Supplementary Information

Tuning Local Hydration Enables a Deeper Understanding of Protein-Ligand Binding: the PP1-Src Kinase Case

Andrea Spitaleri[1,2], Syeda R. Zia[1,3], Patrizio di Micco[4], Bissan Al-Lazikani[4], Miguel A. Soler[1]*, Walter Rocchia[1]*.

[1] CONCEPT Lab, Istituto Italiano di Tecnologia, via Morego 30, I-16163 Genoa, Italy

[2] Center for Omics Sciences, Emerging Bacterial Pathogens Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy

[3] Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi-75270, Pakistan.

[4] Cancer Research UK Cancer Therapeutics Unit at The Institute of Cancer Research, London, SM2 5NG, UK.

Supplementary Methods

Molecular dynamics bias and protocol description

The approach uses classical molecular dynamics to simulate the protein and ligand systems while introducing an external electrostatic-like force acting between a given subset of the solute and the surrounding solvent. The bias acts on a collective variable (CV) that has the form of the interaction electrostatic energy of two systems of point charges in presence of a Debye-Hueckel like screening:

$$CV = \sum_{a\in A}\sum_{b\in B} C\frac{q_a q_b}{r_{ab}}\exp\left(-\kappa r_{ab}\right)$$

where A and B are the two interacting groups of atoms, r_{ab} the distance between atoms a and b, and κ the effective Debye factor.

In the present case, the CV involves on one side the heavy atoms of the biased region of the solute (i.e. ligand, binding site or both) and on the other the Oxygen atoms of nearby water molecules. The charges are equal on value and sign on all the atoms, so that reducing the CV corresponds to applying a repulsion. The actual value of the charge is irrelevant since it becomes part of a global factor that is rescaled when computing the percentage wrt. the unperturbed value. The Debye factor of the effective screening is $(1\text{nm})^{-1}$.

The unbiased hydration of the region of interest, e.g. the ligand, is first calculated and then used as a reference (100% in the x-axis of Fig. 1bcd). Then, a semi-parabolic restraint, shown in Figure 1, is applied to this contrived electrostatic interaction energy of the system formed by the ligand and the surrounding water molecules so as to force repulsion until the target value is reached. If the hydration goes below the predefined threshold, the semi-parabolic form of the bias leads to unbiased MD. In the explorative simulation protocols that we have experimented, the target hydration values differ between the ligand and the binding site (see below).

If the target hydration values are not exceedingly low, they should naturally be reached and passed during natural binding. The entire method envisions the execution of many short (here, 35 ns each) biased simulation replicas. 35ns seemed to be a reasonable trade-off in terms of performance. It allowed to achieve some binding events, and therefore to make a comparison with the unbiased simulations, at a reasonable computational cost.

Plain and biased simulations. We performed 64 plain and 198 biased simulations of Src-PP1 complexes, each of them ~35 ns in length, for a total of ~9.17 μs. Based on the 16 Src-PP1 complexes, we carried out 64 simulations by assigning 4 different initial velocities to each of them. For the application of the repulsive bias, we considered two groups: (1) Group A comprised heavy atoms of the binding pocket or ligand or both as per our adopted protocol (2) Group B comprised the Oxygen atoms of all water molecules. A neighbor list was used to get a better performance. A

short unbiased MD simulation run of 0.02 ps (10 steps) was performed before starting the biased simulation. The output from this aforementioned simulation was used to compute the CV between both groups in the fully solvated well-equilibrated system by means of plumed driver plugin. [1] In the biased simulations, the computed CV value is taken as the initial value, which is then gradually decreased and brought to a predefined target value following a steer plan. We used the "moving harmonic restraints with upper wall" of the PLUMED software.[1] If the current CV is greater than its target value during the simulation, a quadratic bias is applied. The bias will switch off automatically when the current CV becomes lower than the target, thus recovering a plain molecular dynamics situation (see Figure 1 for the dehydration schemes). The plumed.dat file was set with neighbor list (NL) of 0.7nm (NL_CUTOFF=0.7), NL_STRIDE=10, and the bias MOVINGRESTRAINT with VERSE=U, by which the restraint is only acting for CV larger than the threshold. The KAPPA used was set to 10^{-7} $(kJ/mol)^{-1}$ and the restraint speed to 4.1 CV/ps so that the threshold reaches its final target value in the first 10 ns of simulation.

We pursued three different dehydration protocols to increase the likelihood of observing plausible binding events. As expected, the decrease in the CV corresponds to the lowering of the solvation around the targeted system regions (i.e. the binding site and the ligand).

Biased simulation protocol 1 (single bias). In 64 biased simulation runs performed with protocol 1, we repelled water molecules away from the ligand only. The initial value of the CV was gradually decreased to reach the target value of 50% of the initial CV in the first 10 ns of simulation. The 50% restraint was maintained for the remainder of the 25 ns of simulation time. Each simulation terminated at 35 ns.

Biased simulation protocol 2 (double bias). In other 64 biased simulations water molecules were repelled both by the ligand and by the residues of the binding site. The initial values of the CV for the ligand and the pocket residues were gradually decreased to get the target values of 50% and 80%, respectively, in the first 10 ns of the simulation. These target CV values were retained for the remainder of the 25 ns of simulation time. For the rest, the same settings as in protocol 1 were used for the binding site.

Biased simulation protocol 3 (double-delayed bias). Finally, 70 biased simulations were executed with protocol 3. Water molecules were repelled by both the ligand and the residues of the binding pocket. In the first 10 ns of simulation, the initial CV of the ligand was decreased to 40%, while the CV of the binding site was kept at 100% of the respective unbiased values, which we call the "fully solvated" state. From 10 to 25 ns, the target CV of the ligand was further decreased to 20%, while the CV of the binding site was also decreased to 80%. In the last 10 ns of simulation (25-35 ns), the target CV of the ligand and the BS were kept at 20% and 80% of the respective fully solvated conditions. Here, we increased the neighbor list (NL) to 1.2nm to observe possible longer-range effects.

Biased simulation of the KD salt bridge. 32 biased simulation runs were performed, in which we repelled water molecules away from the ligand and from the Lys295-Asp404 residues. The initial values of the CVs were gradually decreased to reach the target value of 50% of the initial CVs in the first 10 ns of simulation. In this biasing approach, if the actual CV is greater than its target value during the simulation, a quadratic bias is applied. The 50% restraint was maintained for the remainder of the 25 ns of simulation time. Each simulation terminated at 35 ns. We used similar parameters for the CV in PP1 hydration as the ones described in the previous bias simulations. For the hydration bias in the salt bridge, we employed a neighbor list (NL) of 0.7nm, which was updated at each step, and a spring constant value of 10^{-5} $(kJ/mol)^{-1}$. Increment of the spring constant was due to the inability of the original value, i.e. 10^{-7}, to reach the target dehydration.

MD simulations of Src and T338M Src isoform. We performed 1 plain simulation of c-Src kinase and one for the T338M isoform, each of them of 1 μs in length. The structural model of the mutated isoform was realized by employing the "Rotamer" tool of Chimera,[2] starting from the Dunbrack 2010 rotamer library.[3] Both Src simulations were prepared and developed as described in the "System set-up for simulation" section.

