Supplementary Information

Estimating the Roles of Protonation and Electronic Polarization in Absolute Binding Affinity Simulations

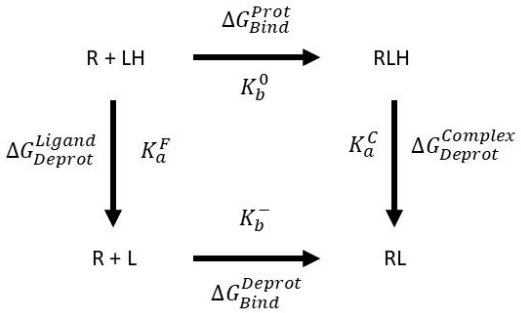
Edward King[1], *Ruxi Qi*[4], *Han Li*[2], *Ray Luo*[1,2,3*], *and Erick Aitchison*[1*]

[1]Departments of Molecular Biology and Biochemistry, [2]Chemical and Biomolecular Engineering, [3]Materials Science and Engineering and Biomedical Engineering, University of California, Irvine, California 92697 (USA)
[4]Cryo-EM Center, Southern University of Science and Technology, Shenzhen, Guangdong 518055 (China)

SI Discussion

Apparent Binding Free Energies with One Titratable Group in the Active Site

The presence of titratable groups in the active site poses an additional challenge for the calculation of binding free energies. The experimentally resolved apparent binding free energies encapsulate the whole physical process, which may not distinguish the contributions of coupled processes to the protein-ligand binding event. Titratable groups in the active site are susceptible to the system pH, and this susceptibility is observed in the pH dependence of receptor-ligand binding[1, 2] and enzymatic catalysis[3]. This can be further complicated when the interaction of the binding ligand shifts the pKa of those titratable groups which can alter protonation states. These interactions and coupled processes need to be considered for the binding free energy calculations. Using a single titratable group as a model coupled process, one can derive the separable contributions from the coupled processes. The coupled binding process can be separated into four separate processes: binding in the protonated form, binding in the deprotonated form, and protonation/deprotonation processes in the free and complex states defined in the illustrated thermodynamic cycle where one of the binding processes is directly calculated.



The apparent binding constant can be expressed with a proton dissociation process for both the complex and free states,

$$K_{app} = \frac{[RLH] + [RL^-]}{[R][LH] + [R][L^-]}$$

This can be further simplified by substitution to give,

$$K_{app} = K_b^0 \frac{\left(1 + (K_a^C)^{-1}[H]\right)}{\left(1 + (K_a^F)^{-1}[H]\right)}$$

where $K_b^0 = \frac{[RLH]}{[R][LH]}$ is the protonated binding equilibrium constant, $K_a^C = \frac{[RL^-][H]}{[RLH]}$ and $K_a^F = \frac{[L^-][H]}{[LH]}$ are the proton dissociation constants in the complex and free states, respectively. The proton dissociation constants can be expressed in pKa and pH units, and the change in free energy can be calculated using

$$\Delta G^o(pH) = -k_b T \left[\ln K_b^0 + \ln \frac{(1 + 10^{pH - pKa^C})}{(1 + 10^{pH - pKa^F})}\right]^{4,5}$$

The charged binding equilibrium constant can be converted to the binding free energy which can be explicitly calculated, where $\Delta G_{bind}^{Prot} = -k_b T \ln K_b^0$, resulting in

$$\Delta G^o(pH) = \Delta G_{bind}^{Prot} - k_b T \ln \frac{(1 + 10^{pH - pKa^C})}{(1 + 10^{pH - pKa^F})}$$

Additionally, this equation then becomes process dependent, where the equation for binding coupled to a proton association process is,

$$\Delta G^o(pH) = \Delta G_{bind}^{Deprot} - k_b T \ln \frac{(1 + 10^{pKa^C - pH})}{(1 + 10^{pKa^F - pH})}$$

Application of the above equation shows that computation of the apparent binding free energy requires the pKa's of the ligand in both the free and bound states in addition to the simulated binding affinity of the ligand in either of the states. However, the application of this simplified single titratable group coupled binding process equation is apparently inadequate in describing the complete binding process for most protein systems where many residues in the active site are also titratable and require proper modeling.

SI Tables

Condition	Epsin	RMSE (kcal/mol)	R
All HIP	1	2.31	0.60
All HID	1	3.75	0.75
Small HIP	1	2.89	0.85
Small HID	1	3.32	0.17

Table 1. MBAR/PBSA binding affinity accuracy with optimized Radiscale and Protscale parameters. Radiscale and Protscale values were scaled to minimize mean absolute error between MBAR/PBSA free energies and explicit solvent free energies.

Condition	Epsin	RMSE (kcal/mol)	R
All-HIP	1.17	1.25	0.74
	2	4.84	0.81
All-HID	1.43	0.89	0.88
	2	2.65	0.88
Small-HIP	1.27	1.00	0.88
	2	3.91	0.87
Small-HID	1.28	1.80	0.32
	2	3.90	0.52

Table 2. Binding affinity prediction accuracy versus solute interior dielectric (Epsin) with MBAR/PBSA. Commonly used Epsin of 2.0 and Epsin resulting in the lowest RMSE to experiment are reported. All-HID condition shows the lowest RMSE and highest Pearson correlation at optimized Epsin. Epsin 2.0 results in improved Pearson correlations, but also higher RMSE's.

Sample	Baseline	Baseline + 150mM salt	Baseline + deprotonated ligands	1DOF All-Hip	6DOF All-Hip	All HID	Small HIP	Small HID	PBSA All HIP	PBSA All HID	PBSA Small HIP	PBSA Small HID	Experiment
1C5X	-13.32	-11.15	-13.32	-11.15	-14.29	-12.68	-11.15	-12.68	-8.81	-8.55	-7.78	-9.68	-9.01
1C5Y	-10.83	-10.37	-10.83	-10.37	-13.91	-11.66	-10.37	-11.66	-7.47	-6.93	-6.35	-8.09	-5.67
1C5Z	-9.89	-6.87	-9.89	-6.87	-8.95	-10.74	-6.87	-10.74	-4.60	-6.60	-3.57	-7.70	-5.42
1GI7	-8.88	-9.68	-8.88	-9.68	-11.39	-10.26	-9.68	-10.26	-7.93	-6.81	-7.21	-7.73	-6.09
1GJ7	-6.87	-7.35	-13.10	-12.08	-17.74	-13.41	-13.41	-12.08	-10.12	-10.49	-11.23	-9.19	-10.86
1GJ8	-8.35	-6.89	-7.92	-7.99	-13.66	-11.51	-11.51	-7.99	-7.13	-7.86	-8.77	-6.71	-9.39
1GJA	-8.49	-8.15	-6.78	-7.82	-9.15	-10.87	-10.87	-7.82	-6.32	-6.38	-7.60	-5.54	-7.32
1GJB	-8.64	-8.44	-9.89	-11.23	-15.75	-12.50	-12.50	-11.23	-8.56	-8.58	-9.59	-7.47	-8.57
1GJD	-4.05	-5.09	-5.47	-5.08	-9.14	-4.85	-4.85	-5.08	-3.60	-3.03	-3.48	-2.73	-7.05
1O3P	-7.66	-10.36	-12.63	-11.41	-16.18	-12.73	-12.73	-11.41	-9.19	-8.81	-9.86	-8.24	-8.99

Table 3. Full binding predictions at all conditions tested compared to experimental values. Absolute binding free energy calculations aggregated from 5 independent replicates with randomized starting velocities. All units reported in kcal/mol.

