Supplementary Information

Methods

Molecular Dynamics

Molecular dynamics simulations were initiated from the crystal structure of Abl kinase bound to Imatinib in the DFG-out state (PDB id: 1OPJ). The residue N336 was mutated back to S336 to generate WT structure. Protonation states of titratable residues were assigned on the basis of pKa calculations performed using PROPKA 3.1[71,72] at pH 7. N368S mutant structure was prepared using the side-chain mutation module of MOE..[73] Using the solution builder module in CHARMM-GUI,[74,75] the Abl kinase-Imatinib crystal structure was solvated with TIP3P water molecules.[76] Finally, Na^+ and Cl^- ions were added, and the system was neutralized with the ionic concentration set to 100 mM. The final simulation box comprising of Abl kinase, Imatinib, water molecules, and ions was $\sim$55 K atoms. All simulations were performed on GROMACS2018 patched with PLUMED version 2.4[77,78] with a 2 fs time step. Simulations were performed utilizing CHARMM36 force field[79] for protein, TIP3P for water molecules[76] and, CHARMM General Force Field (CGenFF)[80] for Imatinib. Temperature and pressure were kept at 300 K and 1 bar using the velocity rescale thermostat[81] and Parrinello–Rahman barostat.[82] The non-bonded interactions were calculated with a 10 Å cutoff, and long-range electrostatics were calculated using the particle-mesh Ewald (PME) method.[83]

AMINO

The basis set of OPs used in this work was determined using AMINO.[29] AMINO reduces a large set of OPs to a minimally-redundant subset using k-medoids clustering with the

following mutual information based distance metric:

$$D(X;Y) = 1 - \frac{I(X;Y)}{H(X,Y)} \tag{1}$$

$$= 2 - \frac{\sum_{x\in X}\sum_{y\in Y} P(x,y) * log(P(x) * P(y))}{\sum_{x\in X}\sum_{y\in Y} P(x,y) * log(P(x,y))} \tag{2}$$

The input trajectory for AMINO was 100 ns of unbiased simulation for N368S and WT Abl bound to Imatinib with an initial set of 84 OPs corresponding to the heavy atom distance between Imatinib and the Cα atoms of Abl. The stationary probabilities for pairs of these OPs were estimated using histograms with 50 bins and clusters of up to 20 OPs were compared to determine the optimal number of outputs. The output of AMINO was a set of 5 OPs for each system.

SGOOP

The RCs used for enhanced sampling were linear combinations of the OPs found using SGOOP.[28,84] This framework uses a maximum caliber model[85,86] to construct a transition matrix along a given low-dimensional projection. The eigenspectrum associated with the transition matrix is then calculated and the largest gap between consecutive eigenvalues, called the spectral gap, is found. This spectral gap represents the timescale separation between slow and fast processes. SGOOP scans different linear combinations in order to find the RC which maximizes the spectral gap and in turn the timescale separation between slow and fast processes. In this work, SGOOP was used in an iterative manner using input trajectories from the most recent simulation.

The first round of SGOOP was performed on a well-tempered metadynamics trajectory using a crude trial RC to learn a new putative RC which was a linear combination of AMINO OPs. Using the learned RC, a well-tempered metadynamics simulation was performed, to accelerate Imatinib dissociation. The resultant biased trajectory was fed back to SGOOP to learn an improved RC. The improved RC was then used to bias the next round of well-

tempered metadynamics, subsequently, SGOOP was performed to further improve the RC. This process was iterated and at each step, the accelerated time and spectral gap were calculated. The iterative process was terminated when an improvement (lowering) in accelerated time was not observed (see SI Fig. 7, 9). Typically, this required 2-3 rounds of iterative SGOOP for both WT and N368S Abl systems.