Comparison with the work of Shan et al. We obtained the trajectories of PP1 binding to Src kinase from the plain MD simulation reported by Shan et al. at D. E. Shaw Research and compared them to those in our study. [4]

Structural and sequence analysis. A complete list of human protein kinases was downloaded from UniProtKB/Swiss-Prot (https://www.uniprot.org). 5655 structures (corresponding to 8655 individual chains) - belonging to 331 genes - have been retrieved using canSAR knowledgebase (http://cansar.icr.ac.uk).[5] For each structure, intramolecular interactions until 10 Å have been calculated taking advantage of Scipy software (https://docs.scipy.org/doc/scipy/reference/index.html). Primary cavities have been selected using canSAR and conserved polar interaction between equivalent Lys295 and Asp404 residues identified using as parameter a distance threshold of <5 Å between any atoms of the two residues. Sequences have been downloaded from UniProtKB/Swiss-

Prot and aligned using clustal Omega (http://www.clustal.org/omega/). In order to correct errors with the sequence alignment, a structural superposition of representative structures has been used and seeded with the sequence alignment. Sequence alignment has then been manually refined based on structural equivalences.

Identification of pocket and ligand placement

We used the NanoShaper (https://concept.iit.it/downloads) tool to identify the binding pocket in our chosen target protein, Src kinase.[6-7] In order to perform the structural analyses, we used the structure of the human HCK-PP1 complex, (PDB code 1QCF). The hematopoietic cell kinase (HCK) protein belongs to the protein kinase superfamily, Tyr-protein kinase family, Src subfamily. Src and HCK have a 63% sequence identity. The residues of the binding site of Src kinase are: Leu273, Val281, Ala293, Lys295, Glu310, Met314, Val323, Ile336, Thr338, Glu339, Tyr340, Met341, Gly344, Leu393, Ala403, Asp404, and Phe405 (the residue numbering follows that of the used HCK model). We used the BiKi Life Sciences toolkit[8] from BiKi Technologies (http://www.bikitech.com/) to position the ligand PP1 (the kinase inhibitor) in front of the binding site with a margin of 1.1 Å, 1.2 Å, 1.3 Å, and 1.4 Å, where the margin is defined as the thickness of the water shell around the ligand. By doing so, the ligand was placed outside the entrance of the binding pocket without clashing or bumping with the protein's side chain residues. After the ligand was placed, it was oriented randomly to get different configurations as the starting point of our simulations. We selected 16 different configurations of the ligand on the basis of the root mean square deviation (RMSD) values of the ligand that lie between 9.73 Å and 12.03 Å with respect to the crystallographic pose of the ligand PP1 in the complex with HCK kinase (PDB code: 1QCF). We chose the initial configurations within this range of RMSD because in the literature it is indicated that there is a trapping region for the ligand due to interaction with water at ~10 Å RMSD in comparison with the native pose[5]. Therefore, we considered the initial poses of the ligand selected by this criterion to be a reasonable starting point for our simulations. The Table 1 shows the RMSD values of the 16 selected ligand configurations with reference to the 1QCF crystal structure. Of the 16 selected configurations, 10 were obtained by placing the ligand at unique positions, and 6 were obtained by inducing random rotation of some of the configurations already obtained by ligand translation at unique positions.

System set-up for simulation

To get the initial coordinates for the protein, we used the co-crystallized structure of human c-Src kinase in its active conformation bound with des-methyl analogue of imatinib, available in the Protein Data Bank (PDB ID: 1Y57).[9] The ligand was removed, and we selected only the kinase domain comprising 259-533 residues for our study. After the ligand placement, each complex was immersed in a cubic box with a margin of 15 Å. All the simulation systems were then solvated with TIP3P water molecules using the LEaP module in AmberTools. The counter ions (i.e. Na^+ and Cl^-) were added to neutralize the system's charge. The Amber99SB-ILDN and the general amber force field (GAFF) were used for protein and ligand parameterization. The partial charges for the ligand were derived with the RESP method in Antechamber.[10] The systems were subjected to energy minimization for 5000 steps, performed by steepest descent method. Each system was equilibrated into four steps with: (1) NVT ensemble (i.e. canonical or constant volume and constant temperature ensemble) at T = 100 K for 100 ps, (2) NVT ensemble at T = 200 K for 100 ps, (3) NVT ensemble at T = 300 K for 100 ps, (4) NPT ensemble (i.e. isothermal–isobaric or constant pressure and constant temperature ensemble) at T = 300 K and P = 1 atm for 1000 ps. The protein's backbone atoms and heavy atoms of the ligand PP1 were restrained during equilibration by applying an isotropic force of 1000 kJ mol^{-1} nm^{-2}. A Parrinello-Rahman barostat was used,[11] and the constant temperature conditions were maintained by velocity-rescaling thermostat.[12] The cut-off values for the short-range electrostatics and van der Waals interactions were set at 1.1 nm. The long-range electrostatic interactions were estimated by the particle mesh Ewald (PME) method.[13] The production dynamics were performed in the NVT (constant volume and constant temperature) ensemble at T = 300 K. All MD simulations were carried out using Gromacs 4.6.1 version[14] on Fermi, Galileo, and Marconi supercomputers at CINECA.

RMSD and PP1-BS distance analysis

We performed the root mean square deviation (RMSD) analysis of the final bound pose (observed in our simulations) with respect to the native pose of the ligand PP1 in the binding site of HCK kinase crystal structure (PDB ID: 1QCF). This was done by the structural alignment of the backbone atoms of HCK and Src of the binding site residues: 273, 274, 281, 293, 295, 323, 336, 338, 339, 340, 341, 344, 345, 393, 403, 404. Likewise, we analyzed the distance in the final bound pose between the centers of mass of ligand PP1 and the binding site residues of Src kinase (PP1-SrcBS).

