

Supporting Information:

Overcoming Challenging Substituent Perturbations with Multisite λ -Dynamics: A Case Study Targeting β -Secretase 1

Jonah Z. Vilseck,[†] Noor Sohail,[†] Ryan L. Hayes,[†] Charles L. Brooks III^{,†,‡}*

[†]Department of Chemistry, [‡]Biophysics Program, University of Michigan, Ann Arbor.

Corresponding Author

*E-mail: brookscl@umich.edu. Address: Department of Chemistry and Biophysics, University of Michigan, 930 N. University Ave., Ann Arbor, MI 48109.

This PDF file includes:

- Computational Details
- Supplementary Text
- Figures S1 to S10
- Tables S1 to S4
- Ligand Partial Atomic Charges
- References

COMPUTATIONAL DETAILS

System Setup. In keeping with the original Janssen Pharmaceutical (JP) publication,¹ initial starting coordinates for BACE1 were obtained from the 3ZOV structure.² Coordinates for missing residues 218-229 and 372-378 were modeled into the former structure utilizing their positions from the 2QK5 structure.³ To relax contacts in the composite structure, the added residues were subjected to 100 steps of steepest descent minimization and 20 ns of molecular dynamics (MD) equilibration with the GBSW implicit solvent model;⁴ all other residues were constrained during the MD simulation to preserve the native 3ZOV structure and fold. Residue flips for histidine, glutamine, and asparagine were assessed using the MolProbity webserver^{5,6} and protonation states of titratable residues were determined with the assistance of PROPKA,⁷ corresponding to a pH of 7.0.¹ The catalytic aspartic acids were kept in their unprotonated states.¹ The protein-ligand complex was solvated using the CHARMM-GUI webserver with a minimum of 10 Å of solvent from each face of the protein-ligand complex,⁸ yielding a cubic water box of 90.0 Å; a sufficient number of Na⁺ and Cl⁻ ions were added to neutralize the net charge of the system and achieve an ionic strength of 150 mM NaCl. The JP ligand was solvated separately using the *convpdb.pl* tool from the MMTSB toolkit with a 12 Å solvent boundary from the ligand.⁹

All simulations were performed using the CHARMM molecular simulation package (developmental version c43a1) with the domain decomposition (DOMDEC) computational kernels on graphic processing units (GPUs).¹⁰⁻¹² Prior to molecular dynamics, each system was

subject to 200-500 steps of steepest decent minimization. MD simulations were run in the isobaric-isothermal (NPT) ensemble at 25° C and 1 atm using a Nose-Hoover thermostat^{13,14} and Langevin pressure piston with a friction coefficient of 20 ps⁻¹.¹⁵ The Leapfrog Verlet integrator was used with an integration time step of 2 fs and trajectory frames were saved every 500–1000 steps. Hydrogen-heavy atom bond lengths were constrained with the SHAKE algorithm.¹⁶ Periodic boundary conditions were employed with nonbonded cutoffs of 12 Å to truncate all long-range interactions and force switching was used to gradually tune these interactions to zero between 10–12 Å.

Force Field Parameterization. A variety of force field parameters were used in this work. The TIP3P water model was used for all explicit solvent calculations.¹⁷ BACE1 was represented with the CHARMM36 and OPLS-AA/M protein force fields.¹⁸⁻²⁰ The JP ligands were represented with the CHARMM General Force Field (CGenFF),²¹⁻²³ the OPLS-AA/CM1A small molecule force field,²⁴⁻²⁶ and a CGenFF/CM1A hybrid model, where CM1A partial atomic charges were substituted for the CGenFF charges but all remaining CGenFF parameters remained unchanged. R1 substituents consisted of the acylguanidinium heterocyclic ring plus the carbon they were bonded to within the N-(4-fluorophenyl)-acetamide core (Figure 1A and the “Ligand Partial Atomic Charges” section below). R2 substituents consisted of the aromatic moieties and did not include the carbonyl carbon to which they were bonded. MSλD utilizes a common core with a single set of force field parameters, most notably partial atomic charges and van der Waals parameters. To facilitate this expedient, and to ensure that all R1 and R2 atomic charges summed to an integer net charge for each ligand end-state, charges for the substituents and the core were

slightly modified. The algorithm for this adjustment will be presented in a forthcoming publication. The new partial atomic charges are reported below in the “Ligand Partial Atomic Charges” section. A root mean square analysis of charge differences (RMSQ) between original CGenFF charges and the adjusted charges revealed that minimal changes were made; an average RMSQ of 0.003 e was observed (Table S4). Furthermore, charge perturbations between charge states for each ligand in an explicit solvent environment suggested that the solvation free energies for these molecules were minimally affected: a mean unsigned difference of 0.12 kcal/mol and a maximum deviation of ~0.3 kcal/mol were observed (Table S4). These deviations are well within the noise for the protein-ligand binding calculations described below, and thus, are not expected to adversely affect the free energy results, ligand ranking, or conclusions drawn therefrom.

Multisite λ -Dynamics Calculations. All 3 R1 and 7 R2 substituents were explored simultaneously with MS λ D using a multiple-topology approach, i.e., explicit atomic representation of all substituents.²⁷⁻³³ The theoretical principles and constraints that facilitate this combinatorial investigation have been described in detail elsewhere.²⁷⁻³³ Unlike prior work, substituents were not harmonically restrained to each other and substituent dihedral angles were scaled by λ ; bonds, angles, and improper dihedral angles were not scaled by λ . This change was found to yield better sampling around the R1–core dihedral angle and improve $\Delta\Delta G_{\text{bind}}$ convergence. When dihedral angles were not scaled by λ or when the R1 substituents were harmonically tethered to each other, the R1 conformation often became trapped in a local energy minimum. Artifacts associated with trapping were minimized by the longer time scales of the MS λ D simulation, which allowed for more global sampling around the R1–core dihedral angle, but were severe for the much shorter TI simulations, described below. To prevent end-point singularities,³⁴ a soft-core potential was used

to scale all nonbonded interactions by λ .³¹ We note that even though force switching functions were used to truncate long range interactions, the particle mesh Ewald (PME) method for treating long-range electrostatic interactions is also available for use with MS λ D.³³ In a recent publication, differences of only 0.1–0.2 kcal/mol were observed in MUEs between MS λ D calculations with and without PME;³³ hence, we do not expect any deficiency in accuracy to be present in this work’s force switched-based results. The adaptive landscape flattening (ALF) algorithm was used to identify appropriate biasing potentials for MS λ D; several stages of ALF were run. First, 40–50 100 ps simulations, followed by 10–15 1 ns simulations were run to identify initial biases.³¹ Due to the more challenging substituent modifications explored in this work, an additional set of 5 duplicate 5 ns calculations were used to further refine the biases. Production simulations commenced with 3–4 ns of equilibration and ran for 20–30 and 40 ns for water and protein-bound simulations, respectively. Five independent production runs were performed using different random seeds for each force field parameter set investigated: CHARMM36+CGenFF, CHARMM36+CGenFF/CM1A, and OPLS-AA/M+OPLS-AA/CM1A. Biasing potential replica exchange (BP-REX MS λ D) was also used with 5 replicas per production run.³⁰ For the CHARMM36+CGenFF parameter set, all ALF and production sampling totaled to 1.84 μ s of MS sampling. Replicas introduced a ± 0.68 kcal/mol offset from the initial fixed biases obtained from ALF between replica neighbors, as described previously.³² The holonomic constraint variable “fnex” was 5.5 for all pre-production simulations run with ALF, and 10.5 for all production calculations.²⁹ End-state populations were binned using a $\lambda > 0.99$ cutoff criteria. Final $\Delta\Delta G_{\text{bind}}$ were calculated by Boltzmann reweighting end-state populations to the original biases and then by

use of equation 1 of the manuscript.³² Uncertainties, σ , for each $\Delta\Delta G_{\text{bind}}$ were calculated as the standard deviation of the mean over 5 independent trials. Equation 3 of the manuscript was then used to convert all relative free energies into absolute free energies for comparison to experiment.¹

