

Comparing the performance of different AMBER protein forcefields, partial charge assignments,
and water models for absolute binding free energy calculations

David J. Huggins^{a,b}

^aTri-Institutional Therapeutics Discovery Institute, New York, NY 10021, USA

^bDepartment of Physiology and Biophysics, Weill Cornell Medical College of Cornell
University, New York, NY 10065, USA

Corresponding Author: David J. Huggins

Email: dhuggins@tritdi.org

Updated: 10 Jan 2022

Abstract

Identifying chemical starting points is a vital first step in small molecule drug discovery and can take significant time and money. For this reason, computational approaches to virtual screening are of great interest as they can lower cost and shorten timeframes. However, simple approaches such as molecular docking and pharmacophore screening are of limited accuracy and provide a low probability of success. Alchemical binding free energies represent a promising approach for virtual screening as they naturally incorporate the key effects of water molecules, protein flexibility, and binding entropy. However, the calculations are technically very challenging, with performance depending on the specific forcefield used. For this reason, it is important that the community has access to benchmark test sets to assess prediction accuracy. In this paper we present an approach to alchemical binding free energies using OpenMM. We identify effective simulation parameters using an existing BRD4(1) test set and present two new benchmark sets (cMET and PDE2A) that can be used in the community for validation purposes. Our findings also highlight the effectiveness of some AMBER forcefields, in particular AMBER ff15ipq.

1. Introduction

Absolute binding free energy calculations are increasingly being exploited for virtual screening in drug discovery.¹ The methods use molecular dynamics (MD) simulations with ligands being decoupled from their interactions in solution and coupled to their interactions in complex. A number of named methods and approaches have been published for calculation of absolute binding free energies.²⁻¹² Named packages include BEDAM,¹³ BFEE,^{14, 15} CHARMM-GUI,¹⁶ XFEP,¹⁷ and GA-FEP.¹⁸ The performance of absolute binding free energy calculations depends on the protein and ligand forcefields used and, to date, there is a dearth of work on identifying the most effective protein and ligand forcefields. To enable effective development and application of absolute binding free energy calculations in virtual screening it is important that well curated benchmark sets are available. Community challenges such as SAMPL¹⁹ have focused on small supramolecular receptors such as octa-acid, cyclodextrin, and cucurbiturils such as CB7.²⁰⁻²² Simple proteins such as the T4 Lysozyme mutants L99A and L99A/M102Q²³⁻²⁶ have been suggested but do not effectively mirror the desired use case. More recently, Mobley et al identified BRD4(1) as a useful and realistic test case.²⁷ Recent publications have applied absolute binding free energies to a number of test cases developed using congeneric series for relative binding free energies.^{6, 17} However, additional test cases with diverse ligands that are appropriate to drug discovery are needed to assess protein and ligand forcefields for virtual screening effectively.

In this study we assess the accuracy of three AMBER protein forcefields (ff14SB,²⁸ ff15ipq,²⁹ AMBER-FB15³⁰), five water models (SPC/E,³¹ TIP3P,³² TIP3P-FB,³³ TIP4P-Ewald,³⁴ and TIP4P-FB³³), and two partial charge models for the ligand (AM1-BCC³⁵ and RESP³⁶) for absolute binding free energy calculations. All are available in the widely used OpenMM³⁷⁻⁴¹ software package such

that the results provide a useful guide to the community. To achieve this, we have curated two new benchmark sets in addition to BRD4(1) and have made these available to the community.

2. Methods

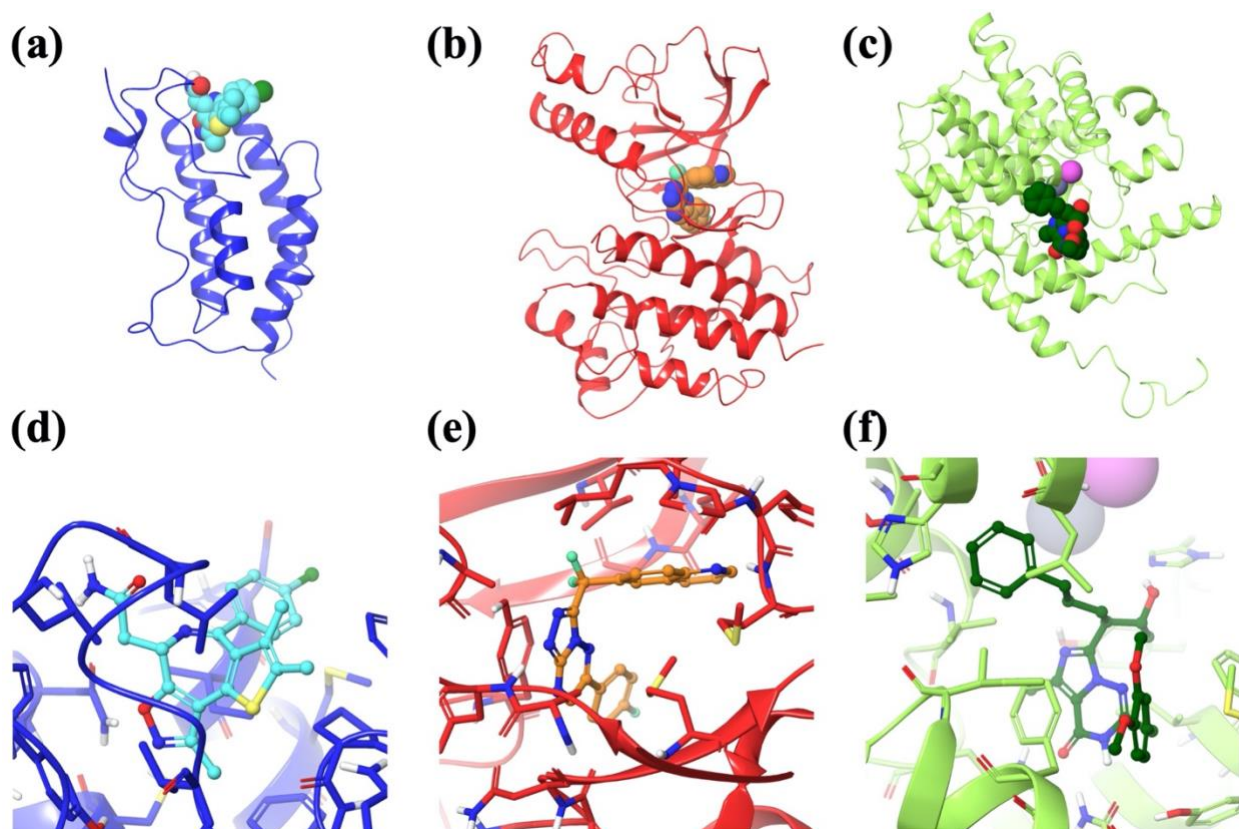
The methods section is comprised of the following subsections. First, subsection 2.1 describes how the test sets were assembled and details the proteins and ligands used. Then, subsection 2.2 describes the process of protein and ligand preparation. Subsection 2.3 lists the different protein, ligand, and water molecule forcefields that were tested. Subsection 2.4 describes the system setup, simulation details, and alchemical protocols used for the absolute binding free energy calculations. Finally, subsection 2.5 specifies the simulation details for the extended molecular dynamics simulations performed on each protein test case.

2.1 Test Set Selection

Validation of absolute binding free energy calculations requires thorough validation of overall accuracy on a dataset of diverse protein-ligand complexes. We used an existing benchmark set from BRD4(1)²⁷ and curated two additional benchmark sets from cMET kinase and PDE2A. These additional benchmark sets represent protein families (kinases and esterases) that have different folds and different binding sites from the bromodomain BRD4(1) and each other. They are also well studied proteins with a large number of crystal structures and published inhibitors. In addition, PDE2A has two metal ions and a number of interfacial water molecules in the binding site. This provides challenges that are not present in BRD4(1) and cMET. The structures of the three protein test cases and their binding sites are shown in Figure 1.

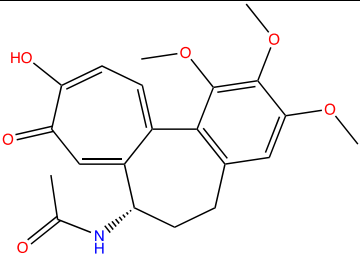
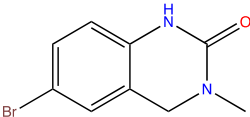
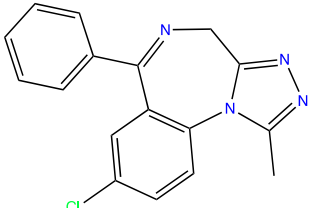
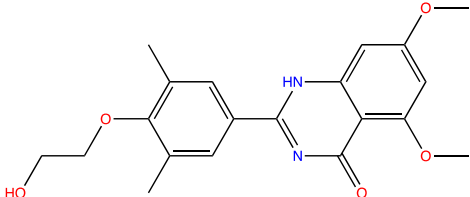
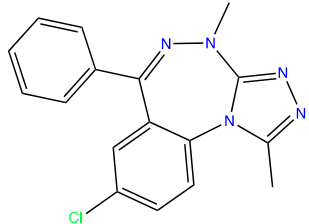
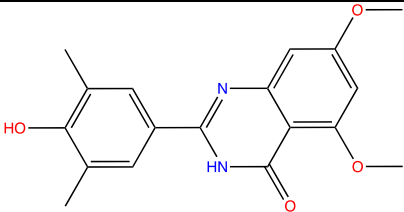
Figure 1 – Crystal structures of the three test systems and their binding sites. Representative holo crystal structures of (a) BRD4(1) from PDBID 4LRG, (b) cMET from PDBID 5EOB, and (c) PDE2A from PDBID 5U7D. Proteins are shown as blue, red, and green cartoon ribbons with

ligands in space filling with cyan, orange, and dark green carbons for BRD4(1), cMET, and PDE2A respectively. The ligands and their binding sites are also shown for (d) BRD4(1), (e) cMET, and (f) PDE2A. Proteins are shown as blue, red, and green cartoon ribbons, ligands are shown in atom-colored balls and sticks with cyan, orange, and dark green carbons, and binding site residues plus ions are shown in atom-colored balls and sticks with blue, red, and green carbons for BRD4(1), cMET, and PDE2A respectively.



For the ligands, we selected compounds from published works with differing scaffolds across a range of affinities.

Table 1 – BRD4(1) test set. The ligand IDs used in this work, the 2D structures of the ligands, the PDBIDs, the experimental IC₅₀/K_d values, the corresponding (approximate) binding free energies, and the ligand IDs in the original literature references.

