

Supporting Information for Publication: Leveraging a separation of states method for relative binding free energy calculations in systems with trapped waters

Swapnil Wagle,[†] Pascal Merz,^{‡,⊥} Yunhui Ge,[¶] Christopher I. Bayly,[§] and
David L. Mobley^{*,†,||}

[†]*Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697,
United States*

[‡]*PM Scientific Consulting, 4000 Basel, Switzerland*

[¶]*Alkermes, Inc. 900 Winter St, Waltham, MA 02451, United States*

[§]*OpenEye Scientific, Cadence Molecular Sciences, 9 Bisbee Ct Suite D, Santa Fe, NM
87508, United States*

^{||}*Department of Chemistry, University of California, Irvine, CA 92697, United States*

[⊥]*Current affiliation: D. E. Shaw Research, 120 W. 45th St., 39th Fl., New York, NY
10036, United States*

E-mail: dmobley@mobleylab.org

S1. Methods

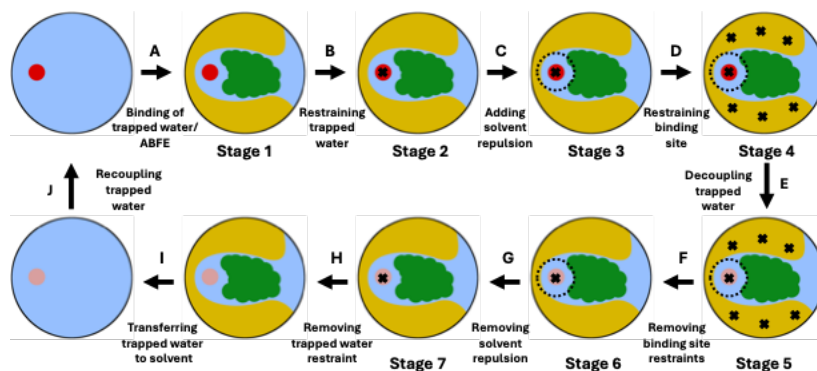
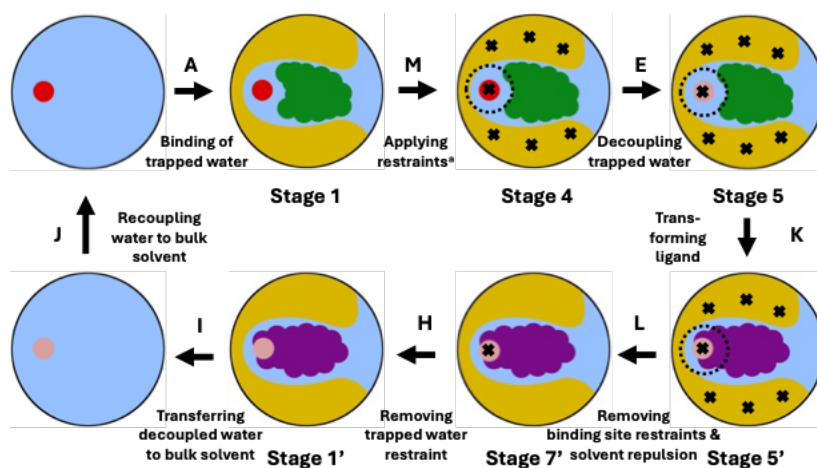


Figure S1. Thermodynamic cycle to calculate the absolute binding free energy (ABFE) of a trapped water in a protein-ligand complex. The red circle in stages 1 through 4 and in the starting stage of edge A represents an interacting trapped water, whereas the faded red/ magenta circle in stages 5 through 7 and in the end stages of edge I represents a non-interacting water. The dashed black circle around the trapped water in stages 3 through 6 represents solvent repulsion. The black crosses in stages 2 through 7 represent position restraints applied to different portions of the system– the trapped water and the protein binding site.



*position restraint on the trapped water, solvent repulsion, and position restraints on the binding site

Figure S2. Shortened thermodynamic cycle to calculate the complex leg of the relative binding free energy (RBEF) between two ligands, where one ligand (green) binds to the protein with a trapped water, while the other ligand (purple) displaces it. The red circle in stages 1 and 4 and in the starting stage of edge A represents an interacting trapped water, whereas the faded red/ magenta circle in stages 5, 5', 7', 1' and in the end stage of edge I represents a non-interacting water. The dashed black circle around the trapped water in stages 4, 5 and 5' represents solvent repulsion. The black crosses in stages 4, 5, 5' and 7' represent position restraints applied to different parts of the system – the trapped water and the protein binding site. For an easy nomenclature, the stages and edges for the purple ligand are denoted with ' superscripts for the corresponding stages and edges for the green ligand in the ABFE thermodynamic cycle (Figure S1). We have named the end state of edge H as stage 1' because it represents the unrestrained binding mode of the purple ligand and its corresponding stage in the ABFE thermodynamic cycle is denoted by stage 1, *i.e.*, the unrestrained binding mode of the green ligand.