TI/MBAR Calculations. Pairwise perturbations were explored with thermodynamic integration and the multistate Bennett acceptance ratio (TI/MBAR) using a dual topology approach.^{35,36}

Utilizing the same MS λ D software described above, the *fix* keyword was used to prevent propagating the λ variables dynamically and independent MD simulations were run at each discrete λ state. Figure S2 describes the R1 and R2 site perturbations, with direction specific arrows representing the transformation pathways used in this work. Redundant calculations and closed cycles were used to eliminate hysteresis and reduce error propagation among subsequent chains of relative perturbations.^{37,38} Because transformations are pairwise and site specific, and to be comprehensive in our comparison between MS λ D and TI/MBAR, each set of perturbations, R1 and R2, were performed with every alternate site substituent, R2 and R1 respectively (Figure S2). In total, 45 perturbations were performed to calculate all 21 $\Delta\Delta G_{\text{bind}}$. These perturbations were run in triplicate in both water and protein-bound states of the thermodynamic cycle (Figure S1) to determine statistical uncertainties. Each perturbation consisted of 11 discrete λ states, 0 \rightarrow 1 in steps of $\Delta\lambda = 0.1$, and MD simulations were run for 5 ns at each λ state. A total of 14.85 μ s of MD sampling was expended. Longer simulations could have been performed for each λ -window to further improve convergence. For example, Tresadern and co-workers found notable improvements for several relative FEP+ perturbations when sampling was increased to 20 ns or greater per λ -window.¹ However, this would only amplify efficiency differences between MS λ D

and TI/MBAR. To mimic largescale FEP benchmarking efforts in the literature,³⁹ this work used 5 ns simulations. MD trajectories were then postprocessed to calculate the necessary energies for MBAR to obtain the final $\Delta\Delta G_{\text{bind}}$ results. Due to the high expense of running these calculations, TI/MBAR was performed for only the CHARMM36+CGenFF force field set to illustrate the anticipated correlation between MS λ D and TI/MBAR. The excellent agreement observed between MS λ D and TI/MBAR (Figure 2) is expected for the other force field parameter sets.

Hydration Free Energies. To assess the accuracy of the partial atomic charges (see the “Supplementary Text” section below), free energies of hydration were also calculated for the 21 JP ligands using CGenFF and CGenFF/CM1A force fields with MS λ D. Using a traditional thermodynamic cycle, alchemical transformations can be performed in water and vacuum to calculate relative free energies of hydration.⁴⁰⁻⁴² The free energy results from the water calculations described above were retained with new gas phase calculations. System setup, force field parameterization, general MD parameters, and bias determination with ALF all mimicked the aqueous phase calculations. Production runs were equilibrated for 2 ns, and then 5 independent production runs of 20 ns each were performed. For consistency, BP-REX was also employed. Excellent convergence was observed for the free energy results and relative free energies of hydration are plotted in Figure S7, where 6A is set as the reference ligand.

SUPPLEMENTARY TEXT

Hydration Free Energies. Hydration free energies (ΔG_{hyd}) have become a standard test for evaluating the quality of partial atomic charges in condensed phase molecular mechanics

simulations.⁴⁰⁻⁴² Relative hydration free energies were computed for all 21 JP ligands using CGenFF parameters with native CGenFF charges and with CM1A charges. Although the experimental ΔG_{hyd} are unknown, as shown in Figure S7, substantial differences of 12–16 kcal/mol in $\Delta\Delta G_{\text{hyd}}$ were observed for the 5-membered rings when CGenFF charges were employed. Experimentally, the addition of a CH or CH₂ group into a nitrogen containing ring is unlikely to cause such a large shift (Figure S8).^{40,41} For example, a $\Delta\Delta G_{\text{hyd}}$ of ~0.5 kcal/mol is observed by increasing azetidine’s ring size up to piperidine. Large hydration free energies are observed experimentally, such as for N-propylguanidine or 4-methyl-2H-imidazole, and the $\Delta\Delta G_{\text{hyd}}$ between aromatic molecules 4-methyl-2H-imidazole and 2-methylpyrazine is not small, 4.68 kcal/mol, but this difference is still a small fraction of what is observed for the 5-membered R1 ring with CGenFF charges. In contrast, comparable $\Delta\Delta G_{\text{hyd}}$ results are observed for all JP ligands with CM1A charges, with maximum differences of ± 1.8 kcal/mol. Collectively, this analysis suggests a force field parameterization inconsistency may be behind the observed CGenFF $\Delta\Delta G_{\text{hyd}}$ and $\Delta\Delta G_{\text{bind}}$ deviations.

To understand how the CGenFF partial atomic charges for the 5-membered R1 ring might be incorrect, a detailed charge comparison was performed. Figure S9 maps out R1 atomic charges in the 5, 6, and 7-membered rings with each charge set, CGenFF and CM1A. Though many atoms show consistent charging across each ring with each charge set (Figure S9A), the carbonyl C and the cationic N atoms in the 5-membered ring show significant deviations from equivalent atoms in the 6 and 7-membered rings with CGenFF charges (Figure S9B). The cationic N is ~0.52 e more positive and the carbonyl C is ~0.71 e more negative than what is observed for the 6 and 7-

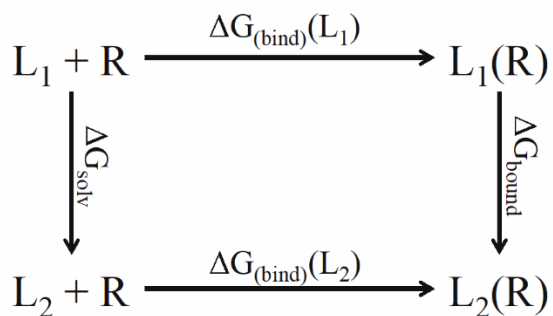
membered R1 rings. This shift likely causes the 5-membered R1 ring to have a stronger hydration free energy than the 6 and 7-membered rings. For example, if the CGenFF charges are modified by arbitrarily moving 0.52 e from the carbonyl C to the cationic N in the 5-membered R1 ring, the $\Delta\Delta G_{\text{hyd}}$ deviations are reduced from -12-16 kcal/mol to +2-6 kcal/mol (“mod-CGenFF” results in Figure S7). A more scientific optimization of partial charges in this 5-membered ring would further improve the $\Delta\Delta G_{\text{hyd}}$ agreement, but this was beyond the scope of this work. This analysis, however, does identify the original CGenFF charges for the 5-membered R1 ring as not optimal, which causes the 5-membered R1 ring containing inhibitors to be over-solvated compared to the other R1 substituents. An overly favorable ΔG_{hyd} for the 5-membered R1 ring creates a large energetic desolvation penalty, and results in too unfavorable binding affinities. Hence, we observe large $\Delta\Delta G_{\text{bind}}$ differences in the initial MS λ D calculations that employed CHARMM36+CGenFF force field parameters. In contrast, no shift is observed with CM1A charges, all $\Delta\Delta G_{\text{hyd}}$ are consistent, and excellent free energies were obtained.