Supplementary Figures

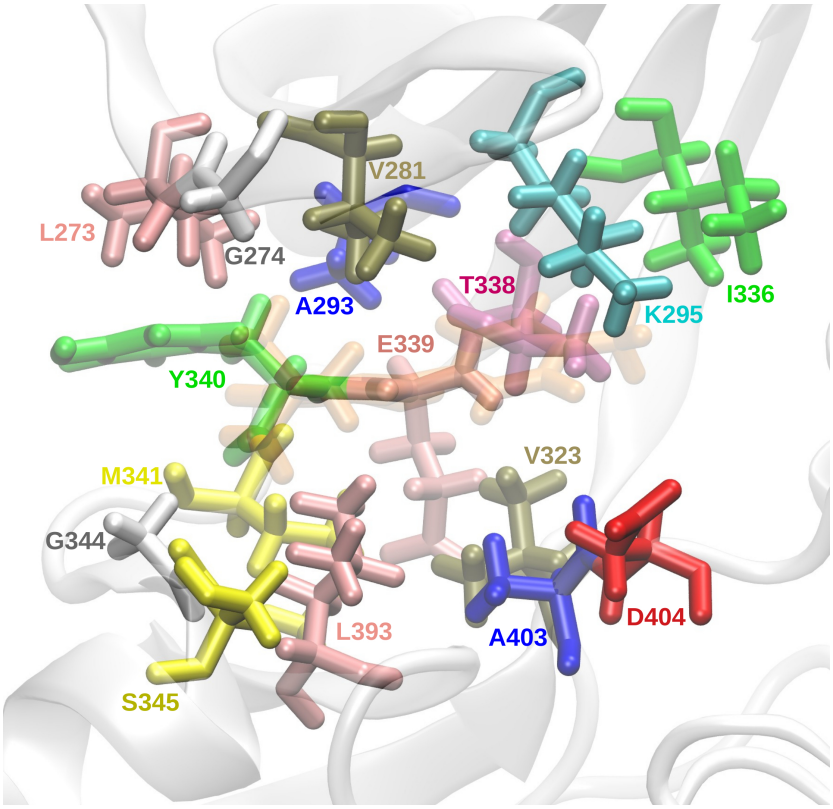

Figure S1. Graphical representation of the residues of the binding site in Src kinase.

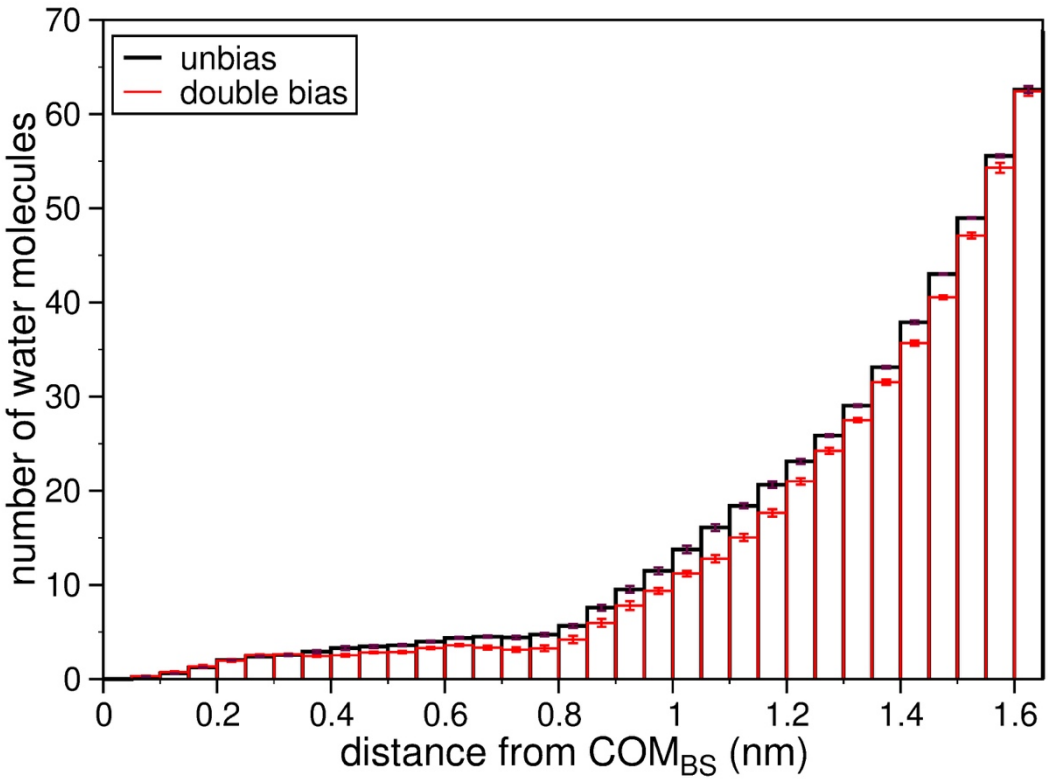

Figure S2. Histogram of the number of water molecules at a given distance interval from the center of mass of the atoms composing the binding site of Src kinase for the unbiased and double biased trajectories.

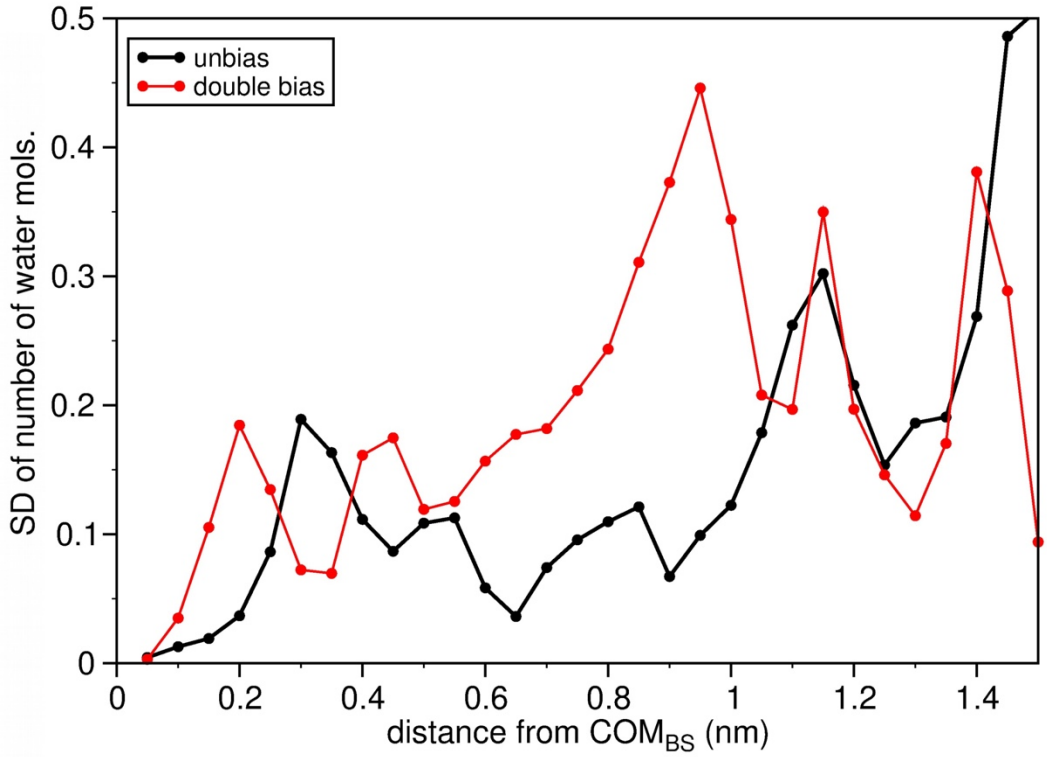

Figure S3. Standard deviation of the number of water molecules at a certain distance interval from the center of mass of the atoms composing the binding site of Src kinase for the unbiased and double biased trajectories.

