# Supporting Information

Wang et al. 10.1073/pnas.1114017109

## SI Text

**Comparison Between OPLS 2005 and OPLS 2.0 Force Fields for Ligands CDA and CDB.** ***Charge distributions on the pyridine ring.*** The charge distributions on atoms of the P3 pyridine ring for ligand 2-(6-chloro-3-{[2,2-difluoro-2(2-pyridinyl)ethyl]amino}-2-oxo-1(2H)-pyrazinyl)-N-[(2-fluoro-6-pyridinyl)methyl]acetamide (CDA) from the two force fields with the atom numbers labeled as in Fig. S1 are given in Table. S1. It is clear that in the Optimized Potentials for Liquid Simulations (OPLS) 2005 force field, the magnitude of the charges on the atoms of the P3 pyridine ring are very large, making it very polar and favoring its pointing into a polar solvent like water, whereas the OPLS 2.0 force field correctly assigns the charges on these atoms and the correct binding pose was sampled. Similar differences of charge distributions on the P1 pyridine ring and on ligand 2-(6-Chloro-3-{[2,2-difluoro-2(2-pyridinyl)ethyl]amino}-2-oxo-1(2H)-pyrazinyl)-N-[(2-fluoro-3-methyl-6-pyridinyl)methyl]acetamide (CDB) were found for these two force fields.

***The distribution of dihedral angle involved in the flipping of the P1 pyridine ring.*** The distribution of the dihedral angle involved in the flipping of the P1 pyridine ring (N-C-C-C labeled in Fig. 1 in main text) determined from a replica exchange with solute tempering (REST) simulation of the thrombin/CDA complex using the OPLS 2005 force field for the CDA where the "hot region" is taken to include the ligand and the 10 residues surrounding the binding pocket is given in Fig. S2 (*Upper*). Two conformations corresponding to the crystal structure are found and an erroneous additional state with even larger probability was observed in the simulation. This erroneous state might be due to deficiencies in the force field for the ligand. This is validated from the distribution of the dihedral angle for ligand CDA in gas phase simulations using OPLS 2005 force field [See Fig. S2 (*Lower*)], where the intrinsic potential energy of the ligand favors the erroneous state. This erroneous state does not appear in the simulation using the OPLS 2.0 force field for the ligand, and then the correct binding pose was sampled.

**Distribution of the Dihedral Angle Sampled by the Middle Lambda Window Using Free Energy Perturbation/Replica Exchange with Solute Tempering (FEP/REST) for the Thrombin System.** The dihedral angle involved in the flipping of the P1 pyridine ring sampled by the middle lambda window using FEP/REST as a function of simulation time is given in Fig. S3. It is clear that in FEP/REST the middle lambda window quickly samples transitions between the two conformations of the ligand. As discussed in the paper, as long as there are a sufficient number of transitions in the middle lambda window, the generalized ensemble equilibrates quickly and the final free energy difference will not be dependent on the starting conformation to do the simulation.

**Distribution of "F-in"/"F-out" Conformations When Protein Heavy Atoms Are Harmonically Restrained.** The distribution of the dihedral angle involved in the flipping of P1 pyridine ring determined from a FEP/REST simulation with the protein heavy atoms harmonically restrained to the initial positions of the crystal structure and starting from different conformations of the ligands for each lambda window (denoted as F-in/out and F-out/in in Table 3 of the main text) is given in Fig. S4. It is clear that the F-out conformation ($\chi = -100$) is the major conformation for the initial state (binding complex of thrombin/CDA) and the F-in conformation ($\chi = 90$) is the major conformation for the final state (binding complex of thrombin/CDB), in agreement with experimental crystal structures. So the different distributions of these two conformations, one in solution, where the protein atoms move freely, and the other in the crystal, where the atoms are constrained, are due to the different physical conditions (in solution vs. in crystal). The different relative binding affinities from FEP/REST, when the protein heavy atoms are allowed to move freely or when they are harmonically restrained suggest that there is a 0.8 kcal/mol difference in protein strain free energy between the two binding complexes.

