PNAS

www.pnas.org

Supplementary Information for

Encounter Complexes and Hidden Poses of Kinase-Inhibitor Binding on the Free-Energy Landscape

Suyong Re, Hiraku Oshima, Kento Kasahara, Motoshi Kamiya, and Yuji Sugita

Yuji Sugita

Email: sugita@riken.jp

This PDF file includes:

Supplementary text

Figures S1 to S12

Tables S1 to S3

Legends for Movies S1 to S4

SI References

Other supplementary materials for this manuscript include the following:

Movies S1 to S4

Supplementary Information Text

Simulation Setup. The system and simulation parameters were set as similar as possible to the previous microsecond-long MD simulation (1). The initial configuration of PP1-Src complex was constructed from the X-ray structure of unphosphorylated c-Src in active conformation (PDBID:1Y57) (2). We replaced a co-crystalized inhibitor with PP1, where the crystal structure of an inactive form of Hck complex with PP1 (PDBID:1QCF) (3) was used as a template. The resultant PP1-Src complex displays nearly identical binding site interactions to those of the templated X-ray structure (Fig. S1). A sulfate ion as well as regions irrelevant to the current study (regions other than the kinase domain of residues 82-258) were removed. The PP1-Src complex was solvated by 7,698 water molecules and neutralized by six sodium cations. The system contains 27,549 atoms in a box with the dimensions of 76 Å × 64 Å × 56 Å.

All simulations were performed using a development version of the GENESIS program package (4, 5). We used the AMBER ff99SB-ILDN (6, 7) and TIP3P (8) parameters for the protein and water molecules, respectively. The ligand parameters were obtained by using GAFF with AM1-BCC (9). All bonds involving hydrogen atoms were constrained using the SHAKE algorithm (10) and water molecules were kept rigid using the SETTLE algorithm (11). Particle-mesh Ewald summation (12, 13) was used to evaluate long-range electrostatic interactions, while non-bonded interactions were truncated at a cutoff distance of 8 Å. The system was initially equilibrated in an NPT ensemble (310 K, 1 atm) for 1 ns by conventional MD simulation. Langevin dynamics with a friction parameter of 1 ps^{-1} was used for temperature and pressure control (14). A time step of 2 fs was used. The rest of the simulations were performed in an NVT ensemble (310 K), where the Langevin dynamics with a friction parameter of 1 ps^{-1} was used for temperature control (15).

Two-Dimensional gREST/REUS Simulation. Multidimensional REMD including replica-exchange umbrella sampling (REUS) and replica-exchange with solute tempering (REST) are described in refs (16, 17) and (18, 19), respectively. A two-dimensional REST/REUS approach was successfully used to predict protein-ligand binding structures (20). Briefly, in the first

dimension a REUS simulation is carried out along the protein-ligand distance. The exchanges of restraining potentials help the ligand escape from the stable bound states, which hardly occurs in the conventional MD simulation. The second dimension further weakens the protein-ligand interaction by temperature replica exchange in a pre-defined solute molecule (typically a ligand molecule) using REST.

gREST/REUS is a simple extension of the original REST/REUS method. We replace the REST with a new scheme, referred to as generalized REST (gREST) (21). The gREST provides a flexible framework for solute definition. The solute can be defined as either part of the molecule and/or part of potential energy terms. This allows us to define the solute region as any desired energy terms (such as dihedral angle and nonbonded energy terms) of the ligand and the binding site residues. This definition not only weakens the protein-ligand interactions but also makes the protein binding site residues more flexible and accelerates water diffusion near the binding pocket. A comparison of the scaling of potential energy terms between the original REST/REUS and the present gREST/REUS is given in Table S1. The gREST/REUS scales the nonbonded interactions between the ligand and the binding site residues (Site-Ligand (C, LJ)) more strongly than REST/REUS. The dihedral angle terms of the binding site residues (Site-Site (D)) are scaled only in gREST/REUS. gREST/REUS also scales the nonbonded interactions between the binding site residues and water molecules (Site-Water (C, LJ)). We show that this simple modification significantly enhances the sampling of binding and unbinding events.

In the present work, we defined the solute region as the dihedral angle and nonbonded energy terms of an inhibitor PP1 and ten binding site residues. The binding site residues were defined as the SITE residues (Leu15, Val23, Ala35, Ile78, Thr80, Glu81, Met83, Ser87, Leu135, and Asp146) in the X-ray structure (PDBID:1QCF). Eight replicas were used to cover the solute temperature range of 310 ~ 3,100 K (T = 310.0, 387.5, 492.1, 632.7, 837.8, 1,148.2, 1,631.6, and 3,100.0 K). The reaction coordinate (ξ) of the umbrella sampling simulation (REUS dimension) was taken as the distance between the centers of mass (COMs) of PP1 and the backbone heavy atoms of two binding site residues (Ala35 and Leu135). The ξ value in the X-ray structure (1QCF) is 3.25 Å. 18 replicas were used to cover the ξ values in the range of 3.0 Å ~ 15.0 Å (ξ = 3.0, 3.5,

4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 9.0, 10.0, 11.0, 12.0, 13.0, 14.0, 15.0). We set the force constants of the restraining potentials to be 4.0 and 2.0 kcal/mol/Å^2 for the first 11 replicas and for the rest of the replicas, respectively.

The total number of replicas in gREST/REUS simulation is $18 \times 8 = 144$. Replica exchange was attempted every 2 ps alternatively for the REUS and gREST dimensions. The initial structures for gREST/REUS simulation were obtained from the preceding umbrella sampling simulations. First PP1 was pulled away from the binding site (from $\xi = 3.5$ Å to $\xi = 15.0$ Å), followed by a reverse simulation (from $\xi = 15.0$ Å to $\xi = 3.0$ Å). The final 18 structures with different ξ values were used for the subsequent gREST/REUS simulation. Following 1 ns of equilibration run for each replica, gREST/REUS for 300 ns per replica (43.2 μs in total) was performed for the purpose of analyzing the PP1-Src binding.

Efficiency of gREST/REUS Simulation. We first compared the performance of the original REST/REUS (20) and gREST/REUS simulations of a 100 ns each per replica independently. In the simulations, the positions of protein atoms that are more than 10 Å away from PP1 position in the X-ray structure were restrained by a harmonic potential with a force constant of 1 kcal/mol/Å in both simulations, for simplify the comparison. We computed the minimum values of heavy atoms root mean-square displacement (RMSD) of the ligand PP1 from the X-ray structure (PDBID: 1QCF) in each replica for the simulations (Fig. S2 and Table S2). The RMSDs for gREST/REUS are significantly smaller than those of REST/REUS in most of replicas. In gREST/REUS, on average, 67% of the replicas have the RMSD below 1.0 Å (Table S2), three times higher than REST/REUS (22%). Fig. S3 shows representative trajectories that demonstrates binding and unbinding events regardless of the initial condition. gREST/REUS simulation samples roughly one hundred binding (and unbinding) events in total (144 replicas $\times$ 0.67).

Fig. S4 shows free-energy profiles at 310 K along the PP1-Src distance (ξ) for the two simulations, obtained using the weighted histogram analysis method (WHAM) (22). The profiles are shown while appending the data every 10 ns to demonstrate the convergence. The free-energy profile of gREST/REUS rapidly converges, while that of REST/REUS slowly changes until 100

ns. REST/REUS heats up only PP1 and likely samples irrelevant conformations. Nonetheless, the two simulations correctly predict the PP1-Src distance in the X-ray structure ($\xi = 3.25$ Å) as the most stable state in the free-energy profiles.

The superior performance of gREST/REUS can be ascribed to the enhanced flexibility of the structures and the interactions in the protein binding site. First, in gREST/REUS, the fluctuations of the binding site residues, which are relevant for drug potency and resistance (23, 24), increase with solute temperature (Fig. S5 and Movies S1 and S2). Second, interactions between the binding site residues and water molecules are weakened at high solute temperatures (Table S1), and thereby water dynamics around the binding site becomes remarkably fast at the temperatures (Fig. S6).