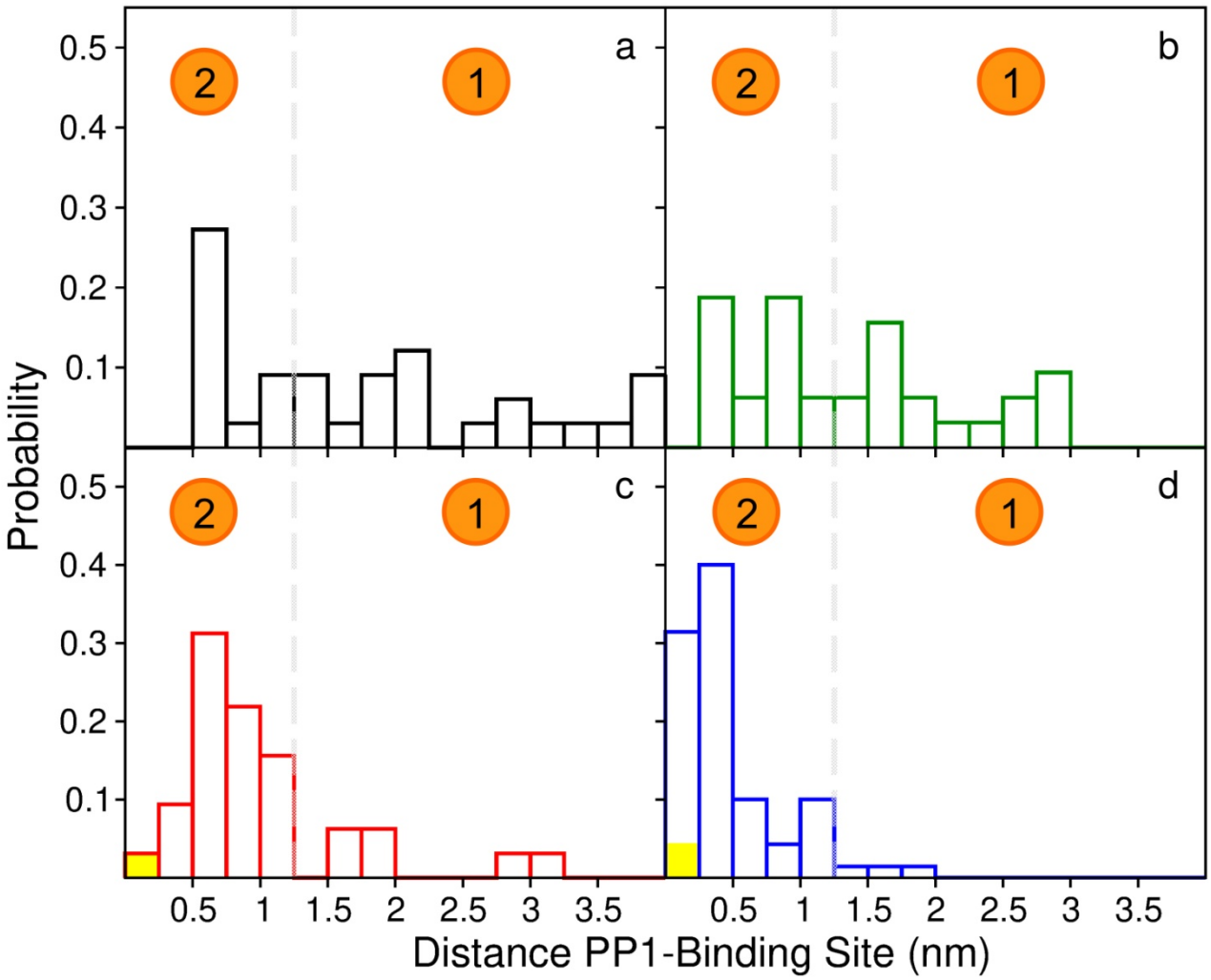

Figure S4. Distribution of the final PP1-Src distances for each set of (a) unbiased, (b) single bias, (c) double bias and (d) double-delayed bias trajectories. The probability of native binding events (case 3) is highlighted in yellow for all set of trajectories.

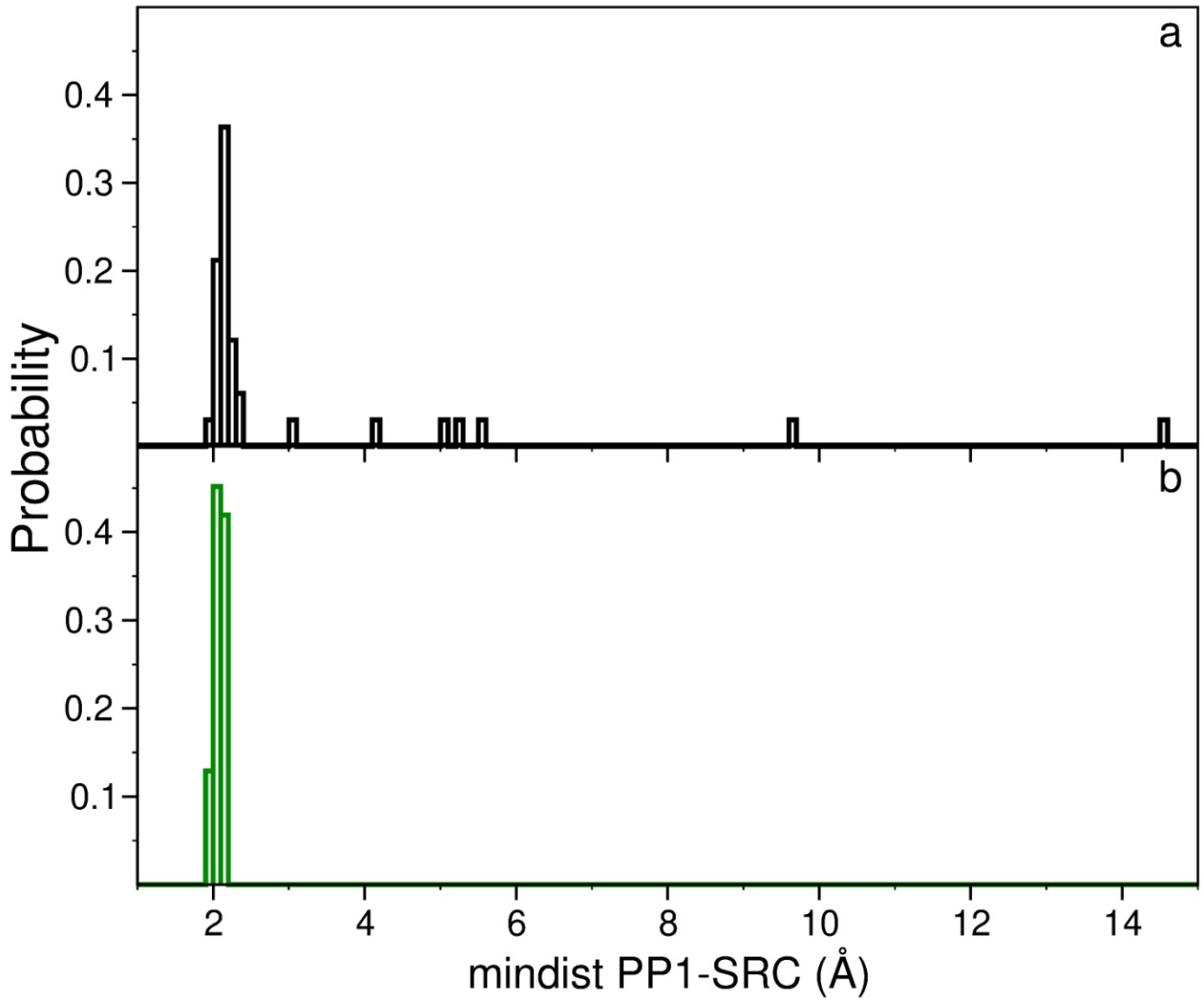

Figure S5. Probability distribution of the average minimum distance between PP1 and Src kinase computed over each (a) unbiased and (b) single bias trajectory set.