SI Figures

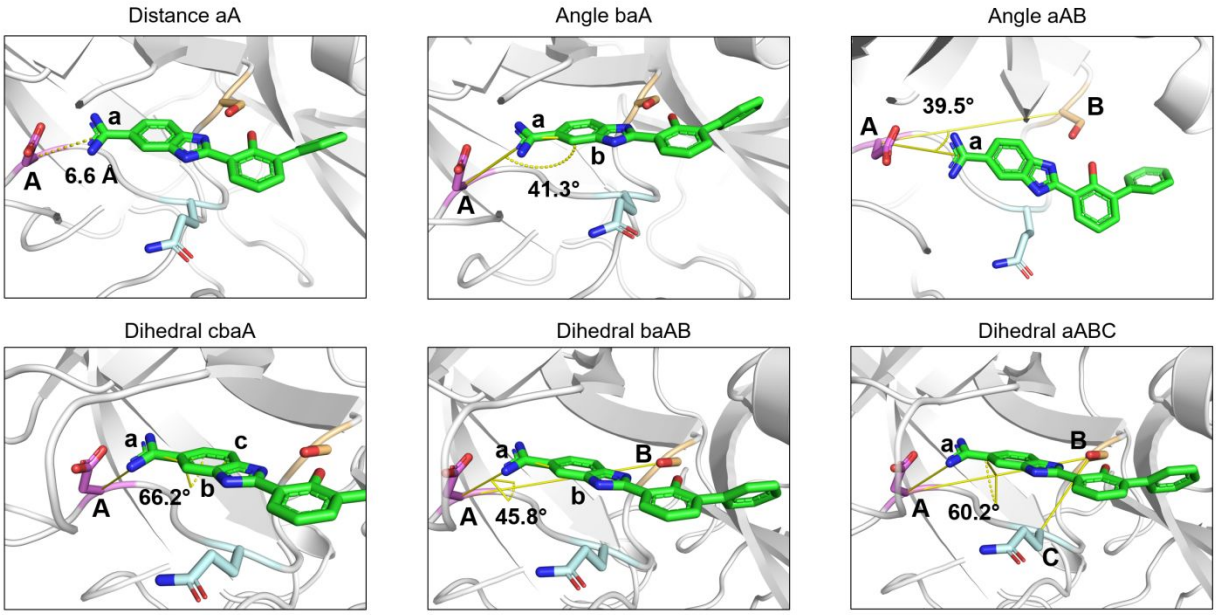


Figure 1. Illustration of Boresch 6DOF orientational restraints. The ligand is constrained by a single distance, two angles, and three dihedrals selected from the end of the equilibration phase to lock the ligand into a target conformation. 1DOF condition involves only the distance restraint, which allows greater exploration of conformational states at the cost of slower convergence.

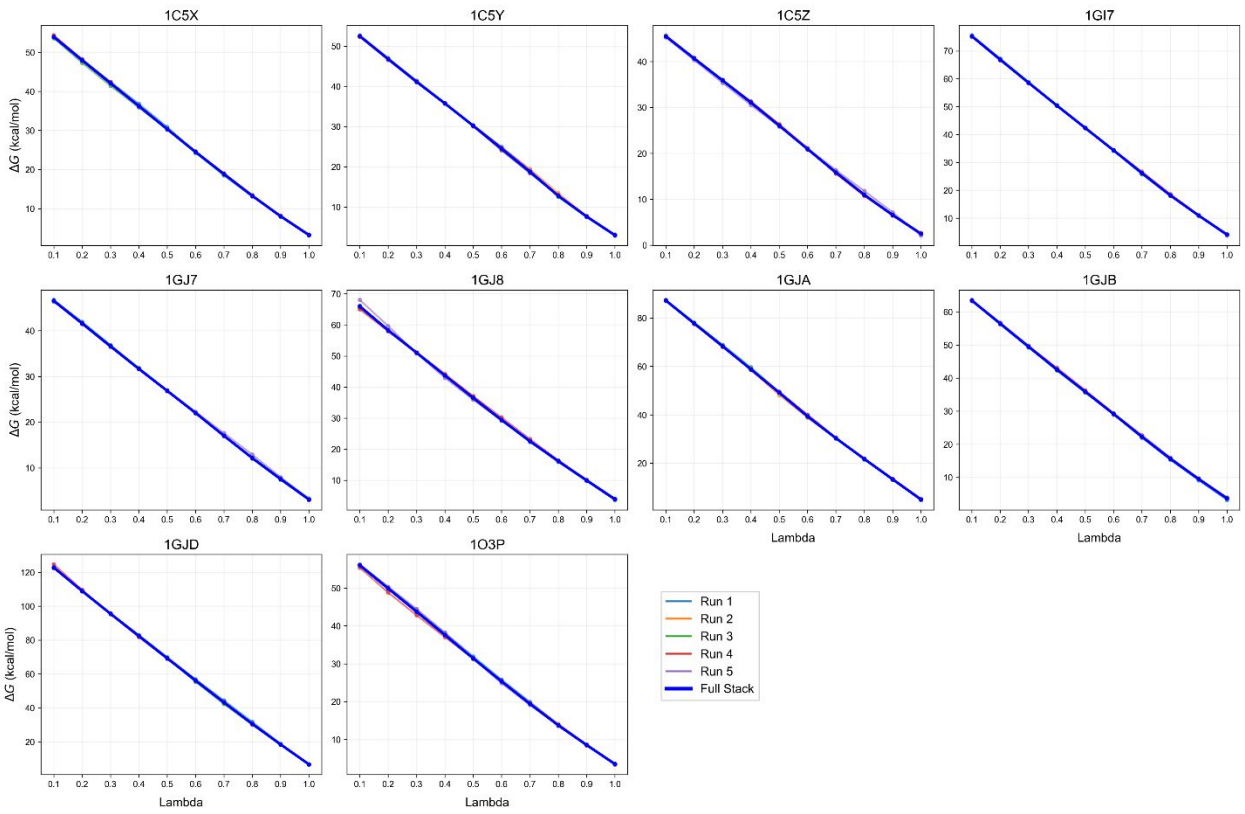


Figure 2. Free energy transitions during the decharging phase for the complex trajectories in the baseline simulation. Individual replicates show only small variation, the aggregated energies show almost complete overlap and smooth, nearly linear transition from full ligand partial charges to zero.

