Supporting Information

A Benchmark of Electrostatic Method Performance in Relative Binding Free Energy Calculations

Yunhui Ge,[†] David F. Hahn,[‡] and David L. Mobley[*,†,¶]

†*Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697, USA*

‡*Computational Chemistry, Janssen Research & Development, Turnhoutseweg 30, Beerse B-2340, Belgium*

¶*Department of Chemistry, University of California, Irvine, CA 92697, USA*

E-mail: dmobley@mobleylab.org

Molecular Dynamics Simulations

The simulations were performed using GROMACS (2021-dev-20200320-89f1227-unknown) with a patch optimizing PME performance on GPU (`https://gerrit.gromacs.org/c/gromacs/+/13382`).

For each perturbation, two sets of simulations were prepared: solvated ligands and ligand-protein complexes. The initial ligand and protein structures were obtained from a previous published work.[1]

The ligand was parameterized using Open Force Field version 1.0.0 (codenamed "Parsley").[2] Hybrid structures and topologies for the ligand pairs were generated using pmx[3,4] following a single topology approach. The workflow established a mapping between atoms

of two ligands based on the maximum common substructure and conformational alignment while minimizing perturbation and stabilizing the system.

The AMBER ff99sb*ILDN force field[5–7] was used for protein parameterization. Dodecahedral boxes were filled with TIP3P explicit solvent model solvated ligand pairs/ligand-protein complexes and counterions (150 mM NaCl).

For each perturbation, two states were prepared for both in-solution/bound state ligands: state A and state B, representing ligand 1 and ligand 2, respectively. An energy minimization was first performed, followed by a 10 ps NVT equilibration at 298K. Then the production equilibrium simulation (in the NPT ensemble) was performed for 6 ns at 298 K and a pressure of 1 bar. 80 snapshots were extracted from the production simulation. For each snapshot, a non-equilibrium transformation from state A to B (and vice versa) was performed during 50 ps. For each perturbation, 3 replicas of the series of simulations described above were performed leading to a total of 60 ns simulation data to calculate the free energy differences for the ligands in their in-solution/bound states.

The stochastic dynamics thermostat was used to control the temperature in the simulations. The Parrinello-Rahman barostat[8] was applied to keep the pressure constant. All bond lengths were constrained using the LINCS algorithm.[9] The van der Waals interactions were smoothly switched off between 1.0 and 1.1 nm. A dispersion correction for energy and pressure was used. The non-bonded interactions for the alchemical transitions were treated with a modified soft-core potential.[10]

Two different methods implemented in GROMACS were used to treat long-range electrostatic interactions: Particle Mesh Ewald (PME) and Reaction Field (RF). For simulations using PME, a direct space cutoff of 1.1 nm and a Fourier grid spacing of 0.12 nm were used. For simulations using RF, a dielectric constant (ϵ_{rf}) of 78.3, the dielectric constant of water at 298 K was used.[11]

Supporting Tables

Table S1: Simulated perturbations of TYK2 and the number of particles.

Ligand1	Ligand 2	No. atoms (ligand-only)	No. atoms (protein-ligand)
ejm-44	ejm-55	5954	62290
ejm-49	ejm-31	6521	62288
ejm-31	ejm-46	5888	62281
jmc-28	jmc-27	5886	62282
ejm-42	ejm-48	6014	62290
ejm-31	ejm-43	5950	62286
ejm-50	ejm-42	5630	62272
ejm-42	ejm-55	5633	62275
jmc-23	ejm-46	5884	62277
ejm-31	ejm-45	5638	62283
ejm-55	ejm-54	5729	62278
ejm-45	ejm-42	5639	62284
ejm-31	ejm-48	5960	62284
ejm-47	ejm-31	5906	62290
ejm-47	ejm-55	5915	62290
ejm-44	ejm-42	5956	62295
jmc-23	jmc-27	5886	62276
ejm-43	ejm-55	5951	62281
jmc-23	jmc-30	5884	62277
jmc-28	jmc-30	5887	62286
ejm-42	ejm-54	5640	62282
ejm-49	ejm-50	6517	62290
jmc-23	ejm-55	5887	62289
ejm-31	jmc-28	5889	62291
jmc-23	jmc-27	5886	62276

Table S2: Simulated perturbations of CDK2 and the number of particles.

Ligand1	Ligand 2	No. atoms (ligand-only)	No. atoms (protein-ligand)
22	1h1r	4978	106892
1oiu	26	5885	106905
26	1h1q	5762	106902
17	1h1q	4987	106904
1oiy	1oi9	5675	106908
17	21	4984	106895

Table S3: Calculated $\Delta\Delta G$ values (kcal/mol) and the experimental measured values of simulated perturbations of TYK2. The uncertainty estimates were performed by 1000 bootstrapping trials and are reported as $x_{x_{low}}^{x_{high}}$ where x is the mean value, x_{high} and x_{low} indicate 95% confidence intervals.

