDB ID are indicated at the bottom of each figure: 2ZM3[25] for IGF-1RK, 1IR3[26] for IRK, 3GQI[27] for FGFR1, 3KRR[28] for JAK2, 2X2K[29] for RETK, 2ZOQ[30] for ERK1, 1QMZ[31] for CDK2, 1ATP[32] for PKA, 4MYG[33] for MAPK13, 4EUU[34] for TBK1K, 3Q53[35] for PAK1K, 2X4Z[36] for PAK4, and 2VAG[37] for CLK.

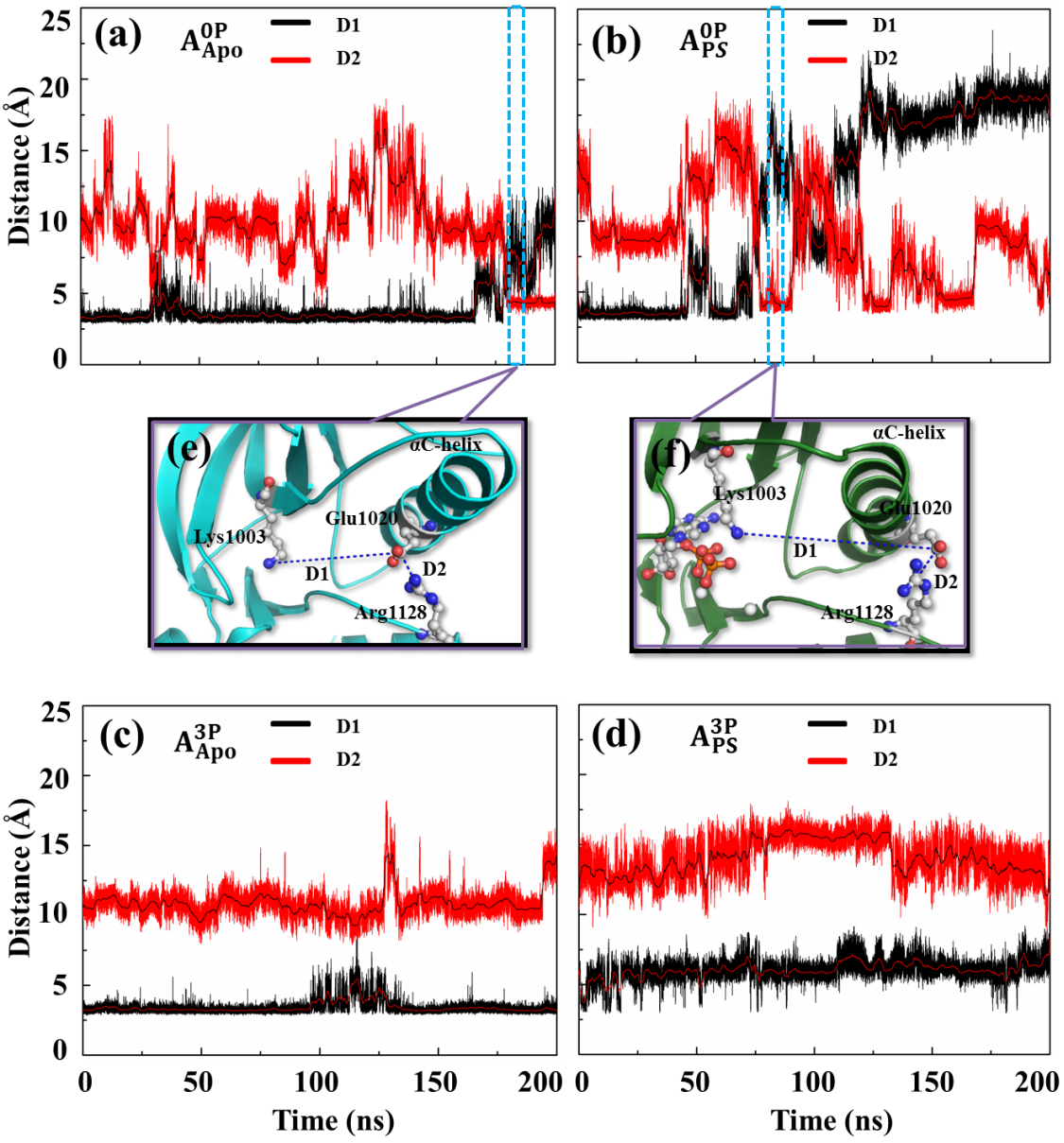


Fig. S21 Switch of the salt-bridge interactions between the Glu1020-Lys1003 and Glu1020-Arg1128 pairs in (a) A^{0P}_{Apo}, (b) A^{0P}_{PS}, (c) A^{3P}_{Apo}, and (d) A^{3P}_{PS} systems. D1 (black) is the distance between the Glu1020 CG and Lys1003 NZ atoms, and D2 (red) is the distance between the Glu1020 and Arg1128 CZ atoms, respectively. In (e) and (f), each snapshot shows the orientation of the αC-helix after the interaction switch, at the time point indicated in the corresponding (a) A^{0P}_{Apo} and (b) A^{0P}_{PS} system figures. All figures were prepared based on the data generated during the 200 ns aMD simulations.