**How the Hot region and the Effective Temperature Profile Were Determined for FEP/REST.** In FEP/REST, in addition to the different energy terms introduced in the alchemical transformation found in normal FEP, the different effective temperatures of the hot region in REST will make the free energy difference between neighboring lambda windows larger, and the precision of the free energy results might be reduced. This is the price paid to get enhanced sampling. The larger the hot region, and the higher the effective temperature of the hot region, the stronger the enhanced sampling effect, but the error bars in the resulting calculated free energy energies between neighboring lambda windows are also increased. So a proper choice of the hot region and effective temperature profile reflects a trade off between the precision of free energy results and the efficiency of the enhanced sampling; consequently the hot region should be as small as possible but still be able to sample structural reorganization effects. In the two systems studied in this article, we know the slow degrees of freedom, so only the residue Val111 or the P1 pyridine ring was included in the hot region. In general, if there is no prior knowledge about the slow degrees of freedom, a proper choice of hot region would include the ligand and the protein residues surrounding the ligand because usually the structural reorganization involves the ligand and the protein residues surrounding the binding pocket.

The free energy difference, $\Delta F$, between neighboring lambda windows depends on the distribution functions $P_0(\Delta E)$ and $P_1(\Delta E)$ of energy differences ($\Delta E$) in forward and backward sampling, respectively (S1), through

$$P_1(\Delta E) = P_0(\Delta E)\exp^{-\beta(\Delta E - \Delta F)}. \quad \textbf{[S1]}$$

The two distributions are equal for the specific energy difference $\Delta E = \Delta F$, and the accuracy of the free energy $\Delta F$ depends on the overlap of the two distributions (S1). At the same time, it is easy to show by imposing detailed balance condition that the acceptance ratio for attempted replica exchanges between neighboring lambda windows also depends on the energy difference from forward and backward sampling: (S2)

$$\Delta_{01} = \beta(E_1(X_0) + E_0(X_1) - E_1(X_1) - E_0(X_0)) \quad \textbf{[S2]}$$

$$= \beta(\Delta E(X_0) - \Delta E(X_1)). \quad \textbf{[S3]}$$

Here, $X_0$ and $X_1$ are the configurations sampled in the forward and backward directions, and $E_0$ and $E_1$ are the potential energy functions for the two states. So both the accuracy of the free energy result and the efficiency of the enhanced sampling are maximized when neighboring lambda windows have regions of overlap in potential energy distribution, and an optimal alchemical lambda schedule and effective temperature schedule will generate equal acceptance ratios for all neighboring lambda win-

dows. In the two systems we studied, with a total number of 16 lambda windows and highest effective temperature of 1,200 K for the T4L/L99A system, and a total number of 23 lambda windows and highest effective temperature of 1,784 K for thethrombin system, and a time interval of 1 ps between attempted exchanges among neighboring replicas, we obtained an average acceptance ratio of 0.54 for T4L/L99A system and 0.59 for the thrombin system.

**Details of the FEP/REST Mutation Path.** A dual topology ideal gas molecule end state method was used to define the mutation path, which facilitates the sampling through the double tunneling mechanism (S3). The electrostatic interactions unique to the initial ligand were turned off before the Lennard Jones (LJ) interactions, and the LJ interactions unique to the final ligand were turned on followed by the electrostatic interactions. The core of the LJ interactions is made softer to avoid the singularities and instabilities in the simulation (S4). The mutation path is symmetric, so mutation from either direction will give identical free energy result. To get more efficient enhanced sampling, the fluorine atom on the P1 pyridine ring was mutated to an identical atom so that the effective volume of P1 pyridine ring was made smaller in the middle lambda window and the transition between the two conformations was faster. The lambda values, the scaling factors for the hot region, and the free energy differences between neighboring lambda windows for the two systems are given in Tables S2 and S3.