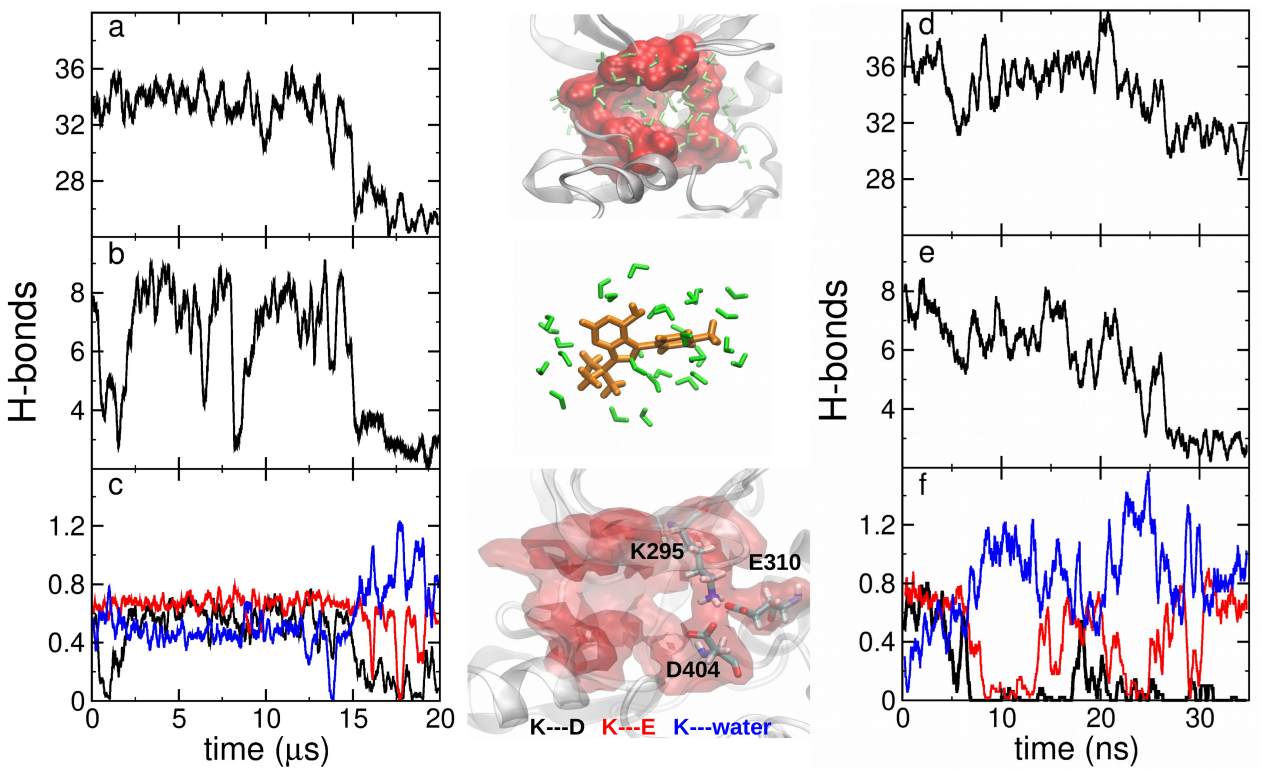

Figure S6. The evolution of the hydration of the targeted regions is evaluated during Simulation 1 from D.E. Shaw Research (a-c) and during the successful double biased MD trajectory (d-f). The application of the hydration bias accelerates the entry of the ligand by almost two orders of magnitude. The behavior of the hydration in Src-BS and PP1 correlates with that observed for the Lys295-Asp404 and Lys295-Glu310 salt bridges, as they must break their interaction (principally the KD salt bridge) to allow the entry of PP1. For clarity, raw values have been smoothed by averaging over 200 ns in (a-c), and over 0.5 ns in (d-f).

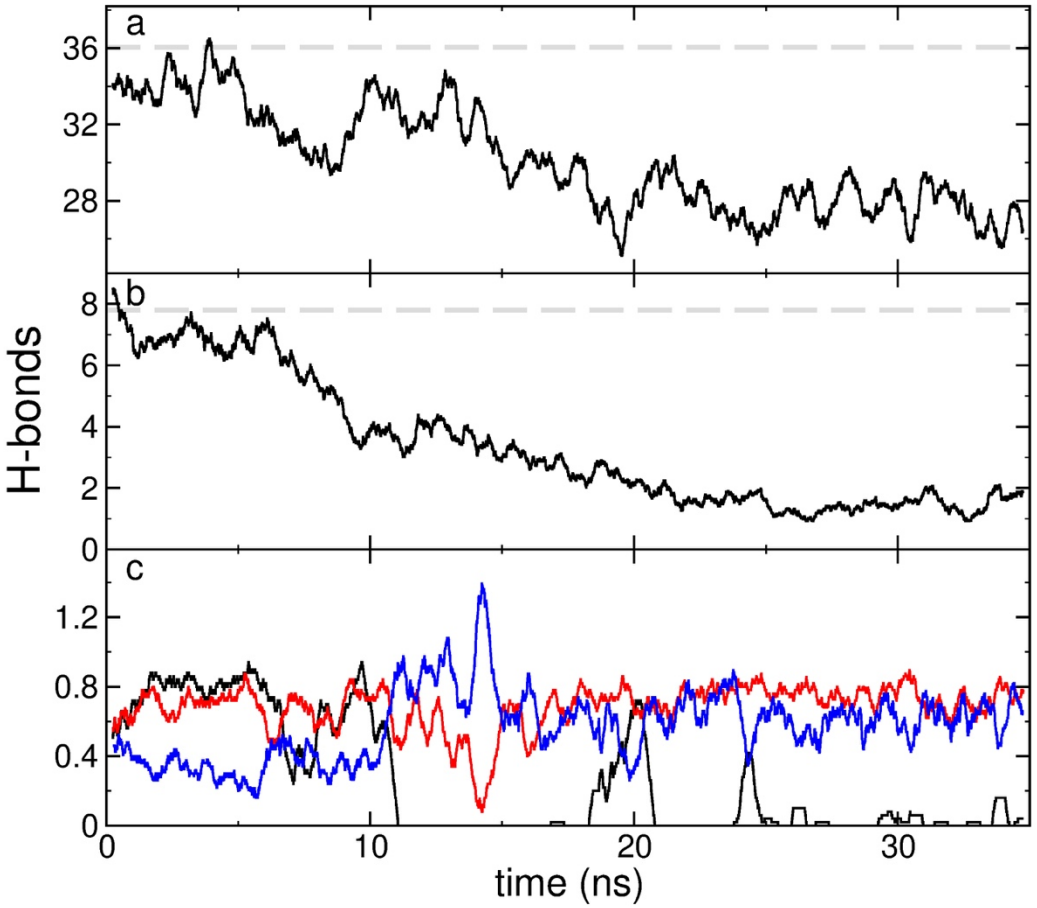

Figure S7. Number of H-bonds (a) on the binding site of Src, (b) around PP1, of formed salt bridges (c) between Lys295-Asp404 (black), Lys295-Glu310 (red), and of Lys295-water interactions (blue) along the double-delayed biased MD trajectory 11, in which is achieved the native binding. Horizontal grey lines indicate the average values in our unbiased simulations.