Ligand ID	Structure	PDBID	IC ₅₀ (nM)	~ΔG _{bind} (kcal/mol)	Lit. ID
B1		4LYS	46,000	-5.95	XD1 ⁴²
B2		4HBV	23,000	-6.36	8 ⁴³
B3		3U5J	2,460	-7.40	Alprazolam ⁴⁴
B4		4MR4	1,142	-7.84	RVX-208 ⁴⁵
B5		3U5L	640	-8.16	BzT-7 ⁴⁴
B6		4MR3	153	-8.99	RVX-OH ⁴⁵

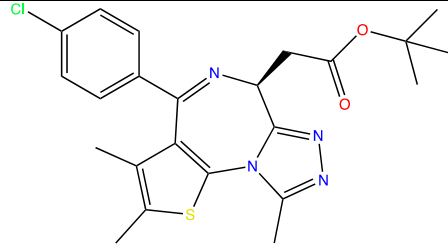
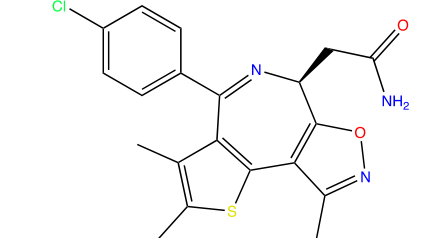
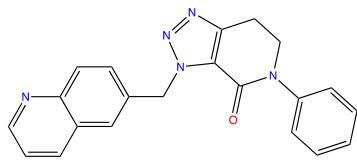
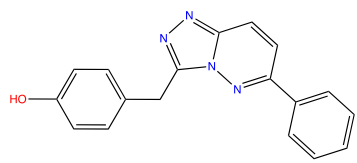
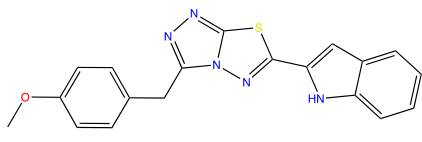
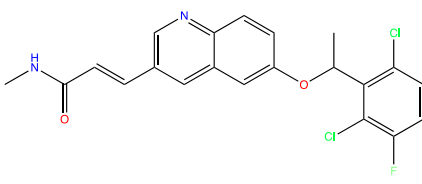
B7		3MXF	49	-9.64	(+)-JQ1 ⁴⁶
B8		4LRG	26	-10.41	3 ⁴⁷

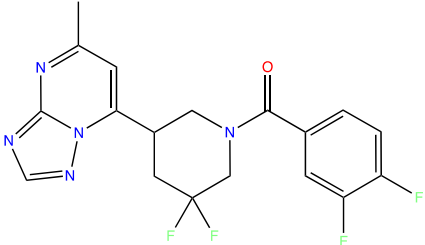
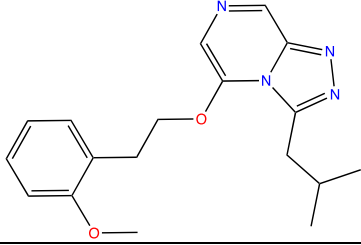
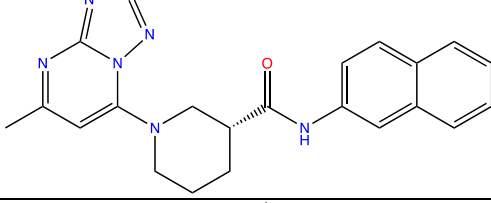
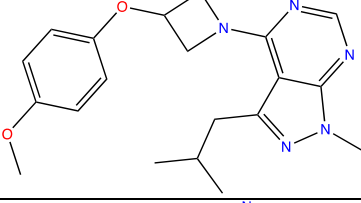
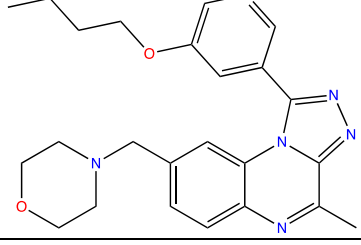
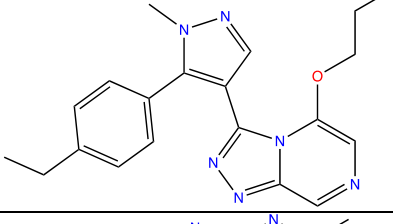
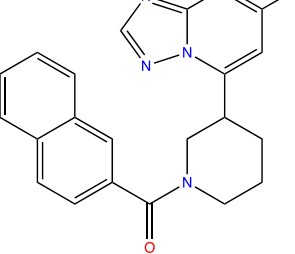
Table 2 – cMET test set. The ligand IDs used in this work, the 2D structures of the ligands, the PDBIDs, the experimental IC₅₀/K_d values, the corresponding (approximate) binding free energies, and the ligand IDs in the original literature references.

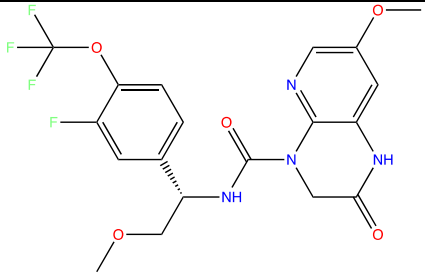
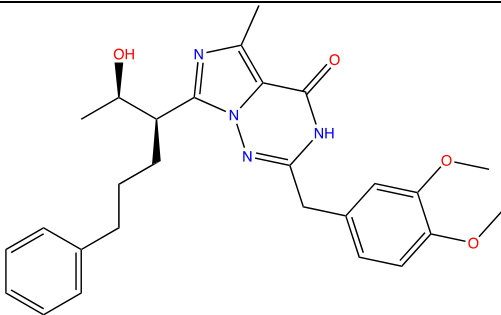
Ligand ID	Structure	PDBID	IC₅₀ (nM)	~ΔG_{bind} (kcal/mol)	Lit. ID
M1		4DEH	612	-8.47	3 ⁴⁸
M2		3CCN	120	-9.44	3a ⁴⁹
M3		5YA5	79	-9.68	HL-11f ⁵⁰
M4		3A4P	28	-10.30	4 ⁵¹

M5		3CD8	9	-10.97	4 ⁴⁹
M6		6SDE	3	-11.62	28 ⁵²
M7		4XYF	1	-12.27	(S)-21f ⁵³
M8		4DEI	0.6	-12.57	3 ⁵⁴
M9		5EOB	0.24	-13.12	23 ⁵⁵

Table 3 – PDE2A test set. The ligand IDs used in this work, the 2D structures of the ligands, the PDBIDs, the experimental IC₅₀/K_d values, the corresponding (approximate) binding free energies, and the ligand IDs in the original literature references.

Ligand ID	Structure	PDBID	IC₅₀ (nM)	~ΔG_{bind} (kcal/mol)	Lit. ID
P1		4C1I	2,460	-7.65	1(EHNA) ⁵⁶

P2		5TZW	987	-8.19	31 ⁵⁷
P3		5U7J	389.6	-8.74	3 ⁵⁸
P4		5TZ3	140	-9.35	7 ⁵⁷
P5		5U7I	66.6	-9.79	4 ⁵⁸
P6		4D08	10.1	-10.90	12 ⁵⁹
P7		5U7K	7.1	-11.11	39 ⁵⁸
P8		5TZA	0.8	-12.40	19 ⁵⁷

P9		5VP0	0.61	-12.57	36 ⁶⁰
P10		5U7D	0.03	-14.35	8 ⁶¹

Protein and ligand structures are available on Github (<https://github.com/djhuggins/Holoware-TestCases>)

2.2 Protein Preparation

The protein structure for BRD4(1) was taken from the benchmark study of Mobley et al (https://github.com/MobleyLab/benchmarksets/tree/master/input_files/BRD4). This structure was generated from PDBID 4LYI.⁴² Protein structures for cMET and PDE2A were downloaded from the protein databank.⁶² We used the protein structure PDBID 4XYF for cMET⁵³ and PDBID 4HTZ (chain D) for PDE2A.⁶³ In all cases, selenomethionines were changed to methionines and missing side-chains were added using Schrödinger's Preparation Wizard, which was also used to evaluate the orientations of the asparagine, glutamine, and histidine residues, as well as the protonation state of all ionizable residues. Heteroatomic species such as buffer solvents were removed. For BRD4(1) and cMET, all water molecules were deleted. For PDE2A, two protein structures were

generated: one where all water molecules were deleted and one that retained 13 active site water molecules. The magnesium and zinc ions were retained in both cases but not restrained in the binding site. The structures were then minimized, converging the heavy-atom RMSD to 0.3 Å. For each test case we aligned all structures to these reference structures and extracted the crystallographic binding modes of the ligands for use in the calculations.

2.3 Forcefields

We used the GAFF 2.11 forcefield for ligand parameters⁶⁴ alongside three protein forcefields, five water models, and two charge models. Not all combinations are appropriate and so the 16 parameter sets used are reported in Table 4.

Table 4 – Forcefield parameters tested. The 16 forcefield parameter sets tested.

Parameter Set	Protein Forcefield	Water Model	Charge Model
1	ff14SB	SPC/E	AM1-BCC
2	ff14SB	SPC/E	RESP
3	ff14SB	TIP3P	AM1-BCC
4	ff14SB	TIP3P	RESP
5	ff14SB	TIP4P-Ewald	AM1-BCC
6	ff14SB	TIP4P-Ewald	RESP
7	FB15	TIP3P-FB	AM1-BCC
8	FB15	TIP3P-FB	RESP
9	FB15	TIP4P-FB	AM1-BCC
10	FB15	TIP4P-FB	RESP
11	ff15ipq	SPC/E	AM1-BCC
12	ff15ipq	SPC/E	RESP
13	ff15ipq	TIP3P	AM1-BCC
14	ff15ipq	TIP3P	RESP
15	ff15ipq	TIP4P-Ewald	AM1-BCC
16	ff15ipq	TIP4P-Ewald	RESP

AM1-BCC partial atomic charges for the ligands were calculated using Antechamber.⁶⁵ RESP partial atomic charges for the ligands were calculated with Jaguar⁶⁶ using the DFT/B3LYP method

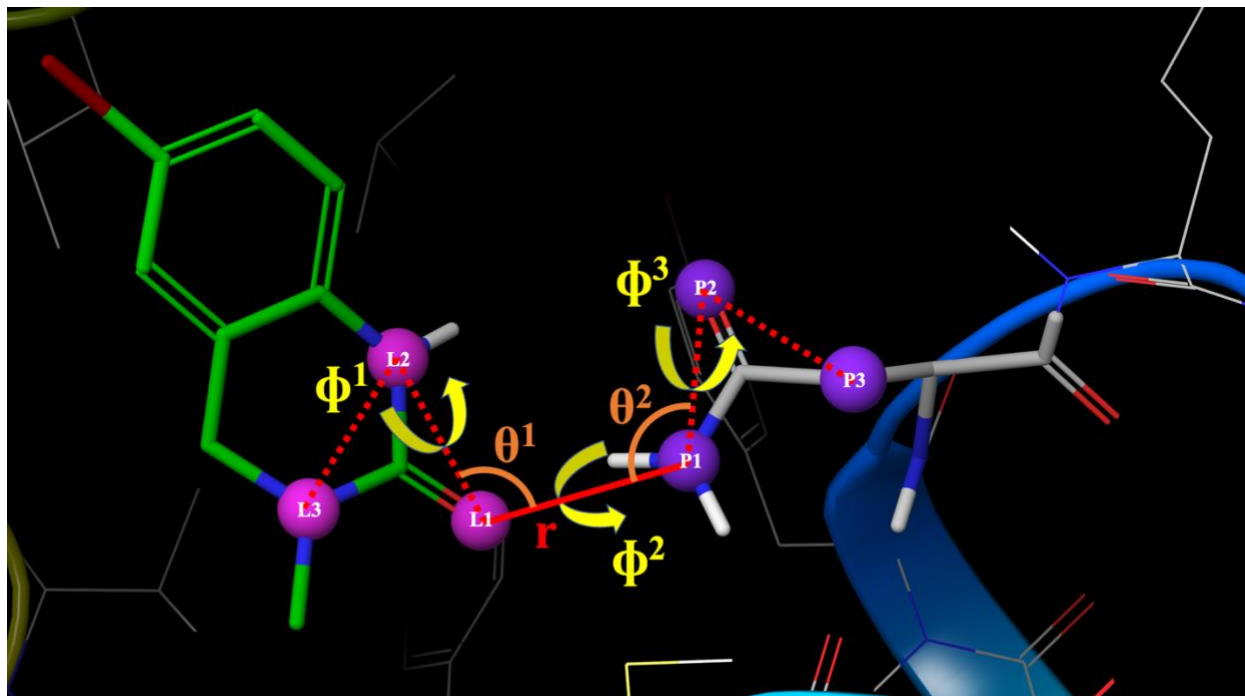
with a Poisson-Boltzmann solver and water as the solvent. Protein partial atomic charges are defined by the protein forcefields. The crystallographic binding modes of the ligands were first subjected to minimization at the 3-21G* level and then charges were fit at the 6-31G** level.

