Electronic Supplementary Material (ESI) for ChemComm.

Supporting Information for:

Modelling Flexible Protein-Ligand Binding in p38α MAP Kinase using the QUBE Force Field

Joshua T. Horton,[†] Alice E. A. Allen,[‡] and Daniel J. Cole[*,†]

†*School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom*

‡*Engineering Laboratory, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom*

E-mail: *daniel.cole@ncl.ac.uk

S1 Computational Methods

S1.1 System Preparation

An x-ray crystal structure of p38α MAP kinase in complex with an inhibitor (PDB ID: 1OUY) structurally similar to ligand **17** was selected as a starting point. This ligand was then truncated and served as the common core substructure used to generate all other compounds via the molecule growing program BOMB.[1] Crystallographic water molecules were removed and the protein and ligand z-matrices were prepared using the chop and pepz utilities of MCPRO 3.2.[2] Any residues within 20 Å of the ligand were retained and a fully flexible region was defined within this region with a radius of 10 Å. It was confirmed that an increase in the radius of the flexible region to 12.5 Å changed the computed relative binding free energy by less than 0.2 kcal/mol for the transformation of **2** to **1** (from +0.36 to +0.20 kcal/mol). The net charge of the system was set to zero via neutralization of distant, titratable residues, and non-bonded energy terms used a 10 Å cutoff. Ligand and key host degrees of freedom were optimized using BOMB. Each protein-ligand complex was solvated in a water cap with radius 25 Å using the JAWS hydration protocol described in detail elsewhere.[3]

S1.2 QUBE Force Field Parametrization

Ligand force fields were parametrized using the QUBEKit software package.[4] Quantum chemistry geometry optimizations and frequency calculations were performed in Gaussian09[5] using the ωB97XD functional and 6-311++G(d,p) basis set. Equilibrium bond lengths and angles were extracted from the QM optimized geometry, and the bond-stretching and angle-bending force constants were derived from the QM Hessian matrix via the modified Seminario method with a vibrational scaling factor of 0.957.[6] Constrained one-dimensional torsional optimizations were also performed using Gaussian09, with the same level of theory and basis set, in 15° increments from 0° to 360°. Torsion parameter optimizations of dihedrals ϕ_1 and

ϕ_2 were performed for each ligand separately using QUBEKit with no Boltzmann weighting or regularization.[4] OPLS atom types were retained during torsion fitting to reduce the parameter search space, while all remaining small molecule torsion parameters were taken from the OPLS force field. Non-bonded parameter assignment was performed for both small molecules and the protein (2961 atoms) using the ONETEP linear-scaling density functional theory code and DDEC AIM analysis (see below). All bonded parameters of the protein were assigned from a transferable library that has been specifically designed to be compatible with the QUBE FF.[7] Water molecules were described using the TIP4P water model.

S1.3 ONETEP Calculations

All ground state electron densities used to derive the non-bonded parameters of both the 18 ligands and the p38 kinase protein (2961 atoms) were computed using the linear-scaling density functional theory code, ONETEP.[8] ONETEP uses a basis set of spatially-truncated nonorthogonal generalized Wannier functions (NGWFs) localized on each atom. Four NGWFs, with radii of 10 Bohr, were used for all atoms with the exception of hydrogen, which used one. NGWFs were expanded in a periodic cardinal sine (psinc) basis, with a grid size ($0.45a_o$), corresponding to a plane wave cutoff energy of 1020 eV. The PBE exchange-correlation functional was used with OPIUM norm-conserving pseudopotentials. The calculation was carried out in an implicit solvent using a dielectric of 4 to model induction effects in the ligands, and 10 in the protein. For several test cases, ligand charges were also computed using a dielectric of 10, but the RMS/maximum differences between the charge sets are just 0.01/0.03 e (see ESI data). The DDEC module implemented in ONETEP was used to partition the electron density and assign atom-centered point charges and atomic volumes. No off-center charges were used in this study.[4] Electron density partitioning was performed using an IH to ISA ratio of 0.02.[9] Lennard-Jones parameters were calculated using the Tkatchenko-Scheffler relations,[10] and protocols described previously.[9]