We also carried out gREST/REUS simulation of a 300 ns each per replica without the weak positional restraints on protein atoms. This simulation result was examined by all analyses in the main text. We find that 78% of the trajectories without the restraints hit the canonical binding pose, compared to 67% with restraints (Table S2). In addition, the simulation without the restraints rapidly converges within 100 ns (Fig. 2*B* in the main text). The free-energy profile without the restraints is slightly changed. First, the well depth becomes shallower, likely because the protein flexibility entropically contributes to the stabilization of the unbound state. Second, the energy barrier around $\xi = 6$ Å disappears. The restraints on the protein atoms likely lead to steric hindrance near the narrow cleft between N- and C-lobes. To check the convergence of other degrees of freedom, we also calculated heavy atoms RMSDs and Cα atoms root mean-square fluctuations (RMSFs) of the protein. The values were calculated for three replicas at 310 K, representing distinct states of the binding process: the first replica with parameter index 1 ($\xi = 3.0$ Å) for bound, the second replica with parameter index 9 ($\xi = 7.0$ Å) for intermediate, and the last with parameter index 18 ($\xi = 15.0$ Å) for encounter states. In all cases, the RMSD values reasonably converges during the 300 ns simulation time (Fig. S7*A*). In Fig. S7*B*, the RMSF values calculated using either the first half (150 ns) or the last half (150 ns) of the simulation are compared with the values from whole 300 ns of the simulation data. The RMSF values well overlap with each other, confirming the convergence, at least within the active state of c-Src

kinase.

Conventional MD Simulations. Conventional MD simulations were also performed for comparison. System setup and simulation parameters are identical to the ones described in Simulation Setup above except for two points. First, the simulation box contains ten PP1 molecules (~ 36 mM concentration, 45,311 atoms in total) in order to enhance the occurrence of binding events. Second, a weak repulsive potential was applied between PP1 molecules to avoid their aggregation. We performed three independent simulations (15 μs in total) and binding events were observed in two out of three simulations (after 4 μs and 3.2 μs in the first and second simulations, respectively). Following Shan et al.,(1) we estimated the k_{on} value to be 4.6 μM^{-1}, where a binding frequency of 0.16 μs^{-1} (2 binding / (4 μs + 3.2 μs + 5 μs)) was used.

Free Energy Analysis. A set of two-dimensional free-energy landscapes were constructed along the reaction coordinate (ξ) and with either PP1 position or orientation with respect to the protein. We define the position and orientation of PP1 using six anchor atoms (P1, P2 and P3 from the protein and L1, L2, and L3 from PP1) as done in the staged binding free energy calculations (25, 26). Each anchor atom is defined using the COM of a group of atoms: (L1) A non-hydrogen PP1 atom, L0, closest to the COM of the PP1 and the atoms bonded to L0. (P1) Protein backbone heavy atoms of the residue closest to the COM of the protein. (P2) Protein backbone heavy atoms of the residue satisfying $30° \leq \angle L1P1P2 \leq 150°$. (P3) Protein backbone heavy atoms of a residue satisfying $30° \leq \angle P1P2P3 \leq 150°$. (L2) A PP1 heavy atom satisfying $30° \leq \angle P1L1L2 \leq 150°$ and the atoms bonded to the heavy atom. (L3) A PP1 heavy atom satisfying $30° \leq \angle L1L2L3 \leq 150°$ and the atoms bonded to the heavy atom. The position and orientation of PP1 are designated by Polar angles, θ (P2-P1-L1) and φ (P3-P2-P1-L1), and Euler angles, α (P1-L1-L2), β (P2-P1-L1-L2) and γ (P1-L1-L2-L3). The fluctuations of the protein anchor atoms in the simulation trajectory are less than 0.4 Å, which are sufficiently stable for the measurements.

Two-dimensional (θ/φ/α/β/γ, ξ) free-energy landscapes were constructed using the multistate Bennett acceptance ratio (MBAR) method (27). First the ξ distribution of each of the 18 replicas

at 310 K were obtained. Then, MBAR was used to calculate the weight factor for each snapshot. The calculated weight factors were applied to the two-dimensional distribution. Free-energy landscapes were constructed by taking negative logarithm of the final distribution and multiplying by RT, where R is the gas constant and T is the temperature.

Clustering Analysis. The trajectory data at 310 K from gREST/REUS simulation was analyzed using standard k-means clustering algorithm in the GENESIS analysis tool, which uses Euclidean distance function to classify the structures into a predefined number of clusters (4, 5). The trajectory data was divided by five sub-regions along the PP1-Src distance (ξ): (1) replica 1-3 (ξ = 3.0 ~ 4.0 Å, blue), (2) replica 4-5 (ξ = 4.5 ~ 5.0 Å, cyan), (3) replica 6-9 (ξ = 5.5 ~ 7.0 Å, green), (4) replica 10-14 (ξ = 7.5 ~ 11.0 Å, magenta), and (5) replica 15-18 (ξ = 12.0 ~ 15.0 Å, red). The clustering analysis was performed for each sub-region. We set the number of clusters equals ten for each analysis. The resultant clusters are summarized in Fig. S8 (the radius values of 2.0 Å on average). According to the cluster distribution, we define the PP1-Src distance of $\xi < 6.0$ Å as the bound region (B), that of 6.0 Å $< \xi < 9.0$ Å as the intermediate region (I), and that of $\xi > 9.0$ Å as the encounter region (E), respectively. Each cluster is further classified by the PP1 orientation (angles β and γ, Fig. S9): $\mathbf{X^{zz}}$ (around $\beta = 0^o$ and $\gamma = 0^o$), $\mathbf{X^{zp}}$ (around $\beta = 0^o$ and $\gamma = 180^o$), $\mathbf{X^{pz}}$ (around $\beta = 180^o$ and $\gamma = 0^o$), and $\mathbf{X^{pp}}$ (around $\beta = 180^o$ and $\gamma = 180^o$) (X = B, I, and E). In the bound region, we find five representative poses, including the canonical binding pose **I** and semi-bound poses **II-V** (Fig. 2*D* in the main text). The poses **I**, **II**, **III**, **IV**, and **V** are respectively assigned as $\mathbf{B^{zz}}$, $\mathbf{B^{pp}}$, $\mathbf{B^{pz}}$, $\mathbf{B^{zz^*}}$, and $\mathbf{B^{zp}}$.

Calculation of Binding Free Energy. The binding free energy of PP1-Src was calculated using the free energy perturbation (FEP) method implemented in the development version of GENESIS program package (4, 5). The binding free energy in solution (ΔG_{bind}) is calculated as follows (28, 29):

P•L (solution) $\rightarrow$ P (solution) + L (solution), $\qquad \Delta G_{bind} = \Delta G_{complex} - \Delta G_{solution}$