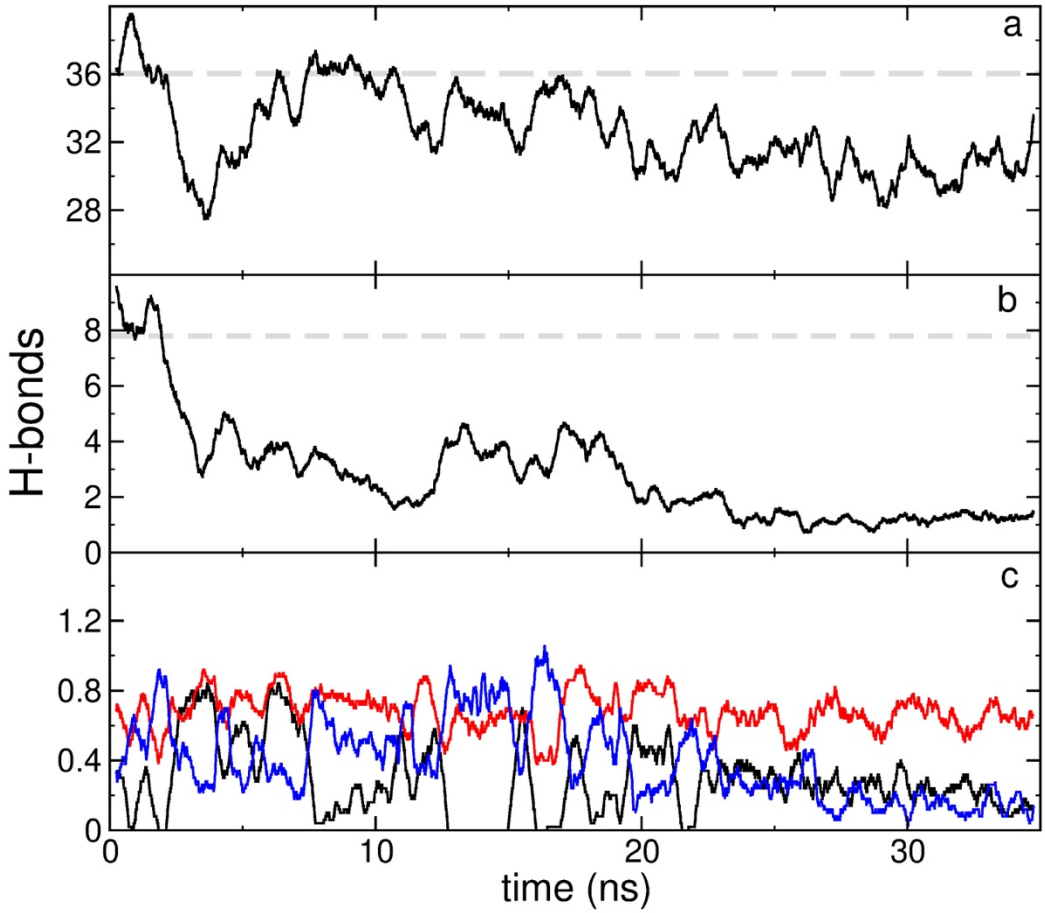

Figure S8. Number of H-bonds (a) on the binding site of Src, (b) around PP1, of formed salt bridges (c) between Lys295-Asp404 (black), Lys295-Glu310 (red), and of Lys295-water interactions (blue) along the double-delayed biased MD trajectory 3, in which is achieved the native binding. Horizontal grey lines indicate the average values in our unbiased simulations.

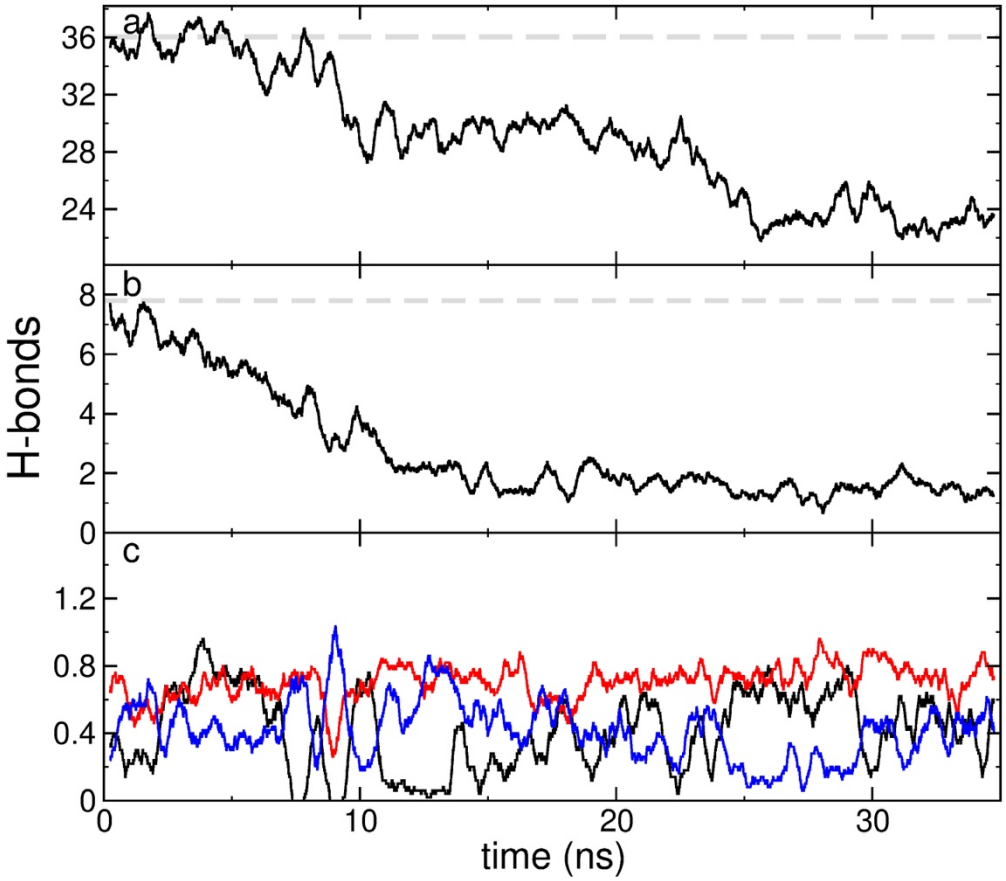

Figure S9. Number of H-bonds (a) on the binding site of Src, (b) around PP1, of formed salt bridges (c) between Lys295-Asp404 (black), Lys295-Glu310 (red), and of Lys295-water interactions (blue) along the double-delayed biased MD trajectory 55, in which is achieved the native binding. Horizontal grey lines indicate the average values in our unbiased simulations.