**Treatment of Bonded Interactions in the Mutation Path.** The bonded interactions connecting the dummy atoms were treated differently from the default method in Desmond (S5). In this section, we give the details about how the bonded interactions involving the dummy atoms are treated in FEP/REST to avoid singularities and instabilities. As mentioned in the paper, this is a problem often not appreciated in the literature on FEP.

In the dual topology FEP method, depending on whether the bonded interactions between the dummy atoms and the rest of the mixed molecule are scaled or not, there are two different methods: "the ideal gas atom end state" method (scaled) and "the ideal gas molecule end state" method (nonscaled). In the ideal gas atom end state method, the dummy atom does not have any bonded interactions with the rest of the molecule and would move freely in the whole simulation volume, making the sampling very difficult. Thus most programs, including Desmond, use the ideal gas molecule end state method, in which the dummy atoms are bonded with the rest of the molecule. However, if there are more than one bonded stretch or bonded angle or bonded dihedral angle interactions between a dummy atom and the rest of the molecule, the distributions sampled for the mixed molecule (molecule with the dummy atoms) will be different from the molecule without the dummy atoms (S5). So in Desmond, only one bonded stretch, bonded angle, and bonded dihedral angle interactions involving a dummy atom are kept whereas all the other bonded interactions are scaled to 0 at the end state (S5). In this way, the contributions of the dummy atoms to the free energies in the binding complex and in pure solvent will be identical and consequently the relative binding affinity will be independent of the dummy atoms. This follows because the relative binding affinity is equal to the difference of these two free energies. For most systems, this method works well, but for some systems, like the sets of ligands studied in this article, it will cause serious problems in the FEP simulation, which is explained in what follows.

Fig. S5*A* displays how the structure of the mixed molecule is mutated from a benzene molecule to a p-xylene molecule. The two dummy hydrogen atoms that are mutated to the methyl groups all have two bonded angle and bonded dihedral angle interactions with the rest of the molecule. Take the dummy hydrogen atom numbered 6 in Fig. S5*A* as an example. It has two bonded angle and bonded dihedral angle interactions with the rest of the molecule ($\theta_1$,$\theta_2$,$\phi_1$,$\phi_2$ labeled in Fig. S5*A*). If all these bonded interactions are kept in the end state, the distribution of the angle formed by atoms numbered 2, 3, and 4 for the mixed molecule with the dummy atom will be different from the distribution of the real molecule without the dummy atom. So in the default setup of Desmond, only one bonded angle and one bonded dihedral angle interaction involving the dummy hydrogen atom ($\theta_1$,$\phi_1$) is kept in the end state, and the others ($\theta_2$,$\phi_2$) are scaled to zero. The energy difference $(U_1 - U_0)$ (where $U_1$ is the potential energy in the previous lambda window and $U_0$ is the potential energy in the end state lambda window) sampled in the end state lambda window using Desmond's default setting is given in Fig. S6. Clearly, the energy difference fluctuate about three different values (approximately −0.5, 21, and 40 kal/mol, respectively).

From the simulated trajectories, we observed that the two dummy hydrogen atoms were located at the correct positions in the initial stage (Fig. S5*A*), then one of them moved to the position that almost overlapped with a carbon atom on the ring (Fig. S5*B*), and at the end the other hydrogen atom moved to the position that almost overlapped with another carbon atom on the ring (Fig. S5*C*). These three configurations for the mixed molecule are located in the potential energy minima for the end state where only one bonded angle and one bonded dihedral angle interaction are kept ($\theta_1$,$\phi_1$) for the dummy hydrogen atoms (the default setup of Desmond). In the previous lambda window, however, when all the bonded interactions involving the dummy atoms (including $\theta_2$,$\phi_2$) are slowly turned on, the last two configurations (b and c) have large bonded angle interactions (also 1–4 pair interactions ) leading to instabilities for the free energy calculation. This is the reason for large jumps in the energy difference profile displayed in Fig. S6 and the three distinct values of energy difference correspond to these three configurations. (In extreme cases, when either dummy atom is located at the same position as the carbon atom on the ring, the 1–4 pair interaction will cause a singularity in the free energy calculation).