S1.4 Free Energy Calculations

Free energy calculations were performed in MCPRO version 3.2 using the single topology approach for both the bound (protein-ligand complex in water) and unbound (ligand in water) simulations as part of a standard thermodynamic cycle. Ligands were transformed over the course of 11 equally spaced λ windows. Simple overlap sampling was employed, with each window comprising 10 million (M) (20M) configurations of equilibration and 30 M (40 M) configurations of averaging for the bound (unbound) simulations. All computed free energy changes (including those presented from previous studies) were computed by aligning the mean energies of the experimental and computed distributions. Replica exchange with solute tempering (REST) was used during each λ window to effectively rescale the non-bonded and dihedral parameters of the ligand, thereby reducing potential energy barriers in "high temperature" replicas of the system.[11,12] Four replicas were run in parallel with REST scaling factors exponentially distributed in the range from 25°C to 250°C (chosen to allow reasonable replica exchange). Exchange attempts between pairs of neighboring replicas were attempted every 10 000 MC steps, with acceptance criteria chosen to maintain detailed balance,[11] and the resulting free energy changes were computed from the room temperature ensemble. The "flip" Monte Carlo (MC) dihedral move modification[13] was also used to encourage crossing between energetically separated poses 1 and 2, with random move sizes that ranged from 60° to 180°. Figure S1 shows the 2D distribution of ϕ_1 and ϕ_2 sampled during MC simulations of compound **1**, and confirms approximately equal populations of poses 1 and 2. Protein conformal sampling employed new protocols, which generate more efficient MC moves specifically targeted at the backbone and side-chains.[14] These moves have been shown to give good agreement with molecular dynamics simulations for the calculation of protein conformational ensembles[14] and protein-ligand binding.[15]

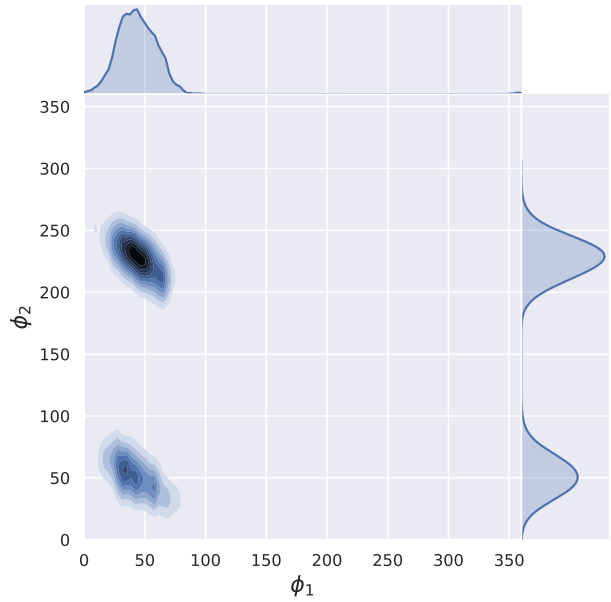

Figure S1. Two-dimensional dihedral distribution observed during the protein-ligand complex MC simulation of ligand **1**.

S2 Transition pathways

All FEP transitions were performed relative to ligand **1** via the following pathways.

$2 \rightarrow 1$

$3 \rightarrow 1$

$4 \rightarrow 9 \rightarrow 10 \rightarrow 7 \rightarrow 1$

$5 \rightarrow 1$

$6 \rightarrow 3 \rightarrow 1$

$7 \rightarrow 1$

$8 \rightarrow 1$

$9 \rightarrow 10 \rightarrow 7 \rightarrow 1$

$10 \rightarrow 7 \rightarrow 1$

$11 \rightarrow 8 \rightarrow 1$

$12 \rightarrow 18 \rightarrow 1$

$13 \rightarrow 3 \rightarrow 1$

$14 \rightarrow 18 \rightarrow 1$

$15 \rightarrow 2 \rightarrow 1$

$16 \rightarrow 17 \rightarrow 18 \rightarrow 1$

$17 \rightarrow 18 \rightarrow 1$

$18 \rightarrow 1$

S3 Torsion fitting

Table S1. Root mean square deviation between QM and QUBE torsional energy profiles for rotation of ϕ_1 and ϕ_2 for each of the 18 molecules.

	RMSE (kcal/mol)	
Compound	ϕ_1	ϕ_2
1	0.038	0.107
2	0.061	0.063
3	0.057	0.061
4	0.132	0.230
5	0.046	0.259
6	0.073	0.192
7	0.284	0.188
8	0.372	0.287
9	0.147	0.141
10	0.116	0.158
11	0.215	0.363
12	0.407	0.453
13	0.381	0.140
14	0.341	0.053
15	0.475	0.303
16	0.161	0.144
17	0.103	0.438
18	0.112	0.539
Average	**0.196**	**0.229**

S4 Correlation between QUBE and QM Energetics

By deriving the QUBE force field directly from QM, our goal is to provide accurate and automated molecule-specific parameters that reproduce as closely as possible the full QM potential energy surface. Figure S2 shows the correlation between QUBE and QM relative energies of structures **3** and **10** extracted from Monte Carlo simulations. The correlation between QUBE and QM energetics is similar to that previously reported,[4] and significantly QUBE does not predict any physically unreasonable structures (either bound to the protein or in water) whilst retaining the fixed MM functional form that provides us with a practical method for deployment in free energy predictions. Table S2 further reports correlations for all 18 ligands studied here.