P•L (solution) $\rightarrow$ P (solution) + L (gas), $\qquad \Delta G_{complex}$

L (solution) $\rightarrow$ L (gas), $\qquad \Delta G_{solution}$

P and L stand for the protein (Src) and the ligand (PP1), respectively. Two simulations each were performed for the solvated PP1-Src complex (for $\Delta G_{complex}$) and the solvated PP1 (for $\Delta G_{solution}$) system. In both simulations, the intermolecular interactions of the ligand (PP1) to the other molecules (Src and water molecules) are gradually turned off using the coupling parameter λ:

$$U(\lambda^{el}, \lambda^{LJ}) = \sum_{nonbond} \lambda^{LJ} \varepsilon \left[\left(\frac{R^2_{min,ij}}{r^2_{ij} + \alpha^{sc}(1 - \lambda^{LJ})} \right)^6 - 2 \left(\frac{R^2_{min,ij}}{r^2_{ij} + \alpha^{sc}(1 - \lambda^{LJ})} \right)^3 \right]$$

$$+ \sum_{nonbond} \lambda^{el} \frac{q_i q_j \text{erfc}\left(\alpha \sqrt{r^2_{ij} + \beta^{sc}(1 - \lambda^{el})} \right)}{\varepsilon_1 \sqrt{r^2_{ij} + \beta^{sc}(1 - \lambda^{el})}} + \text{(PME reciprocal + self)}$$

where ε and R_{min} are the well depth and minimum distance of a Lennard-Jones (LJ) potential, q_i is the atomic charge, ε_1 is the effective dielectric constant, α is the exponent of the complementary error function, and r_{ij} is the distance between two atoms. λ^{LJ} and λ^{el} are the coupling parameters for LJ and Coulomb terms, respectively (λ=1 for full interaction and λ=0 for no interaction). α^{sc} and β^{sc} are scaling factors of the soft-core potentials for the LJ and Coulomb terms, respectively (30, 31). In this work, we use α^{sc} and β^{sc} values of 5.0 and 0.5, respectively. The intermolecular interactions are turned off through 32 windows in the order of Coulomb terms followed by LJ terms: 11 λ^{el}(1.0, 0.9, 0.75, 0.6, 0.45, 0.35, 0.275, 0.200, 0.125, 0.050 and 0) and 21 λ^{LJ}(1.0, 0.9, 0.8, 0.7, 0.6, 0.525, 0.450, 0.400, 0.350, 0.325, 0.300, 0.275, 0.250, 0.225, 0.200, 0.175, 0.150, 0.125, 0.100, 0.050 and 0).

FEP calculations were performed for the five bound poses of Fig. 2*D* in the main text (Pose **I**, **II**, **III**, **IV**, and **V**). For $\Delta G_{solution}$ calculation, the solvated PP1 system was independently constructed (PP1 solvated by 4,191 water molecules containing 12,613 atoms in a box of dimensions 50 Å $\times$ 54 Å $\times$ 53 Å). Each system was first minimized and then gradually heated to 310 K in an NVT ensemble followed by 100 ps of equilibration in an NPT ensemble. To preserve the binding pose during these steps, the positions of the protein backbone and PP1 atoms were restrained by a harmonic restraint potential with a force constant of 10 kcal/mol/Å. Finally, the system was equilibrated for 20 ns in an NVT ensemble with and without the restraints depending on the system (no restraints for pose **I**, 1 kcal/mol/Å for pose **IV**, **V** and **III**, 10 kcal/mol/Å for **II**). Final configurations were used for subsequent FEP calculations. We conducted 12 independent FEP calculations (NVT ensemble, 310 K) with different initial momentum

distributions. In each calculation, the simulation was run for 1 ns per each λ (1 ns $\times$ 32 windows $\times$ 12 calculations = 384 ns in total). Finally, the free energy differences, $\Delta G_{complex}$ and $\Delta G_{solution}$, are estimated using Bennett's Acceptance Ratio (BAR) method (32). The initial 100 ps of data was discarded from the final averaging. The results are summarized in Table S3. Although the binding free energy value of the canonical pose (pose **I**) agrees with the value obtained by converting the experimental IC$_{50}$ (170 nM, assuming K_d = IC$_{50)}$ (33), we consider this a coincidence considering the crude assumption. In pinciple, the observed K_d involves contributions from all possible binding poses. For a precise estimation of the binding free energy, we consider multiple binding poses as follows (34):

$$\Delta G_{mbp} = -RT\ln\left(\frac{1}{n}\sum_{i=1}^{n}\exp(-\beta\Delta G_i)\right)$$

We considered the canonical pose **I** and semi-bound poses **II**, **III**, **IV** and **V**, thus n = 5. The obtained value of -8.5 kcal/mol is in reasonable agreement with the experimental IC$_{50}$.
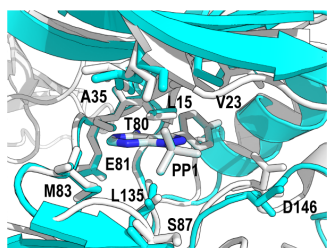
Information of the SITE residues in the X-ray structure (PDBID: 1QCF)

```
REMARK 800 SITE
REMARK 800 SITE_IDENTIFIER: AC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE PP1 A 532

SITE     1 AC1 12 LEU A 273  VAL A 281  ALA A 293  ILE A 336
SITE     2 AC1 12 THR A 338  GLU A 339  MET A 341  SER A 345
SITE     3 AC1 12 LEU A 393  ASP A 404  HOH A 582  HOH A 774
```