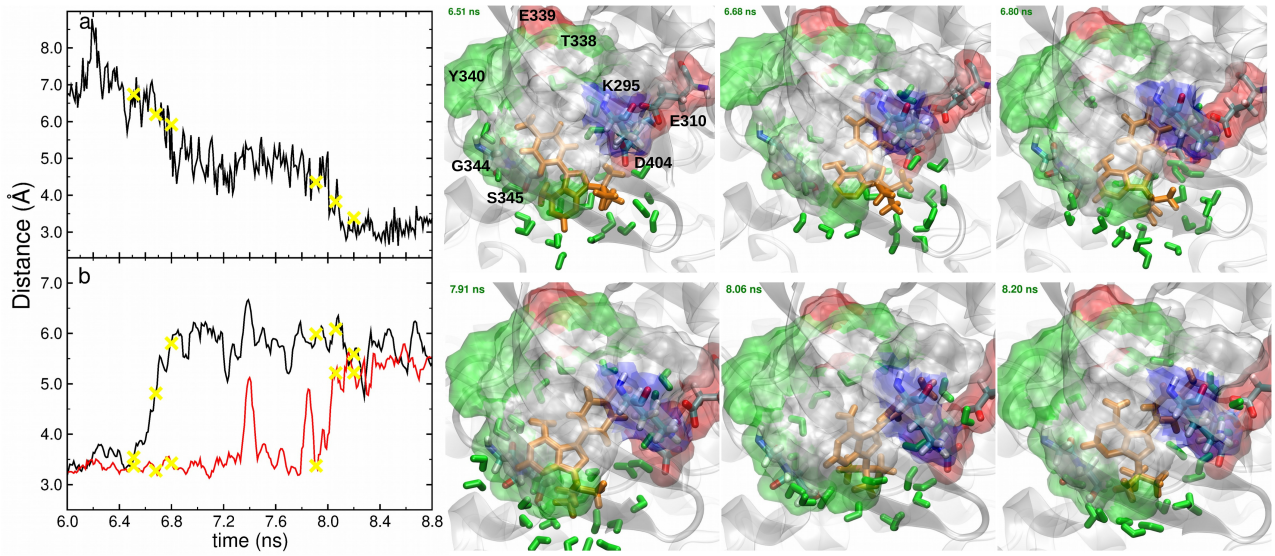

Figure S10. Distance between (a) PP1 and Src binding site and (b) Lys295-Asp404 (black) and Lys295-Glu310 (red) along the interval in which PP1 achieves the binding site in the successful double biased MD trajectory. In snapshots, alternative orientation of PP1 and its hydration shell as well as the polar (green) and charged (red for negative charged and blue for positive charged) residues of the Src binding site are represented in licorice by employing VMD software.

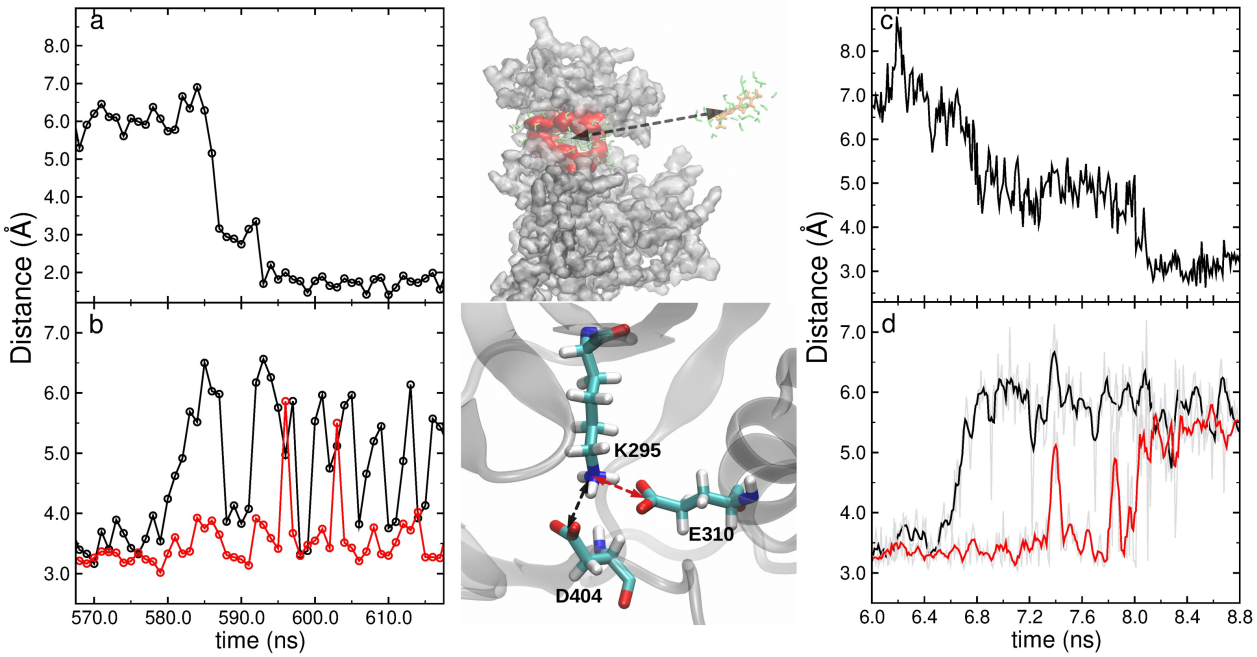

Figure S11. Comparative analysis of the binding of PP1 into the site of Src kinase between simulation 3 from D.E. Shaw Research (on the left) and the successful double biased trajectory (on the right). Distances between PP1 and Src binding site center of masses (panels a, c) and Lys295-Asp404 (black) and Lys295-Glu310 (red) (panels b,d) are compared along the process.