S1.1. Input parameters for the equilibrium free energy simulations. In the ABFE and RBEF thermodynamic cycles (Figures S1 and S2, respectively), we calculated the free energy to apply the restraints to different parts of the system–the trapped water, the solvent, and the binding site (see Section 2.2 of the main text for details), using an equilibrium free energy calculation.

Then, we decoupled the trapped water using an NES free energy calculation and removed the relevant restraints using a second equilibrium free energy calculation. In the RBFEE thermodynamic cycle, we transformed the ligand after the trapped water was decoupled, using a second NES free energy calculation. To run these equilibrium and NES free energy calculations, we used the free energy calculation workflow implemented in the GROMACS 2022.1 simulation package.¹ Before running the free energy calculations, we made certain changes to the GROMACS topology files of the protein-ligand complexes, which are described in Section S1.3, defining different states (interacting/on, non-interacting/off, and switching) of the restraints, the trapped water, and the ligand. The relevant states for the restraints, the trapped water, and the ligand for any free energy calculation were selected using the *define* input parameter in the mdp files for the calculations. In this section, we describe the preparation of the mdp files to run the free energy calculations.

All free energy calculations were run in the NPT ensemble, for which the simulation parameters are described in Section 2.6.2.

In the equilibrium free energy calculation to apply the restraints (see Section 2.6.3 of the main text for more details), the harmonic restraint on the trapped water, the solvent repulsion, and the position restraints on the binding site were scaled using the input parameters *bonded-lambdas*, *vdw-lambdas*, and *restraint-lambdas*, respectively, as described in Table S1. For the ABFE calculations, the equilibrium free energy simulation with 4 λ states, corresponding to stages 1 through 4, respectively, was run. For the RBFEE calculations, the equilibrium free energy simulation with only the λ states 0 and 3, corresponding to stages 1 and 4, respectively, was run. In the mdp files for the λ states, the switching states of the restraints, the interacting state of the trapped water, and ligand A (only for the RBFEE calculations) were selected from the topology file using the *define* input parameter.

Similarly, in the equilibrium free energy calculation for removing the restraints, the harmonic restraint on the non-interacting trapped water, the solvent repulsion, and the position restraints on the binding site were scaled using the input parameters *bonded-lambdas*, *vdw-lambdas*, and *restraint-lambdas*, respectively, as described in Table S2. For the ABFE calculations, the equilibrium free energy simulation with 10 λ states was run. The λ states 0, 1, and 9 represented stages 7, 6, and 5, respectively, and the λ states 2 through 8 were simulated as the 7 intermediate λ states for removing the position restraints on the binding site (see Section 2.3.3 of the main text for details).

For the RBFEE calculations, one equilibrium free energy simulation was run with only the λ states 0 and 9, corresponding to stages 7' and 5', respectively. In the mdp files for the λ states of the free energy calculation, the switching states of the restraints, the non-interacting state of the trapped water, and ligand B were selected from the topology file using the *define* input parameter. A second equilibrium free energy simulation was run with a single λ state 9 corresponding to the stage 5; in the mdp file for this λ state, the switching states of the restraints, the non-interacting state of the trapped water, and ligand A were selected from the topology file using the *define* input parameter.

Table S1. λ schedules for stages 1 through 4

| Stage | 1 | 2 | 3 | 4 |
|--------------------------|---|---|---|---|
| λ state | 0 | 1 | 2 | 3 |
| <i>bonded-lambdas</i> | 0 | 1 | 1 | 1 |
| <i>vdw-lambdas</i> | 0 | 0 | 1 | 1 |
| <i>restraint-lambdas</i> | 0 | 0 | 0 | 1 |

In the NES switches to couple/decouple the trapped water (edge E, Figures S1 and S2), the vdW parameters and the partial charges of the trapped water were scaled (see Section 2.6.4 of the main

Table S2. λ schedules for stages 5/5' through 7/7'

| Stage | 7/7' | 6/6' | intermediate λ states | | | | | | | 5/5' |
|--------------------------|------|------|-------------------------------|------|------|------|------|------|------|------|
| λ state | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| <i>bonded-lambdas</i> | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| <i>vdw-lambdas</i> | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| <i>restraint-lambdas</i> | 0 | 0 | 0.01 | 0.02 | 0.04 | 0.08 | 0.16 | 0.32 | 0.64 | 1 |

text for more details). In the mdp files for the NES switches, the switching state of the trapped water, the on states of the restraints, and the topology of ligand A were selected from the topology file using the *define* input parameter.