Fig. S1. Binding site interactions of PP1-Src complex (cyan) superimposed on the X-ray structure (white, PDBID: 1QCF). The information of the SITE residues in the X-ray structure is also shown.
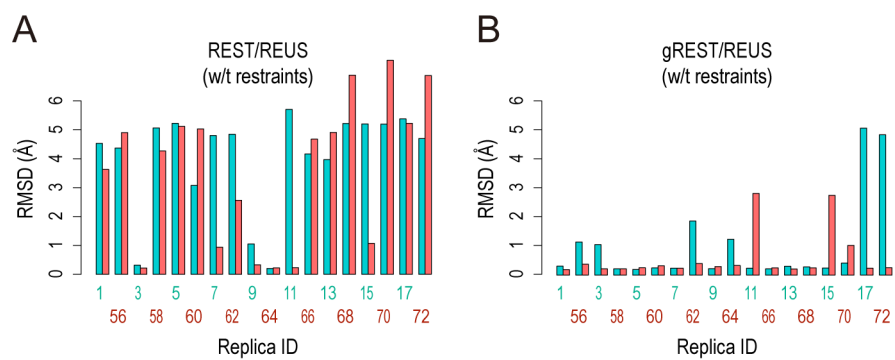
Fig. S2. Efficiency of simulations. (*A*) Minimum values of heavy atoms RMSD of the ligand PP1 from the X-ray structure (1QCF) for selected replicas (replica indexes of 1-18 and 55-72) for REST/REUS and (*B*) for gREST/REUS.
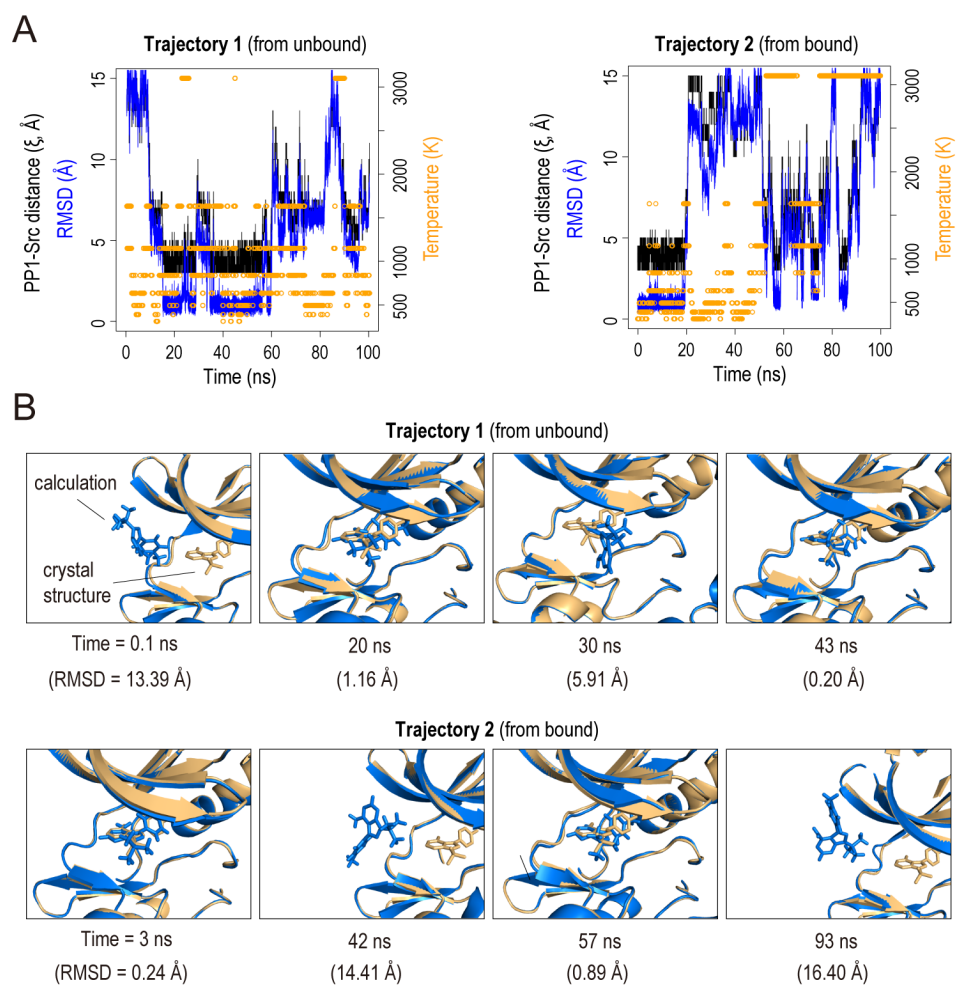
Fig. S3. Binding and unbinding trajectories of gREST/REUS simulations. (*A*) Time courses of the PP1-Src distance ξ (Å, black line), heavy atoms RMSD of the ligand PP1 from the X-ray structure (Å, blue line), solute temperature T (K, orange circle) for arbitrary selected two replicas, replica 5 (trajectory 1) and replica 7 (trajectory 2). Random walks in both PP1-Src distance and temperature spaces are recognized. (*B*) Representative snapshots from trajectories 1 and 2 (blue) superimposed on the X-ray structure (gold).
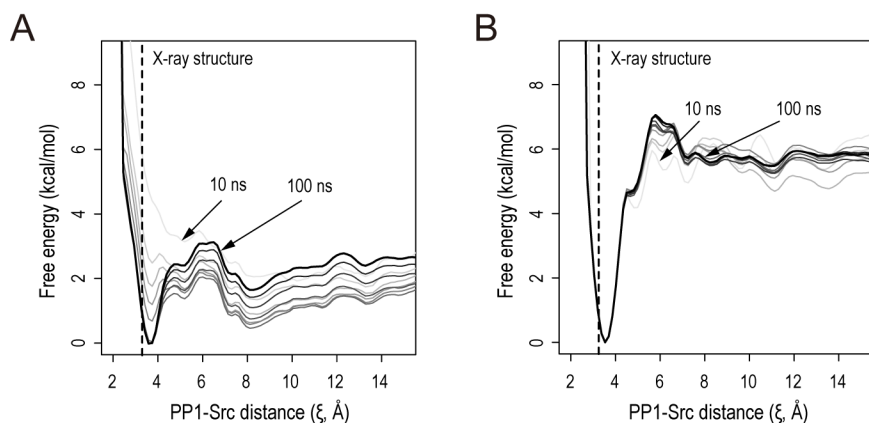
Fig. S4. Convergence of simulations. (*A*) Free-energy profiles at 310 K along the PP1-Src distance (ξ) for REST/REUS and (*B*) for gREST/REUS. The profiles are drawn while appending the data every 10 ns (from light gray to black) to show convergence. The PP1-Src distance (ξ) in the X-ray structure (1QCF, ξ = 3.25 Å) is shown as a dotted line.
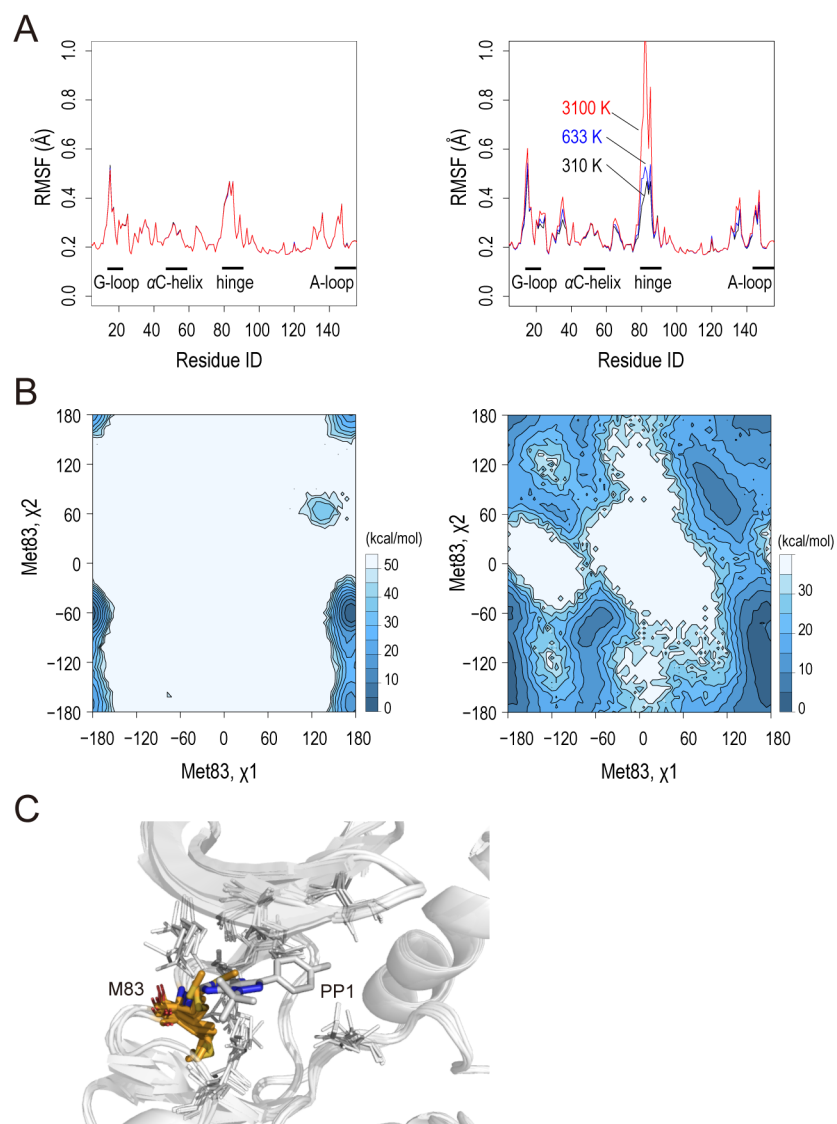
Fig. S5. Fluctuations of binding site residues. (*A*) The root mean-square fluctuations (RMSFs) of the binding site including the gatekeeper (Thr80) and hinge (Glu81, and Met83) residues for REST/REUS (left) and gREST/REUS (right). In both cases, a 100 ns trajectory at $\xi = 15.0$ Å was used. RMSF is shown for different solute temperatures: 310 K (black), 633 K (blue), and 3,100 K (red). (*B*) Free-energy profile of the side chain orientation (χ1 and χ2) of Met83 at a solute temperature of 3,100 K for REST/REUS (left) and gREST/REUS (right). χ1 and χ2 are defined as the dihedral angles C-CA-CB-CG and CA-CB-CG-SD, respectively. (*C*) A collection of snapshots at a solute temperature of 3,100K taken from the gREST/REUS simulation. Met83 is shown in orange. The bound PP1 position is also shown for comparison.