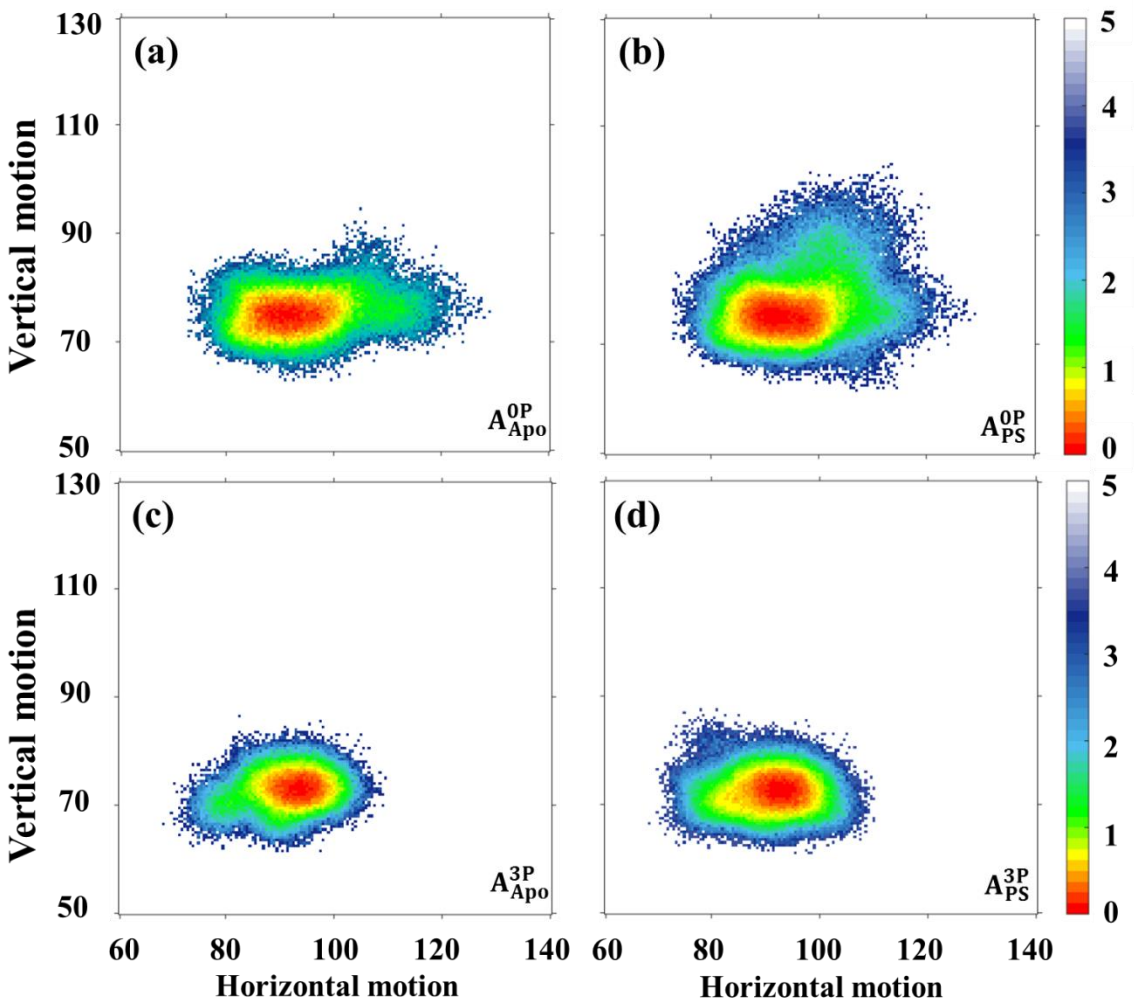


Fig. S22 Un-reweighted potentials of mean force of the αC-helix dynamics from the aMD simulations for the (a) A^{0P}_{Apo}, (b) A^{0P}_{PS}, (c) A^{3P}_{Apo}, and (d) A^{3P}_{PS} systems, presented in **Fig. S21**. In all systems, the unphosphorylated kinase showed a wider αC-helix orientation distribution than that of the phosphorylated kinase.