For the RBFE calculations, in the NES switches to transform the ligand (edge K, Figure S2), the vdW parameters, the partial charges, and the bonded parameters of the ligands were transformed (see Section 2.6.4 of the main text for more details). In the mdp files for the NES switches, the non-interacting state of the trapped water, the on states of the restraints, and the hybrid topology of ligands were selected from the topology file using the *define* input parameter.

S1.2. Free energy calculations of edges H, I and J. Edge H is the free energy of restraining a non-interacting water at a point in space. In general, the free energy contribution of a radial restraint is given by equation 1

$$\Delta G_{\text{Restraint}} = RT \log \left(C^o 4\pi \int_0^\infty dr r^2 \exp -\beta U_r(r) \right) \quad (1)$$

where R is the molar gas constant, T is the simulation temperature, C^o is the standard molarity, and $U_r(r)$ is the radial restraint potential. In this work, we restrained the non-interacting water using a harmonic restraint with zero equilibrium distance. The potential U_r can therefore be written as

$$U_r(r) = \frac{1}{2} k r^2 \quad (2)$$

where k is the force constant of the harmonic restraint and r is the instantaneous distance of the trapped water from its restraint site. Substituting this expression of U_r into equation 1 results in

$$\Delta G_{\text{Restraint}} = RT \log \left(C^o \left(\frac{2\pi RT}{k} \right)^{3/2} \right) \quad (3)$$

The free energy for edge H was calculated to be $-0.229 \text{ kcal mol}^{-1}$ for a force constant of $0.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for the harmonic restraint on the non-interacting water and a standard concentration of water of 55 molar.

Edge I represents the transfer of the non-interacting water to bulk solvent. There is no free energy associated with this transfer, therefore, the free energy contribution of edge I will always be zero.

The free energy contribution of edge J is independent of the protein-ligand complex, and it only depends on the water model used. For TIP3P water model, used in our study, we calculated this free energy value to be $-6.12 \pm 0.03 \text{ kcal mol}^{-1}$, which agrees well with previous computational² as well as experimental³ studies.

S1.3. Changes we made in GROMACS topology files to simulate the thermodynamic cycles. To simulate the different stages of the thermodynamic cycle in Figures S1 and S2, we used the free energy calculation workflow implemented in GROMACS 2022.1 simulation package.¹ In

our thermodynamic cycles, we implemented several different lambda schedules (see Section S1.1 for details), for three processes– 1. to apply and remove restraints, 2. to couple/decouple the trapped water, and 3. to transform the ligands. The three processes are performed sequentially in our thermodynamic cycles. To ensure that a lambda schedule implemented for one process does not alter the any of the other two processes, we decoupled the other processes from the lambda schedule for one process by defining various *ifdef* sections in the topology. The *ifdef* sections were used to describe an interacting (or on), a non-interacting (or off), and a switching (or turning on to off, or vice versa) state of different parts of the system (*e.g.*, the trapped water, the ligand, and the restraints). In a free energy calculation for a process, the switching state of the relevant parts of the system would be selected using the *define* input parameter in the mdp files of the calculation, while each of the other parts of the system would be in the interacting or the non-interacting state, *i.e.*, decoupled from the lambda schedules.

To restrain and decouple the trapped water in the thermodynamic cycle, we renamed the trapped water residue to distinguish it from the bulk water in the GROMACS topology (which would otherwise treat all waters as equivalent). The bonded and nonbonded parameters of the trapped water residue were set identically to those for the bulk water residue. Because GROMACS allows only one declaration of the *settles* section in the topology file (which we used for the bulk water residue; see Section 2.6 of the main text for details), we defined constraints on the hydrogen-oxygen and hydrogen-hydrogen distances in the trapped water residue definition, to constrain its geometry. To model a non-interacting trapped water, we used the same bonded parameters and hydrogen bond constraints as defined for the trapped water described above, however, using atom types for which the vdW parameters and partial charges were all set to 0. To couple/decouple the trapped water in NES switches, we defined a switching state of the trapped water residue, in which, states A and B were defined as the interacting and non-interacting trapped water, respectively. We defined three *ifdef* sections for the trapped water, one for each of the interacting, non-interacting, and switching states of the trapped water.