Structural Analysis. Molecular dynamics (MD) simulations and free energy calculations can not only provide insights into the thermodynamics of protein-ligand binding and ligand ranking, but structural analysis of MD trajectories can help explain observed trends and guide future SBDD efforts. For example, the R2 substituent G has the best experimental potency, while A and F substituents yielded the least potent inhibitors.¹ Structurally, all R2 substituents occupy the P3 pocket beneath the 10s loop (Figure 1B). The A substituent is not long enough to penetrate deeply into the pocket, resulting in weaker interactions and activity. Conversely, while the F group is long enough to fill the P3 pocket, its ethoxy methoxy tail features many rotatable bonds.

Reorganizing these many degrees of freedom to bind likely introduces a large entropic penalty.¹ The alkyne group on G, however, has both a rigid preorganized conformation and can readily stick into the P3 pocket to interact with Ala396, Thr293, and Tyr75 residues (Figure S10A). The MS λ D trajectories also suggest that the alkyne can rotate 120° to become sandwiched between Tyr75 and the 10s backbone (Figure S10B). This conformation could be particularly attractive for further optimization of the R2 substituents. For example, if the alkyne moiety of G were replaced with an aromatic ring, strengthened π - π interactions with Tyr75 may result.

Figure Acknowledgments. Figures in this work were created with the help of PyMOL and MSMEexplorer.^{43,44}

FIGURES



$$\Delta\Delta G_{\text{bind}}(L_1 \rightarrow L_2) = \Delta G_{\text{bind}}(L_2) - \Delta G_{\text{bind}}(L_1) = \Delta G_{\text{bound}} - \Delta G_{\text{solv}}$$

Figure S1. The thermodynamic cycle for the computation of a relative free energies of binding between two ligands (L_1 and L_2) to a protein receptor (R). The vertical arms represent the alchemical transformations investigated in unbound-solvent and protein-bound states of the chemical system.

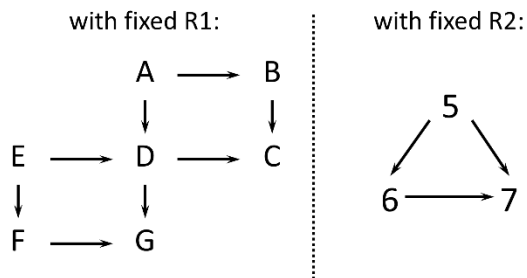


Figure S2. Two types of closed TI perturbation cycles employed in this work. A→G R2 perturbations were performed with a fixed R1 substituent; 5→7 R1 ring perturbations were performed with a fixed R2 substituent.

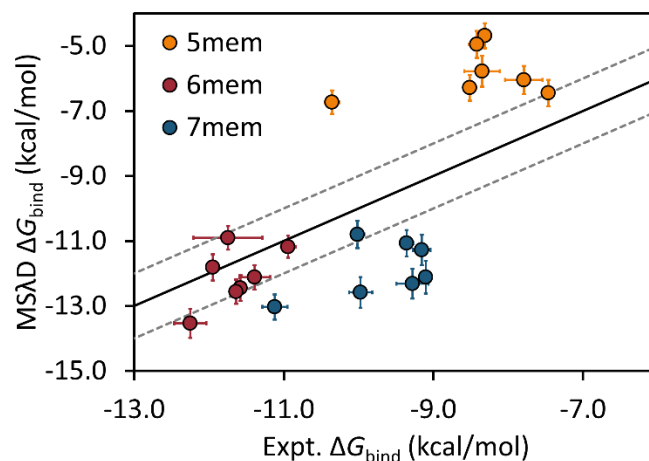


Figure S3. Correlation between MSλD computed and experimental free energies of binding (kcal/mol) for the JP ligands. These results were obtained with CHARMM36 and CGenFF force field parameters. R1 substituents are colored according to the ring size: orange (5-membered), red (6-membered), and blue (7-membered). The solid black line represents $y=x$, and grey dashed lines represent $y = x \pm 1$.

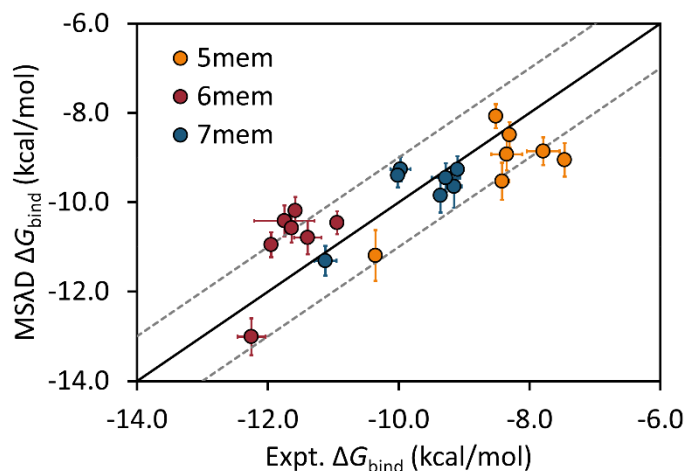


Figure S4. Correlation between MSλD computed and experimental free energies of binding (kcal/mol) for the JP ligands. These results were obtained with OPLS-AA/M and OPLS-AA/CM1A force field parameters. R1 substituents are colored according to the ring size: orange (5-membered), red (6-membered), and blue (7-membered). The solid black line represents $y=x$, and grey dashed lines represent $y = x \pm 1$.

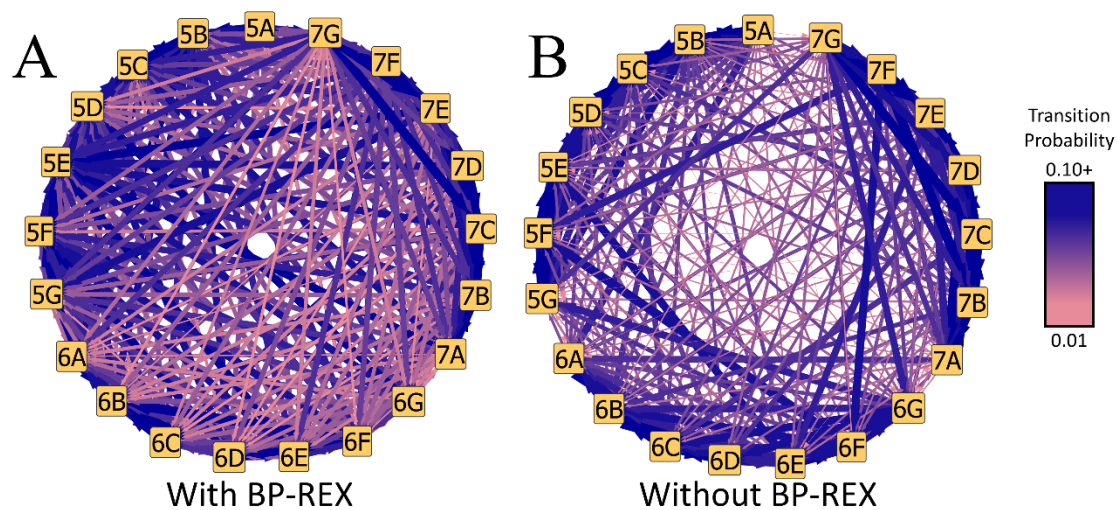


Figure S5. Transition probability pathways for alchemically perturbing between ligand end-states with MS λ D for the unbound ligand in water. (A) Significantly more transitions are observed when the BP-REX algorithm is employed. (B) Fewer transitions are observed without replica exchange. Arrow thickness and color correlate to high (blue, thick) to low (pink, thin) transition probabilities.