Infrequent Metadynamics

We performed 11 independent infrequent metadynamics simulations starting from the crystal structure of WT and N368S Abl. Metadynamics was performed using the PLUMED implementation of well-tempered metadynamics with a bias factor of 10, an initial hill height of 1.5 kJ/mol, and bias deposited every 20 picoseconds.[77,78,87] Biases were deposited on the RC learned from the iterative SGOOP calculation as described previously. The sigma value of the Gaussian bias kernel was estimated by calculating the standard deviation of the biased RC from the 100 ns of equilibrium MD simulation of WT and N368S Abl. The choice of 11 simulations was made following results from other metadynamics drug unbinding studies that showed approximately 10 simulations provided sufficient statistics.[43,70]

P-value Analysis

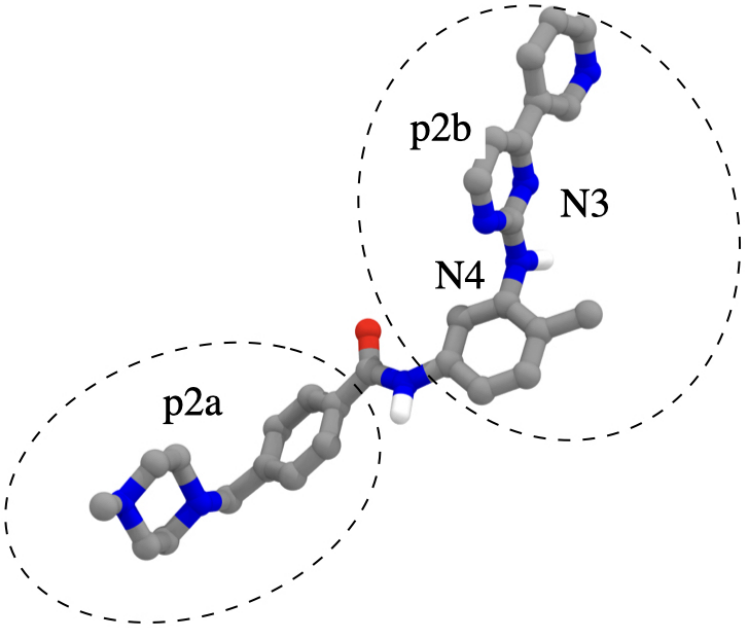
P-value analysis was performed using the Kolmogorov-Smirnoff test. This test compares an empirical cumulative distribution to a theoretical cumulative distribution by calculating the maximum of their difference. The resulting p-value measures the likelihood that the empirical data was not drawn from the theoretical distribution. In this work, we compare the distribution of residence times to the expected distribution of a homogeneous Poisson process shown below.[43,88]

$$F(t) = 1 - \exp\left(-\frac{t}{\tau}\right) \tag{3}$$

The characteristic times of the Poisson processes, τ, were selected using a least-squares fit to the wild-type and N368S Abl residence time distributions. In cases where a poor reaction coordinate is used for enhanced sampling, the empirical distribution will not reflect a Poisson process and will therefore have a high p-value.

Alchemical Free Energy Perturbation

As a precusor to Alchemical Free Energy Perturbation (FEP) simulation, a 5 ns equilibrium MD simulation of the solvated Abl-imatinib bound and apo system (generated after removing imatinib bound to Abl crystal structure) was performed. These equilibrium simulations were performed using NAMD2.12[89]in NPT ensemble, utilizing CHARMM36[79] all-atom forcefield for proteins, TIP3P[76] for water and CGenFF[80] for imatinib. In order to study effect of the mutation N368S three independent FEP simulations in the dual topology regime were initiated from three frames extracted from the aforementioned equilibrated apo and bound trajectories. Starting from the equilibrated apo and bound structures, dual topology of the hybrid molecule was prepared using the Mutator plugin and the psfgen module in VMD[90]. Simulations were performed at 300 K using Langevin dynamics[91] with a damping constant of 0.5 ps^{-1}.Pressure was maintained at 1 atm using the Langevin piston method[91]. The cutoff used for the short-range interactions were 12 Å with the switching applied at 10 Å. A soft-core potential[92] was applied on vander Waals interactions to avoid numerical instabilities at the end points of the alchemical transformation. Bidirectional alchemical transformation was performed over 20 Lambda (λ) windows for both apo and imatinib bound systems. In each λ window, system was equilibrated for 5 ps followed by 100 ps of data collection to compute the ensemble averaged ΔG for N368S mutation corresponding to apo and bound systems. Free energy difference and the associated error from the bidirectional transformation was calculated using BAR[93] estimator available in ParseFEP plugin in VMD[90].