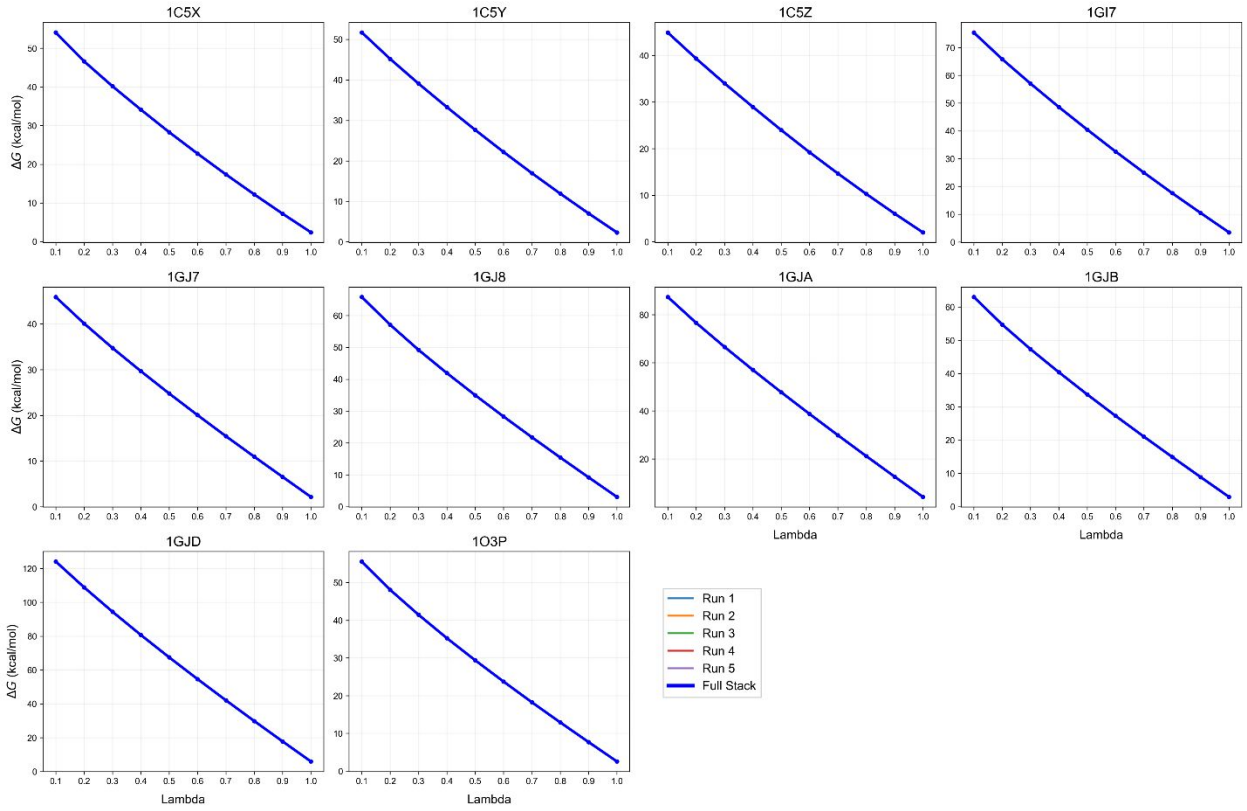


Figure 3. Free energy transition during the decharging phase for the ligand trajectories in the baseline alchemical simulation. The same pattern of small variation and linear transition from full ligand partial charges to zero as the complex is observed.

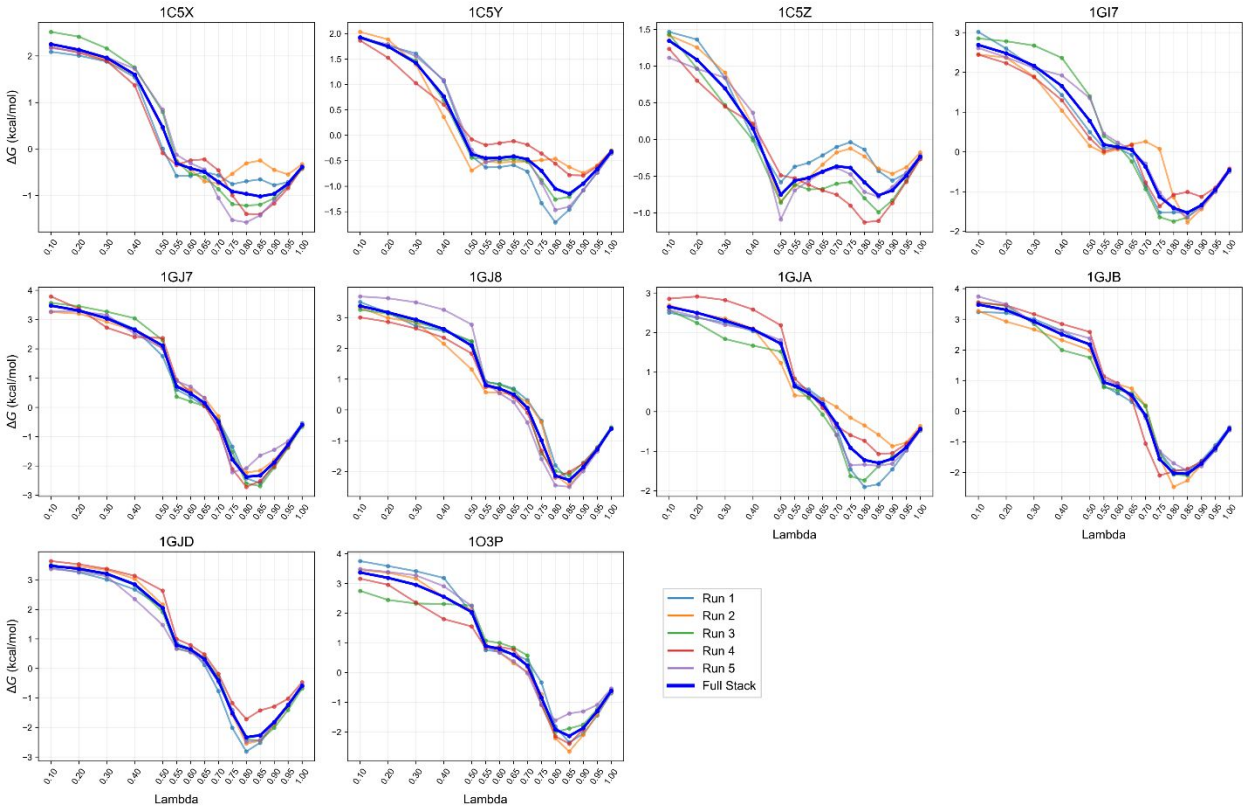


Figure 4. Free energy transition during the VDW phase for the complex trajectories in the baseline simulation. High variance is observed between replicates, highlighting the sampling difficulties associated with decoupling VDW interactions.