To implement the harmonic restraint on the trapped water, we added a virtual site of type 2 at the target position of the trapped water in an equilibrated structure of the complex (see Section 2.6.2 of the main text for details). The virtual site was defined as a linear combination of two of the protein’s backbone atoms. We restricted the search of the two backbone atoms defining the virtual site to as close to the trapped water’s position as possible. We selected all amino acid residues that had any atom within a cutoff distance of 5 Å from the trapped water. If less than three amino acids were selected, we increased the cut-off distance by 1 Å iteratively until at least three amino acid residues were selected. From the backbone atoms of the selected residues, we selected all the pairs of two atoms and calculated the distance of the trapped water’s oxygen from the line joining the two atoms. The pair of atoms resulting in the smallest distance was selected for defining the virtual site. We defined the harmonic restraint on the trapped water as a bonded interaction between the virtual site and the trapped water’s oxygen atom. We defined two *ifdef* sections, one defining an on (or interacting) state and the other defining a switching state, for this bonded interaction. In the on state, a force constant of $0.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ was defined for the bond, and in the switching state, the force constant of the bonded interaction was changed from $0.5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ (state A) to $0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ (state B).

To implement solvent repulsion at the target position of the trapped water, we introduced a new atom type for the virtual site defined to restrain the trapped water. We defined a nonbonded repulsive interaction between the new atom type and the atom type of the bulk water oxygen. A repulsive potential can be defined in the GROMACS topology in the *nonbond_params* section using a negative sigma value between two atom types. We set the repulsive potential between the new atom type of the virtual site and the atom type of bulk water’s oxygen to be roughly equivalent to

the repulsive potential between two water molecules. Therefore, we defined a sigma of -0.315365 nm and an epsilon parameter of 0.648520 kJ mol $^{-1}$ for the repulsive interaction. We defined three *ifdef* sections in the topology file for the repulsive potential. In the first *ifdef* section, the atom type of the virtual site particle was set to the one without the repulsive potential (the off state). In the section *ifdef* section, the atom type of the virtual site particle was set to the atom type for which the repulsive potential was defined (the on state). In the third *ifdef* section, the atom type of the virtual particle was changed from the atom type without the repulsive potential to the one with the repulsive potential (the switching state).

To restrain the binding site, we selected all the heavy atoms of the amino acid residues that had any atom within a cutoff distance of 5 Å from the trapped water and applied a positional restraint on each of these atoms. If less than three amino acids were selected, we increased the cutoff distance by 1 Å iteratively until at least three amino residues were selected for the restraint. We defined two *ifdef* sections for the binding site restraints. In the first *ifdef* section, each of the restraints was applied with a force constant of 0.5 kcal mol $^{-1}$ Å $^{-2}$ (the on state). In the second *ifdef* section, the force constant of each of the restraints was changed from 0.5 kcal mol $^{-1}$ Å $^{-2}$ to 0 kcal mol $^{-1}$ Å $^{-2}$ (the switching state).

Only for the RBE calculations, we simulated our systems using a ligand topology that combined the two ligands into a hybrid (see Section 2.6.1 of the main text for details). The ligand topology was generated using the PMX software package,⁴ where ligand A was written on the left and ligand B on the right. We extracted the individual topologies of the two ligands from the hybrid topology and we defined the ligand topology using three *ifdef* sections—the first *ifdef* section for the topology of ligand A, the second *ifdef* section for the topology of ligand B, and the third *ifdef* section for the hybrid topology where ligand A was changed into ligand B.