System	OP	Abl kinase residue (Cα)	Imatinib atoms
Wild-type	d1	T315	N3-N4
Wild-type	d1a	Y232	p2a
Wild-type	d1b	Y253	p2a
Wild-type	d2a	S229	p2b
Wild-type	d2b	E238	p2b
N368S	d'1	T315	N3-N4
N368S	d'1a	K274	p2a
N368S	d'1b	I313	p2a
N368S	d'2a	Y253	p2b
N368S	d'2b	F283	p2b

Table 2: Summary of order parameters for wild-type and N368S simulation set. Order parameters are defined as the distance between the Cα atom of Abl kinase and the center of mass of Imatinib atoms.

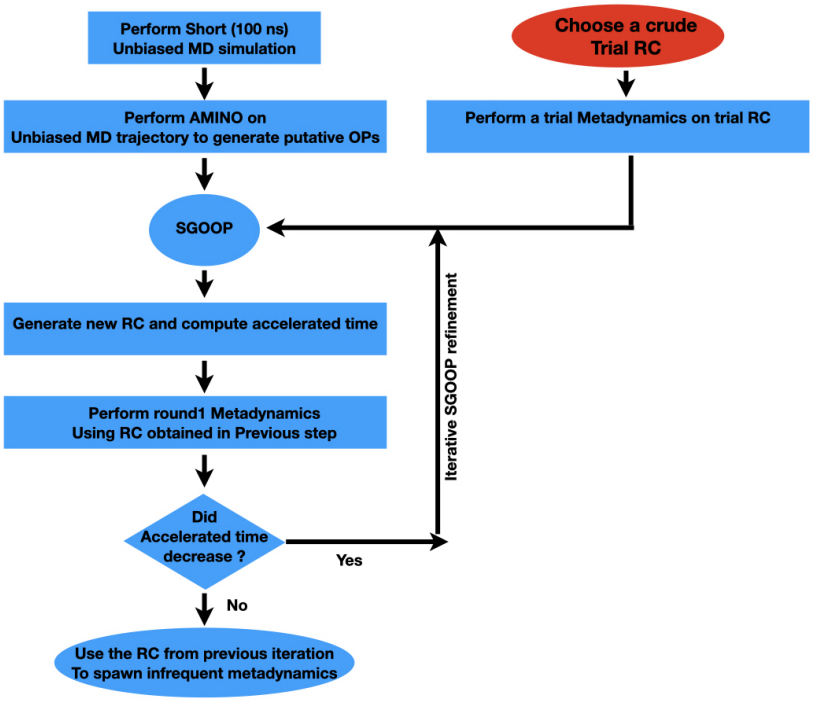


Figure 7: Schematic representing our protocol to generate a suitable RC to be used in biasing infrequent metadynamics simulation. Starting from a trial RC and an unbiased MD simulation, we perform AMINO to generate a list of OPs. Subsequently, we systematically iterate between sampling and RC optimization in a nearly automated manner to obtain the RC to used to bias infrequent metadynamics simulations.