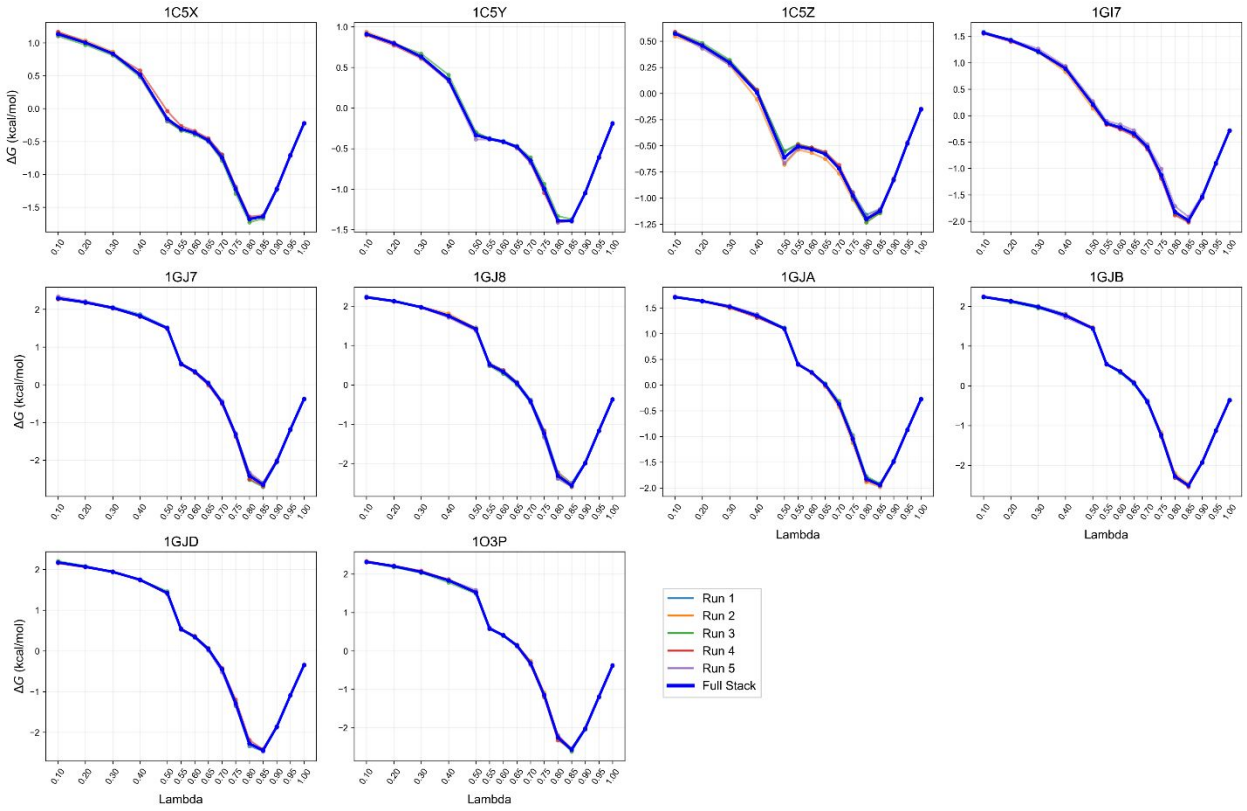


Figure 5. Free energy transition during the VDW phase for the ligand trajectories in the baseline simulation. Replicates show high agreement over the course of the highly non-linear transitions.

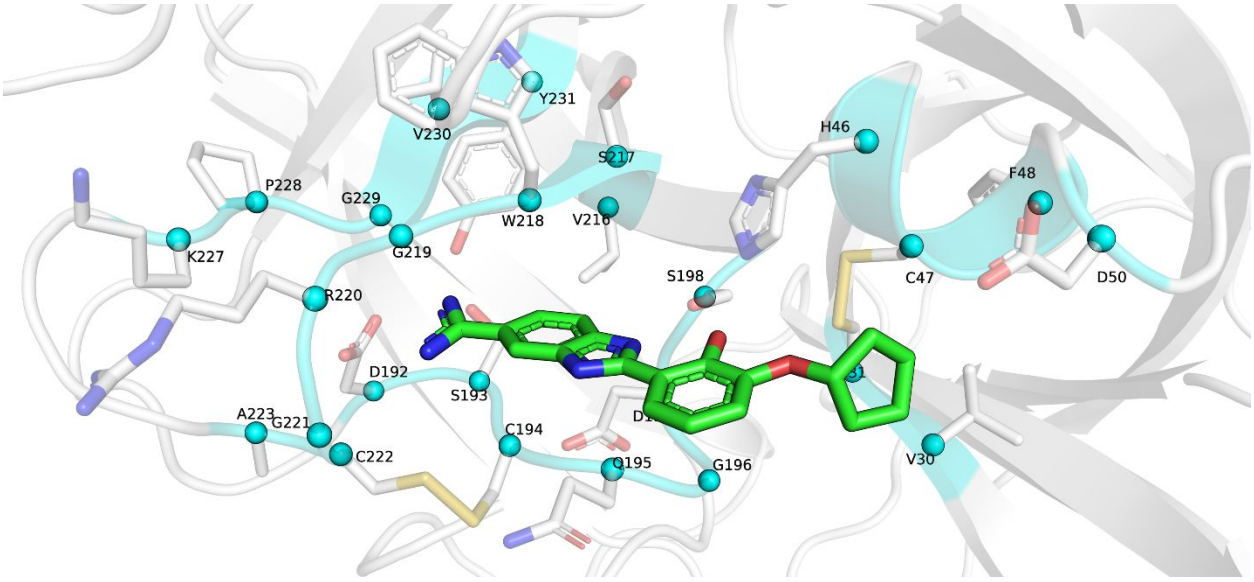


Figure 6. Illustration of the UPA binding pocket with all residues within 6 Å of the ligand highlighted. Notable residues include His-46 which is titratable and observed to form a hydrogen bond with the ligand phenol. Asp-192 is located at the base of the binding pocket and forms salt bridges with the positively charged amidine. Sample ligand 1O3P is highlighted in green.

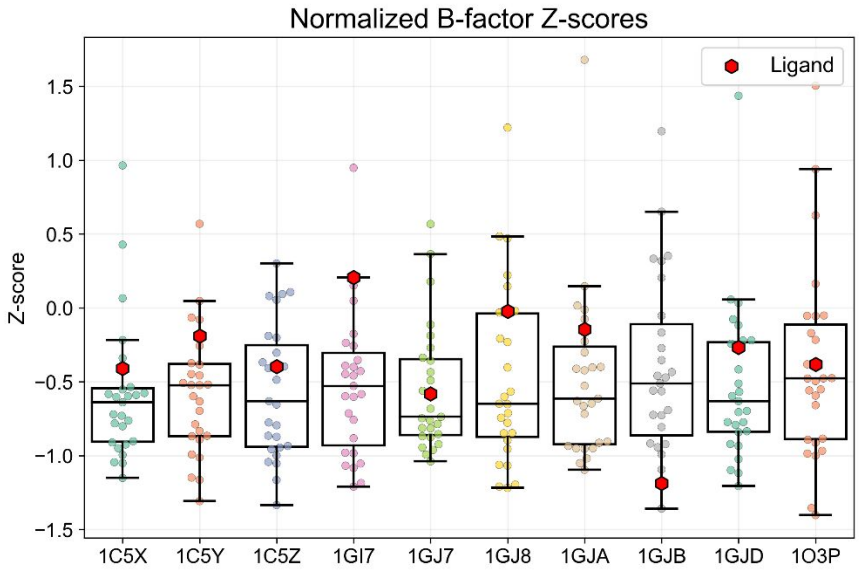


Figure 7. Analysis of binding pocket flexibility through normalized B-factor Z-scores. All structures show similar binding pocket flexibility, with higher than average rigidity relative to the rest of the protein. Ligands show varying levels of displacement, notably 1GI7 shows the highest flexibility, which is larger in size but unable to form a hydrogen bond to Ser-198. 1GJB shows the highest stability, potentially due to its hydrophobic benzene groups and internal hydrogen bond between the ligand phenol and nitrogen. Each marker represents the Z-score per residue with all atoms averaged.