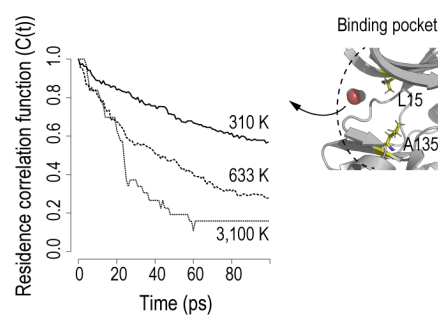
Fig. S6. Hydration dynamics of the binding pocket. The binding pocket is defined as the region within 10 Å from the Cα atoms of Leu15 and Ala135. A single replica 10-ns gREST simulation was performed at three temperatures (310 K, 633 K, and 3,100 K) for the analysis. For each simulation trajectory, water molecules initially inside the pocket were selected and probed to construct the function θ (θ =1 if water is inside the pocket, θ =0 otherwise). Residence correlation functions of water molecules inside the binding pocket are calculated for each temperature, 310 K (solid line), 633 K (dashed line), and 3,100 K (dotted line), using $C(t) = \frac{1}{N_h}\sum_{i=1}^{N_h}[\theta_i(0)\theta_i(t)]$, where N_h is the number of water molecules inside the pocket.
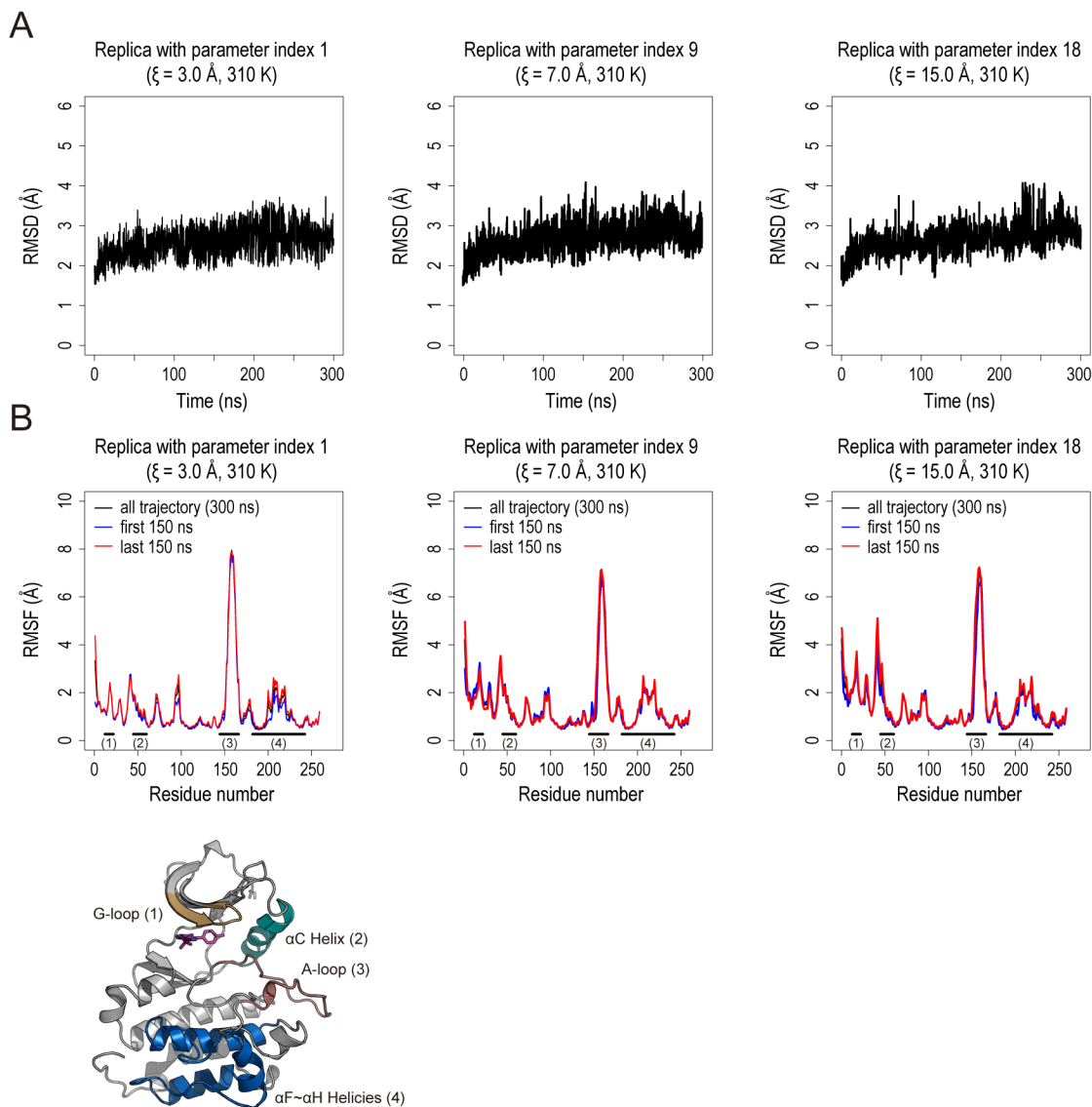
Fig. S7. Convergence of simulations (protein structure): (*A*) Time course of heavy atoms RMSD of the protein during the 300 ns simulation. The values were calculated for three replicas at 310 K, representing distinct states of the binding process: the first replica with parameter index 1 (ξ = 3.0 Å) for bound, the second replica with parameter index 9 (ξ = 7.0 Å) for intermediate, and the last with parameter index 18 (ξ = 15.0 Å) for encounter states. (*B*) Cα atoms RMSFs of the protein calculated for the three replicas (indices 1, 9, and 18) at 310 K using either the first half (150 ns) (blue) or the last half (150 ns) (red) of the simulation in comparison with the values from the whole 300 ns of the simulation data (black). In both RMSD and RMSF calculations, C-terminus residues (residue index 261–275) were excluded for clarity.