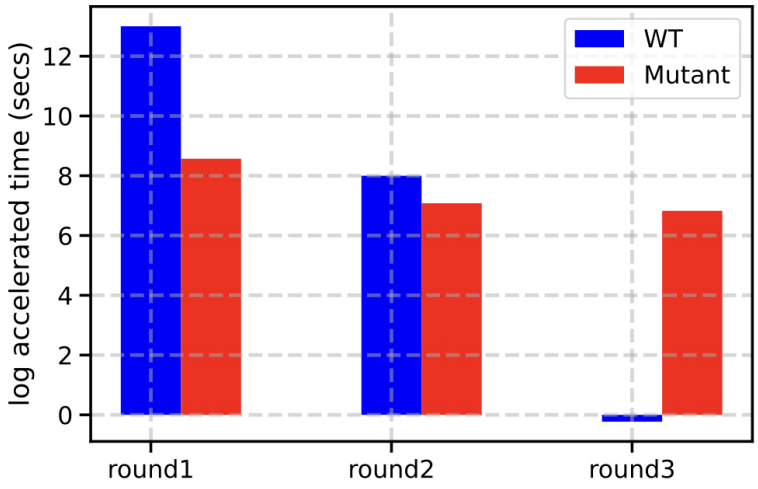


Figure 8: Variation in accelerated time (logarithmic scale) through rounds of preliminary metadynamics (i.e. frequent biasing as opposed to infrequent) is used to assess the quality of RC. Bar plot showing the evolution of log accelerated time in WT (blue bar) and mutant (red bar) upon 3 stages of iterative SGOOP refinement. The reaction coordinate chosen for infrequent metadynamics run was obtained from the refinement round that corresponded to minimum accelerated time.

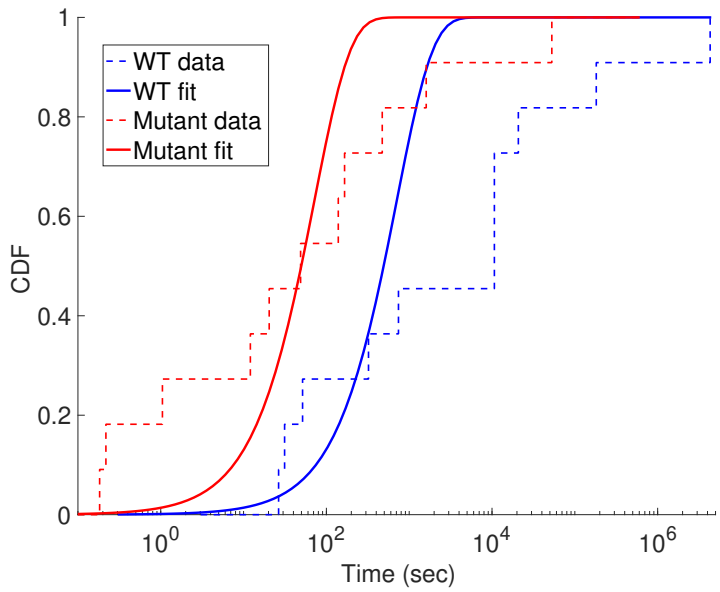


Figure 9: Empirical (dashed line) and fitted cumulative distribution functions (solid line) for WT (blue curves) and mutant (red curves) corresponding to kinetics and p-values reported in main text in Table 3.