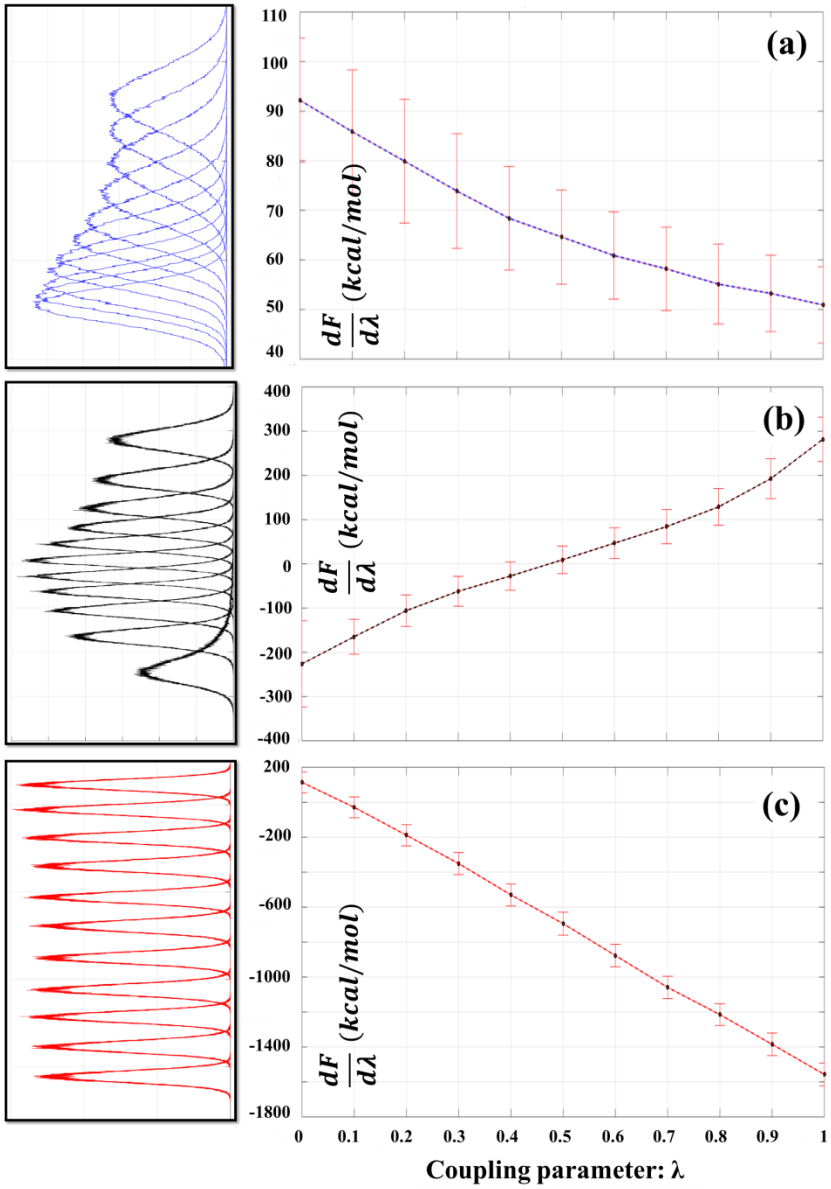


Fig. S23 TI integrands and energy gap distribution for the $A_{Apo}^{0P\rightarrow 3P}$ transformation: (a) uncharging of each tyrosine, (b) transforming the van der Waals interactions of tyrosine to those of pTyr, and (c) charging of each pTyr. The error bar on the integrand shows the standard deviation of the TI integrand at each λ value.

References

1. D. Hamelberg, C. A. de Oliveira and J. A. McCammon, *J. Chem. Phys.*, 2007, **127**, 155102.
2. D. Hamelberg, J. Mongan and J. A. McCammon, *J. Chem. Phys.*, 2004, **120**, 11919-11929.
3. Y. L. Miao, S. E. Nichols and J. A. McCammon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 6398-6406.
4. B. R. Brooks, D. Janezic and M. Karplus, *J. Comput. Chem.*, 1995, **16**, 1522-1542.
5. H. F. Lou and R. I. Cukier, *J. Phys. Chem. B*, 2006, **110**, 12796-12808.
6. A. Sethi, J. Eargle, A. A. Black and Z. Luthey-Schulten, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 6620-6625.
7. A. B. Pyrkosz, J. Eargle, A. Sethi and Z. Luthey-Schulten, *J. Mol. Biol.*, 2010, **397**, 1350-1371.
8. R. W. Floyd, *Commun. Acm.*, 1962, **5**, 345-345.
9. T. H. Cormen and T. H. Cormen, *Introduction to algorithms*, MIT Press, Cambridge, Mass., 2nd edn., 2001.
10. M. E. J. Newman, *Phys. Rev. E*, 2004, **70**, 056131.
11. M. Girvan and M. E. J. Newman, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 7821-7826.
12. M. E. J. Newman and M. Girvan, *Phys. Rev. E*, 2004, **69**, 026113.
13. J. Lee and T. L. Hendrickson, *J. Mol. Biol.*, 2004, **344**, 1167-1174.
14. N. M. Glykos, *J. Comput. Chem.*, 2006, **27**, 1765-1768.
15. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph. Model.*, 1996, **14**, 33-38.
16. S. Munshi, D. L. Hall, M. Kornienko, P. L. Darke and L. C. Kuo, *Acta. Crystallogr. D Biol. Crystallogr.*, 2003, **59**, 1725-1730.