2.4 Absolute Binding Free Energy Calculations

Absolute binding free energy calculations were performed using OpenMM³⁷⁻⁴¹ version 7.2⁶⁷ with the openmmtools toolkit for Hamiltonian replica exchange (HREX).⁶⁸ Simulations were conducted at 300 K and 1 atm using a Monte Carlo barostat. Three calculations were performed to estimate each absolute binding free energy: decoupling of the ligand in solution, coupling of the ligand in complex whilst restrained to the protein, and releasing of these restraints. Free energy changes were calculated using MBAR^{69, 70} with 37 lambda states for the solution, 37 lambda states for the complex, and 9 lambda states for the restraints. For decoupling in solution, the electrostatics were first entirely decoupled before decoupling the sterics. For coupling in complex, the sterics were first entirely coupled before coupling the electrostatics.^{71, 72} The exact sequence of lambda windows with the specified lambda values for the electrostatics, sterics, and restraints is described by the lambda schedule. We found that sigmoidal lambda schedules tapered around $\Lambda=0.15$ for the sterics, $\Lambda=0.5$ for the electrostatics, and $\Lambda=0.0$ for the restraints enhanced the mixing of the alchemical states for the HREX and improved convergence of the free energy (data not shown). However, this may be specific to the ligands, proteins, forcefields, and softcore functions used. The exact lambda schedules used are reported in the supplementary information tables S1 and S2. Best practices dictate that the ligand should be restrained to the protein to prevent the ligand leaving the binding site and exploring the entire simulation box when all protein-ligand interactions are decoupled. For this reason, ligands were restrained to the protein using the six

restraints identified by Boresch.^{8, 12, 73} The restraints were selected and calculated using an automated protocol that is described in the SI methods. We are careful to choose restraining atoms that are part of a relatively rigid framework for both the protein and the ligand such that the thermodynamic cycle is close to exact. An example of the restraints is shown in Figure 2.

Figure 2 – Illustration of the bond, angle, and dihedral restraints connecting the protein and ligand together. The six restraints used in the free-energy calculations for ligand B2. The six atoms involved in the restraint are displayed as balls and labelled L1, L2, L3, P1, P2, and P3. The bond (r) is colored and labelled in red. The angles (θ^1 and θ^2) are colored and labelled in orange. The dihedrals (ϕ^1 , ϕ^2 , and ϕ^3) are colored and labelled in yellow. The protein is shown as cartoon ribbons and atom-colored wires with grey carbons. Residue Asn140 is shown as atom-colored sticks with grey carbons. The ligand B2 is shown as atom-colored sticks with green carbons.



Force constants from 0.25-8.0 kcal/mol/Å² for the bonds, 2.5-80.0 kcal/mol/rad² for the angles, and 5.0-160.0 kcal/mol for the dihedrals were tested (data not shown) in line with literature values.^{7-11, 71, 73} The final force constants used were 1.0 kcal/mol/Å² for the bonds, 5.0 kcal/mol/rad² for the angles, and 5.0 kcal/mol for the dihedrals. Simulations were performed with a timestep of 4.0 fs using hydrogen mass repartitioning and a hydrogen mass of 4 AMU.⁷⁴ Based on the scoping calculations presented in the results, the final simulations were performed with 500 MD steps per iteration for 200 equilibration iterations and 1000 production iterations. Solvent systems were generated with a 9.0 Å buffer between the solute and the edge of the cubic periodic box. Complex systems were generated with a 5.0 Å buffer between the solute and the edge of the cubic periodic box. Systems were neutralized as appropriate and the ionic strength was set to 150 mM with Na⁺ and Cl⁻ ions. Electrostatics were modelled with PME⁷⁵ and van der Waals were modelled using a nonbonded cutoff of 10.0 Å. Bonds to hydrogen were constrained and water molecules were modelled as rigid. To avoid the numerical instabilities referred to as end point catastrophes that occur when ligands approach the fully decoupled state, OpenMMTools employs a softcore function.⁷⁶ Default parameters were used for softcore_alpha (0.5), softcore_a (1), softcore_b (1), softcore_c (6), softcore_beta (0.0), softcore_d (1), softcore_e (1), and softcore_f (2). Python input files are available on Github (<https://github.com/djhuggins/Holoware-TestCases>)

2.5 Molecular Dynamics Simulations

MD simulations were performed on the initial states for each protein test case to explore structural variations in the absence of the ligands. Calculations were performed using the setup protocol and with the same simulation parameters as the absolute binding free energy calculations. RMSD values and system measurements were performed with MDTraj.⁷⁷

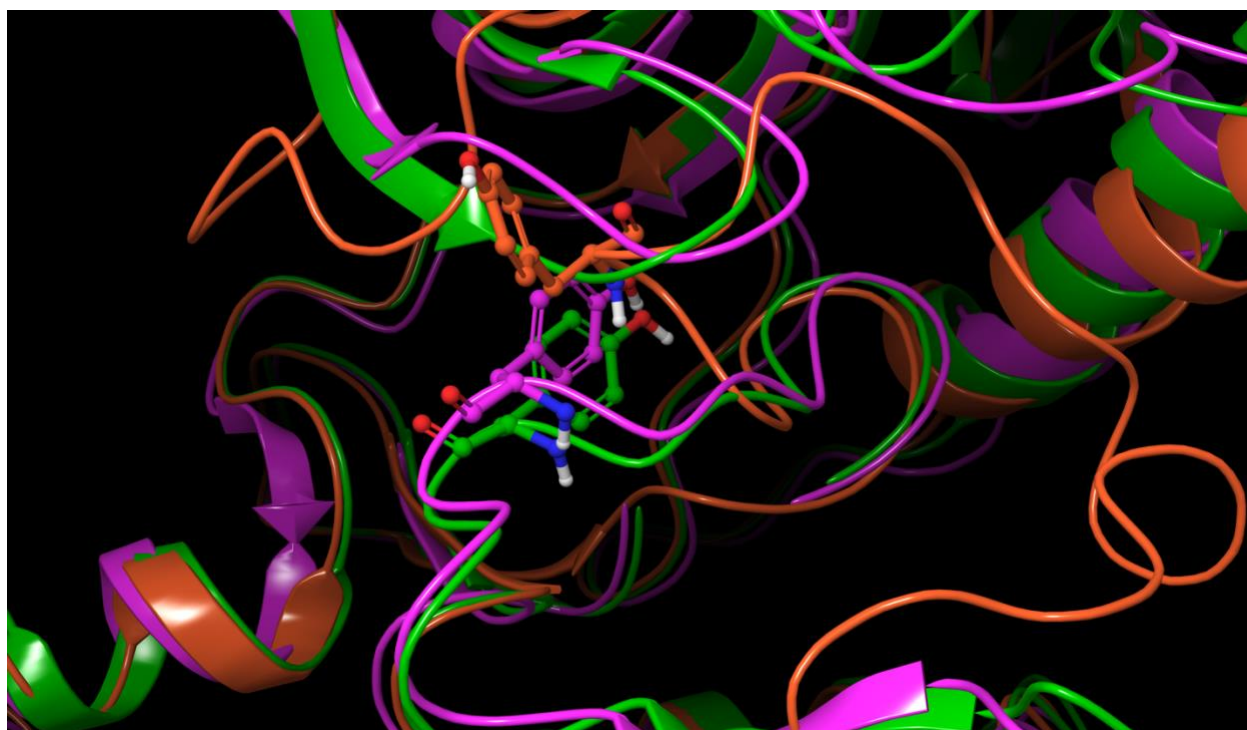
3. Results and Discussion

3.1 Scoping Calculations

Scoping calculations were performed for both MD simulations and absolute binding free energy calculations in order to refine the OpenMM parameters and protocols. MD simulations were performed on the initial states for each protein test case (five independent simulations of 24 ns). In the case of BRD4(1), a transition to a more favorable “open” conformation of the protein’s ZA-loop (Pro86 to Tyr98) has been noted in molecular dynamics simulations in the absence of ligands.⁹ By monitoring the Asp88 and Asp96 ϕ and ψ backbone torsions and using ff14SB and TIP3P parameters, Heinzelmann et al noted a transition after approximately 50 ns and the ZA-loop remained in the new conformation for an additional 250 ns of simulation. By monitoring the same backbone torsions and using ff14SB and SPC/E parameters, we note the transition in one of the five simulations of BRD4(1) and it occurs after approximately 12 ns. The transition does not occur in the other four simulations. The data is shown in supplementary figure S1. Heinzelmann et al calculated that the transition between the two states is associated with free energy changes of -2.54 , -3.72 , and -1.67 kcal/mol, for TIP3P, TIP4P-Ewald, and SPC/E, respectively (using ff14SB parameters for the protein) and adjusted their binding free energy predictions accordingly.⁹ In the case of cMET, protein conformational change is also important. The activation loop of the kinase adopts a specific conformation in the presence of the inhibitors in Table 2, all of which make key contacts with Tyr1230 in the activation loop.^{50, 55, 78} In the apo state, the activation loop and the P-loop have markedly different conformations.⁷⁹ For this reason the apo state is not suitable for free energy calculations with these ligands. The five independent simulations of the initial state illustrate that the activation loop retains essentially the same conformation on the timescale of the simulation. Figure 3 shows the apo state from PDBID 2G15,⁷⁹ the initial state used for the

simulations (a ligand bound structure), and an example of the conformation after 24 ns of MD with the ligand absent.

Figure 3 – Relevant conformations of cMET. The crystallographically observed apo conformation of cMET from PDBID 2G15 (orange), the crystallographically observed ligand bound conformation of cMET from PDBID 4XYF used for the simulations (green), and an example conformation after 24 ns of MD simulation in the absence of the ligand (magenta). The protein is shown as cartoon ribbons and residue Tyr1230 is shown in atom-colored balls and sticks.