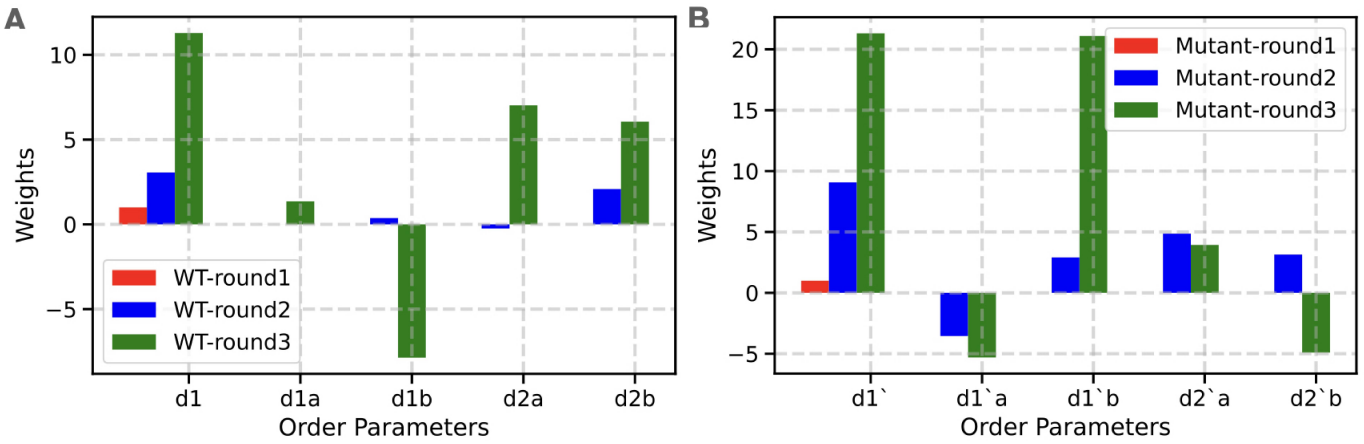


Figure 10: RC constructed for A) WT B) mutant after multiple rounds of SGOOP.

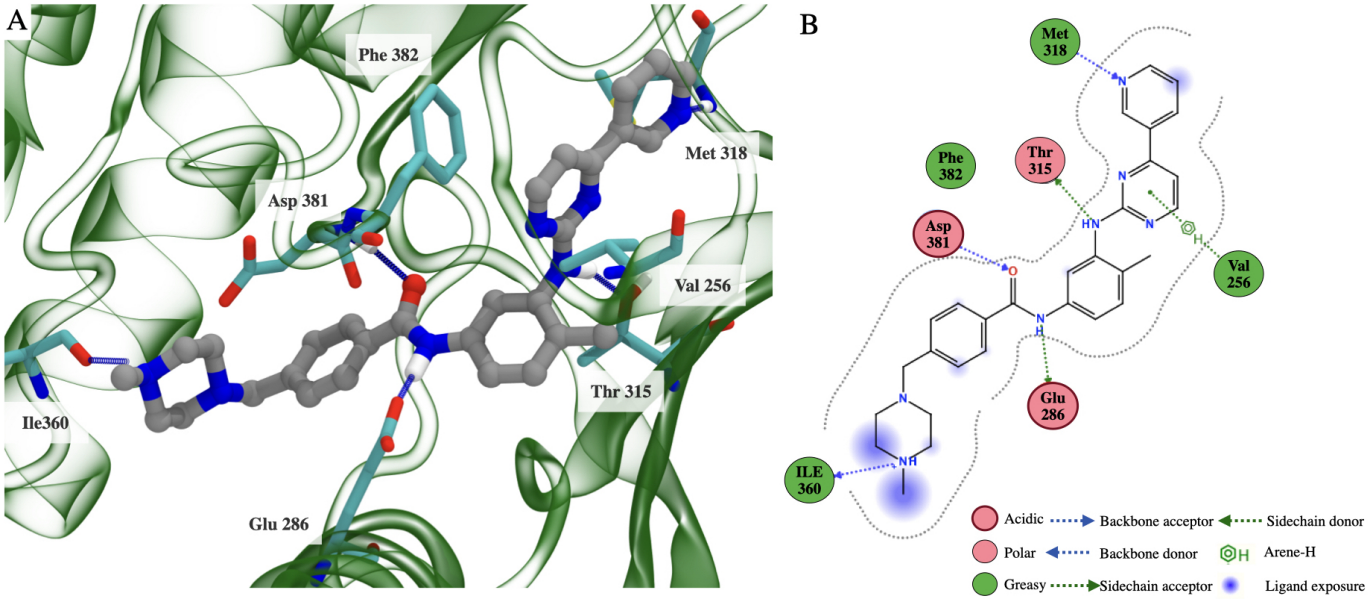


Figure 11: Imatinib binding site in Abl Kinase A) Atomic representation of the binding site interaction. Imatinib is depicted in gray stick and binding site residues is shown as cyan sticks. Hydrogen bonds are depicted by blue dashed lines. B) Schematic representation of the ligand interaction diagram.