To avoid instabilities and singularities in the free energy calculations caused by the bonded interactions involving dummy atoms, we treated these bonded interactions in the end state differently in this article. We choose to keep all of the bonded stretch and bonded angle interactions involving the dummy atoms in the end state. For the dihedral angle interactions, only one bonded dihedral angle interaction involving a dummy atom is kept whereas the other dihedral angle interactions are scaled to zero in the end state. Take the dummy hydrogen atom numbered 6 in Fig. S5, for example. We include two bonded angle terms ($\theta_1$,$\theta_2$) and one bonded dihedral angle term ($\phi_1$) in the end state, whereas the other bonded dihedral angle term ($\phi_2$) is scaled to 0 in the end state. Thus in the end state configurations b and c in Fig. S5 are located in the high energy region in phase space and thus will not be sampled. In this way, the instability and singularity problems encountered for the bonded interactions in FEP are eliminated. Although the distributions sampled for the "mixed molecule" (molecule including the dummy atoms) might be a bit different from the distributions for the molecule without the dummy atoms, the error introduced in this treatment is negligible because the fluctuations of bond angle are very small for the two sets of ligands studied in this article. In addition, the error in the relative binding affinity, which is the difference between the free energies in the binding complex and in free solvent, will be very small because there is an excellent cancellation of errors in these two free energies.

1. Pohorille A, Jarzynski C, Chipot C (2010) Good practices in free-energy calculations. *J Phys Chem B* 114:10235–10253.
2. 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* 115:9431–9438.
3. Min D, Li H, Li G, Bitetti-Putzer R, Yang W (2007) Synergistic approach to improve alchemical free energy calculation in rugged energy surface. *J Chem Phys* 126:144109.
4. 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* 222:529–539.
5. Shobana, Roux B, Andersen OS (2000) Free energy simulations: Thermodynamic reversibility and variability. *J Phys Chem B* 104:5179–5190.