17. S. R. Hubbard, L. Wei, L. Ellis and W. A. Hendrickson, *Nature*, 1994, **372**, 746-754.
18. S. C. Artim, J. M. Mendrola and M. A. Lemmon, *Biochem. J.*, 2012, **448**, 213-220.
19. T. Bertrand, M. Kothe, J. Liu, A. Dupuy, A. Rak, P. F. Berne, S. Davis, T. Gladysheva, C. Valtre, J. Y. Crenne and M. Mathieu, *J. Mol. Biol.*, 2012, **423**, 439-453.
20. J. H. Till, M. Becerra, A. Watty, Y. Lu, Y. Ma, T. A. Neubert, S. J. Burden and S. R. Hubbard, *Structure*, 2002, **10**, 1187-1196.
21. P. Canning, L. Tan, K. Chu, S. W. Lee, N. S. Gray and A. N. Bullock, *J. Mol. Biol.*, 2014, **426**, 2457-2470.
22. C. D. Mol, D. R. Dougan, T. R. Schneider, R. J. Skene, M. L. Kraus, D. N. Scheibe, G. P. Snell, H. Zou, B. C. Sang and K. P. Wilson, *J. Biol. Chem.*, 2004, **279**, 31655-31663.
23. B. Nagar, O. Hantschel, M. A. Young, K. Scheffzek, D. Veach, W. Bornmann, B. Clarkson, G. Superti-Furga and J. Kuriyan, *Cell*, 2003, **112**, 859-871.
24. W. Q. Xu, A. Doshi, M. Lei, M. J. Eck and S. C. Harrison, *Mol. Cell.*, 1999, **3**, 629-638.
25. S. C. Mayer, A. L. Banker, F. Boschelli, L. Di, M. Johnson, C. H. Kenny, G. Krishnamurthy, K. Kutterer, F. Moy, S. Petusky, M. Ravi, D. Tkach, H. R. Tsou and W. Xu, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 3641-3645.
26. S. R. Hubbard, *EMBO J.*, 1997, **16**, 5572-5581.
27. J. H. Bae, E. D. Lew, S. Yuzawa, F. Tome, I. Lax and J. Schlessinger, *Cell*, 2009, **138**, 514-524.
28. F. Baffert, C. H. Regnier, A. De Pover, C. Pissot-Soldermann, G. A. Tavares, F. Blasco, J. Brueggen, P. Chene, P. Drueckes, D. Erdmann, P. Furet, M. Gerspacher, M. Lang, D. Ledieu, L. Nolan, S. Ruetz, J. Trappe, E. Vangrevelinghe, M. Wartmann, L. Wyder, F. Hofmann and T. Radimerski, *Mol. Cancer Ther.*, 2010, **9**, 1945-1955.