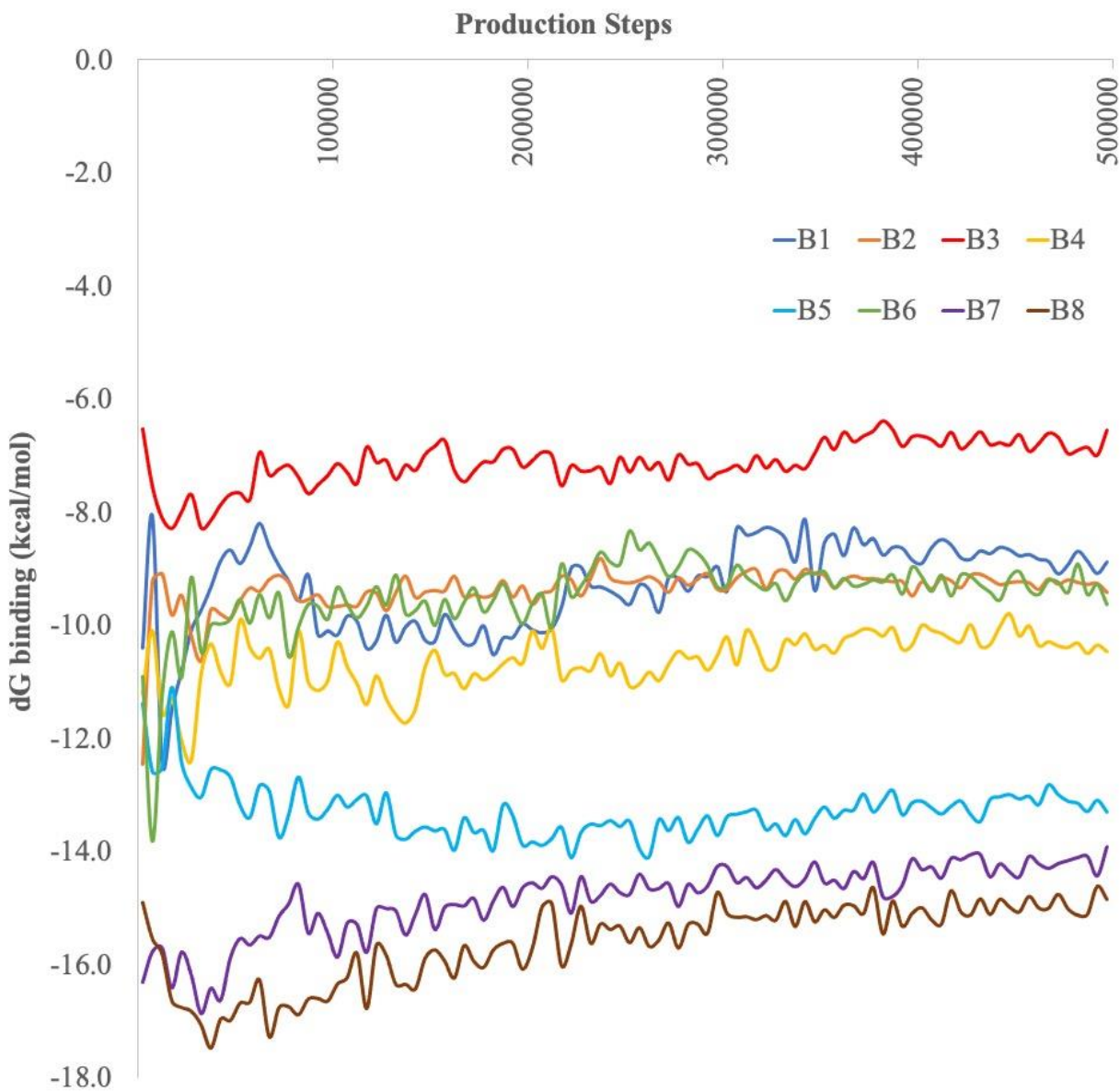
Whilst the protein clearly does not shift to the apo state conformation in any of the five independent simulations, the activation loop does show some flexibility and explores different conformations on the timescale of the simulations. This is illustrated by the time series data for backbone torsion angles of Met122, Tyr1230, Asp1231, and Lys1232 shown in supplementary figure S2. For

PDE2A, the simulation was monitored to check that the zinc and magnesium ions do not leave the binding site when no ligand is present. For all five independent simulations the ions remain in the site at a distance of approximately 4.7 Å from one another (the distance in the crystal structure is approximately 3.7 Å). The data is shown in supplementary figure S3.

Moving on to the alchemical simulations, we studied the effect of the frequency of HREX attempts on the results for the BRD4(1) test case with ligand B6. We used the ff14SB forcefield with the SPC/E water model and AM1-BCC charges. For a fixed number of total steps (50,000 equilibration plus 250,000 production) we varied the frequency of HREX attempts. Increasing the frequency of HREX attempts from one every 1000 steps to one every 50 steps lengthens the simulation time proportionally (as expected) but does not improve the convergence or the results. The data is shown in supplementary figure S4.

We then considered the number of equilibration steps using ligands B6, M8, and P10 for the three test cases. For a fixed number of production steps (200000) and frequency of HREX attempts (one every 1000 steps) we varied the number of equilibration steps. Decreasing the number of equilibration steps from 50000 to 5000 does not worsen the convergence or the results. However, the data suggests that more than 200000 production steps may be required for more flexible ligands such as P10. The data is shown in supplementary figures S5, S6, and S7 for BRD4(1), cMET, and PDE2A respectively. We also considered the convergence of the predictions as the number of production steps was increased. The number of equilibration steps was fixed at 100,000 and the frequency of HREX attempts was fixed at one every 500 steps. The convergence of the total binding free energy for all eight ligands in the BRD4(1) test set is shown in Figure 4.

Figure 4 – Convergence of the total binding free energy for all eight ligands in the BRD4(1) test set. The sum of the complex, solvent, restraint, and correction terms for ligands B1-B8 using the ff14SB forcefield with the SPC/E water model and AM1-BCC charges. Free energies were estimated every 5000 steps.



The total binding free energy estimates show little variation after 300,000 steps, stabilizing to within ± 1.0 kcal/mol of the value after 500,000 steps. In fact, if one monitors the correlation

between experiment and prediction as the number of steps is increased, the R^2 stabilizes around 0.6 after only 100,000 steps. After analyzing these results, we chose the parameters for the test set calculations noted in Table 5.

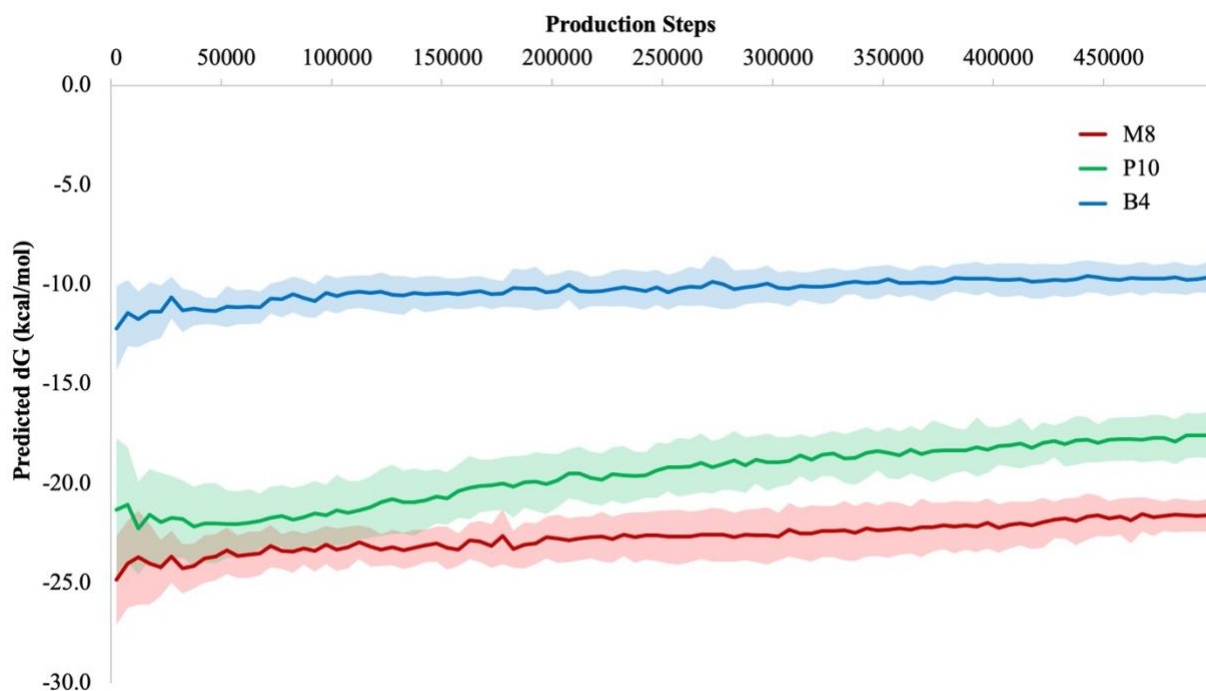
Table 5 – Final parameters. The simulation parameters used for the three test sets.

Parameter Name	Value
HREX attempt frequency	500
Equilibration steps	100,000
Production steps	500,000

These parameters correspond to 2.0 nanoseconds of production simulation per lambda window. This is much smaller than the timescale observed for the BRD4(1) conformational change, with only one transition observed after 12.0 nanoseconds in only one of five independent simulations. Thus with 37 lambda windows we use 74 nanoseconds per transformation for the solvent/complex calculations and with 9 lambda windows we use 18 nanoseconds per transformation for the restraint calculation. Importantly, the MD simulations of the apo state suggest that the BRD4(1) protein is likely to remain in the ligand bound state on the timescale of the simulations. For cMET, the MD simulations of the initial state suggest that the activation loop remains in the state that is competent for ligand binding when the ligand is fully decoupled. However, conformations of the activation loop may be differentially sampled when the ligand is fully decoupled and this may affect convergence and reproducibility. In the case of PDE2A, the MD simulations of the apo state suggest that the zinc and magnesium ions remain in the binding site on the timescale of the simulations. As a final check of the convergence and reproducibility using these parameters, we performed five repeats for each protein test case using the ff14SB forcefield with the SPC/E water model and AM1-BCC charges. Ligands B4, M8, and P10 were selected as they have more rotatable

bonds and are likely to present the greatest challenge for sampling. A plot of the mean and standard deviation of the total binding free energy is shown in Figure 5.

Figure 5 – Convergence plots for ligands B4, M8, and P10 using the optimized set of parameters. Means and standard deviations of the free energy values for five independent simulations of ligand B4, M8, and P10 using the ff14SB forcefield with the SPC/E water model and AM1-BCC charges as the number of production steps is increased. The means are shown as a solid lines and the standard deviations are shown as shaded areas in blue, red, and green for B4, M8, and P10 respectively.



After 500,000 production steps the standard deviations across the five independent simulations are 0.78, 0.83, and 1.15 kcal/mol for B4, M8, and P10 respectively. The complex leg is the main source of the variance, with standard deviations of 0.76, 0.81, and 1.06 kcal/mol respectively. As

expected, the variance is highest for ligand P10 which is the most flexible with eleven non-methyl rotatable torsion angles. The variance is expected to be much lower for molecules with lower flexibility. For example, ligand B2 with zero non-methyl rotatable torsion angles has a standard deviation of 0.35 kcal/mol across five independent simulations (data shown in Figure S8). Whilst the data suggests that flexible molecules do present a challenge for sampling, we consider these parameters acceptable for assessing performance across the different test cases and ligands.

3.2 BRD4(1) Performance

The results for BRD4(1) are presented in Table 6.

Table 6 – BRD4(1) results. The results for the BRD4(1) test set with the 16 forcefield parameter sets.