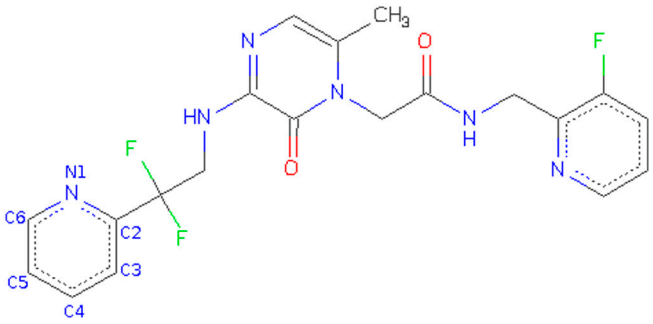


**Fig. S1.** The atom numbering on the P3 pyridine ring for ligand CDA.

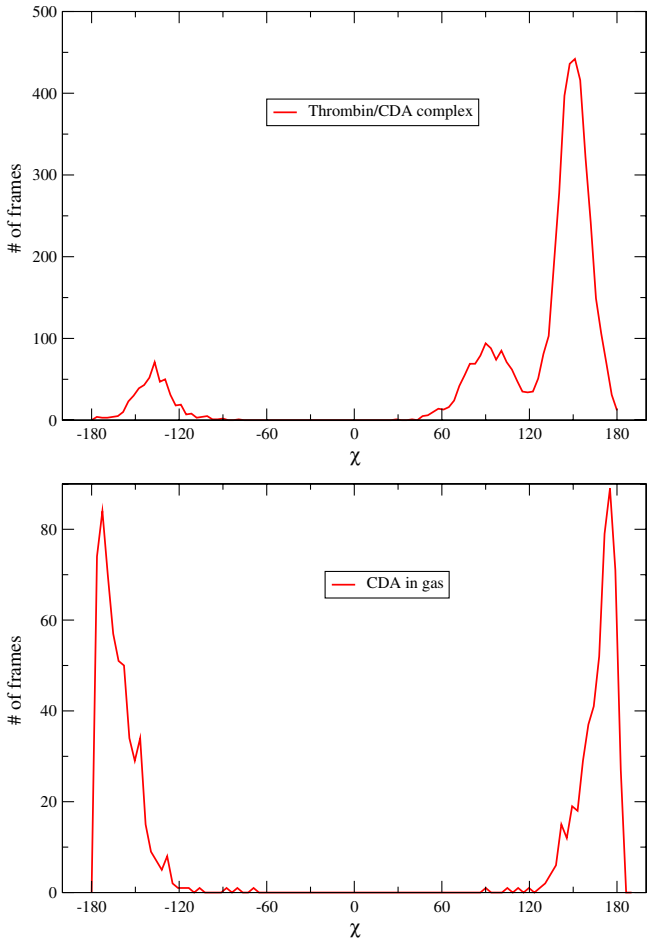


**Fig. S2.** The distribution of the dihedral angle involved in the flipping of P1 pyridine ring (N-C-C-C labeled in Fig. 1 in main text) for thrombin/CDA complex using OPLS 2005 force field for the ligand (*Upper*), and the same angle distribution for the ligand in gas phase (*Lower*).

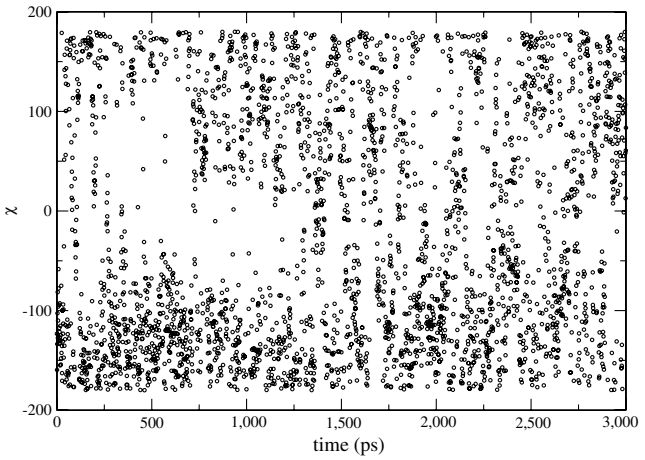


**Fig. S3.** The dihedral angle involved in the flipping of the P1 pyridine ring (N-C-C-C labeled in Fig. 1 in main text) of the ligands as a function of simulation time sampled in the middle lambda window using FEP/REST.

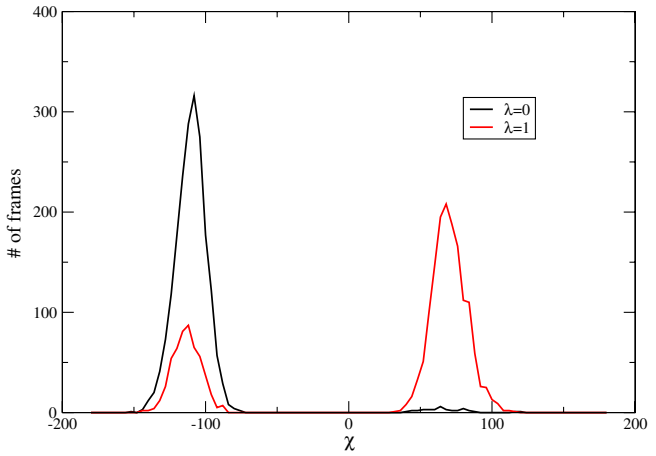


**Fig. S4.** The distribution of the dihedral involved in the flipping of P1 pyridine ring from a FEP/REST simulation with the protein heavy atoms harmonically restrained to the initial position.