A

B

(1) replica1-3
(ξ = 3.0~4.0 Å)

(2) replica4-5
(ξ = 4.5~5.0 Å)

(3) replica6-9
(ξ = 5.5~7.0 Å)

(4) replica10-14
(ξ = 7.5~11.0 Å)

(5) replica15-18
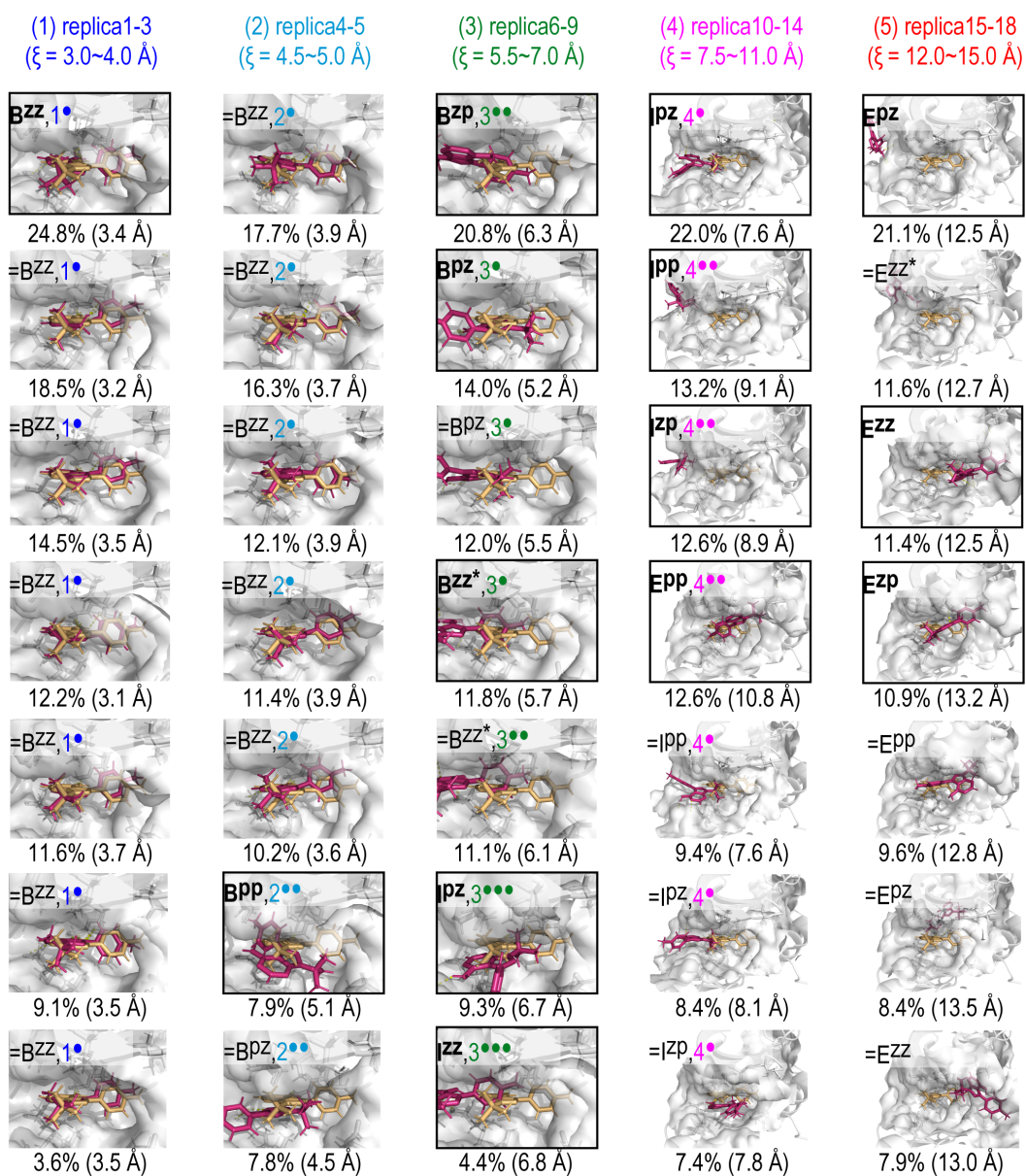**(ξ = 12.0~15.0 Å)**

B^{zz},1●	=B^{zz},2●	**B^{zp}**,3●●	**I^{pz}**,4●	**E^{pz}**
24.8% (3.4 Å)	17.7% (3.9 Å)	20.8% (6.3 Å)	22.0% (7.6 Å)	21.1% (12.5 Å)
=B^{zz},1●	=B^{zz},2●	**B^{pz}**,3●	**I^{pp}**,4●●	=E^{zz*}
18.5% (3.2 Å)	16.3% (3.7 Å)	14.0% (5.2 Å)	13.2% (9.1 Å)	11.6% (12.7 Å)
=B^{zz},1●	=B^{zz},2●	=B^{pz},3●	**I^{zp}**,4●●	**E^{zz}**
14.5% (3.5 Å)	12.1% (3.9 Å)	12.0% (5.5 Å)	12.6% (8.9 Å)	11.4% (12.5 Å)
=B^{zz},1●	=B^{zz},2●	**B^{zz*}**,3●	**E^{pp}**,4●●	**E^{zp}**
12.2% (3.1 Å)	11.4% (3.9 Å)	11.8% (5.7 Å)	12.6% (10.8 Å)	10.9% (13.2 Å)
=B^{zz},1●	=B^{zz},2●	=B^{zz*},3●●	=I^{pp},4●	=E^{pp}
11.6% (3.7 Å)	10.2% (3.6 Å)	11.1% (6.1 Å)	9.4% (7.6 Å)	9.6% (12.8 Å)
=B^{zz},1●	**B^{pp}**,2●●	**I^{pz}**,3●●●	=I^{pz},4●	=E^{pz}
9.1% (3.5 Å)	7.9% (5.1 Å)	9.3% (6.7 Å)	8.4% (8.1 Å)	8.4% (13.5 Å)
=B^{zz},1●	=B^{pz},2●●	**I^{zz}**,3●●●	=I^{zp},4●	=E^{zz}
3.6% (3.5 Å)	7.8% (4.5 Å)	4.4% (6.8 Å)	7.4% (7.8 Å)	7.9% (13.0 Å)

17

Figure S8. Results of clustering analysis. (*A*) The cluster distribution of five sub-regions along the PP1-Src distance (ξ): (1) replica 1-3 ($\xi = 3.0 \sim 4.0$ Å, blue), (2) replica 4-5 ($\xi = 4.5 \sim 5.0$ Å, cyan), (3) replica 6-9 ($\xi = 5.5 \sim 7.0$ Å, green), (4) replica 10-14 ($\xi = 7.5 \sim 11.0$ Å, magenta), and (5) replica 15-18 ($\xi = 12.0 \sim 15.0$ Å, red). (*B*) The snapshots of top seven clusters (occupy ~80%) in each sub-region. The X-ray structure (1QCF, gold) is superposed on the snapshots as a reference. The relative population of each cluster and the representative ξ value (in parenthesis) are also written. Each snapshot is classified by the PP1-Src distance (Bound region (B, $\xi < 6.0$ Å), intermediate region (I, 6.0 Å $< \xi < 9.0$ Å), encounter region (E, $\xi > 9.0$ Å)) and the PP1 orientation ($\mathbf{X^{zz}}$ (around $\beta = 0^{o}$ and $\gamma = 0^{o}$), $\mathbf{X^{zp}}$ (around $\beta = 0^{o}$ and $\gamma = 180^{o}$), $\mathbf{X^{pz}}$ (around $\beta = 180^{o}$ and $\gamma = 0^{o}$), and $\mathbf{X^{pp}}$ (around $\beta = 180^{o}$ and $\gamma = 180^{o}$) (X = B, I, and E)). Representative structures are highlighted with black border.