Protein Forcefield	Water Model	Charge Model	MSE (kcal/mol)	MUE (kcal/mol)	RMSE (kcal/mol)	R²	τ
ff14SB	SPC/E	AM1-BCC	-2.78	3.00	3.36	0.54	0.71
ff14SB	SPC/E	RESP	-1.38	1.78	2.02	0.46	0.40
ff14SB	TIP3P	AM1-BCC	-1.51	1.67	2.20	0.48	0.43
ff14SB	TIP3P	RESP	-1.01	1.54	1.78	0.31	0.36
ff14SB	TIP4P-Ewald	AM1-BCC	-2.65	2.79	3.17	0.53	0.50
ff14SB	TIP4P-Ewald	RESP	-0.91	1.37	1.44	0.56	0.50
FB15	TIP3P-FB	AM1-BCC	-2.85	2.86	3.18	0.67	0.71
FB15	TIP3P-FB	RESP	-1.10	1.62	1.85	0.37	0.36
FB15	TIP4P-FB	AM1-BCC	-2.04	2.34	2.55	0.54	0.57
FB15	TIP4P-FB	RESP	-0.27	1.36	1.62	0.33	0.29
ff15ipq	SPC/E	AM1-BCC	-2.05	2.50	2.99	0.57	0.57
ff15ipq	SPC/E	RESP	-0.73	1.03	1.26	0.66	0.71
ff15ipq	TIP3P	AM1-BCC	-0.88	1.25	1.47	0.60	0.50
ff15ipq	TIP3P	RESP	-0.30	1.13	1.40	0.31	0.43
ff15ipq	TIP4P-Ewald	AM1-BCC	-2.51	2.55	3.01	0.47	0.50
ff15ipq	TIP4P-Ewald	RESP	-1.28	1.70	2.07	0.50	0.50

The mean signed error (MSE) is negative in all cases indicating a bias to predict the binding free energy as more favorable than experiment (between 0.27 and 2.85 kcal/mol). However, the mean unsigned error (MUE) and root mean square error (RMSE) are reasonable in most cases, with the combination of ff15ipq, SPC/E, RESP yielding the best performance for both of 1.03 kcal/mol and 1.26 kcal/mol respectively. The R^2 and Kendall τ in this case are 0.66 and 0.71 respectively. It should be noted that these predictions are based on simulations that are not expected to sample the “open” state predicted to be favored by Heinzelmann et al.⁹ One cannot strictly apply the free energy changes for transitions to the “open” state calculated by Heinzelmann et al to our calculations because they performed the binding free energy calculations with protein torsional restraints and then calculated the free energy of releasing them whilst no protein torsional restraints were used in the current study. However, one can consider how the errors would be affected for parameter sets 1-6 which correspond to those used by Heinzelmann et al (-2.54, -3.72, and -1.67 kcal/mol for TIP3P, TIP4P-Ewald, and SPC/E, respectively using ff14SB parameters for the protein). For example, the MUEs for the parameter sets ff14SB, TIP3P, AM1BCC and ff14SB, TIP3P, RESP would improve by over 1.0 kcal/mol to 2.00 and 0.97 kcal respectively, while the MUE for the parameter set ff14SB, TIP4P-Ewald, RESP would worsen by over 1.0 kcal/mol to 2.89 kcal. Whilst informative, these offsets cannot be strictly used as a full accounting of the thermodynamic cycle is required separately for each ligand. The full set of results for all molecules and all parameter sets are in supplementary table S3.

3.3 *cMET Performance*

The results for cMET are presented in Table 7.

Table 7 – cMET results. The results for the cMET test set with the 16 forcefield parameter sets.

Protein Forcefield	Water Model	Charge Model	MSE (kcal/mol)	MUE (kcal/mol)	RMSE (kcal/mol)	R²	τ
ff14SB	SPC/E	AM1-BCC	-8.71	8.71	8.85	0.47	0.50
ff14SB	SPC/E	RESP	-7.83	7.83	8.09	0.43	0.56
ff14SB	TIP3P	AM1-BCC	-6.24	6.24	6.63	0.61	0.56
ff14SB	TIP3P	RESP	-5.77	5.77	6.00	0.57	0.72
ff14SB	TIP4P-Ewald	AM1-BCC	-7.92	7.92	8.12	0.77	0.78
ff14SB	TIP4P-Ewald	RESP	-7.46	7.46	7.63	0.65	0.65
FB15	TIP3P-FB	AM1-BCC	-10.08	10.08	10.21	0.71	0.72
FB15	TIP3P-FB	RESP	-9.79	9.79	9.92	0.57	0.67
FB15	TIP4P-FB	AM1-BCC	-9.33	9.33	9.60	0.45	0.61
FB15	TIP4P-FB	RESP	-8.09	8.09	8.22	0.69	0.82
ff15ipq	SPC/E	AM1-BCC	-7.66	7.66	7.72	0.73	0.72
ff15ipq	SPC/E	RESP	-7.29	7.29	7.63	0.25	0.50
ff15ipq	TIP3P	AM1-BCC	-4.97	4.97	5.32	0.75	0.83
ff15ipq	TIP3P	RESP	-4.66	4.88	5.24	0.31	0.50
ff15ipq	TIP4P-Ewald	AM1-BCC	-7.79	7.79	8.05	0.34	0.50
ff15ipq	TIP4P-Ewald	RESP	-6.91	6.91	7.35	0.24	0.33

The MSE is again negative in all cases indicating a bias to predict the binding free energy as more favorable than experiment. In this case the trend is more severe than for BRD4(1), with MSE values between 4.66 and 10.08 kcal/mol. This leads to poorer MUE and RMSE values compared with the BRD4(1) test case. We posit that the bias in this case is caused by the need for a significant protein rearrangement to allow ligand binding. As noted above, the residue Tyr1230 forms a pi-stacking interaction with all the ligands and this requires a significant rearrangement of the A loop.⁷⁸ This rearrangement is so major that attempts to quantify the free energy change computationally would be extremely challenging. Despite this bias, the correlation with experiment is higher than 0.7 in a number of cases. The best R² of 0.77 is found for ff14SB, TIP4P-Ewald, AM1-BCC and the best Kendall τ of 0.83 is found for ff15ipq, TIP3P, AM1-BCC. Notably, the ff15ipq, TIP3P, AM1-BCC parameter set correlates well with experiment for both the BRD4(1)

and cMET test cases, with R^2 values of 0.60 and 0.75. The full set of results for all molecules and all parameter sets are in supplementary table S4.

3.4 PDE2A Performance

For PDE2A we first studied predictions with and without water molecules in the binding site of the input protein structure. There are many water molecules bridging between the ligand and the protein in addition to coordinating the active site magnesium and zinc ions. Alongside this we studied predictions with the original AMBER parameters for divalent cations and a set of parameters rationally designed to be compatible with PME simulations in explicit solvent (the HFE set).⁸⁰ These preliminary calculations were performed for only 200 production iterations with 20 equilibration iterations and only the ff14SB forcefield with the SPC/E or TIP3P water model. For both water models and both divalent cation parameters we found that the results without water molecules in the binding site of the input protein structure were always poorer (mean unsigned error ranging from 5.15 to 7.70 kcal/mol) than the results with water molecules present in the input protein structure (mean unsigned error ranging from 3.24 to 5.10 kcal/mol). To expand on this, using a structure without water molecules in the binding site of the input protein structure led to a much larger bias to predict the binding free energy as more favorable than experiment. We also found that using the HFE divalent cation parameters improved accuracy over the original divalent cation parameters in all cases (the mean unsigned error improved by between 0.07 and 2.55 kcal/mol). The full results for these calculations are in supplementary table S5. The full results for PDE2A (with water molecules in the binding site of the input protein structure and the HFE divalent cation parameters) are presented in Table 8.

Table 8 – PDE2A results. The results for the PDE2A test set with the 16 forcefield parameter sets (with water molecules in the binding site of the input protein structure and the HFE divalent cation parameters).

Protein Forcefield	Water Model	Charge Model	MSE (kcal/mol)	MUE (kcal/mol)	RMSE (kcal/mol)	R²	τ
ff14SB	SPC/E	AM1-BCC	-0.40	1.44	1.67	0.65	0.47
ff14SB	SPC/E	RESP	1.16	2.32	2.94	0.49	0.56
ff14SB	TIP3P	AM1-BCC	1.99	1.99	2.61	0.48	0.42
ff14SB	TIP3P	RESP	2.08	2.21	2.71	0.62	0.56
ff14SB	TIP4P-Ewald	AM1-BCC	-1.11	1.82	2.29	0.51	0.56
ff14SB	TIP4P-Ewald	RESP	0.30	1.29	1.96	0.66	0.60
FB15	TIP3P-FB	AM1-BCC	-1.38	1.78	2.51	0.71	0.64
FB15	TIP3P-FB	RESP	-0.15	2.28	2.71	0.70	0.69
FB15	TIP4P-FB	AM1-BCC	-1.27	1.77	2.40	0.81	0.81
FB15	TIP4P-FB	RESP	-0.72	1.83	2.10	0.71	0.64
ff15ipq	SPC/E	AM1-BCC	-0.52	1.65	2.12	0.61	0.51
ff15ipq	SPC/E	RESP	0.36	1.40	1.90	0.67	0.51
ff15ipq	TIP3P	AM1-BCC	0.87	1.45	1.81	0.59	0.51
ff15ipq	TIP3P	RESP	1.80	2.05	2.89	0.64	0.60
ff15ipq	TIP4P-Ewald	AM1-BCC	-2.51	2.64	3.01	0.73	0.60
ff15ipq	TIP4P-Ewald	RESP	-1.15	2.01	2.57	0.61	0.60

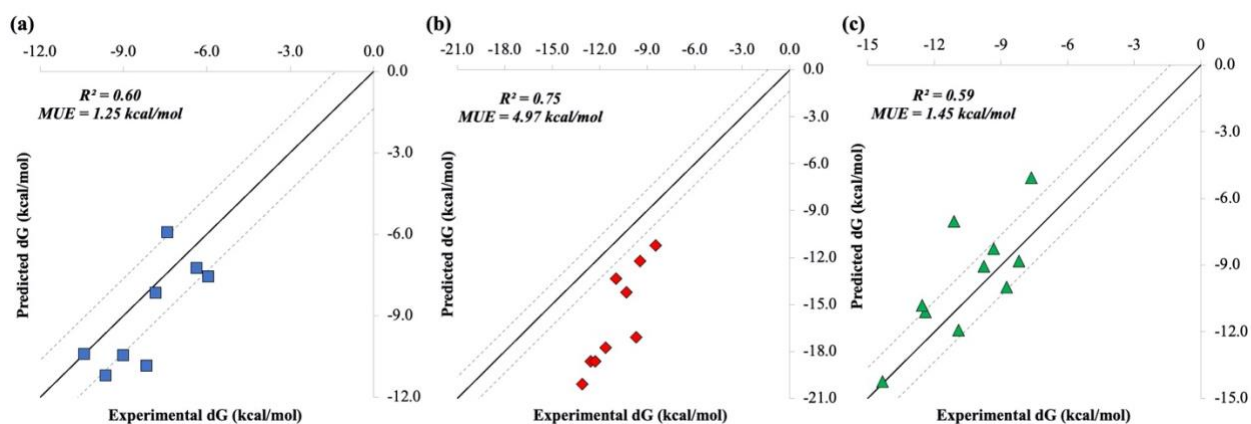
Many of the parameter sets perform well, without a clear bias to predict the binding free energy as more favorable than experiment as seen for the other two test cases. The majority of the parameter sets yield R² values greater than 0.6 and some of the parameter sets yield MUE values below 1.5 kcal/mol. Again, there is a slight disconnect between the best performance in terms of error and the best performance in terms of correlation. Again, the combination of ff15ipq, TIP3P, AM1-BCC yields a good performance relative to the other parameter sets, with an MUE and RMSE of 1.45 kcal/mol and 1.81 kcal/mol respectively. The R² and Kendall τ in this case are 0.59 and 0.51. Conversely, the ff15ipq, TIP4P-Ewald, AM1-BCC parameter set shows R² and Kendall τ values that are higher than this at 0.81 and 0.81 but MUE and RMSE values that are also higher 1.77

kcal/mol and 2.40 kcal/mol respectively. The full set of results for all molecules and all parameter sets are in supplementary table S6.