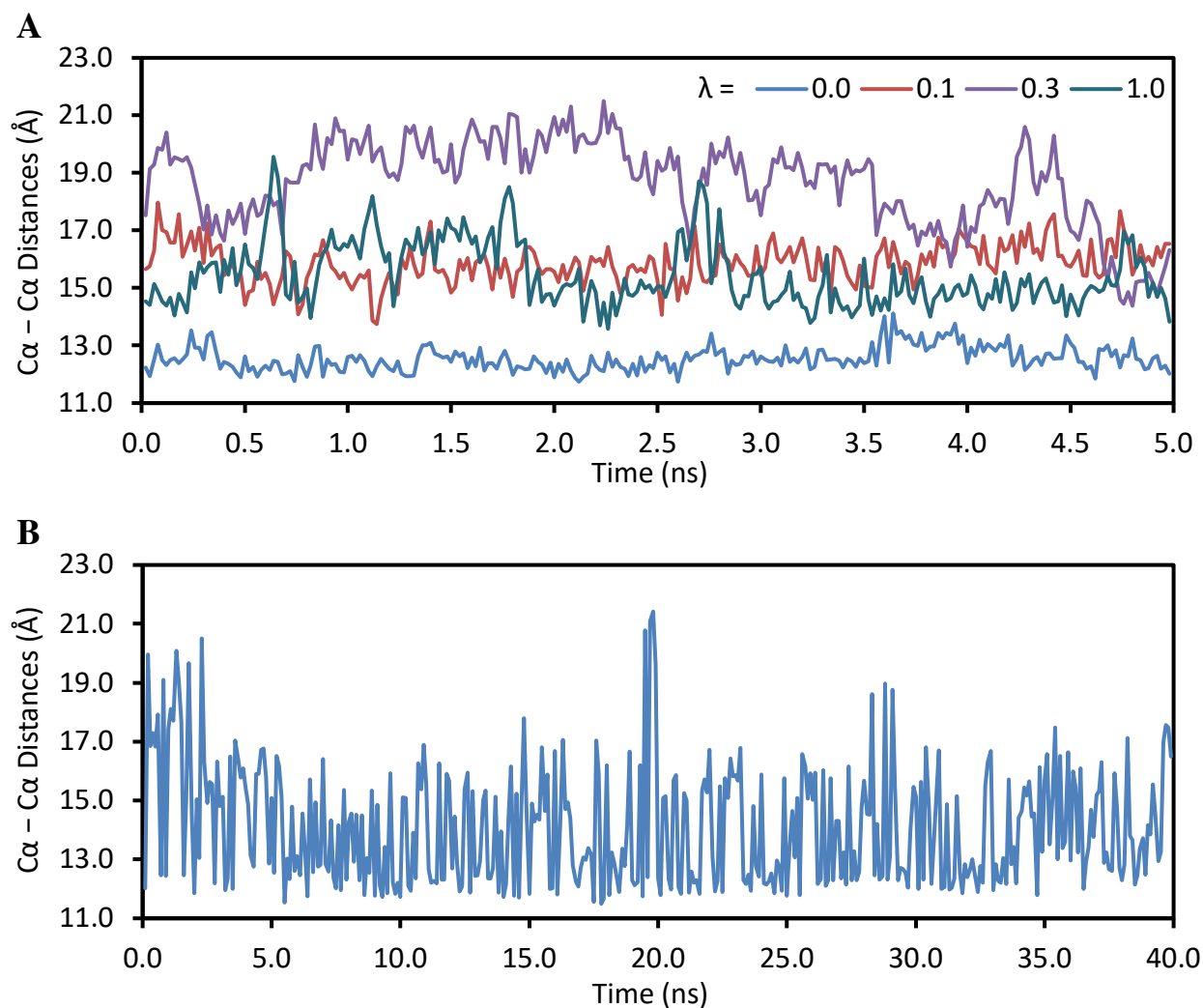


Figure S6. Representative flap loop movements quantified as Cα(Asp93)–Cα(Gln134) distances as a function of time (ns). (A) Four TI windows from the 5B→6B perturbation, including 5B ($\lambda = 0.0$), 6B ($\lambda = 1.0$), and two intermediate ($\lambda = 0.1$ and 0.3) states. The $\lambda = 0.0$ simulation is trapped in a closed conformation and the $\lambda = 0.3$ simulation is trapped in an open conformation. For each λ state, very few or no transitions are observed between open and closed flap conformations. (B) In contrast, in one 40 ns long MSλD simulation, multiple transitions between open and closed flap conformations are observed.

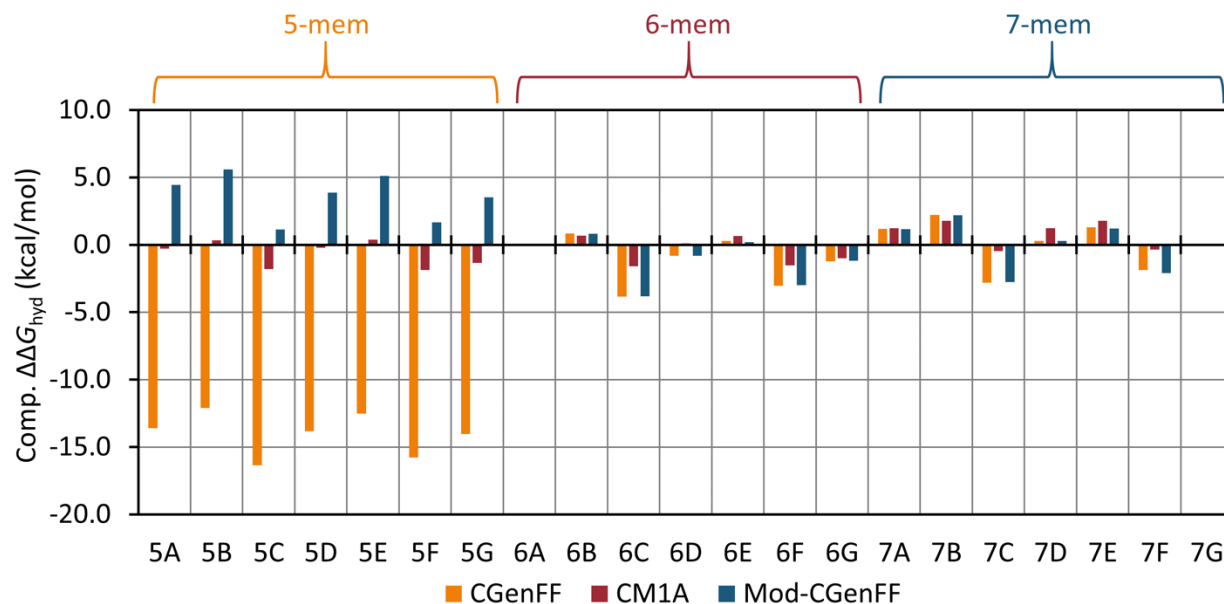


Figure S7. Computed relative free energies of hydration (kcal/mol) for all 21 JP ligands using CGenFF, CM1A, or modified-CGenFF ligand parameters. Ligand 6A is set as the reference ligand with $\Delta\Delta G_{\text{hyd}} = 0.00$.