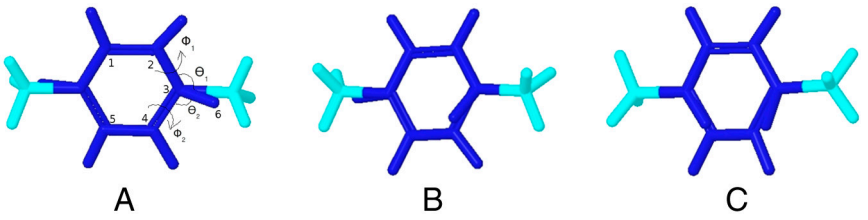


**Fig. S5.** The structure of the mixed molecule to mutate from benzene to p-xylene. Three different configurations are sampled in the end state when only one bonded angle and bonded dihedral angle interactions are kept for the dummy hydrogen atoms.

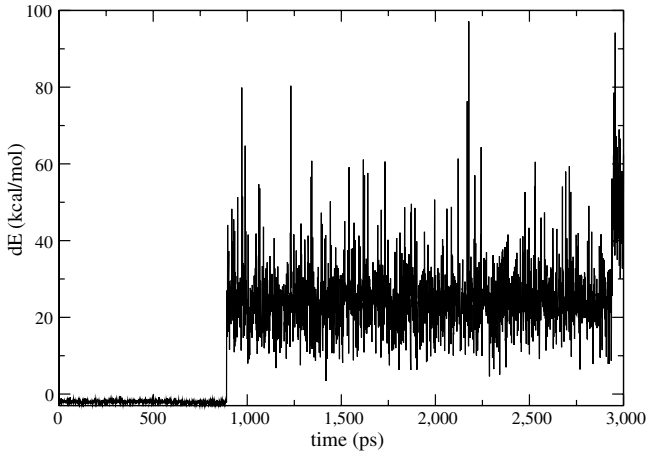


**Fig. S6.** The energy difference between the previous lambda window and the end state lambda window sampled by the end state lambda window for mutation from benzene to p-xylene when only one bonded angle and bonded dihedral angle interactions involving the dummy hydrogen atoms are kept in the end state.

**Table S1. Charge distributions on atoms of the P3 pyridine ring for ligand CDA for two force fields**

| Atom number | OPLS 2.0 | OPLS 2005 |
|---|---|---|
| N1 | −0.431 | −0.678 |
| C2 | 0.045 | 0.370 |
| C3 | −0.138 | −0.447 |
| C4 | −0.091 | 0.227 |
| C5 | −0.164 | −0.447 |
| C6 | 0.089 | 0.473 |

**Table S2. Lambda values, scaling factors, and free energy differences between neighboring lambda windows for the T4L/L99A system**

| $\lambda$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| bondedA | 1.0 | 0.933 | 0.867 | 0.8 | 0.733 | 0.667 | 0.6 | 0.533 |
| bondedB | 0.0 | 0.067 | 0.133 | 0.2 | 0.267 | 0.333 | 0.4 | 0.467 |
| chargeA | 1.0 | 0.75 | 0.5 | 0.25 | 0.0 | 0.0 | 0.0 | 0.0 |
| chargeB | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| vdwA | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.857 | 0.714 | 0.571 |
| vdwB | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.143 | 0.286 | 0.429 |
| scaling | 1.00 | 0.8464 | 0.7056 | 0.5776 | 0.4624 | 0.3721 | 0.3025 | 0.25 |
| $\Delta G_{FEP}$ | −2.4997 | −2.5577 | −2.6721 | −2.7345 | −0.7043 | 0.3626 | 0.8722 | 0.2075 |
| $\Delta G_{FEP/REST}$ | 1.6669 | 1.6307 | 1.5967 | 1.5675 | 3.0860 | 3.6672 | 3.4928 | −0.0463 |
| $\lambda$ | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| bondedA | 0.467 | 0.4 | 0.333 | 0.267 | 0.2 | 0.133 | 0.067 | 0.0 |
| bondedB | 0.533 | 0.6 | 0.667 | 0.733 | 0.8 | 0.867 | 0.933 | 1.0 |
| chargeA | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| chargeB | 0.0 | 0.0 | 0.0 | 0.0 | 0.25 | 0.5 | 0.75 | 1.0 |
| vdwA | 0.429 | 0.286 | 0.143 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| vdwB | 0.571 | 0.714 | 0.857 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| scaling | 0.25 | 0.3025 | 0.3721 | 0.4624 | 0.5776 | 0.7056 | 0.8464 | 1.00 |
| $\Delta G_{FEP}$ | −0.5328 | −0.9243 | −1.1963 | 2.4196 | 2.3307 | 2.2174 | 2.0848 | |
| $\Delta G_{FEP/REST}$ | −3.3633 | −4.2287 | −4.9902 | −1.9003 | −1.9433 | −1.9780 | −2.0394 | |