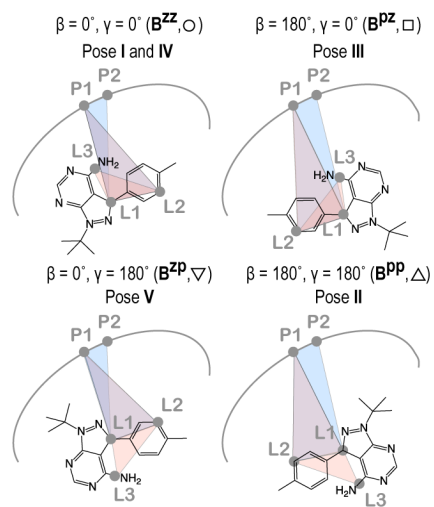Fig. S9. Four typical bound orientations: $\mathbf{B^{zz}}$ (β, $\gamma = 0^o$, 0°: $\bigcirc$), $\mathbf{B^{zp}}$ (β, $\gamma = 0^o$, 180°: $\bigtriangledown$), $\mathbf{B^{pz}}$ (β, $\gamma = 180^o$, 0°: $\square$) and $\mathbf{B^{pp}}$ (β, $\gamma = 180^o$, 180°: $\triangle$). The planes defining the two dihedral angles, β and γ, are shown in blue and pink, respectively.
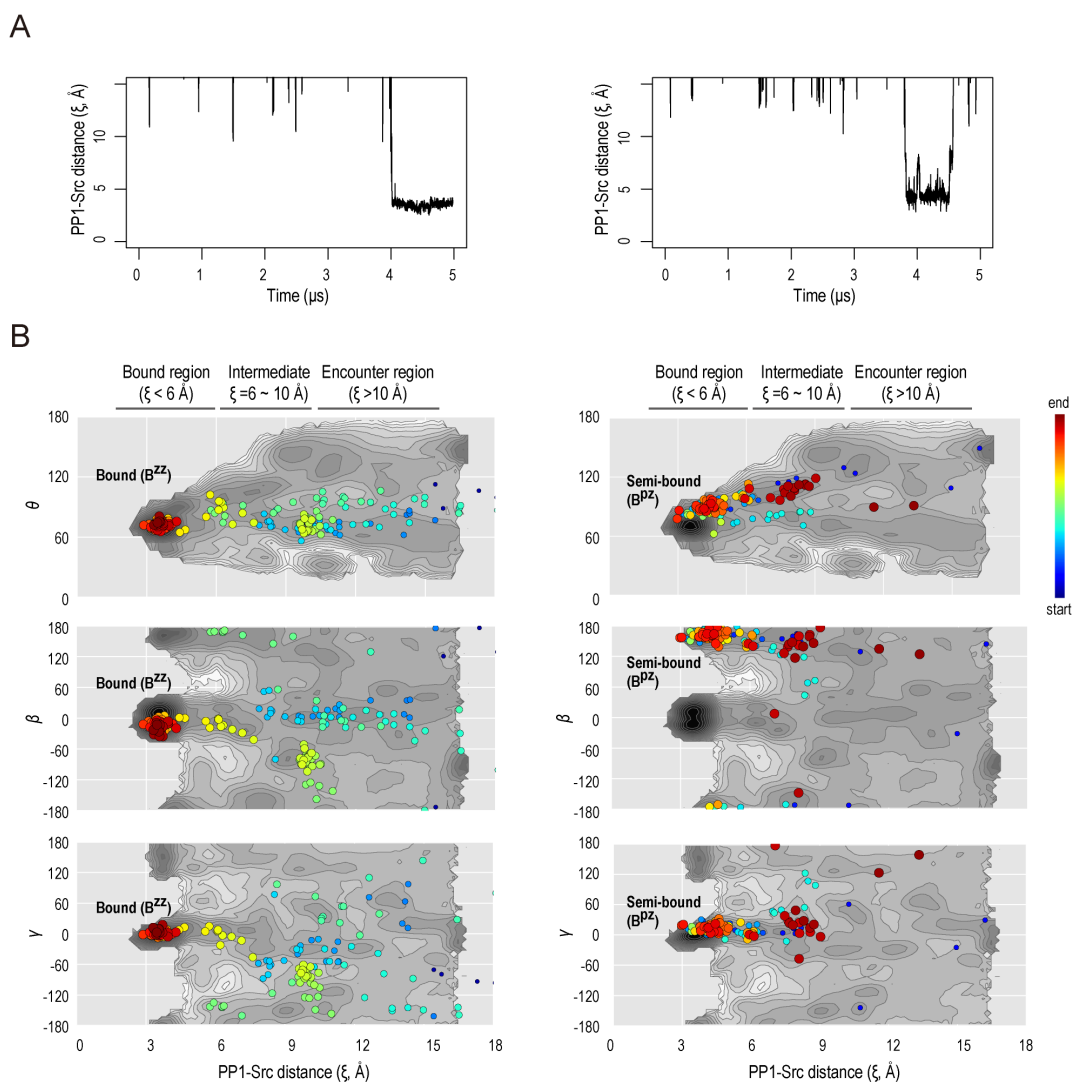
Fig. S10. Trajectories from conventional MD simulations. (*A*) Time courses of the PP1-Src distance ξ (Å, black line) for trajectories leading to the binding pose $\mathbf{B^{zz}}$ (left) and the semi-bound pose $\mathbf{B^{pz}}$ (right). (*B*) Superposition of the corresponding trajectories (toward $\mathbf{B^{zz}}$ (left) and $\mathbf{B^{pz}}$ (right)) on the free-energy landscapes of the PP1 position (θ, ξ) and orientation (β/γ, ξ). The color of each trajectory gradually changes from blue (small circle) to red (large circle) along the course of the simulation.

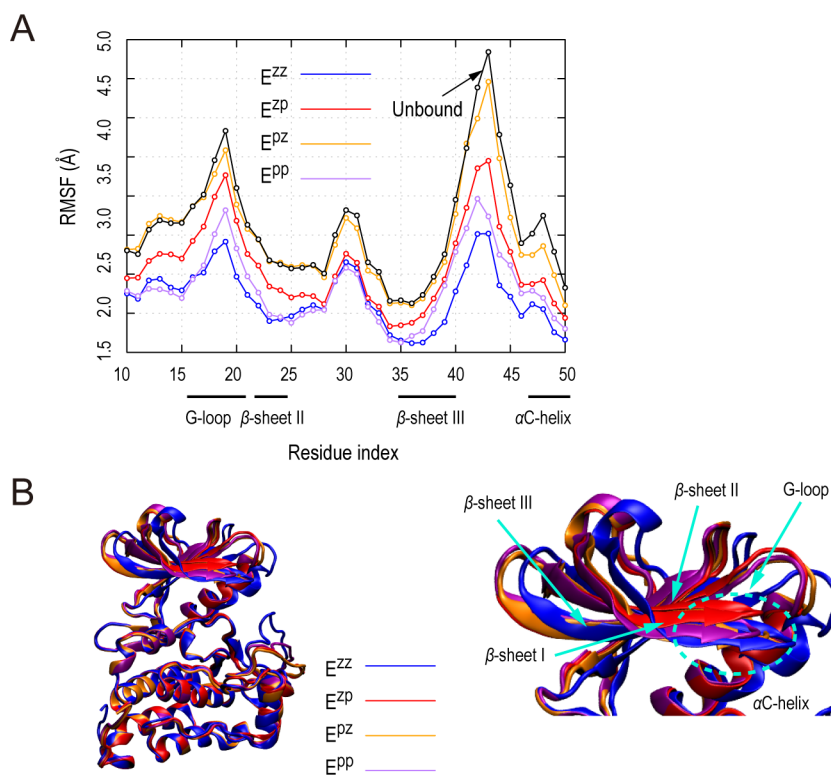Fig. S11. G-loop fluctuations in the encounter complexes. (a) RMSF values of the G-loop residues for four encounter complexes, $\mathbf{E^{zz}}$, $\mathbf{E^{zp}}$, $\mathbf{E^{pz}}$, and $\mathbf{E^{pp}}$, for gREST/REUS simulation. The RMSF value of the unbound state were calculated using the trajectory of replica for $\xi = 15$Å. RMSF values are calculated after fitting to a reference frame by the portion of C-lobe. (b) A superposition of the four encounter complex structures.
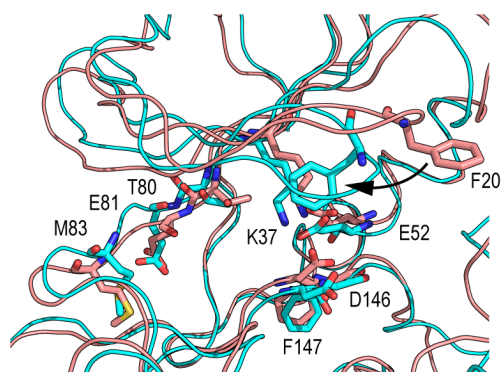
Fig. S12. Changes in the binding site during the PP1 intrusion step: Projection of two structures corresponding to the encounter complex ($\mathbf{E^{zz}}$, pink) and the bound state ($\mathbf{B^{zz}}$, cyan) along the major pathway. Key residues are represented as stick. In the encounter state, PP1 is located at the position of Phe20 in the bound state. Upon PP1 intrusion, Phe20 reorients to fill the space originally occupied by PP1.