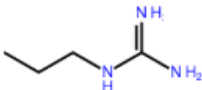
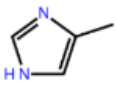
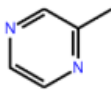
Piperidine	Pyrrolidine	Azetidine
$\Delta G_{\text{hyd}} = -5.11$	$\Delta G_{\text{hyd}} = -5.48$	$\Delta G_{\text{hyd}} = -5.56$
		
N-propyl-guanidine	4-methyl-2H-imidazole	2-methyl pyrazine
$\Delta G_{\text{hyd}} = -10.92$	$\Delta G_{\text{hyd}} = -10.25$	$\Delta G_{\text{hyd}} = -5.57$
		

Figure S8. Experimental free energies of hydration (kcal/mol) for six nitrogen containing molecules that exemplify some aspect of the 5, 6, and 7-membered R1 rings.^{40,41}

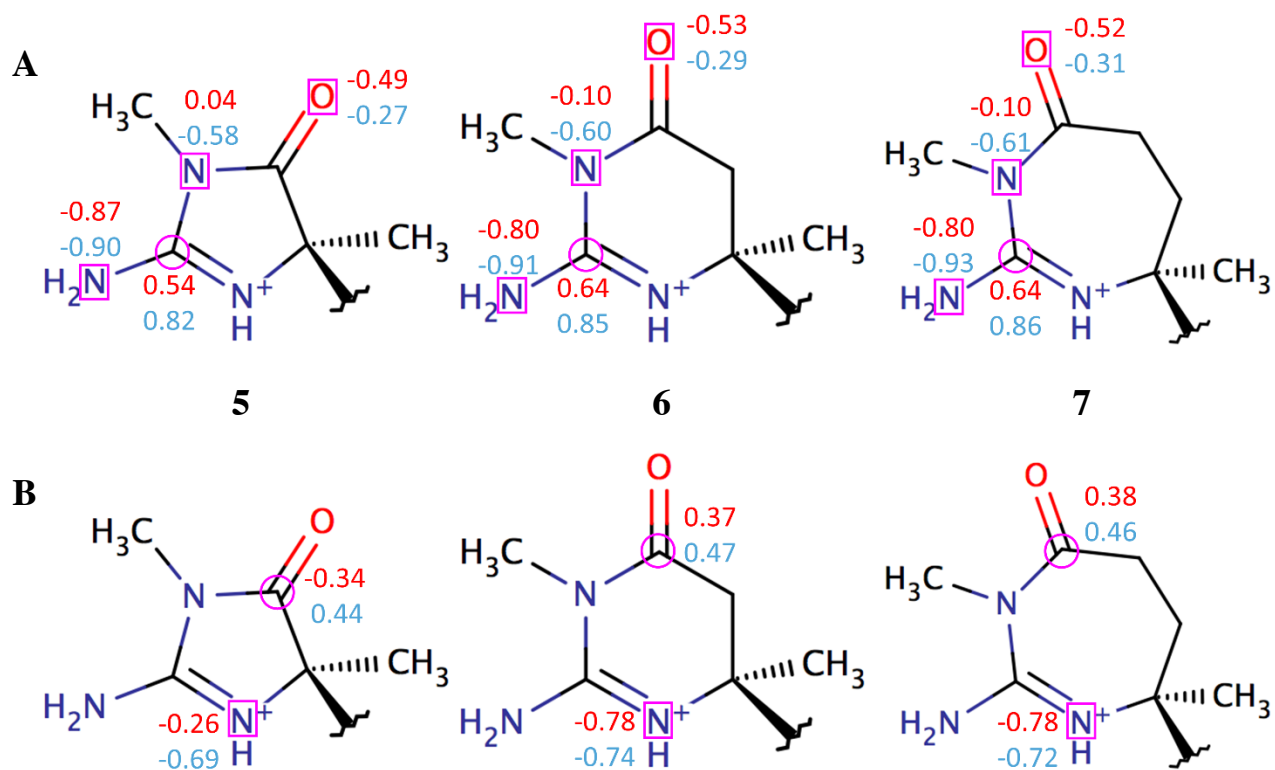


Figure S9. Partial atomic charges for R1 substituents: 5, 6, and 7-membered rings. CGenFF charges are in red, CM1A charges are in blue. Magenta boxes or circles indicate which atoms have their atomic charges labeled. (A) Atomic charges for these atoms are mostly consistent across the 5, 6, and 7-membered rings. (B) Notable CGenFF charge discrepancies are observed for 5-membered ring atoms compared to equivalent atoms in the 6 and 7-membered rings.

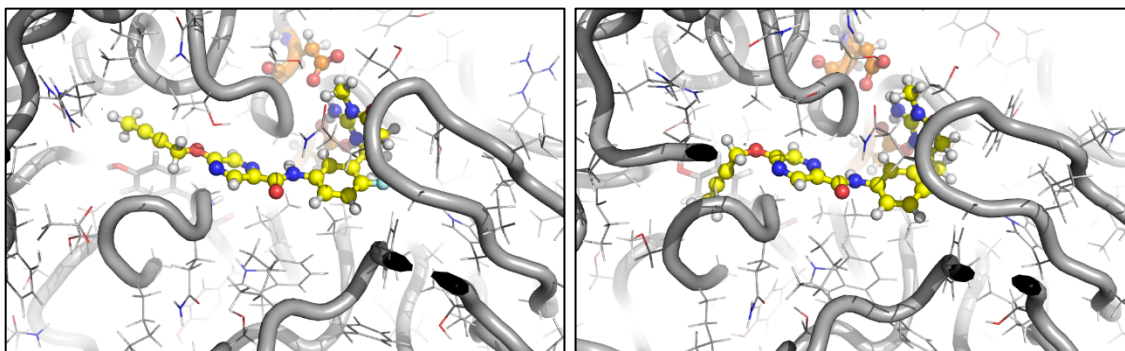


Figure S10. Structural snapshots of 6G (yellow) bound to BACE1 (grey ribbon with side chains sticks; the catalytic aspartic acids are colored orange). (A) The alkyne moiety of R2 = G sticks straight into the P3 pocket. (B) The alkyne moiety of R2 = G is rotated 120° to sit beneath the 10s loop.

TABLES

Table S1. Computed Free Energies of Binding (kcal/mol) from MS λ D and TI/MBAR Calculations

Obtained with CGenFF Ligand Parameters.

Index	Expt.^a	$\pm 1\sigma$	TI/MBAR^b	$\pm 1\sigma$	MSλD^b	$\pm 1\sigma$
5A	-7.79	0.25	-6.19	0.36	-6.05	0.44
5B	-8.51	0.09	-6.58	0.41	-6.29	0.40
5C	-8.31	0.08	-5.11	0.42	-4.69	0.38
5D	-8.35	0.24	-6.40	0.57	-5.78	0.48
5E	-8.42	0.10	-4.65	0.84	-4.95	0.42
5F	-7.46	0.03	-5.88	0.78	-6.45	0.41
5G	-10.35	0.10	-7.07	0.73	-6.73	0.36
6A	-10.94	0.10	-11.58	0.00	-11.18	0.34
6B	-11.58	0.05	-12.49	0.07	-12.45	0.39
6C	-11.75	0.46	-11.01	0.22	-10.90	0.37
6D	-11.39	0.21	-12.55	0.29	-12.12	0.38
6E	-11.95	0.03	-11.05	0.36	-11.81	0.40
6F	-11.64	0.09	-11.98	0.48	-12.56	0.37
6G	-12.25	0.21	-14.10	0.61	-13.53	0.45
7A	-9.15	0.12	-11.25	0.19	-11.28	0.47
7B	-9.97	0.15	-12.10	0.36	-12.59	0.47
7C	-10.01	0.09	-10.68	0.45	-10.80	0.42
7D	-9.28	0.22	-11.86	0.49	-12.32	0.46
7E	-9.36	0.09	-10.47	0.51	-11.07	0.40
7F	-9.10	0.02	-12.14	0.86	-12.11	0.50
7G	-11.12	0.17	-13.54	0.79	-13.03	0.38
MUE to Expt.			1.80		1.83	
Pearson R to Expt.			0.77		0.75	
MUE to TI/MBAR					0.37	
Pearson R to TI/MBAR					0.99	

^a Ref 1. ^b Protein force field is CHARMM36.