Note: vdwA and vdwB are the lambda values of the van der Waals interaction to mutate from molecule A to molecule B.

Wang et al. www.pnas.org/cgi/doi/10.1073/pnas.1114017109

**Table S3. Lambda values, scaling factors, and free energy differences between neighboring lambda windows for the thrombin system**

| $\lambda$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| bondedA | 1.00 | 0.95 | 0.91 | 0.86 | 0.82 | 0.77 | 0.73 | 0.68 | 0.63 | 0.59 | 0.54 | 0.50 |
| bondedB | 0.00 | 0.05 | 0.09 | 0.14 | 0.18 | 0.23 | 0.27 | 0.32 | 0.37 | 0.41 | 0.46 | 0.50 |
| chargeA | 1.00 | 0.75 | 0.50 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| chargeB | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| dwA | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.68 | 0.46 | 0.33 | 0.25 | 0.19 | 0.12 | 0.00 |
| vdwB | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| scaling | 1.00 | 0.9216 | 0.8464 | 0.7569 | 0.6724 | 0.5776 | 0.49 | 0.4096 | 0.3364 | 0.2704 | 0.2116 | 0.1681 |
| $\Delta G_{FEP}$ | 2.5092 | 2.2140 | 1.9203 | 1.6569 | 0.1517 | −0.3940 | −0.3947 | −0.2190 | −0.1103 | −0.0743 | 0.0912 | −0.3519 |
| $\Delta G_{FEP/REST}$ | 1.3099 | 0.7003 | −1.1062 | −0.5334 | −1.7920 | −1.9750 | −1.9259 | −1.7625 | −1.6065 | −1.4625 | −1.0676 | 0.9690 |
| $\lambda$ | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | |
| bondedA | 0.46 | 0.41 | 0.37 | 0.32 | 0.27 | 0.23 | 0.18 | 0.14 | 0.09 | 0.05 | 0.00 | |
| bondedB | 0.54 | 0.59 | 0.63 | 0.68 | 0.73 | 0.77 | 0.82 | 0.86 | 0.91 | 0.95 | 1.00 | |
| chargeA | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| chargeB | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.50 | 0.75 | 1.00 | |
| vdwA | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| vdwB | 0.12 | 0.19 | 0.25 | 0.33 | 0.46 | 0.68 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| scaling | 0.2116 | 0.2704 | 0.3364 | 0.4096 | 0.49 | 0.5776 | 0.6724 | 0.7569 | 0.8464 | 0.9216 | 1.00 | |
| $\Delta G_{FEP}$ | −0.0413 | 0.0475 | 0.0895 | 0.1379 | −0.0907 | −1.2740 | −0.0423 | −0.8294 | −1.5706 | −2.4305 | | |
| $\Delta G_{FEP/REST}$ | 1.4394 | 1.6101 | 1.7749 | 1.8948 | 1.6985 | 0.8331 | 1.5105 | 0.8022 | −0.3918 | −1.8040 | | |

Note: vdwA and vdwB are the lambda values of the van der Waals interaction to mutate from molecule A to molecule B.