S2. Results

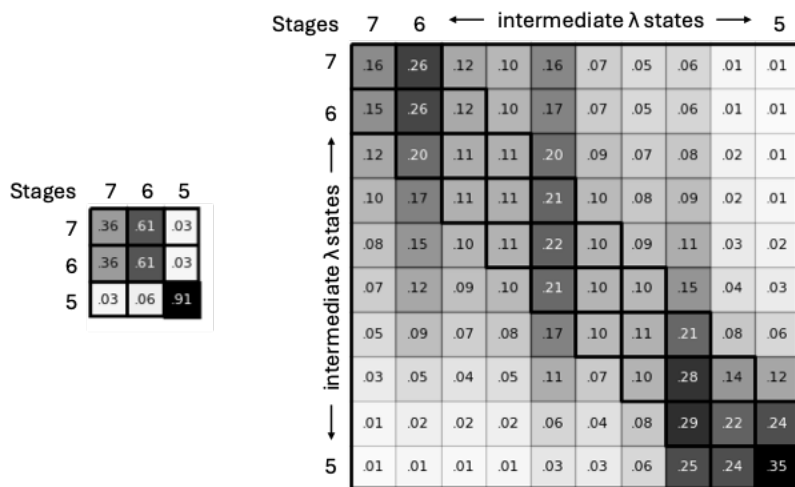


Figure S3. Phase space overlap matrices for the equilibrium free energy calculation for removing the solvent repulsion and binding site restraints for the BPTI system. In the equilibrium free energy calculation, the trapped water was in a non-interacting state. In the simulation of the state without the binding site restraints (stage 6, Figure S1), the binding site collapsed, whereas in the simulation of the restrained state (stage 5, Figure S1), the binding site was restrained to its original shape. (Left) Phase space overlap matrix for the equilibrium calculation without the intermediate λ states between the unrestrained and the restrained states of the system. (Right) Phase space overlap matrix for the equilibrium calculation with the 7 intermediate λ states between the unrestrained and the restrained states of the system.

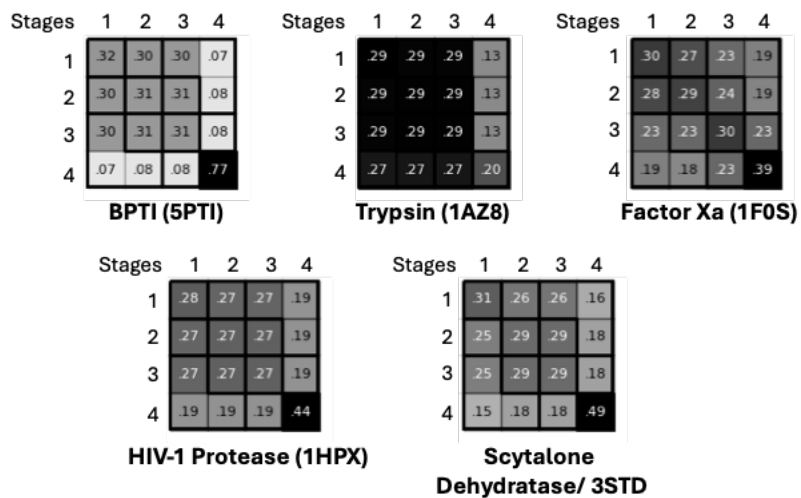


Figure S4. Phase space overlap matrices for the equilibrium free energy calculations for applying the positional restraint on the trapped water, solvent repulsion and binding site restraints (Figure S1). In the equilibrium free energy calculations, the trapped waters were in the interacting state.