Table S2. Welch's Two Sample Unequal Variance t -test Comparing ΔG_{bind} Results Between MS λ D and TI/MBAR.^{45,46} All 21 ligands are statistically equivalent at the 99% confidence interval. Only 1 ligand is statistically different between methods at the 95% confidence interval.

Ligand	t^a	ν^b	$p(t)^c$	Reject Null Hypothesis?	
				$p < 0.025^d$	$p < 0.005^e$
5A	0.489	5.131	0.3305	FALSE	FALSE
5B	0.977	4.244	0.2210	FALSE	FALSE
5C	1.418	3.968	0.1353	FALSE	FALSE
5D	1.578	3.727	0.1114	FALSE	FALSE
5E	-0.577	2.616	0.2926	FALSE	FALSE
5F	-1.172	2.681	0.1700	FALSE	FALSE
5G	0.754	2.599	0.2544	FALSE	FALSE
6A	2.631	4.000	0.0304	FALSE	FALSE
6B	0.223	4.416	0.3658	FALSE	FALSE
6C	0.527	5.962	0.3265	FALSE	FALSE
6D	1.802	5.386	0.0844	FALSE	FALSE
6E	-2.771	4.756	0.0238	TRUE	FALSE
6F	-1.797	3.460	0.0854	FALSE	FALSE
6G	1.405	3.341	0.1352	FALSE	FALSE
7A	-0.127	5.639	0.3782	FALSE	FALSE
7B	-1.658	5.373	0.1021	FALSE	FALSE
7C	-0.374	4.080	0.3446	FALSE	FALSE
7D	-1.315	4.101	0.1531	FALSE	FALSE
7E	-1.742	3.510	0.0913	FALSE	FALSE
7F	0.055	2.835	0.3651	FALSE	FALSE
7G	1.048	2.569	0.1922	FALSE	FALSE

$$^a t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}; ^b \nu = \frac{\left(\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}\right)^2}{\frac{\sigma_1^4}{N_1^2 \nu_1} + \frac{\sigma_2^4}{N_2^2 \nu_2}}; \nu_i = N_i - 1; ^c p(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

^d Corresponds to a 95% confidence level; ^e Corresponds to a 99% confidence level.

Table S3. Computed Free Energies of Binding (kcal/mol) from MS λ D Obtained with CGenFF/CM1A and OPLS-AA/CM1A Ligand Parameters.

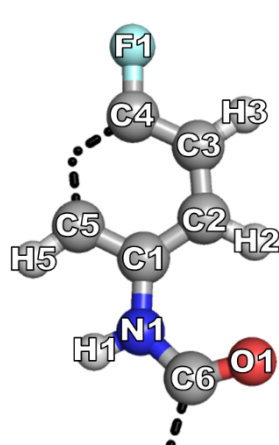
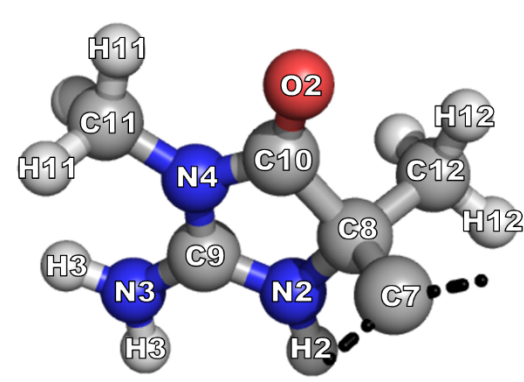
Index	Expt. ^a	$\pm 1\sigma$	MS λ D		MS λ D	
			CGenFF/CM1A ^b	$\pm 1\sigma$	OPLS-AA/CM1A ^c	$\pm 1\sigma$
5A	-7.79	0.25	-7.75	0.42	-8.86	0.31
5B	-8.51	0.09	-8.02	0.32	-8.08	0.27
5C	-8.31	0.08	-8.04	0.36	-8.49	0.29
5D	-8.35	0.24	-8.61	0.34	-8.93	0.36
5E	-8.42	0.10	-8.60	0.37	-9.53	0.42
5F	-7.46	0.03	-9.04	0.25	-9.05	0.37
5G	-10.35	0.10	-10.35	0.35	-11.19	0.57
6A	-10.94	0.10	-10.59	0.23	-10.46	0.26
6B	-11.58	0.05	-11.10	0.26	-10.19	0.30
6C	-11.75	0.46	-11.30	0.25	-10.42	0.35
6D	-11.39	0.21	-12.11	0.32	-10.79	0.38
6E	-11.95	0.03	-11.86	0.28	-10.95	0.28
6F	-11.64	0.09	-12.09	0.29	-10.57	0.33
6G	-12.25	0.21	-13.83	0.38	-13.01	0.41
7A	-9.15	0.12	-8.91	0.30	-9.65	0.48
7B	-9.97	0.15	-9.04	0.26	-9.26	0.28
7C	-10.01	0.09	-9.57	0.25	-9.40	0.27
7D	-9.28	0.22	-9.01	0.27	-9.45	0.32
7E	-9.36	0.09	-8.79	0.35	-9.85	0.38
7F	-9.10	0.02	-9.23	0.36	-9.27	0.30
7G	-11.12	0.17	-10.88	0.32	-11.31	0.33
MUE to Expt.			0.47		0.73	
Pearson R to Expt.			0.92		0.83	

^a Ref 1. ^b Protein force field is CHARMM36. ^c Protein force field is OPLS-AA/M.

Table S4. Root Mean Square Charge Analysis (e) and Unsigned Relative Free Energies of Hydration (kcal/mol) Between Original and Adjusted CGenFF Partial Atomic Charges.

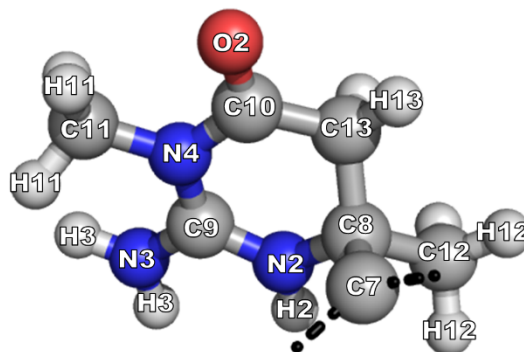
Ligand	RMSQ	Unsigned $\Delta\Delta G_{\text{hyd}}$
5A	0.00867	0.28
5B	0.00363	0.05
5C	0.00357	0.04
5D	0.00241	0.00
5E	0.00234	0.03
5F	0.00224	0.02
5G	0.00229	0.01
6A	0.00816	0.24
6B	0.00296	0.17
6C	0.00291	0.23
6D	0.00147	0.09
6E	0.00142	0.07
6F	0.00137	0.10
6G	0.00140	0.08
7A	0.00790	0.29
7B	0.00287	0.24
7C	0.00282	0.26
7D	0.00143	0.06
7E	0.00138	0.04
7G	0.00133	0.05
7G	0.00136	0.09
Average	0.00304	0.12

LIGAND PARTIAL ATOMIC CHARGES

Atom Name	CGenFF	CM1A	Structure ^a
Core			
C1	0.15931	0.19977	
C2	-0.17369	-0.08533	
H2	0.11531	0.19277	
C3	-0.12769	-0.15943	
H3	0.18631	0.17397	
C4	0.12531	0.11907	
C5	-0.17669	-0.14793	
H5	0.11531	0.11077	
C6	0.44030	0.57266	
F1	-0.22369	-0.02543	
N1	-0.49470	-0.78533	
H1	0.32531	0.43137	
O1	-0.44170	-0.37063	
R1 = 5-membered ring			
C7	0.14485	-0.17438	
C8	0.66584	0.19973	
C9	0.54484	0.81943	
C10	-0.34316	0.44963	
C11	0.14284	-0.02087	
H11	0.08984	0.13042	
C12	-0.27316	-0.22427	
H12	0.08984	0.12463	
N2	-0.25516	-0.69397	
H2	0.37985	0.46893	
N3	-0.86716	-0.89997	
H3	0.45984	0.47662	
N4	0.03984	-0.57877	
O2	-0.49016	-0.26548	