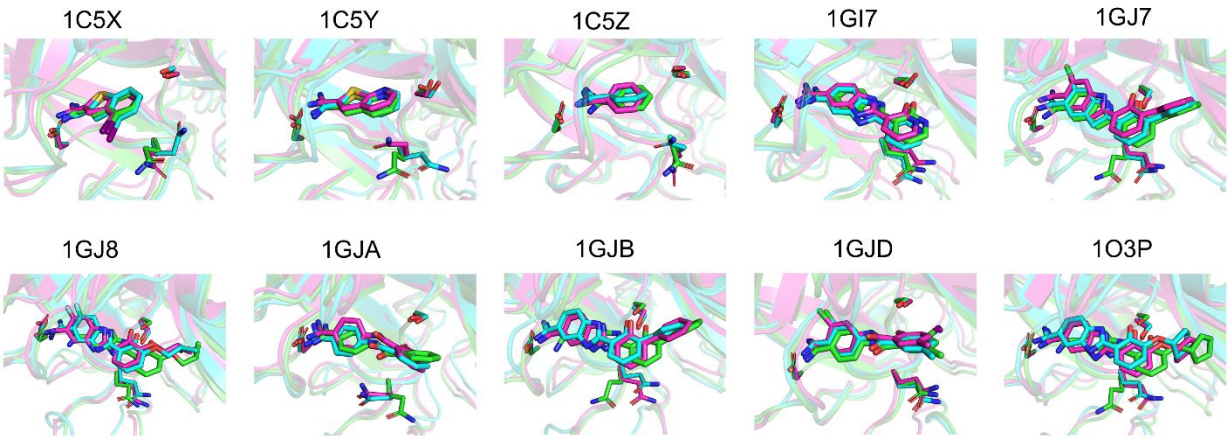


Figure 8. Inhibitor equilibration poses from GAFF and GAFF2 compared to starting crystal poses. GAFF and GAFF2 trajectories show similar trends, with the ligands moving further into the binding pocket to more tightly interact with Asp-192, and outward twisting of the phenol tail to relieve steric clash. Structures were generated from identifying the frame with the lowest RMSD to the average structure from the last 10 ns of equilibration. The starting crystal structure models are colored green, GAFF samples are colored cyan, and GAFF2 samples are colored purple.

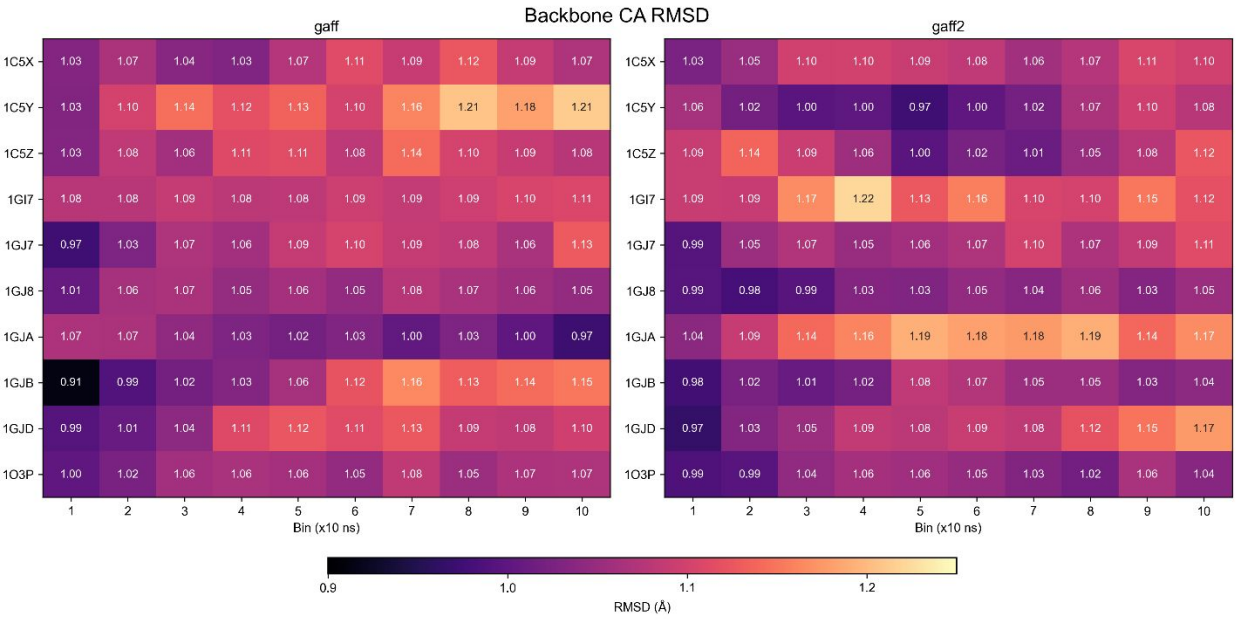


Figure 9. Backbone CA RMSD development over equilibration with GAFF and GAFF2 force fields. No clear pattern emerges, all proteins drift away from the starting ligand pose and show a maximum divergence of ~1.2 Å RMSD, indicating that minor conformational changes occur.