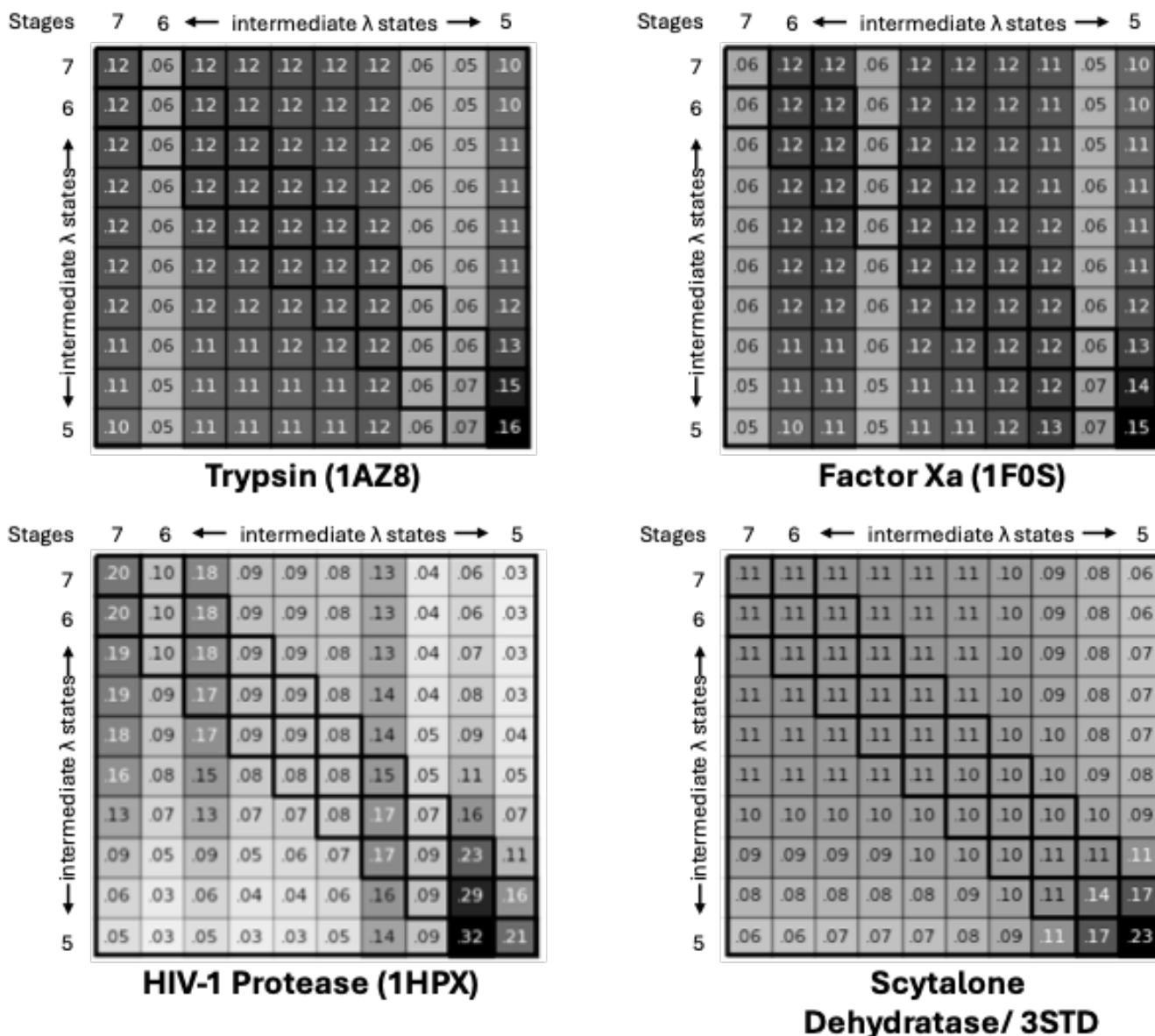


Figure S5. Phase space overlap matrices for the equilibrium free energy calculations for removing the solvent repulsion and binding site restraints (Figure S1). In the equilibrium free energy calculations, the tapped waters were in the non-interacting state. The overlap matrix for the system BPTI (5PTI) is shown in Figure S3.

Table S3. Relative binding free energies (RBFEs) (in kcal mol⁻¹) for ligand pairs studied here

| System | Ligands | RBF E_{calc}^a | RBF E_{calc}^b | RBF $E_{exp.}$ |
|-----------------------|---------|------------------|------------------|----------------|
| thrombin | B5-B1a | -0.83 ± 0.2 | -0.97 ± 0.4 | -0.61 |
| scytalone dehydratase | C3d-C5d | -2.28 ± 0.7 | -1.59 ± 0.5 | -1.98 |
| factor Xa | IID-IIE | -3.09 ± 0.3 | -2.46 ± 0.5 | -1.98 |
| BACE1 | C4j-C4b | 0.032 ± 0.8 | 0.32 ± 0.6 | -0.61 |
| BTK | S8-S11 | -4.25 ± 0.6 | -4.08 ± 0.7 | -1.91 |

^a RBFEs are calculated using the bigger thermodynamic cycle shown in Figure 3; uncertainties are calculated over three simulation repeats. ^b RBFEs are calculated using the smaller thermodynamic cycle shown in Figure 4; uncertainties are calculated over three simulation repeats.

References

- (1) Bauer, P.; Hess, B.; Lindahl, E. GROMACS 2022 Source code. 2022; <https://zenodo.org/records/6103835>.
- (2) Shirts, M. R.; Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *The Journal of Chemical Physics* **2005**, *122*, 134508.
- (3) Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of nonionic solutes. *The Journal of Chemical Physics* **1984**, *81*, 2016–2027.
- (4) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. pmx: Automated protein structure and topology generation for alchemical perturbations. *Journal of Computational Chemistry* **2015**, *36*, 348–354.