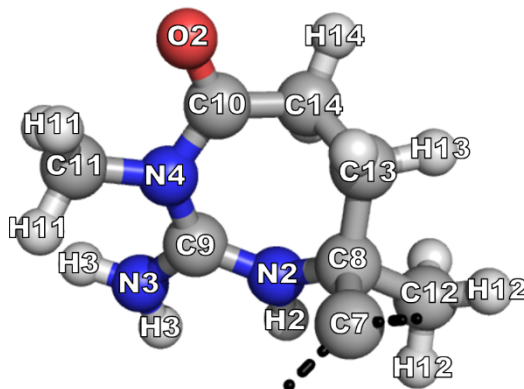
R1 = 6-membered ring

C7	0.18209	-0.18396
C8	0.4511	0.24133
C9	0.64209	0.85203
C10	0.37209	0.46673
C11	0.07509	-0.01557
H11	0.09009	0.12643
C12	-0.2639	-0.23787
H12	0.09009	0.11304
C13	-0.17091	-0.20747
H13	0.09009	0.16173
N2	-0.78891	-0.74297
H2	0.44009	0.45543
N3	-0.80091	-0.91397
H3	0.46009	0.47283
N4	-0.09991	-0.60537
O2	-0.53091	-0.29747



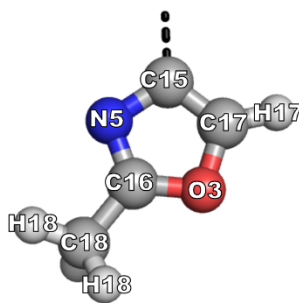
R1 = 7-membered ring

C7	0.18008	-0.17525
C8	0.43508	0.23694
C9	0.64208	0.86225
C10	0.37608	0.45955
C11	0.07508	-0.01805
H11	0.09008	0.12545
C12	-0.26492	-0.24286
H12	0.09008	0.10835
C13	-0.15692	-0.18086
H13	0.09008	0.13315
C14	-0.18892	-0.19306
H14	0.09008	0.14085
N2	-0.78592	-0.72546
H2	0.44008	0.43855
N3	-0.80092	-0.93285
H3	0.46008	0.47185
N4	-0.09992	-0.61285
O2	-0.52392	-0.31075



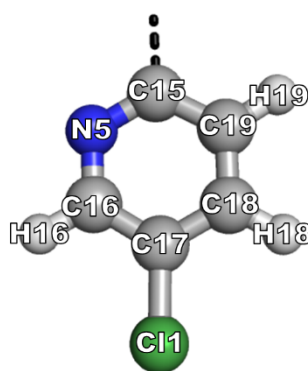
R2 = A

C15	0.2458	-0.1262
C16	0.2418	0.1914
C17	0.1298	0.0436
H17	0.1228	0.2283
C18	-0.1292	-0.1593
H18	0.0928	0.1196
N5	-0.6202	-0.3828
O3	-0.2462	-0.1785



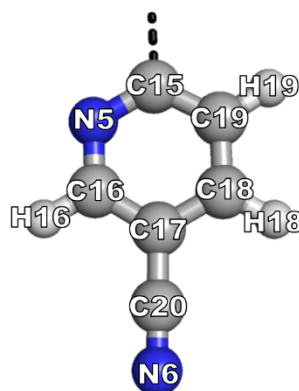
R2 = B

C15	0.2297	0.00162
C16	0.2047	0.03012
H16	0.1057	0.17902
C17	0.0447	-0.06048
C18	-0.0963	-0.07988
H18	0.1627	0.17162
C19	-0.0723	-0.09738
H19	0.1127	0.18402
N5	-0.5203	-0.34898
Cl1	-0.1483	-0.00438



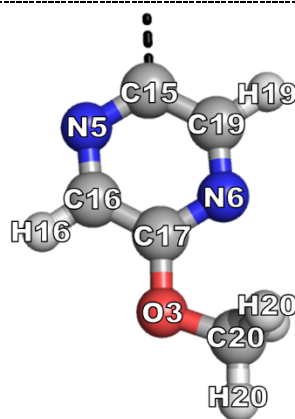
R2 = C

C15	0.2329	0.02853
C16	0.14591	0.06553
H16	0.11991	0.17822
C17	0.09791	-0.01967
C18	-0.08909	-0.04597
H18	0.11291	0.17013
C19	-0.07609	-0.11117
H19	0.11291	0.18742
C20	0.35691	0.13883
N5	-0.52609	-0.36778
N6	-0.46509	-0.24877



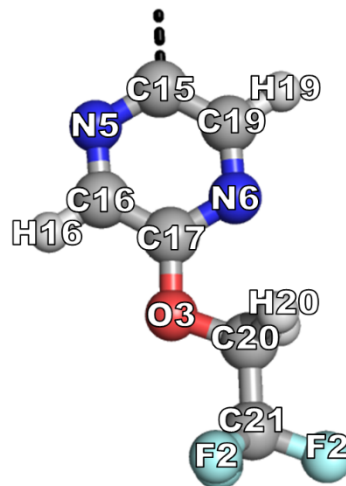
R2 = D

C15	0.20839	-0.1377
C16	0.11839	-0.06291
H16	0.11738	0.19049
C17	0.53039	0.26399
C19	0.19739	0.1187
H19	0.11738	0.2097
C20	-0.09962	-0.03201
H20	0.09038	0.10509
N5	-0.47662	-0.22731
N6	-0.57061	-0.41331
O3	-0.39061	-0.24961



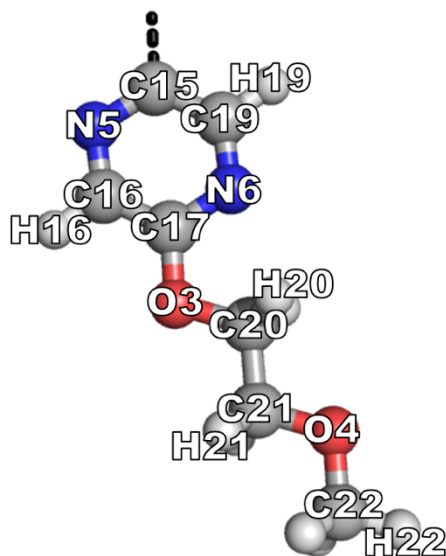
R2 = E

C15	0.20832	-0.12359
C16	0.11831	-0.04879
H16	0.11731	0.19791
C17	0.53032	0.25611
C19	0.19732	0.11481
H19	0.11731	0.21471
C20	0.05631	-0.04728
H20	0.09031	0.14601
C21	0.35931	0.40712
N5	-0.47669	-0.22309
N6	-0.57068	-0.40639
O3	-0.39569	-0.23949
F2	-0.13969	-0.13958