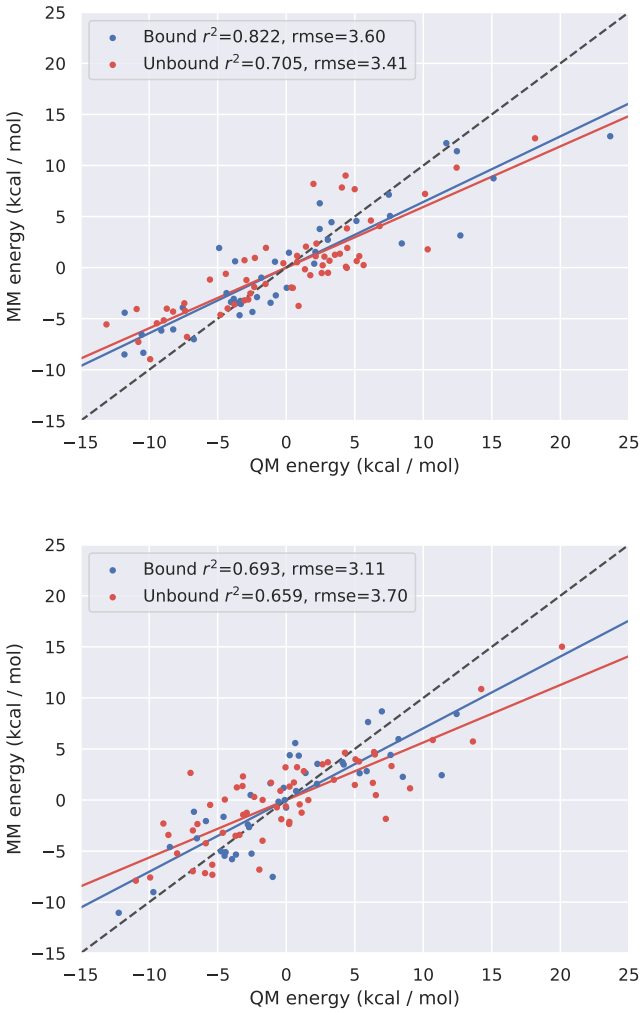

Figure S2. Comparison between QUBE and QM single point energies of structures of **3** (top) and **10** (bottom) extracted from bound and unbound (in water) MC simulations. The mean energies of each distribution have been shifted to zero. Also shown are the correlation (r^2) and root mean square errors (rmse, kcal/mol) between the two distributions.

Table S2. The correlation between the single point energies calculated using the QUBE FF and QM on structures extracted from MC simulations in the bound (protein-ligand complex in water) and unbound (ligand in water) states. Note that the correlation is relatively low for **14** in the bound state, but this appears to be due to the limited variability of structures, and hence energies, sampled.

	correlation (r^2)		rmse (kcal/mol)	
Compound	Bound	Unbound	Bound	Unbound
1	0.665	0.507	3.08	3.61
2	0.694	0.643	3.26	3.89
3	0.822	0.705	3.60	3.41
4	0.682	0.735	3.02	3.60
5	0.860	0.681	2.82	3.80
6	0.714	0.697	2.97	3.63
7	0.480	0.571	3.06	3.90
8	0.460	0.574	4.50	3.31
9	0.651	0.713	4.10	3.49
10	0.693	0.659	3.11	3.70
11	0.593	0.651	3.51	3.40
12	0.660	0.615	3.74	4.06
13	0.684	0.722	3.94	3.58
14	0.220	0.643	3.63	3.88
15	0.570	0.675	4.37	3.33
16	0.591	0.623	3.11	3.73
17	0.583	0.622	2.79	3.51
18	0.631	0.759	3.22	3.61
Average	**0.625**	**0.655**	**3.44**	**3.64**