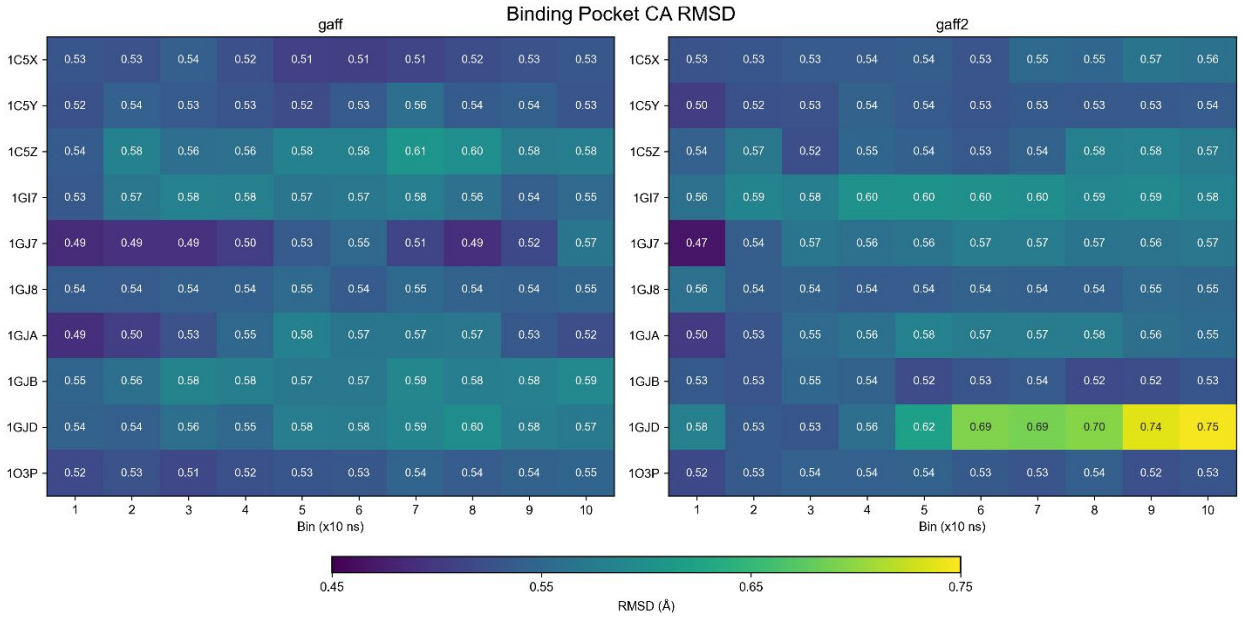


Figure 10. Binding pocket CA RMSD development over equilibration with GAFF and GAFF2 force fields. All GAFF samples show stability and do not change noticeably from the crystal over the course of equilibration. In GAFF2, 1GJD shows larger divergence from the crystal pose reaching 0.75 A RMSD.

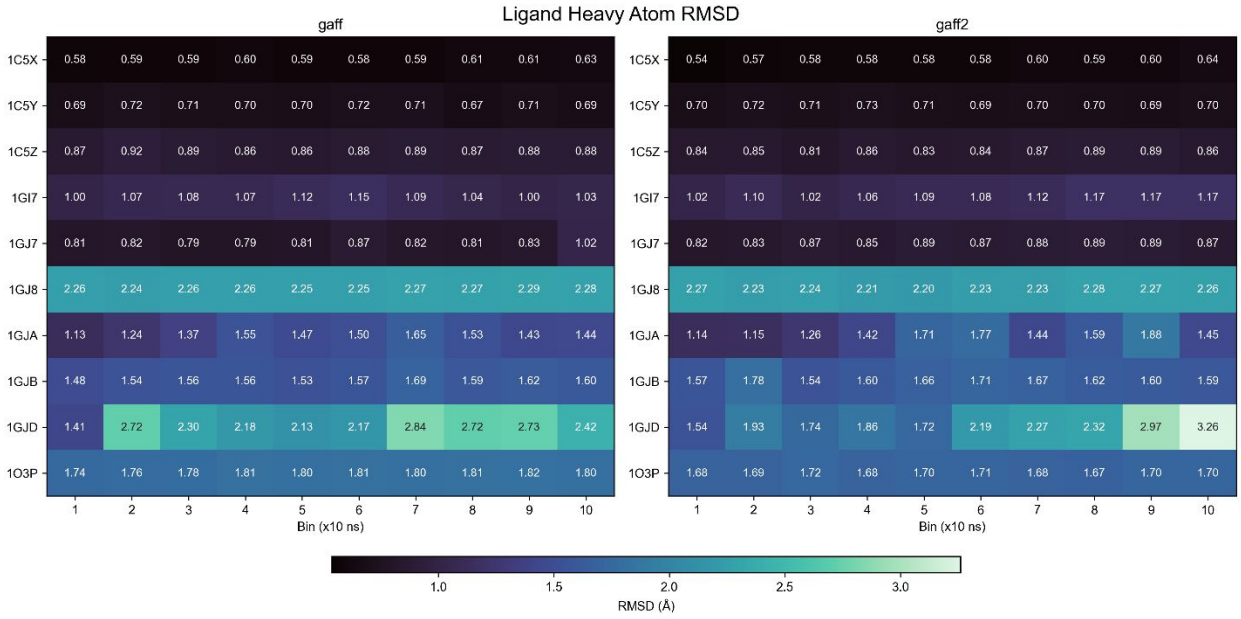


Figure 11. Ligand heavy atom RMSD development over the equilibration with GAFF and GAFF2 force fields. Small ligands (1C5X, 1C5Y, 1C5Z, and 1GI7) show minimal changes in positioning. 1GJ8 shows consistent departure from the crystal pose, the ligand moves further into the binding pocket to maximize hydrophobic interactions and polar interactions with Asp-192. 1GJD shows dissimilarity with crystal as well, from the rotation of the phenol group outward leading to the loss of the hydrogen bond. The aberration is more substantial with GAFF2.

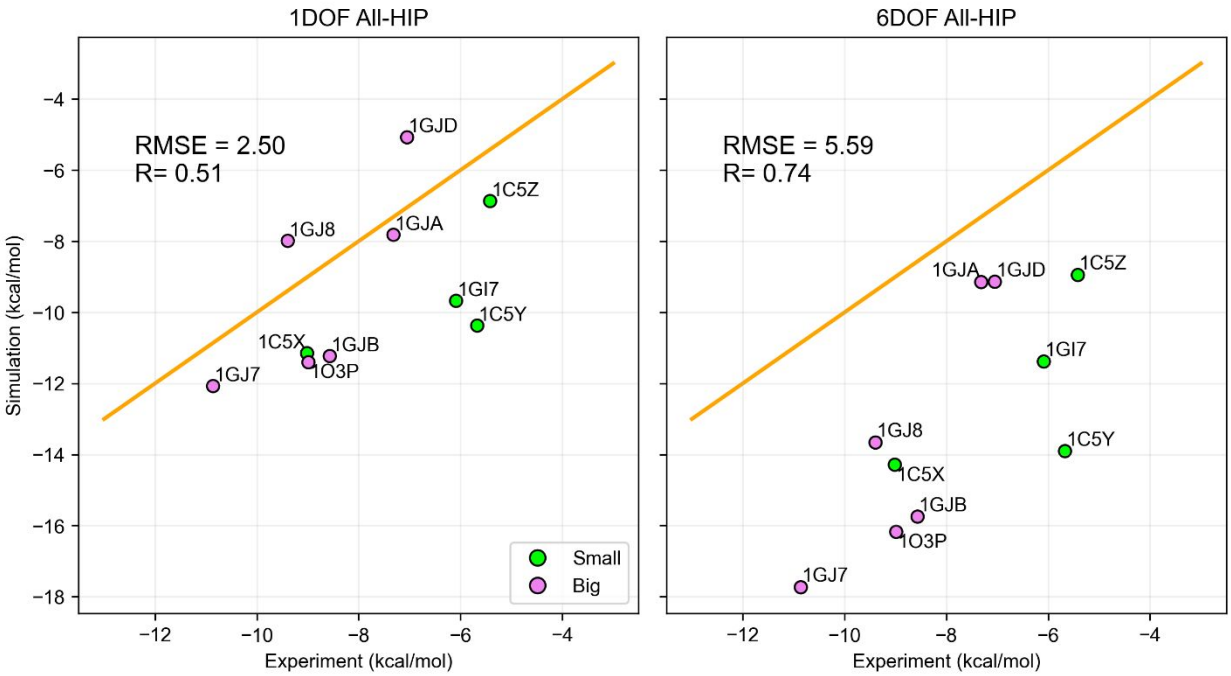


Figure 12. Comparison of 1DOF and 6DOF restraint schemes. The 1DOF single distance restraint showed lower error, but worse Pearson correlation than the 6DOF (Boresch) method. Samples with the 6DOF restraint showed excessively negative free energy predictions, indicating potential over-stabilization in a favorable pose.