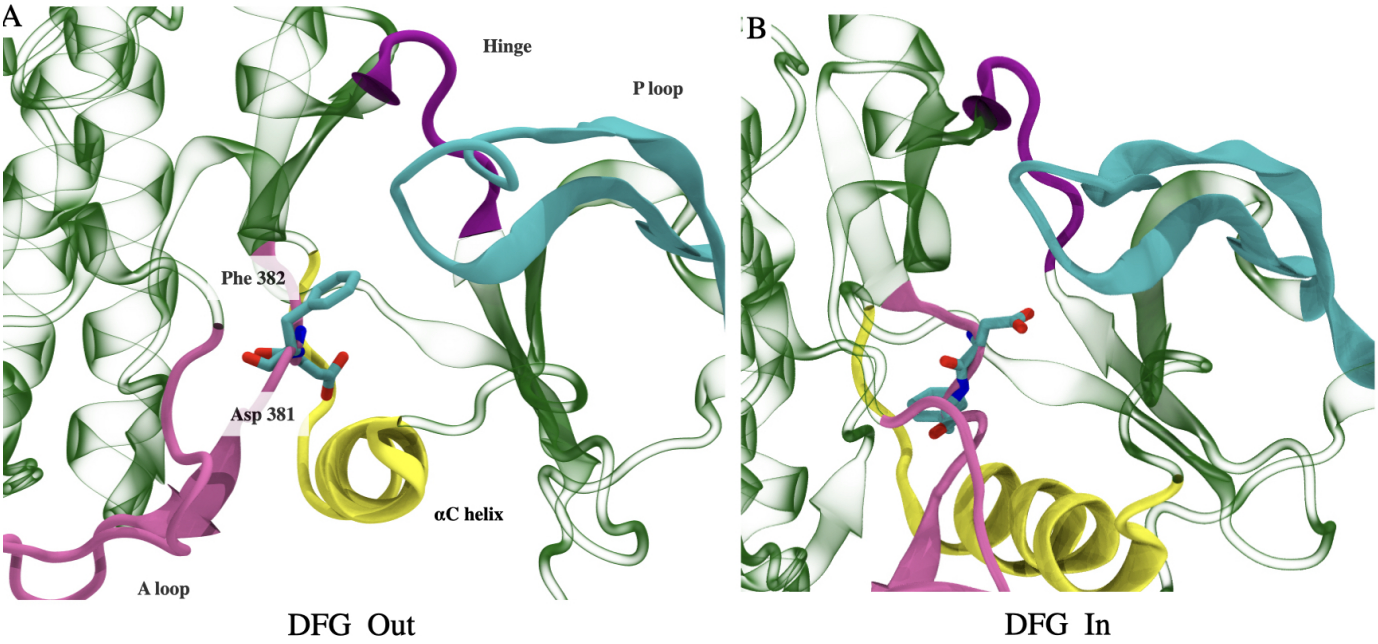


Figure 12: The active DFG-in (right) and inactive DFG-out (left) conformations are illustrated here. In the DFG-in conformation, D381 points toward the binding site where it can coordinate with ATP-bound magnesium.[94,95] Imatinib, a type II kinase inhibitor, selectively binds to the DFG-out conformation.