Table S1. Comparison of potential energy scaling between REST/REUS and gREST/REUS.
The potential energy of a replica having solute temperature index of *m* in gREST is given as a
sum of solute-solute (E_{uu}), solute-solvent (E_{uv}), solvent-solvent (E_{vv}) energies.

$$E_m^{gREST} = \frac{\beta_m}{\beta_0} E_{uu} + \sum_i \left(\frac{\beta_m}{\beta_0}\right)^{\frac{k_i}{l_i}} E_{uv,i} + E_{vv}$$

Here, β_m and β_0 are the inverse of the solute temperature T_m for solute temperature
index *m* and of the simulation temperature T_0, respectively. k_i is the number of particles
involved in the solute region of the *i*-th solute-solvent interaction, while l_i is the maximum
number of particles involved in the *i*-th solute-solvent interaction. In REST/REUS, only the ligand
is treated as solute, whereas in gREST/REUS both the ligand and binding site residues (Site) are
treated as solute. In the latter, only dihedral-angle (D) and nonbonded energy terms (Coulomb
(C), and Lennard-Jones (LJ)) are treated as the solute.

	REST/REUS	gREST/REUS
Protein-Protein (D, C, LJ)	1	1
Protein-Site (D)	1	$(\beta_m/\beta_0)^{1/4 \sim 3/4}$
Protein-Site (C, LJ)	1	$(\beta_m/\beta_0)^{1/2}$
Protein-Ligand (C, LJ)	$(\beta_m/\beta_0)^{1/2}$	$(\beta_m/\beta_0)^{1/2}$
Protein-Water (C, LJ)	1	1
Site-Site (D, C, LJ)	1	β_m/β_0
Site-Ligand (C, LJ)	$(\beta_m/\beta_0)^{1/2}$	β_m/β_0
Site-Water (C, LJ)	1	$(\beta_m/\beta_0)^{1/2}$
Ligand-Water (C, LJ)	$(\beta_m/\beta_0)^{1/2}$	$(\beta_m/\beta_0)^{1/2}$
Water-Water (C, LJ)	1	1

Table S2. RMSD from the X-ray structure. The ratio of heavy atoms RMSD of the ligand PP1 values below 1 Å and the minimum value in each replica of the REST/REUS and gREST/REUS simulations are listed. The values are summarized for every 18 replicas.

Replica ID	REST/REUS (w/t restraints)		gREST/REUS (w/t restraints)		gREST/REUS (w/o restraints)	
	< 1 Å [%]	Min. [Å]	< 1 Å [%]	Min. [Å]	< 1 Å [%]	Min. [Å]
1-18	11	0.201	67	0.195	67	0.229
19-36	22	0.225	56	0.181	83	0.231
37-54	17	0.229	78	0.203	78	0.228
55-72	28	0.216	83	0.187	67	0.251
73-90	17	0.319	78	0.201	72	0.219
91-108	22	0.265	67	0.228	83	0.178
109-126	39	0.183	56	0.216	83	0.191
127-144	22	0.235	56	0.208	89	0.207
Average	22	0.234	67	0.202	78	0.217

Table S3. Binding free energy. The calculated absolute binding free energy of each pose (ΔG_{bind}) and the weighted average (ΔG_{mbp}) of multiple poses in kcal/mol together with the experimental IC$_{50}$.

	I	II	III	IV	V
ΔG_{bind} (kcal/mol)	-9.5	-2.8	-3.8	-5.1	-3.8
ΔG_{mbp} (kcal/mol)	-8.5				
Experimental IC$_{50}$(33)	-9.6				

Movie S1 (separate file). Flexibility of binding site residues (gREST/REUS)

Movie S2 (separate file). Flexibility of binding site residues (REST/REUS)

Movie S3 (separate file). Parallel binding pathway (distance vs ligand orientation)

Movie S4 (separate file). Parallel binding pathway (distance vs ligand position)

References

1. Shan Y, et al. (2011) How does a drug molecule find its target binding site? *J Am Chem Soc* 133(24):9181–9183.

2. Cowan-Jacob SW, et al. (2005) The Crystal Structure of a c-Src Complex in an Active Conformation Suggests Possible Steps in c-Src Activation. *Structure* 13(6):861–871.

3. Schindler T, et al. (1999) Crystal structure of Hck in complex with a Src family-selective tyrosine kinase inhibitor. *Mol Cell* 3(5):639–648.

4. Kobayashi C, et al. (2017) GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J Comput Chem* 38(25):2193–2206.

5. Jung J, et al. (2015) GENESIS: A hybrid-parallel and multi-scale molecular dynamics simulator with enhanced sampling algorithms for biomolecular and cellular simulations. *Wiley Interdiscip Rev Comput Mol Sci* 5(4):310–323.

6. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct Funct Bioinforma* 78(8):1950–1958.

7. Hornak V, et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct Funct Bioinforma* 65(3):712–725.

8. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926.

9. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general Amber force field. *J Comput Chem* 25(9):1157–1174.

10. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23(3):327–341.

11. Miyamoto S, Kollman PA (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* 13(8):952–962.

12. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19):8577–8593.

13. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N ·log(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089–10092.

14. Quigley D, Probert MIJ (2004) Langevin dynamics in constant pressure extended systems. *J Chem Phys* 120(24):11432–11441.

15. Adelman SA (1976) Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *J Chem Phys* 64(6):2375.

16. Sugita Y, Kitao A, Okamoto Y (2000) Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys* 113(15):6042–6051.

17. Fukunishi H, Watanabe O, Takada S (2002) On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J Chem Phys* 116(20):9058–9067.

18. Liu P, Kim B, Friesner RA, Berne BJ (2005) Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc Natl Acad Sci* 102(39):13749–13754.

19. Wang L, Friesner RA, Berne BJ (2011) Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (REST2). *J Phys Chem B* 115(30):9431–9438.

20. Kokubo H, Tanaka T, Okamoto Y (2013) Two-dimensional replica-exchange method for predicting protein-ligand binding structures. *J Comput Chem* 34(30):2601–2614.

21. Kamiya M, Sugita Y (2018) Flexible selection of the solute region in replica exchange with solute tempering: Application to protein-folding simulations. *J Chem Phys* 149(7):072304.

22. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13(8):1011–1021.

23. Buschbeck M (2006) Strategies to Overcome Resistance to Targeted Protein Kinase Inhibitors in the Treatment of Cancer. *Drugs R D* 7(2):73–86.

24. Getlik M, et al. (2009) Hybrid compound design to overcome the gatekeeper T338M mutation in cSrc. *J Med Chem* 52(13):3915–3926.

25. Wang J, Deng Y, Roux B (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys J* 91(8):2798–2814.

26. Jo S, Jiang W, Lee HS, Roux B, Im W (2013) CHARMM-GUI ligand binder for absolute binding free energy calculations and its application. *J Chem Inf Model* 53(1):267–277.

27. Shirts MR, Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* 129(12):124105.

28. Fujitani H, et al. (2005) Direct calculation of the binding free energies of FKBP ligands. *J Chem Phys* 123(8):084108.

29. Fujitani H, Tanida Y, Matsuura A (2009) Massively parallel computation of absolute binding free energy with well-equilibrated states. *Phys Rev E - Stat Nonlinear, Soft Matter Phys* 79(2):021914.

30. Zacharias M, Straatsma TP, McCammon JA (1994) Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J Chem Phys* 100(12):9025–9031.

31. Beutler TC, Mark AE, van Schaik RC, Gerber PR, van Gunsteren WF (1994) Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem Phys Lett* 222(6):529–539.

32. Bennett CH (1976) Efficient estimation of free energy differences from Monte Carlo data. *J Comput Phys* 22(2):245–268.

33. Hanke JH, et al. (1996) Discovery of a Novel, Potent, and Src Family-selective Tyrosine Kinase Inhibitor. *J Biol Chem* 271(2):695–701.

34. Kaus JW, et al. (2015) How to deal with multiple binding poses in alchemical relative protein-ligand binding free energy calculations. *J Chem Theory Comput* 11(6):2670–2679.