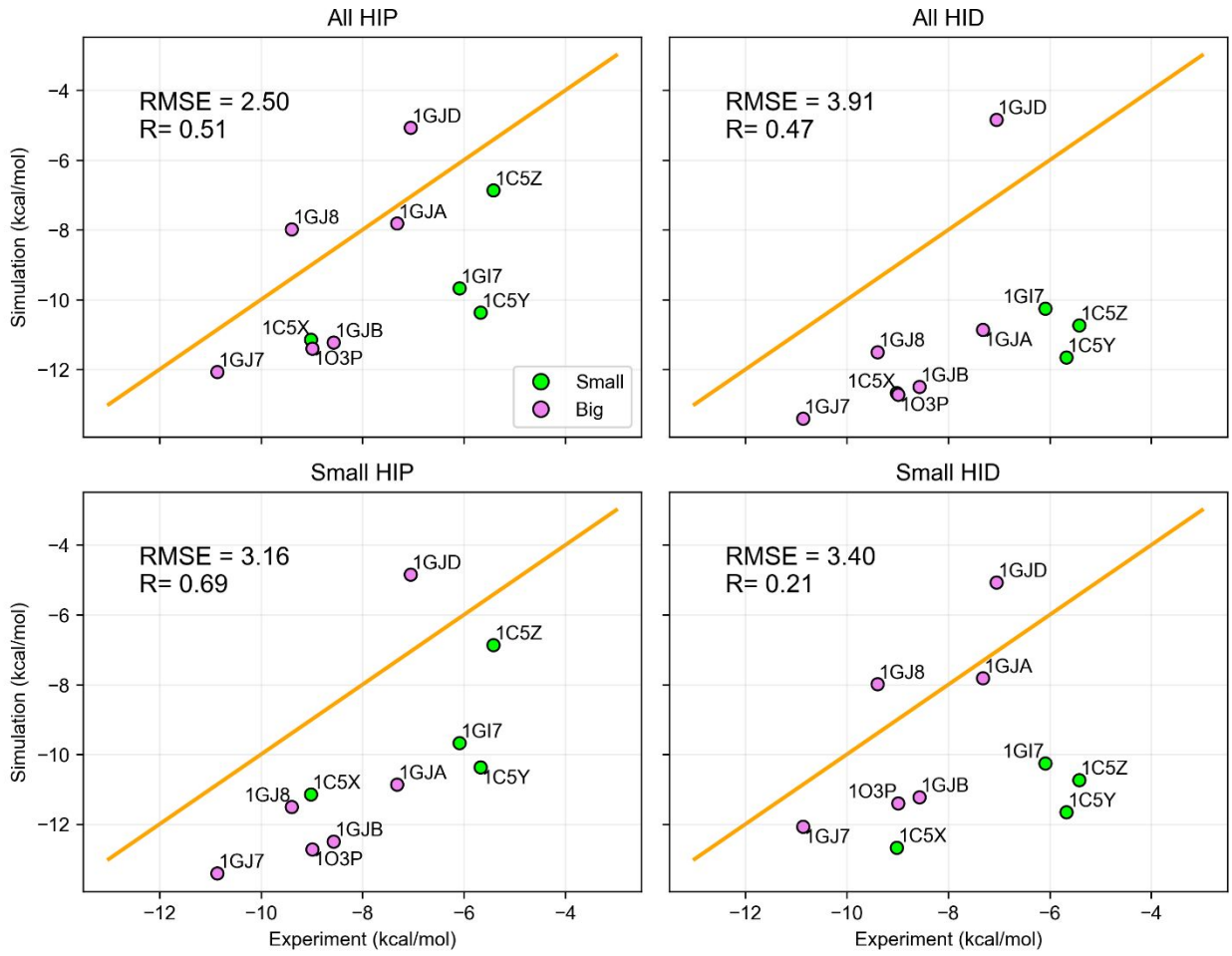


Figure 13. Binding affinity predictions with standard alchemical simulation with different protonation states. In general, binding affinities are predicted to be more negative than expected, possibly due to exaggeration of favorable charge-charge interactions typical of the point-charge models used. 1GJD is shown to be an outlier, with free energies far more positive than the cluster of other tested ligands, this is likely related to the issues in sampling incorrect binding poses recognized during equilibration where the phenol swings outward such that the native hydrogen bond to Ser-198 is not maintained.