3.5 Forcefield Performance

The results are reasonable for all three test cases, with many parameter sets yielding an R^2 correlation with experiment greater than 0.6. However, the use of RESP charges in place of AM1-BCC charges does not tend to improve the accuracy of the predictions, if anything leading to poorer performance on average. The TIP4P-Ewald model also does not improve performance relative to the TIP3P and SPC/E models, which show similar performance. Overall, it is interesting to note the good performance of the ff15ipq forcefield in combination with the TIP3P water model relative to the other parameter sets, especially at accurately reproducing the free energy values. In this case, results are similar for both AM1-BCC and RESP charges. Figure 6 shows plots of the correlation between experiment and prediction for the ff15ipq, TIP3P, AM1-BCC parameter set.

Figure 6 - Correlation between experiment and prediction for the ff15ipq, TIP3P, AM1-BCC parameter set for (a) BRD4(1), (b) cMET, and (c) PDE2A.



The R^2 values indicating correlation are 0.60, 0.75, and 0.59 for BRD4(1), cMET, and PDE2A respectively. The mean unsigned errors for BRD4(1) and PDE2A are 1.25 kcal/mol and 1.45 kcal/mol respectively while the mean unsigned error for cMET is low in comparison (4.97 kcal/mol). As discussed, this is likely due to a significant conformational change required for ligand binding that is not sampled on the timescales employed here.

4. Summary and Conclusions

We have developed a workflow in OpenMM to calculate absolute binding free energies and validated it using the existing test case of BRD4(1) and two newly developed test cases (cMET and PDE2A). The results for many of the parameter sets in the BRD4(1) and PDE2A test cases show MUEs below 1.5 kcal/mol and R^2 values above 0.6. For the cMET test case, many parameter sets yield similar correlations, but relatively poorer MUE/RMSE values as discussed. For BRD4(1), we can compare these results with previous studies. Aldeghi et al applied MBAR calculations to BRD4(1).¹⁰ In Table 1 they report results for 11 ligands, 6 of which are in the test set used here. With the parameter set of the Amber99SB-ILDN protein forcefield,⁸¹ GAFF ligand parameters, the TIP3P water model, and AM1-BCC charges they report an MUE of 0.6 kcal/mol, an RMSE of 0.8 kcal/mol, and a correlation coefficient of 0.84 (R^2 of 0.71) for all 11 ligands. Heinzlmann et al applied attach-pull-release calculations to BRD4(1).⁹ In Table 2 they report results for 7 ligands, 5 of which are in the test set used here. With the optimal parameter set of the ff14SB protein forcefield, GAFF ligand parameters, the SPC/E water model, and AM1-BCC charges they report an RMSE of 1.42 kcal/mol, a correlation coefficient of 0.83 (R^2 of 0.69), and a Kendall τ of 0.49 for all 7 ligands (accounting for the free energy of the protein's closed-to-open conformational change). For our work using the parameter set of the ff15ipq protein forcefield, the

SPC/E water model, and RESP charges the MUE is 1.03 kcal/mol, the RMSE is 1.26 kcal/mol, the correlation coefficient is 0.81 (R^2 of 0.66), and the Kendall τ is 0.71 for all 10 ligands. Results in all three studies are comparable and show a good performance relative to literature values, a performance that would be useful for virtual screening.

For the cMET and PDE2A test cases we cannot compare directly with previous studies but the MUE, RMSE, R^2 , and Kendall τ values can be analyzed. For cMET, the R^2 values are relatively high with a maximum of 0.77 but the errors are relatively high with MSE values between 4.66 and 10.08 kcal/mol. We suggest that this is due to the need for a significant protein rearrangement to facilitate ligand binding for all ligands. The optimal parameter set of the ff15ipq protein forcefield, the TIP3P water model, and AM1-BCC charges yields an MUE of 4.97 kcal/mol, an RMSE of 5.32 kcal/mol, an R^2 of 0.75, and a Kendall τ of 0.83. For PDE2A, preliminary calculations also suggest that structural waters are important for accurate predictions and were not treated appropriately using a standard MD method when they were missing from the starting structure. This reinforces the need for methods such as GCMC that can place water molecules during a simulation.^{82,83} However, when water molecules are included in the starting structure, the accuracy is typically better than 2.0 kcal/mol and the R^2 values are higher than 0.6. Absolute binding free energies may also benefit from the use of enhanced sampling techniques such as solute tempering⁸⁴ or temperature replica exchange⁸⁵ but have not been explored here.

The parameter set of the ff15ipq protein forcefield, the TIP3P water model, and AM1-BCC charges yields an MUE of 1.45 kcal/mol, an RMSE of 1.81 kcal/mol, an R^2 of 0.59, and a Kendall τ of 0.51. As discussed in the introduction, there are a number of published approaches for calculation of alchemical binding free energy and it would be interesting to apply them to the two new test cases presented here and discover whether the ff15ipq protein forcefield is equally effective.

These results are reasonable in the context of absolute binding free energy predictions and in comparison with similar studies. XFEP was applied to a set of test cases developed using congeneric series for relative binding free energies. The authors reported an overall RMSE of 1.11 kcal/mol, an R^2 of 0.47, and a Kendall τ of 0.49.¹⁷ GA-FEP has been applied to both congeneric series and diverse ligands.¹⁸ Across the test cases, the authors reported RMSEs between 0.62 kcal/mol and 1.54 kcal/mol with R^2 values between 0.48 and 0.88. Khalak et al also analysed a set of test cases with congeneric series to calculate absolute binding free energies.⁶ They reported an overall MUE of 1.17 kcal/mol and an R^2 of 0.40.

MD simulations of the apo states indicate that the ligand-bound states of the three proteins are stable on the timescale of the simulations performed here. However, these findings may well be dependent on the forcefield and conformational changes are certainly dependent on the timescale of the relevant simulations. Such considerations are system dependent and will always be important for MD-based methods. Heinzlmann et al noted a transition to a more favored “open” state of BRD4(1) after 50 ns of MD that is unlikely to be sampled on the timescale of standard free energy calculations. We note this transition after 12 ns in one of five independent simulations of BRD4(1). Corrections to binding free energy predictions may be warranted in such cases and they would likely improve results for BRD4(1) using the parameter sets studied here based on simple considerations. Importantly, the cMET test case developed here shows a much more significant protein rearrangement upon binding all of the ligands studied and this likely requires a parameter set dependent correction to binding free energy predictions. However, if all ligands bind to a similar state that is stable on the timescale of the calculations, there should be minimal effect on the correlation and ranking as the corrections will provide a similar offset to all of the predictions. Notably, of the three test cases studied here, the cMET test case shows the best performance in

terms of correlation with experimental binding free energies but the worst performance in terms of error.

For all three cases we note that there is no clear improvement when using RESP charges relative to AM1-BCC charges. However, alternative protocols to generate the RESP charges should be explored in future work. In addition, no clear improvement is seen when the FB15 protein model and/or the four-point water models TIP4P-Ewald/TIP4P-FB are used. Relative to the other parameter sets tested, we also note the good performance of the ff15ipq protein forcefield for all three test cases, particularly when considering the MUE and RMSE. Non-polarizable forcefields typically use an average or specific polarization of molecules and cannot explicitly model or quantify changes in polarization. We posit that this manifests most clearly in absolute binding free energy calculations where the entire ligand changes its environment from the solvent to the complex and polarization changes are most significant.

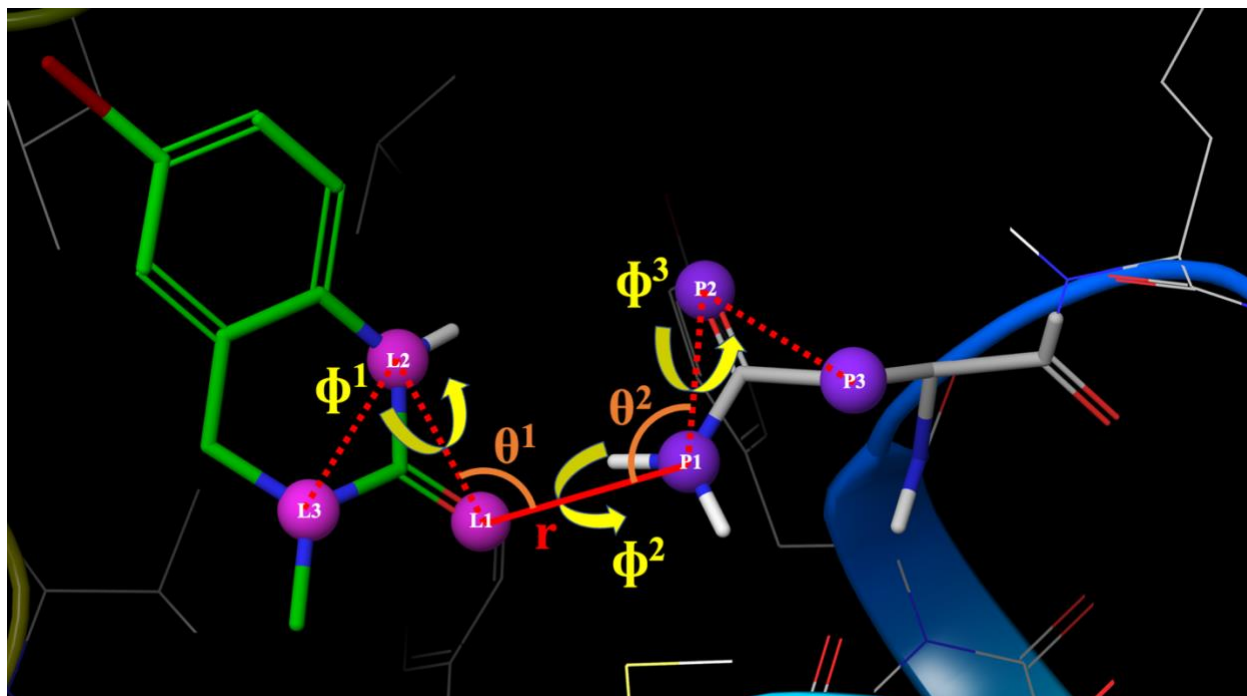
In summary, this work reports the predictive accuracy of 16 parameter sets in calculating absolute binding free energies for 27 ligands across one existing and two newly developed test cases. There are an array of forcefields that could be used for such calculations and deciding which will be accurate in any given case is challenging. We have provided some guidance based on the results for BRD4(1), cMET, and PDE2A but performance varies so the best practice is to validate effectiveness on a case-by-case basis. Unfortunately, appropriate benchmarking data is typically unavailable in the use case where virtual screening is being used to identify novel chemical matter. Benchmark against a similar protein system could also prove useful but, in the absence of such information, the work presented here suggests that the ff15ipq protein forcefield performs very well, with the parameter set ff15ipq, TIP3P, AM1-BCC yielding the best results across the three test sets considered. However, there are a large number of parameter sets that have not been

considered here. For example, the mdgx program from Ambergtools can generate ligand parameters using the IPolQ approach that was used to generate the ff15ipq protein forcefield.⁸⁶ Open Force Field have also recently released their 1.0.0 small molecule force field (Parsley)⁸⁷ and a version 2.0.0 (Sage). In addition, the combination of the OPC water model⁸⁸ and the ff19-SB protein forcefield has been recommended for better predictive power and could be added to openmmforcefields.⁸⁹ This study introduces two new benchmark sets that can be used to identify the most effective parameter sets for absolute binding free energy calculations in future work.