Ligand1	Ligand 2	$\Delta\Delta G$ (CPU-PME)	$\Delta\Delta G$ (GPU-PME)	$\Delta\Delta G$ (CPU-RF)	$\Delta\Delta G$ (GPU-RF)	$\Delta\Delta G$ (EXP)
ejm-44	ejm-55	-4.40±0.44	-4.54±0.73	-3.88±0.55	-4.79±0.70	-1.79
ejm-49	ejm-31	-0.89±0.13	-0.57±0.09	-0.81±0.20	-0.59±0.36	-1.79
ejm-31	ejm-46	-0.25±0.21	-0.96±0.17	-0.71±0.07	-0.68±0.11	-1.77
jmc-28	jmc-27	-0.71±0.06	-0.63±0.06	-0.91±0.11	-1.00±0.08	-0.30
ejm-42	ejm-48	0.83±0.23	1.22±0.26	1.01±0.11	1.25±0.24	0.78
ejm-31	ejm-43	1.87±0.11	1.94±0.12	1.58±0.63	1.58±0.41	1.28
ejm-50	ejm-42	-0.24±0.13	0.07±0.12	-0.16±0.11	0.12±0.13	-0.80
ejm-42	ejm-55	-1.17±0.12	-0.95±0.10	-1.05±0.07	-1.07±0.12	0.57
jmc-23	ejm-46	-0.04±0.18	-0.31±0.19	0.13±0.13	0.30±0.13	0.39
ejm-55	ejm-54	-1.10±0.31	-0.86±0.10	-0.99±0.19	-1.22±0.07	-1.32
ejm-45	ejm-42	0.57±0.07	0.12±0.11	0.45±0.12	0.50±0.18	-0.22
ejm-31	ejm-48	0.63±0.34	1.12±0.43	0.94±0.30	0.65±0.72	0.54
ejm-47	ejm-31	-0.12±0.34	-0.27±0.72	-0.30±0.22	-0.20±0.26	0.16
ejm-47	ejm-55	-0.91±0.11	-0.77±0.24	-0.63±0.17	-0.82±0.24	0.49
ejm-44	ejm-42	-2.22±0.11	-2.12±0.36	-2.39±0.64	-1.86±0.27	-2.36
jmc-23	jmc-27	0.07±0.21	0.13±0.16	0.11±0.19	0.28±0.06	0.42
ejm-43	ejm-55	-2.40±0.30	-2.60±0.26	-2.08±0.33	-2.23±0.37	-0.95
jmc-23	jmc-30	-0.38±0.09	-0.29±0.06	-0.22±0.10	-0.06±0.13	0.76
jmc-28	jmc-30	-1.00±0.17	-1.06±0.48	-1.14±0.23	-1.40±0.19	0.04
ejm-42	ejm-54	-1.33±0.17	-1.36±0.18	-1.43±0.11	-1.54±0.12	-0.75
ejm-49	ejm-50	-1.28±0.28	-1.18±0.32	-1.75±0.11	-1.38±0.41	-1.23
jmc-23	ejm-55	0.35±0.33	-0.13±0.17	0.12±0.38	0.09±0.21	2.49
ejm-31	ejm-45	0.47±0.31	0.54±0.22	0.49±0.29	0.61±0.11	-0.02
ejm-31	jmc-28	-0.58±0.24	-0.56±0.17	-0.15±0.41	0.18±0.30	-1.44
RMSE		$1.09_{0.77}^{1.43}$	$1.16_{0.82}^{1.56}$	$1.04_{0.76}^{1.35}$	$1.20_{0.83}^{1.61}$	
MUE		$0.86_{0.61}^{1.14}$	$0.93_{0.66}^{1.27}$	$0.86_{0.64}^{1.10}$	$0.94_{0.66}^{1.27}$	

Table S4: Simulation wallclock time (min) of TYK2.

Simulation	complex	ligand-only
CPU-PME	2838.2±328.1	288.1±27.6
GPU-PME	185.4±4.12	69.0±3.8
CPU-RF	1903.7±334.2	211.6±20.5
GPU-RF	204.7±5.9	66.0±4.8

Table S5: Simulation wallclock time (min) of CDK2.

Simulation	complex	ligand-only
CPU-PME	4447.3±423.7	252.8±24.7
GPU-PME	326.8±6.5	83.1±3.8
CPU-RF	3143.0±204.5	210.2±21.9
GPU-RF	325.9±11.9	77.2±4.3

References

(1) Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Tresadern, G.; de Groot, B. L. Large Scale Relative Protein Ligand Binding Affinities Using Non-Equilibrium Alchemy. *Chem. Sci.* **2020**, *11*, 1140–1152.

(2) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Lucas, X.; Tjanaka, B.; 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. *ChemRxiv* **2020**,

(3) Seeliger, D.; de Groot, B. L. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophys. J.* **2010**, *98*, 2309–2316.

(4) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. Pmx: Automated Protein Structure and Topology Generation for Alchemical Perturbations. *J. Comput. Chem.* **2015**, *36*, 348–354.

(5) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.

(6) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.

(7) 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: Improved Protein Side-Chain Potentials. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950–1958.

(8) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(9) Hess, B.; Bekker, H.; Berendsen, H. J. C. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(10) Gapsys, V.; Seeliger, D.; de Groot, B. L. New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2373–2382.

(11) Malmberg, C. G.; Maryott, A. A. Dielectric Constant of Water From 0-Degrees-C to 100-Degrees-C. *J. Res. Natl. Bur. Stand. (U. S.)* **1956**, *56*, 1–8.