S5 Analysis of QUBE and OPLS Binding Free Energies

Figure S3 shows the correlation between the QUBE and OPLS predictions of the binding free energies of the 17 inhibitors to p38α MAP kinase. Although both force fields have similar errors relative to experiment, as demonstrated by several statistical measures (Table S3), there are some quite large differences in individual predictions. For example, there are differences between QUBE and OPLS in excess of 2 kcal/mol in the computed binding free energies for compounds **2**, **12** and **13**. The latter two are perhaps not surprising given the sampling and force field difficulties discussed in the main text. Compound **2** has a F substituent at the R_3 position with QUBE non-bonded parameters: $q = -0.21$ e, $\sigma = 2.89$ Å, $\epsilon = 0.066$ kcal/mol. The corresponding OPLS/CM1A parameters are: $q = -0.08$ e, $\sigma = 2.90$ Å, $\epsilon = 0.060$ kcal/mol. The difference in the charge sets may be sufficient to explain the difference in binding prediction for compound **2**, but larger datasets involving fluorinated compounds will be required to investigate further.

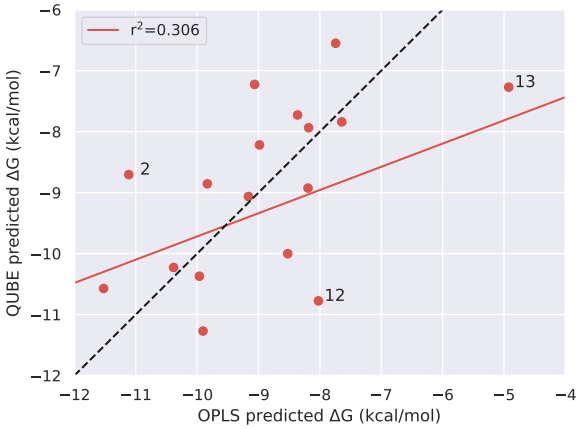

Figure S3. Correlation between QUBE and OPLS predictions of the binding free energy for the 17 inhibitors.

Table S3. Comparison between force field methods and experiment. Mean unsigned error (MUE, kcal/mol), root mean square error (RMSE, kcal/mol) and Spearman's rank correlation coefficient for each theoretical method are shown. OPLS data are taken from the previous literature.[16]

Force Field	MUE	RMSE	Spearman's rho
OPLS	0.88	1.30	0.46
QUBE	0.98	1.14	0.40

References

(1) Barreiro, G.; Kim, J. T.; Guimarães, C. R.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. From docking false-positive to active anti-HIV agent. *J. Med. Chem.* **2007**, *50*, 5324–5329.

(2) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26*, 1689–1700.

(3) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2009**, *113*, 13337–13346.

(4) Horton, J. T.; Allen, A. E.; Dodda, L. S.; Cole, D. J. QUBEKit: Automating the Derivation of Force Field Parameters from Quantum Mechanics. *J. Chem. Inf. Model.* **2019**, *59*, 1366–1381.

(5) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, .; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

(6) Allen, A. E.; Payne, M. C.; Cole, D. J. Harmonic force constants for molecular mechanics force fields via Hessian matrix projection. *J. Chem. Theory Comput.* **2017**, *14*, 274–281.

(7) Allen, A. E.; Robertson, M. J.; Payne, M. C.; Cole, D. J. Development and Validation of the Quantum Mechanical Bespoke Protein Force Field. *ACS Omega* **2019**, *4*, 14537–14550.

(8) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.* **2005**, *122*, 084119.

(9) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular force field parameterization via atoms-in-molecule electron density partitioning. *J. Chem. Theory Comput.* **2016**, *12*, 2312–2323.

(10) Tkatchenko, A.; Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

(11) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *P. Natl. Acad. Sci. USA* **2005**, *102*, 13749–13754.

(12) Wang, L.; Berne, B.; Friesner, R. A. On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities. *P. Natl. Acad. Sci. USA* **2012**, *109*, 1937–1942.

(13) Cole, D. J.; Tirado-Rives, J.; Jorgensen, W. L. Enhanced Monte Carlo sampling through replica exchange with solute tempering. *J. Chem. Theory Comput.* **2014**, *10*, 565–571.

(14) Cabeza de Vaca, I.; Qian, Y.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. Enhanced Monte Carlo Methods for Modeling Proteins Including Computation of Absolute Free Energies of Binding. *J. Chem. Theory Comput.* **2018**, *14*, 3279–3288.

(15) 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. *J. Phys. Chem. B* **2019**, *123*, 8675–8685.

(16) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Effects of water placement on predictions of binding affinities for p38α MAP kinase inhibitors. *J. Chem. Theory Comput.* **2010**, *6*, 3850–3856.