Acknowledgements

The authors thank Alex Wade, Andrea Rizzi, and Peter Eastman for helpful discussions. The authors gratefully acknowledge the generous support to this project provided by the Tri-Institutional Therapeutics Discovery Institute (TDI), a 501(c)(3) organization. TDI receives financial support from Takeda Pharmaceutical Company, TDI's parent institutes (Memorial Sloan Kettering Cancer Center, The Rockefeller University and Weill Cornell Medicine) and from a generous contribution from Mr. Lewis Sanders and other philanthropic sources.

TOC



References

1. Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W., Rigorous Free Energy Simulations in Virtual Screening. *J Chem Inf Model* **2020**, *60* (9), 4153-4169.
2. Woo, H.-J.; Roux, B., Calculation of absolute protein–ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (19), 6825-6830.
3. Wang, J.; Deng, Y.; Roux, B., Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophysical Journal* **2006**, *91* (8), 2798-2814.
4. Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A., Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. *Journal of Molecular Biology* **2007**, *371* (4), 1118-1134.
5. Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S., Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *The Journal of chemical physics* **2006**, *125* (8), 084901.
6. Khalak, Y.; Tresadern, G.; Aldeghi, M.; Baumann, H. M.; Mobley, D. L.; de Groot, B. L.; Gapsys, V., Alchemical absolute protein–ligand binding free energies for drug design. *Chemical Science* **2021**, *12* (41), 13958-13971.
7. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Predictions of ligand selectivity from absolute binding free energy calculations. *Journal of the American Chemical Society* **2017**, *139* (2), 946-957.

8. Heinzlmann, G.; Gilson, M. K., Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Scientific reports* **2021**, *11* (1), 1-18.
9. Heinzlmann, G.; Henriksen, N. M.; Gilson, M. K., Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *J Chem Theory Comput* **2017**, *13* (7), 3260-3275.
10. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical science* **2016**, *7* (1), 207-218.
11. Qian, Y.; Cabeza de Vaca, I.; Vilseck, J. Z.; Cole, D. J.; Tirado-Rives, J.; Jorgensen, W. L., Absolute Free Energy of Binding Calculations for Macrophage Migration Inhibitory Factor in Complex with a Druglike Inhibitor. *The Journal of Physical Chemistry B* **2019**, *123* (41), 8675-8685.
12. Irwin, B. W.; Huggins, D. J., Estimating atomic contributions to hydration and binding using free energy perturbation. *Journal of chemical theory and computation* **2018**, *14* (6), 3218-3227.
13. Gallicchio, E.; Lapelosa, M.; Levy, R. M., Binding energy distribution analysis method (BEDAM) for estimation of Protein– Ligand binding affinities. *Journal of Chemical Theory and Computation* **2010**, *6* (9), 2961-2977.
14. Fu, H.; Gumbart, J. C.; Chen, H.; Shao, X.; Cai, W.; Chipot, C., BFEE: A user-friendly graphical interface facilitating absolute binding free-energy calculations. *Journal of chemical information and modeling* **2018**, *58* (3), 556-560.

15. Fu, H.; Chen, H.; Cai, W.; Shao, X.; Chipot, C., BFEE2: Automated, Streamlined, and Accurate Absolute Binding Free-Energy Calculations. *Journal of Chemical Information and Modeling* **2021**, *61* (5), 2116-2123.
16. Kim, S.; Oshima, H.; Zhang, H.; Kern, N. R.; Re, S.; Lee, J.; Roux, B.; Sugita, Y.; Jiang, W.; Im, W., CHARMM-GUI free energy calculator for absolute and relative ligand solvation and binding free energy simulations. *Journal of chemical theory and computation* **2020**, *16* (11), 7207-7218.
17. Lin, Z.; Zou, J.; Liu, S.; Peng, C.; Li, Z.; Wan, X.; Fang, D.; Yin, J.; Gobbo, G.; Chen, Y., A Cloud Computing Platform for Scalable Relative and Absolute Binding Free Energy Predictions: New Opportunities and Challenges for Drug Discovery. *Journal of Chemical Information and Modeling* **2021**, *61* (6), 2720-2732.
18. Li, Z.; Huang, Y.; Wu, Y.; Chen, J.; Wu, D.; Zhan, C.-G.; Luo, H.-B., Absolute binding free energy calculation and design of a subnanomolar inhibitor of phosphodiesterase-10. *Journal of medicinal chemistry* **2019**, *62* (4), 2099-2111.
19. Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K., Overview of the SAMPL5 host-guest challenge: Are we doing better? *J Comput Aided Mol Des* **2017**, *31* (1), 1-19.
20. Skillman, A. G., SAMPL3: blinded prediction of host-guest binding affinities, hydration free energies, and trypsin inhibitors. *J Comput Aided Mol Des* **2012**, *26* (5), 473-4.
21. Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D., The SAMPL6

SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations.

Journal of Computer-Aided Molecular Design **2020**, *34* (5), 601-633.

22. Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D., Overview of the SAMPL6 host-guest binding affinity prediction challenge. *Journal of Computer-Aided Molecular Design* **2018**, *32* (10), 937-963.

23. Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K., Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J Mol Biol* **2009**, *394* (4), 747-63.

24. Xie, B.; Nguyen, T. H.; Minh, D. D. L., Absolute Binding Free Energies between T4 Lysozyme and 141 Small Molecules: Calculations Based on Multiple Rigid Receptor Configurations. *J Chem Theory Comput* **2017**, *13* (6), 2930-2944.

25. Wang, K.; Chodera, J. D.; Yang, Y.; Shirts, M. R., Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *Journal of Computer-Aided Molecular Design* **2013**, *27* (12), 989-1007.

26. Mobley, D. L.; Gilson, M. K., Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu Rev Biophys* **2017**, *46*, 531-558.

27. Mobley, D. L.; Heinzemann, G.; Henriksen, N. M.; Gilson, M. K., Predicting binding free energies: Frontiers and benchmarks (a perpetual review). **2017**.

28. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **2015**, *11* (8), 3696-713.

29. Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T., Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J Chem Theory Comput* **2016**, *12* (8), 3926-47.
30. Wang, L. P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martinez, T. J.; Pande, V. S., Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J Phys Chem B* **2017**, *121* (16), 4023-4039.
31. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The missing term in effective pair potentials. *The Journal of Physical Chemistry* **1987**, *91* (24), 6269-6271.
32. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **1983**, *79* (2), 926-935.
33. Wang, L. P.; Martinez, T. J.; Pande, V. S., Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J Phys Chem Lett* **2014**, *5* (11), 1885-91.
34. Horn, H. W.; Swope, W. C.; Pitara, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* **2004**, *120* (20), 9665-78.
35. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **2002**, *23* (16), 1623-41.
36. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. *Journal of Physical Chemistry* **1993**, *97* (40), 10269-10280.

37. Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S., Accelerating molecular dynamic simulation on graphics processing units. *Journal of computational chemistry* **2009**, *30* (6), 864-872.
38. Eastman, P.; Pande, V., OpenMM: A hardware-independent framework for molecular simulations. *Computing in science & engineering* **2010**, *12* (4), 34-39.
39. Eastman, P.; Pande, V. S., Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. *Journal of computational chemistry* **2010**, *31* (6), 1268-1272.
40. Eastman, P.; Pande, V. S., Constant constraint matrix approximation: a robust, parallelizable constraint method for molecular simulations. *Journal of chemical theory and computation* **2010**, *6* (2), 434-437.
41. Chodera, J. D.; Shirts, M. R., Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *The Journal of chemical physics* **2011**, *135* (19), 194110.
42. Lucas, X.; Wohlwend, D.; Hügler, M.; Schmidtkunz, K.; Gerhardt, S.; Schüle, R.; Jung, M.; Einsle, O.; Günther, S., 4-Acyl Pyrroles: Mimicking Acetylated Lysines in Histone Code Reading. *Angewandte Chemie International Edition* **2013**, *52* (52), 14055-14059.
43. Fish, P. V.; Filippakopoulos, P.; Bish, G.; Brennan, P. E.; Bunnage, M. E.; Cook, A. S.; Federov, O.; Gerstenberger, B. S.; Jones, H.; Knapp, S.; Marsden, B.; Nocka, K.; Owen, D. R.; Philpott, M.; Picaud, S.; Primiano, M. J.; Ralph, M. J.; Sciammetta, N.; Trzupek, J. D., Identification of a Chemical Probe for Bromo and Extra C-Terminal Bromodomain Inhibition through Optimization of a Fragment-Derived Hit. *Journal of Medicinal Chemistry* **2012**, *55* (22), 9831-9837.

44. Filippakopoulos, P.; Picaud, S.; Fedorov, O.; Keller, M.; Wrobel, M.; Morgenstern, O.; Bracher, F.; Knapp, S., Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. *Bioorganic & Medicinal Chemistry* **2012**, *20* (6), 1878-1886.
45. Picaud, S.; Wells, C.; Felletar, I.; Brotherton, D.; Martin, S.; Savitsky, P.; Diez-Dacal, B.; Philpott, M.; Bountra, C.; Lingard, H.; Fedorov, O.; Müller, S.; Brennan, P. E.; Knapp, S.; Filippakopoulos, P., RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proceedings of the National Academy of Sciences* **2013**, *110* (49), 19754-19759.
46. Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.; McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E., Selective inhibition of BET bromodomains. *Nature* **2010**, *468* (7327), 1067-1073.
47. Gehling, V. S.; Hewitt, M. C.; Vaswani, R. G.; Leblanc, Y.; Côté, A.; Nasveschuk, C. G.; Taylor, A. M.; Harmange, J.-C.; Audia, J. E.; Pardo, E.; Joshi, S.; Sandy, P.; Mertz, J. A.; Sims, R. J.; Bergeron, L.; Bryant, B. M.; Bellon, S.; Poy, F.; Jayaram, H.; Sankaranarayanan, R.; Yellapantula, S.; Bangalore Srinivasamurthy, N.; Birudukota, S.; Albrecht, B. K., Discovery, Design, and Optimization of Isoxazole Azepine BET Inhibitors. *ACS Medicinal Chemistry Letters* **2013**, *4* (9), 835-840.
48. Bode, C. M.; Boezio, A. A.; Albrecht, B. K.; Bellon, S. F.; Berry, L.; Broome, M. A.; Choquette, D.; Dussault, I.; Lewis, R. T.; Lin, M.-H. J.; Rex, K.; Whittington, D. A.; Yang,