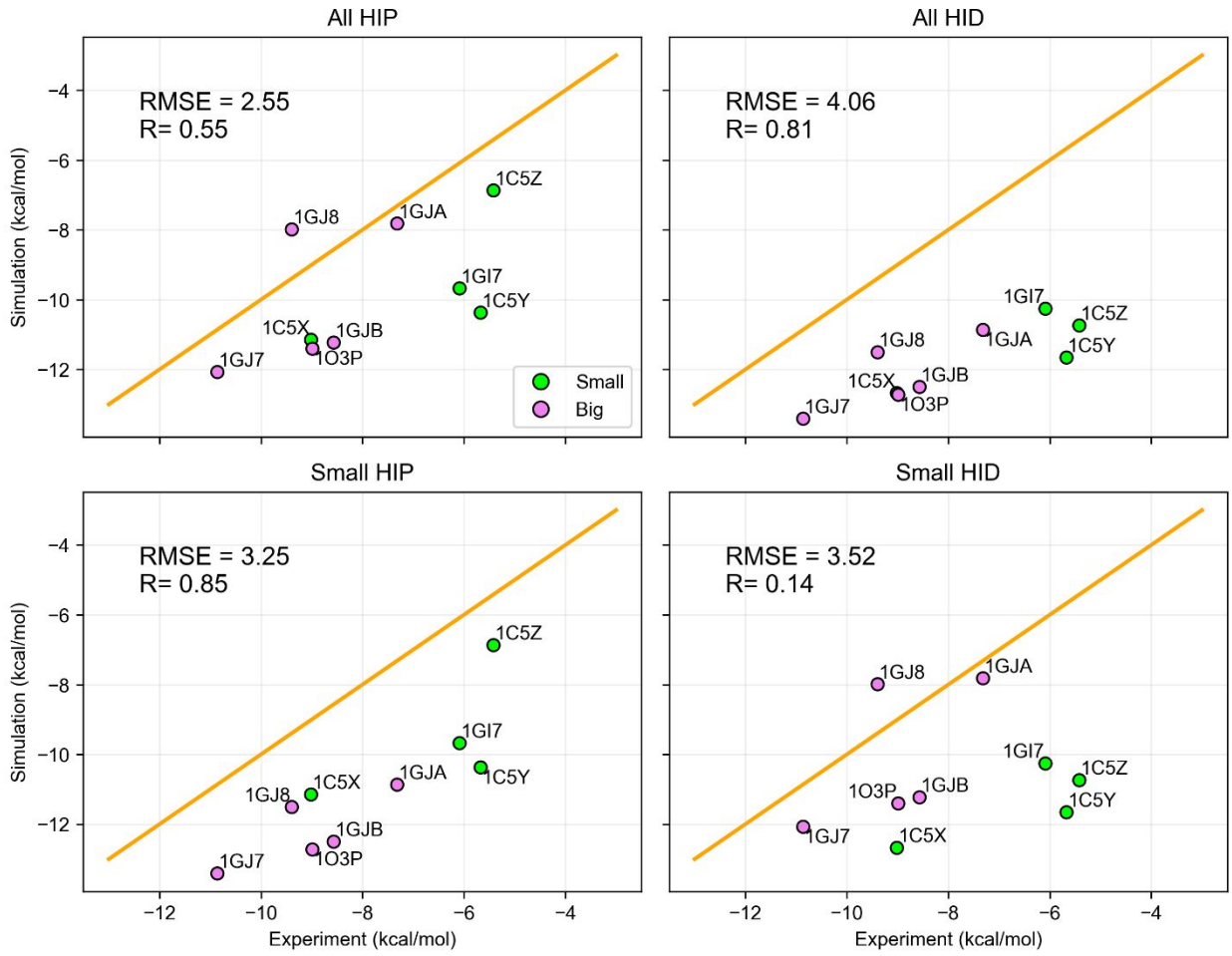


Figure 14. Binding affinity predictions with outlier 1GJD removed for standard alchemical simulation with different protonation states. In the standard alchemical simulation, minimal change is seen in RMSE for all conditions. However, Pearson correlation is found to improve dramatically for both All-HID and Small-HIP conditions.

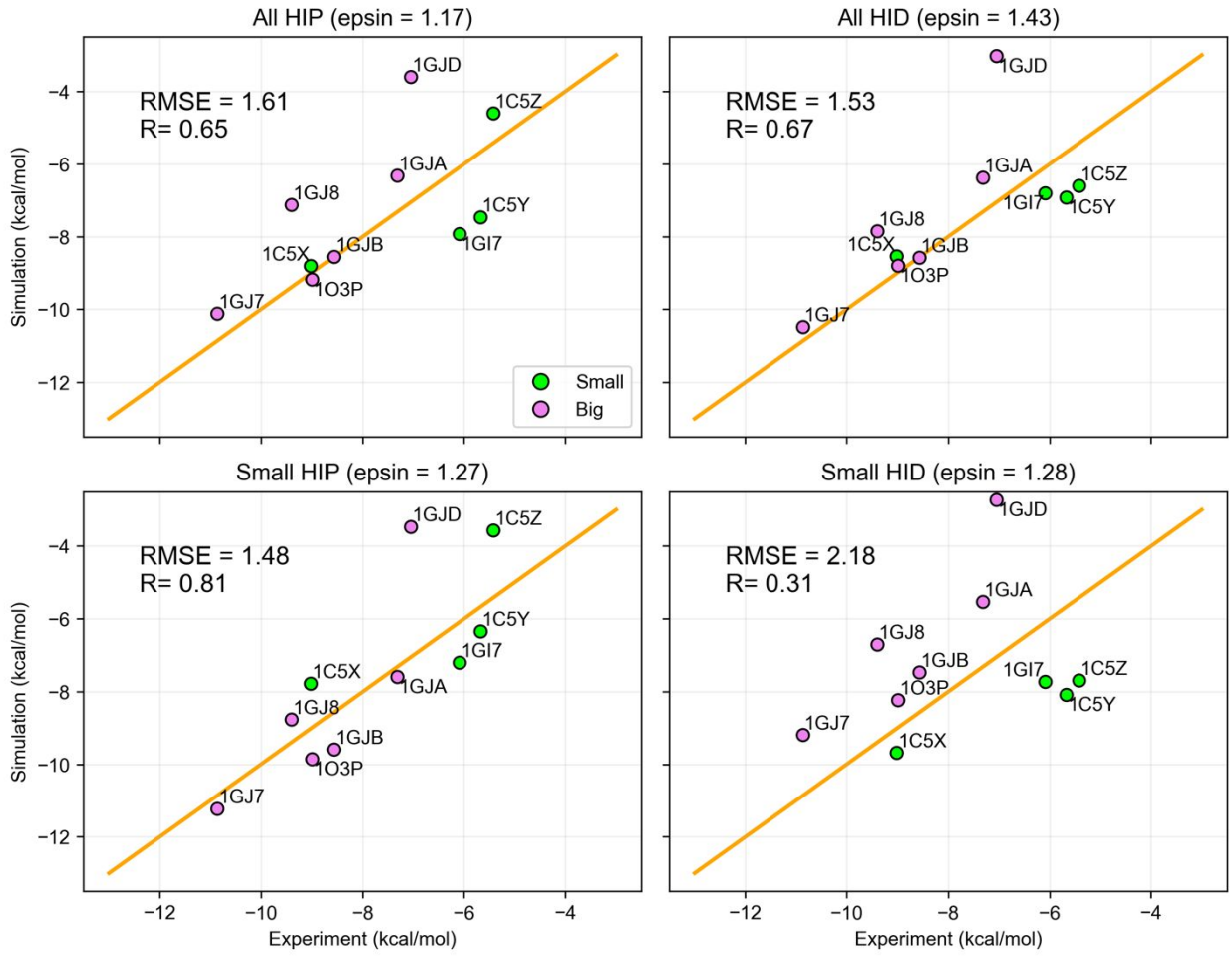


Figure 15. MBAR/PBSA binding affinity calculations including the outlier 1GJD. All metrics are found to worsen with the outlier pushing the trend toward overly positive values.

SI References

1. Maiti, A.; Drohat, A. C., Dependence of substrate binding and catalysis on pH, ionic strength, and temperature for thymine DNA glycosylase: Insights into recognition and processing of G.T mispairs. *DNA Repair (Amst)* **2011,** *10* (5), 545-53.
2. Hansen, M. J.; Olsen, J. G.; Bernichtein, S.; O'Shea, C.; Sigurskjold, B. W.; Goffin, V.; Kragelund, B. B., Development of prolactin receptor antagonists with reduced pH-dependence of receptor binding. *J Mol Recognit* **2011,** *24* (4), 533-47.
3. Talley, K.; Alexov, E., On the pH-optimum of activity and stability of proteins. *Proteins* **2010,** *78* (12), 2699-706.
4. Jensen, J. H., Calculating pH and salt dependence of protein-protein binding. *Curr Pharm Biotechnol* **2008,** *9* (2), 96-102.
5. Mason, A. C.; Jensen, J. H., Protein-protein binding is often associated with changes in protonation state. *Proteins* **2008,** *71* (1), 81-91.