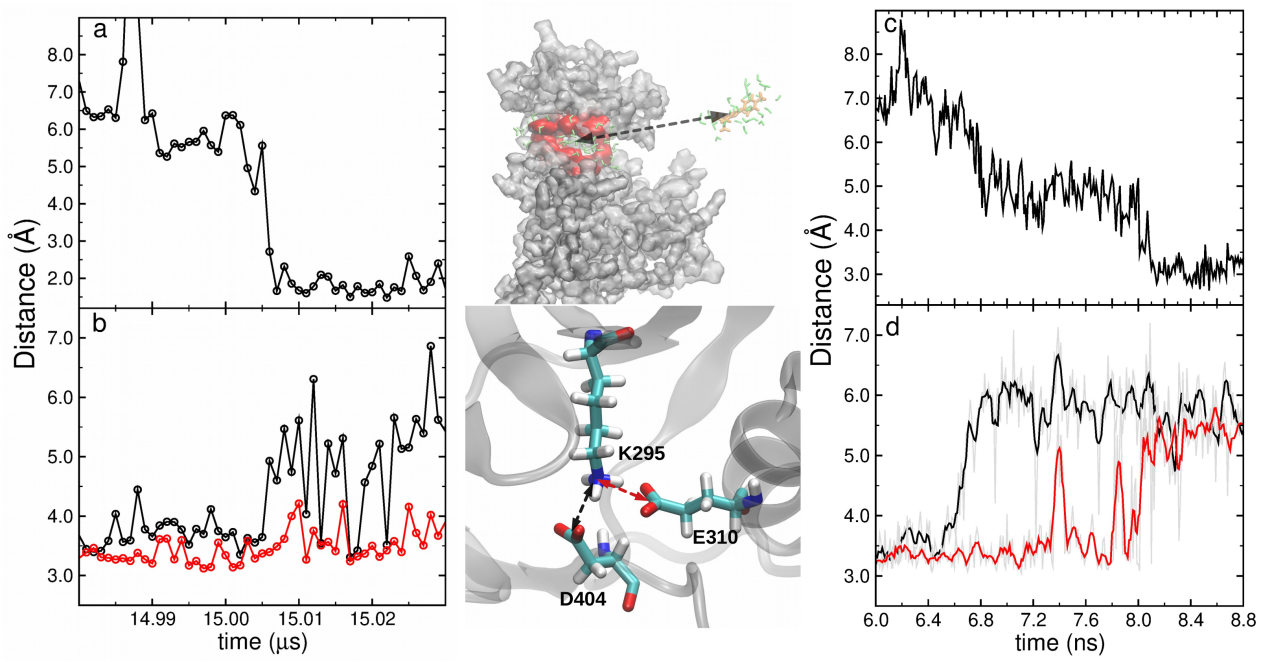

Figure S12. Comparative analysis of the binding of PP1 into the site of Src kinase between simulation 1 from D.E. Shaw Research (on the left) and the successful double biased trajectory (on the right). Distances between PP1 and Src binding site center of masses (panels a, c) and Lys295-Asp404 (black) and Lys295-Glu310 (red) (panels b,d) are compared along the process.

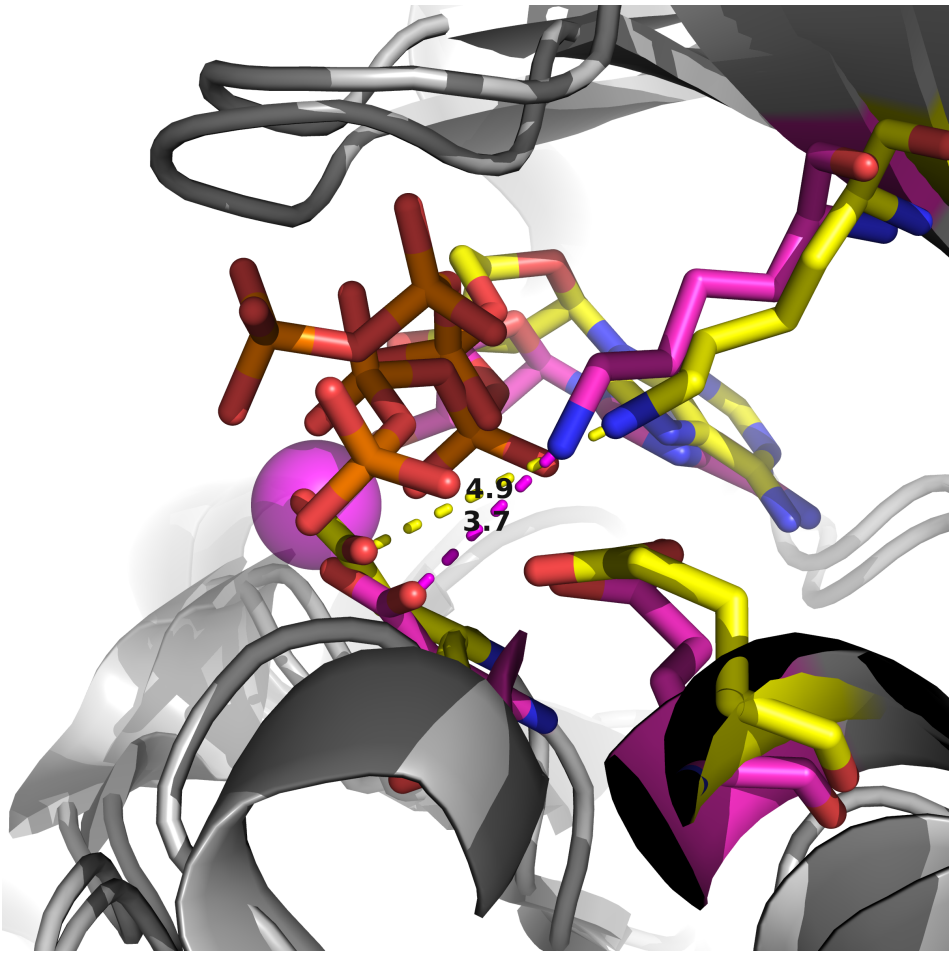

Figure S13. Structural conformation of Lys295, Glu310, and Asp404 in two ATP-Src crystal complexes in presence (violet) and without (yellow) Mg^{2+} cation. Pdb codes: 3a8w and 5drd, respectively.

References

1. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014,** *185* (2), 604-613.
2. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **2004,** *25* (13), 1605-12.
3. Shapovalov, M. V.; Dunbrack, R. L., Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **2011,** *19* (6), 844-58.
4. Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How does a drug molecule find its target binding site? *J Am Chem Soc* **2011,** *133* (24), 9181-3.
5. Bulusu, K. C.; Tym, J. E.; Coker, E. A.; Schierz, A. C.; Al-Lazikani, B. canSAR: updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res* **2014,** *42* (Database issue), D1040-7.
6. Decherchi, S.; Rocchia, W. A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale. *PLoS One* **2013,** *8* (4), e59744.
7. Decherchi, S.; Spitaleri, A.; Stone, J.; Rocchia, W. NanoShaper-VMD interface: computing and visualizing surfaces, pockets and channels in molecular systems. *Bioinformatics* **2019,** *35* (7), 1241-1243.
8. Decherchi, S.; Bottegoni, G.; Spitaleri, A.; Rocchia, W.; Cavalli, A. BiKi Life Sciences: A New Suite for Molecular Dynamics and Related Methods in Drug Discovery. *J Chem Inf Model* **2018,** *58* (2), 219-224.
9. Cowan-Jacob, S. W.; Fendrich, G.; Manley, P. W.; Jahnke, W.; Fabbro, D.; Liebetanz, J.; Meyer, T. The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. *Structure* **2005,** *13* (6), 861-71.
10. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* **2006,** *25* (2), 247-60.
11. Parrinello, M. R., A.; . Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981,** *52* (12).
12. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **2007,** *126* (1), 014101.
13. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995,** *103* (19), 8577-8593.
14. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29* (7), 845-54.