29. L. Mologni, R. Rostagno, S. Brussolo, P. P. Knowles, S. Kjaer, J. Murray-Rust, E. Rosso, A. Zambon, L. Scapozza, N. Q. McDonald, V. Lucchini and C. Gambacorti-Passerini, *Bioorg. Med. Chem.*, 2010, **18**, 1482-1496.
30. T. Kinoshita, I. Yoshida, S. Nakae, K. Okita, M. Gouda, M. Matsubara, K. Yokota, H. Ishiguro and T. Tada, *Biochem. Bioph. Res. Co.*, 2008, **377**, 1123-1127.
31. N. R. Brown, M. E. Noble, J. A. Endicott and L. N. Johnson, *Nat. Cell Biol.*, 1999, **1**, 438-443.
32. J. Zheng, E. A. Trafny, D. R. Knighton, N. H. Xuong, S. S. Taylor, L. F. Ten Eyck and J. M. Sowadski, *Acta. Crystallogr. D Biol. Crystallogr.*, 1993, **49**, 362-365.
33. Z. Yurtsever, S. M. Scheaffer, A. G. Romero, M. J. Holtzman and T. J. Brett, *Acta. Crystallogr. D Biol. Crystallogr.*, 2015, **71**, 790-799.
34. X. Ma, E. Helgason, Q. T. Phung, C. L. Quan, R. S. Iyer, M. W. Lee, K. K. Bowman, M. A. Starovasnik and E. C. Dueber, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 9378-9383.
35. J. Wang, J. W. Wu and Z. X. Wang, *Structure*, 2011, **19**, 1752-1761.
36. B. W. Murray, C. X. Guo, J. Piraino, J. K. Westwick, C. Zhang, J. Lamerdin, E. Dagostino, D. Knighton, C. M. Loi, M. Zager, E. Kraynov, I. Popoff, J. G. Christensen, R. Martinez, S. E. Kephart, J. Marakovits, S. Karlicek, S. Bergqvist and T. Smeal, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 9446-9451.
37. O. Fedorov, K. Huber, A. Eisenreich, P. Filippakopoulos, O. King, A. N. Bullock, D. Szklarczyk, L. J. Jensen, D. Fabbro, J. Trappe, U. Rauch, F. Bracher and S. Knapp, *Chem. Biol.*, 2011, **18**, 67-76.