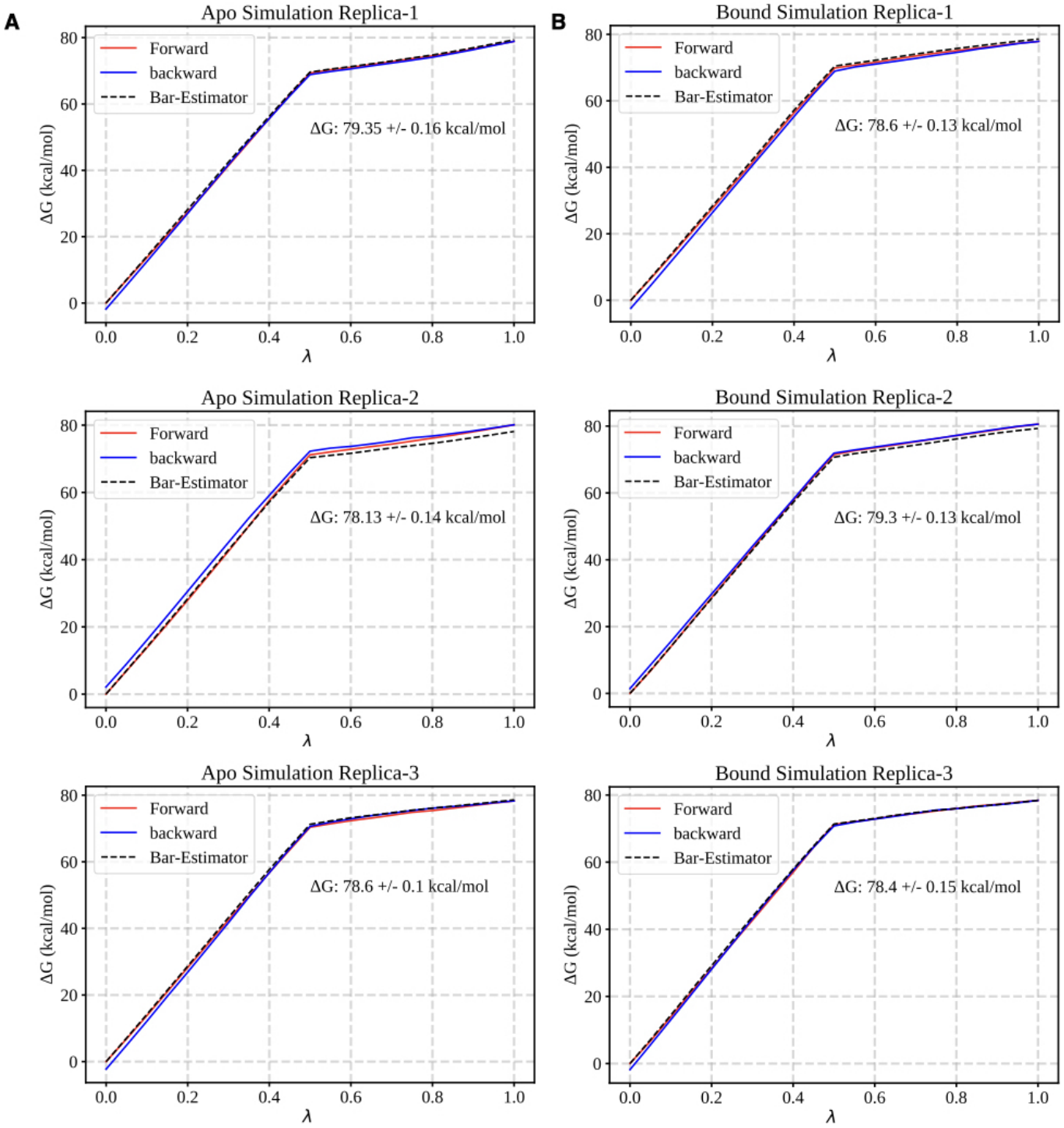


Figure 13: Three independent ΔG (kcal/mol) estimates of the N368S mutation for A) apo and B) bound simulations. ΔG estimate from the forward and backward transtions is shown in red and blue solid line. ΔG estimate from the Bar-Estimator is depicted by black dashed lines. X and Y-axis represents λ window and the cumulative ΔG (kcal/mol) respectively.

Table 3: Summary of change in imatinib binding affinity as a result of N368S mutation. Binding affinity change ($\Delta\Delta$G(kcal/mol)) was computed using three independent alchemical FEP simulations. Mean $\Delta\Delta$G and the standard error of the mean from all three independent replicates is denoted as the footnote

Replicates	ΔG_{apo} (kcal/mol)	ΔG_{bound} (kcal/mol)	$\Delta\Delta$G(kcal/mol)
Replica 1	79.35±0.16	78.6±0.13	-0.75
Replica 2	78.13±0.14	79.30±0.13	1.17
Replica 3	78.6±0.12	78.4±0.15	-0.20

*Mean $\Delta\Delta$G=0.07± 0.81(kcal/mol)