- Y.; Harmange, J.-C., Discovery and optimization of a potent and selective triazolopyridinone series of c-Met inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2012**, *22* (12), 4089-4093.
49. Albrecht, B. K.; Harmange, J.-C.; Bauer, D.; Berry, L.; Bode, C.; Boezio, A. A.; Chen, A.; Choquette, D.; Dussault, I.; Fridrich, C.; Hirai, S.; Hoffman, D.; Larrow, J. F.; Kaplan-Lefko, P.; Lin, J.; Lohman, J.; Long, A. M.; Moriguchi, J.; O'Connor, A.; Potashman, M. H.; Reese, M.; Rex, K.; Siegmund, A.; Shah, K.; Shimanovich, R.; Springer, S. K.; Teffera, Y.; Yang, Y.; Zhang, Y.; Bellon, S. F., Discovery and Optimization of Triazolopyridazines as Potent and Selective Inhibitors of the c-Met Kinase. *Journal of Medicinal Chemistry* **2008**, *51* (10), 2879-2882.
50. Zhang, L.; Zhao, J.; Zhang, B.; Lu, T.; Chen, Y., Discovery of [1,2,4]triazolo[3,4-b][1,3,4]thiadiazole derivatives as novel, potent and selective c-Met kinase inhibitors: Synthesis, SAR study, and biological activity. *Eur J Med Chem* **2018**, *150*, 809-816.
51. Nishii, H.; Chiba, T.; Morikami, K.; Fukami, T. A.; Sakamoto, H.; Ko, K.; Koyano, H., Discovery of 6-benzyloxyquinolines as c-Met selective kinase inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2010**, *20* (4), 1405-1409.
52. Jia, H.; Dai, G.; Weng, J.; Zhang, Z.; Wang, Q.; Zhou, F.; Jiao, L.; Cui, Y.; Ren, Y.; Fan, S.; Zhou, J.; Qing, W.; Gu, Y.; Wang, J.; Sai, Y.; Su, W., Discovery of (S)-1-(1-(Imidazo[1,2-a]pyridin-6-yl)ethyl)-6-(1-methyl-1H-pyrazol-4-yl)-1H-[1,2,3]triazolo[4,5-b]pyrazine (Volitinib) as a Highly Potent and Selective Mesenchymal–Epithelial Transition Factor (c-Met) Inhibitor in Clinical Development for Treatment of Cancer. *Journal of Medicinal Chemistry* **2014**, *57* (18), 7577-7589.
53. Peterson, E. A.; Teffera, Y.; Albrecht, B. K.; Bauer, D.; Bellon, S. F.; Boezio, A.; Boezio, C.; Broome, M. A.; Choquette, D.; Copeland, K. W.; Dussault, I.; Lewis, R.; Lin,

M.-H. J.; Lohman, J.; Liu, J.; Potashman, M.; Rex, K.; Shimanovich, R.; Whittington, D. A.; Vaida, K. R.; Harmange, J.-C., Discovery of Potent and Selective 8-Fluorotriazolopyridine c-Met Inhibitors. *Journal of Medicinal Chemistry* **2015**, *58* (5), 2417-2430.

54. Boezio, A. A.; Copeland, K. W.; Rex, K.; K. Albrecht, B.; Bauer, D.; Bellon, S. F.; Boezio, C.; Broome, M. A.; Choquette, D.; Coxon, A.; Dussault, I.; Hirai, S.; Lewis, R.; Lin, M.-H. J.; Lohman, J.; Liu, J.; Peterson, E. A.; Potashman, M.; Shimanovich, R.; Teffera, Y.; Whittington, D. A.; Vaida, K. R.; Harmange, J.-C., Discovery of (R)-6-(1-(8-Fluoro-6-(1-methyl-1H-pyrazol-4-yl)-[1,2,4]triazolo[4,3-a]pyridin-3-yl)ethyl)-3-(2-methoxyethoxy)-1,6-naphthyridin-5(6H)-one (AMG 337), a Potent and Selective Inhibitor of MET with High Unbound Target Coverage and Robust In Vivo Antitumor Activity. *Journal of Medicinal Chemistry* **2016**, *59* (6), 2328-2342.

55. Zhan, Z.; Peng, X.; Liu, Q.; Chen, F.; Ji, Y.; Yao, S.; Xi, Y.; Lin, Y.; Chen, T.; Xu, Y.; Ai, J.; Geng, M.; Duan, W., Discovery of 6-(difluoro(6-(4-fluorophenyl)-[1,2,4]triazolo[4,3-b][1,2,4]triazin-3-yl)methyl)quinoline as a highly potent and selective c-Met inhibitor. *Eur J Med Chem* **2016**, *116*, 239-251.

56. Qiu, X.; Huang, Y.; Wu, D.; Mao, F.; Zhu, J.; Yan, W.; Luo, H.-B.; Li, J., Discovery of novel purine nucleoside derivatives as phosphodiesterase 2 (PDE2) inhibitors: Structure-based virtual screening, optimization and biological evaluation. *Bioorganic & Medicinal Chemistry* **2018**, *26* (1), 119-133.

57. Gomez, L.; Massari, M. E.; Vickers, T.; Freestone, G.; Vernier, W.; Ly, K.; Xu, R.; McCarrick, M.; Marrone, T.; Metz, M.; Yan, Y. G.; Yoder, Z. W.; Lemus, R.; Broadbent, N. J.; Barido, R.; Warren, N.; Schmelzer, K.; Neul, D.; Lee, D.; Andersen, C. B.; Sebring, K.; Aertgeerts, K.; Zhou, X.; Tabatabaei, A.; Peters, M.; Breitenbucher, J. G., Design and

Synthesis of Novel and Selective Phosphodiesterase 2 (PDE2a) Inhibitors for the Treatment of Memory Disorders. *Journal of Medicinal Chemistry* **2017**, *60* (5), 2037-2051.

58. Helal, C. J.; Arnold, E. P.; Boyden, T. L.; Chang, C.; Chappie, T. A.; Fennell, K. F.; Forman, M. D.; Hajos, M.; Harms, J. F.; Hoffman, W. E.; Humphrey, J. M.; Kang, Z.; Kleiman, R. J.; Kormos, B. L.; Lee, C.-W.; Lu, J.; Maklad, N.; McDowell, L.; Mente, S.; O'Connor, R. E.; Pandit, J.; Piotrowski, M.; Schmidt, A. W.; Schmidt, C. J.; Ueno, H.; Verhoest, P. R.; Yang, E. X., Application of Structure-Based Design and Parallel Chemistry to Identify a Potent, Selective, and Brain Penetrant Phosphodiesterase 2A Inhibitor. *Journal of Medicinal Chemistry* **2017**, *60* (13), 5673-5698.

59. Buijnsters, P.; De Angelis, M.; Langlois, X.; Rombouts, F. J. R.; Sanderson, W.; Tresadern, G.; Ritchie, A.; Trabanco, A. A.; VanHoof, G.; Roosbroeck, Y. V.; Andrés, J.-I., Structure-Based Design of a Potent, Selective, and Brain Penetrating PDE2 Inhibitor with Demonstrated Target Engagement. *ACS Medicinal Chemistry Letters* **2014**, *5* (9), 1049-1053.

60. Mikami, S.; Nakamura, S.; Ashizawa, T.; Nomura, I.; Kawasaki, M.; Sasaki, S.; Oki, H.; Kokubo, H.; Hoffman, I. D.; Zou, H.; Uchiyama, N.; Nakashima, K.; Kamiguchi, N.; Imada, H.; Suzuki, N.; Iwashita, H.; Taniguchi, T., Discovery of Clinical Candidate N-((1S)-1-(3-Fluoro-4-(trifluoromethoxy)phenyl)-2-methoxyethyl)-7-methoxy-2-oxo-2,3-dihydropyrido[2,3-b]pyrazine-4(1H)-carboxamide (TAK-915): A Highly Potent, Selective, and Brain-Penetrating Phosphodiesterase 2A Inhibitor for the Treatment of Cognitive Disorders. *Journal of Medicinal Chemistry* **2017**, *60* (18), 7677-7702.

61. Obach, R. S.; Walker, G. S.; Sharma, R.; Jenkinson, S.; Tran, T. P.; Stepan, A. F., Lead Diversification at the Nanomole Scale Using Liver Microsomes and Quantitative Nuclear

Magnetic Resonance Spectroscopy: Application to Phosphodiesterase 2 Inhibitors. *Journal of Medicinal Chemistry* **2018**, *61* (8), 3626-3640.

62. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000**, *28* (1), 235-42.

63. Zhu, J.; Yang, Q.; Dai, D.; Huang, Q., X-ray Crystal Structure of Phosphodiesterase 2 in Complex with a Highly Selective, Nanomolar Inhibitor Reveals a Binding-Induced Pocket Important for Selectivity. *Journal of the American Chemical Society* **2013**, *135* (32), 11708-11711.

64. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004**, *25* (9), 1157-74.

65. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc* **2001**, *222*, U403.

66. Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A., Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *International Journal of Quantum Chemistry* **2013**, *113* (18), 2110-2142.

67. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S., OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **2017**, *13* (7), e1005659.

68. Wade, A. D.; Rizzi, A.; Wang, Y.; Huggins, D. J., Computational Fluorine Scanning Using Free-Energy Perturbation. *Journal of Chemical Information and Modeling* **2019**, *59* (6), 2776-2784.

69. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* **2008**, *129* (12), 124105.
70. Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976**, *22* (2), 245-268.
71. Mobley, D. L.; Chodera, J. D.; Dill, K. A., On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J Chem Phys* **2006**, *125* (8), 084902.
72. Rocklin, G. J.; Mobley, D. L.; Dill, K. A., Separated topologies--a method for relative binding free energy calculations using orientational restraints. *J Chem Phys* **2013**, *138* (8), 085104.
73. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M., Absolute binding free energies: a quantitative approach for their calculation. *The Journal of Physical Chemistry B* **2003**, *107* (35), 9535-9551.
74. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **2015**, *11* (4), 1864-1874.
75. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **1993**, *98* (12), 10089-10092.
76. Pham, T. T.; Shirts, M. R., Identifying low variance pathways for free energy calculations of molecular transformations in solution phase. *The Journal of chemical physics* **2011**, *135* (3), 034114.
77. McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S., MDTraj: a modern

- open library for the analysis of molecular dynamics trajectories. *Biophysical journal* **2015**, *109* (8), 1528-1532.
78. Tang, Q.; Aronov, A. M.; Deininger, D. D.; Giroux, S.; Lauffer, D. J.; Li, P.; Liang, J.; McGinty, K.; Ronkin, S.; Swett, R.; Waal, N.; Boucher, D.; Ford, P. J.; Moody, C. S., Discovery of Potent, Selective Triazolothiadiazole-Containing c-Met Inhibitors. *ACS Medicinal Chemistry Letters* **2021**, *12* (6), 955-960.
79. Wang, W.; Marimuthu, A.; Tsai, J.; Kumar, A.; Krupka, H. I.; Zhang, C.; Powell, B.; Suzuki, Y.; Nguyen, H.; Tabrizizad, M.; Luu, C.; West, B. L., Structural characterization of autoinhibited c-Met kinase produced by coexpression in bacteria with phosphatase. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103* (10), 3563-3568.
80. Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz Jr, K. M., Rational design of particle mesh Ewald compatible Lennard-Jones parameters for+ 2 metal cations in explicit solvent. *Journal of chemical theory and computation* **2013**, *9* (6), 2733-2748.
81. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78* (8), 1950-8.
82. Samways, M. L.; Bruce Macdonald, H. E.; Essex, J. W., grand: A Python Module for Grand Canonical Water Sampling in OpenMM. *Journal of Chemical Information and Modeling* **2020**, *60* (10), 4436-4441.
83. Ross, G. A.; Russell, E.; Deng, Y.; Lu, C.; Harder, E. D.; Abel, R.; Wang, L., Enhancing Water Sampling in Free Energy Calculations with Grand Canonical Monte Carlo. *Journal of Chemical Theory and Computation* **2020**, *16* (10), 6061-6076.

84. Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J., Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (39), 13749-13754.
85. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999**, *314* (1-2), 141-151.
86. Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A., Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization. *The Journal of Physical Chemistry B* **2013**, *117* (8), 2328-2338.
87. Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P., Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *Journal of Chemical Theory and Computation* **2021**, *17* (10), 6262-6280.
88. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters* **2014**, *5* (21), 3863-3871.
89. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migués, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020**, *16* (1), 528-552.