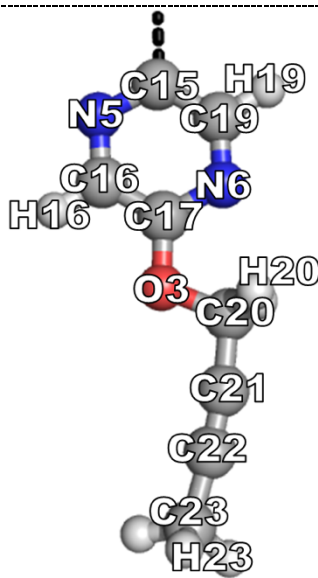
R2 = F

C15	0.20825	-0.14031
C16	0.11825	-0.06641
H16	0.11725	0.18949
C17	0.53025	0.27219
C19	0.19725	0.12099
H19	0.11725	0.20989
C20	-0.01075	0.02539
H20	0.09025	0.11979
C21	-0.01075	-0.00521
H21	0.09025	0.08889
C22	-0.09975	-0.04551
H22	0.09025	0.08239
N5	-0.47675	-0.22521
N6	-0.57075	-0.41411
O3	-0.38975	-0.26311
O4	-0.33875	-0.34731



R2 = G

C15	0.20828	-0.143
C16	0.11828	-0.0655
H16	0.11728	0.1908
C17	0.53028	0.27619
C19	0.19728	0.1274
H19	0.11728	0.2087
C20	0.06728	0.15159
H20	0.09028	0.12389
C21	-0.08472	-0.21921
C22	-0.07872	-0.07591
C23	-0.19072	-0.12181
H23	0.09027	0.10339
N5	-0.47672	-0.2245
N6	-0.57072	-0.4366
O3	-0.38272	-0.25081



^a Black dashed lines represent bonds between different components of the ligand.

REFERENCES

- (1) Keränen, H.; Pérez-Benito, L.; Ciordia, M.; Delgado, F.; Steinbrecher, T. B.; Oehrich, D.; van Vlijmen, H. W. T.; Trabanco, A. A.; Tresadern, G. Acylguanidine Beta Secretase 1 Inhibitors: A Combined Experimental and Free Energy Perturbation Study. *J. Chem. Theory Comput.* **2017**, *13*, 1439-1453.
- (2) Banner, D. W.; Gsell, B.; Benz, J.; Bertschinger, J.; Burger, D.; Brack, S.; Cuppuleri, S.; Debulpaep, M.; Gast, A.; Grabulovski, D.; Hennig, M.; Hilpert, H.; Huber, W.; Kuglstatter, A.; Kuszniir, E.; Laeremans, T.; Matile, H.; Miscenic, C.; Rufer, A.; Schlatter, D.; Steyeart, J.; Stihle, M.; Thoma, R.; Weber, M.; Ruf, A. Mapping the Conformational Space Accessible to BACE2 Using Surface Mutants and Co-Crystals with Fab-Fragments, Fynomers, and Xaperones. *Acta Crystallogr. Sect.D* **2013**, *69*, 1124-1137.
- (3) Iserloh, U.; Wu, Y.; Cumming, J. N.; Pan, J.; Wang, L. Y.; Stamford, A. W.; Kennedy, M. E.; Kuvelkar, R.; Chen, X.; Parker, E. M.; Strickland, C.; Voigt, J. Potent pyrrolidine- and piperidine-based BACE-1 inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 414-417.
- (4) Im, W.; Lee, M. S.; Brooks, C. L., III Generalized Born Model with a Simple Smoothing Function. *J. Comput. Chem.* **2003**, *24*, 1691-1702.
- (5) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W.B., III; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35*, W375-83.

- (6) Chen, V.B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12-21.
- (7) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem. Theory Comput.* **2011**, *7*, 525–537.
- (8) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859-1865.
- (9) Feig, M.; Karanicolas, J.; Brooks, C. L., III MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J. Mol. Graph. Model.* **2004**, *22*, 377-395.
- (10) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D. Jr; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545-1614.
- (11) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, B. D.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187-217.

- (12) Hynninen, A. P.; Crowley, M. F. New Faster CHARMM Molecular Dynamics Engine. *J. Comput. Chem.* **2014**, *35*, 406-413.
- (13) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511-519.
- (14) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Review A* **1985**, *31*, 1695-1697.
- (15) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613-4621.
- (16) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
- (17) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.
- (18) Best, R. B.; Mittal, J.; Feig, M.; MacKerell, A. D. Jr. Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of α -Helix And β -Hairpin Formation. *Biophys. J.* **2012**, *103*, 1045-1051.
- (19) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ , and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257-3273.

- (20) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499-3509.
- (21) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. Jr. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* 2009, *31*, 671-690.
- (22) Vanommeslaeghe, K.; MacKerell, A. D. Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* 2012, *52*, 3144-3154.
- (23) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155-3168.
- (24) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: An automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017**, *45*, W331-W336; DOI:10.1093/nar/gkx312.
- (25) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6665-6670.
- (26) Udier-Blagovic, M.; Morales de Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. Accuracy of Free Energies of Hydration from CM1 and CM3 Atomic Charges. *J. Comput. Chem.* **2004**, *25*, 1322-1332.

- (27) Knight, J. L.; Brooks, C. L., III λ Dynamics Free Energy Simulation Methods. *J. Comput. Chem.* **2009**, *30*, 1692-1700.
- (28) Knight, J. L.; Brooks, C. L., III Multisite λ Dynamics for Simulated Structure-Activity Relationship Studies. *J. Chem. Theory Comput.* **2011**, *7*, 2728-2739.
- (29) Knight, J. L.; Brooks, C. L., III Applying Efficient Implicit Constraints in Alchemical Free Energy Simulations. *J. Comput. Chem.* **2011**, *32*, 3423-3432.
- (30) Armacost, K. A.; Goh, G. B.; Brooks, C. L., III Biasing Potential Replica Exchange Multisite λ Dynamics for Efficient Free Energy Calculations. *J. Chem. Theory Comput.* **2015**, *11*, 1267-1277.
- (31) Hayes, R. L.; Armacost, K. A.; Vilseck, J. Z.; Brooks, C. L., III Adaptive Landscape Flattening Accelerates Sampling of Alchemical Space in Multisite λ Dynamics. *J. Phys. Chem. B* **2017**, *121*, 3626-3635.
- (32) Vilseck, J. Z.; Armacost, K. A.; Hayes, R. L.; Goh, G. B.; Brooks, C. L., III Predicting Binding Free Energies in a Large Combinatorial Chemical Space Using Multisite λ Dynamics. *J. Phys. Chem. Lett.* **2018**, *9*, 3328-3332.
- (33) Hayes, R. L.; Vilseck, J. Z.; Brooks, C. L., III Approaching Protein Design with Multisite λ Dynamics: Accurate and Scalable Mutational Folding Free Energies in T4 Lysozyme. *Protein Sci.* **2018**, *27*, 1910-1922.

- (34) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (35) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300-313.
- (36) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
- (37) Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L. Lead Optimization Mapper: Automating Free Energy Calculations for Lead Optimization. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 755-770.
- (38) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. *J. Chem. Theory Comput.* **2013**, *9*, 1282–1293.
- (39) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695-2703.

- (40) Matos, G. D. R.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *J. Chem. Eng. Data* **2017**, *62*, 1559-1569.
- (41) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1509–1519.
- (42) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121*, 3864-3870.
- (43) The PyMOL Molecular Graphics System, Version 1.8, Schrodinger LLC, New York, New York, United States.
- (44) Hernandez, C. X.; Harrigan, M. P.; Sultan, M. M.; Pande, V. S. MSMExplore: Data Visualizations for Biomolecular Dynamics. *Journal of Open Source Software* **2017**, *2*, 188.
- (45) Welch, B. L. The generalization of student's problems when several different population variances are involved. *Biometrika* **1947**, *34*, 28-35.
- (46) Satterthwaite, F. E. An approximate distribution of estimates of variance components. *Biometrics* **1946**, *2*, 110-114.

FULL MANUSCRIPT REFERENCES

- (7) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695-2703.
- (29) Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D. Jr; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545-1614.