Implementation of the QUBE force field in SOMD for high-throughput alchemical free energy calculations

Lauren Nelson,[a†] Sofia Bariami,[‡] Chris Ringrose,[†] Joshua T. Horton,[†] Vadiraj Kurdekar,[†] Antonia S. J. S. Mey,[‡] Julien Michel,[‡] and Daniel J. Cole[*,†]

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

‡*EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, United Kingdom*

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

[a]L.N. and S.B. contributed equally to this work.

Abstract

The quantum mechanical bespoke (QUBE) force field approach has been developed to facilitate the automated derivation of potential energy function parameters for modelling protein-ligand binding. To date the approach has been validated in the context of Monte Carlo simulations of protein-ligand complexes. We describe here the implementation of the QUBE force field in the alchemical free energy calculation molecular dynamics simulation package SOMD. The implementation is validated by demonstrating the reproducibility of absolute hydration free energies computed with the QUBE force field across the SOMD and GROMACS software packages. We further demonstrate, by way of a case study involving two series of non-nucleoside inhibitors of HIV-1 reverse transcriptase, that the availability of QUBE in a modern simulation package that makes efficient use of GPU acceleration will facilitate high-throughput alchemical free energy calculations.

Introduction

Accurate prediction of protein-ligand binding affinity is invaluable in the early stages of drug discovery. High-throughput alchemical free energy calculations are an attractive tool for this task, enabling rigorous calculation of binding free energies. However, accuracy remains limited by the description of interatomic interactions and sampling of conformational space.[1–5] The potential energy surfaces of protein-ligand complexes are almost always described by molecular mechanics (MM) force fields, of which AMBER,[6] OPLS,[7] CHARMM[8] and GROMOS[9] are popular examples. These transferable biological force fields employ similar functional forms and their parameters are typically fit to reproduce the quantum mechanical (QM) and/or experimental properties of small organic molecules. Some alternative force fields, such as AMOEBA[10] and QMDFF,[11] have a greater focus on fitting to QM data, including symmetry-adapted intermolecular perturbation theory models,[12] but typically have a more complex functional form and their use in protein-ligand binding studies are not routine.

Recently, an alternative approach to biomolecular force field design has been proposed, named the QUantum mechanical BEspoke (QUBE) force field, in which virtually all force field parameters are derived specifically for the molecule under study directly from a small number of QM calculations. QUBE shares its functional form with the OPLS force field, and so is rapid to evaluate in the context of alchemical free energy calculations. Full details may be found elsewhere.[13,14] In brief, non-bonded (charge and Lennard-Jones) parameters of the QUBE force field are derived from atoms-in-molecule partitioning of the ground state QM electron density,[15,16] in particular, employing the Tkatchenko-Scheffler relations for van der Waals interactions.[17] QUBE bond and angle parameters are derived from the QM Hessian matrix, using the modified Seminario method,[18,19] while anharmonic dihedral parameters are fit to relaxed QM torsion scans.[13] Small molecule QUBE force fields may be derived using the QUBEKit python package, and they have been extensively validated against experimental liquid properties.[13] Atoms-in-molecule protocols are available as part of the ONETEP

linear-scaling density functional theory (DFT) software,[20] and hence QUBE non-bonded parameters may be readily derived for entire proteins comprising thousands of atoms. These parameters have been supplemented by compatible libraries of bonded parameters, and simulations of protein dynamics using the resulting QUBE force fields have been validated against experimental NMR observables.[14]

To date, the QUBE force field has been used to compute absolute binding free energies of a series of benzene derivatives to the L99A mutant of T4 lysozyme[21] and relative binding free energies of several flexible inhibitors of p38α MAP kinase.[22] In both cases, mean unsigned errors in protein-ligand binding free energies under 1 kcal/mol were reported. Such accuracy was shown to be broadly similar to that of the OPLS-AA force field in these cases.[21,22] In the initial development of QUBE, the OPLS functional form has been retained for compatibility with existing MM software (for example, these studies used the MCPRO software[23]). However, now that a baseline accuracy has been established, future development will target rapid and systematic evolution of the force field functional form. These strategies involve identifying key mappings between QM observables (such as the electron density) and force field parameters. Examples include automated addition of off-center charges to account for anisotropic electron density,[13] which has not yet been thoroughly explored in protein–ligand binding studies,[15] and higher-order dispersion terms to move beyond the dipole–dipole r^{-6} interaction.[24,25]

For rapid testing of new force fields, it is desirable to interface with free energy software that can be readily adapted to new force field functional forms whilst achieving efficient performance on modern computing hardware. Options for high-throughput alchemical free energy simulations include Schrödinger's commercial FEP+ package,[26] as well as AMBER[6] and GROMACS.[9] However it is not trivial to modify functional forms in such packages that feature highly optimized routines for efficient evaluation of forces. For this reason, here we implement QUBE in the Sire molecular simulation framework,[27] which includes the SOMD molecular dynamics engine for free energy calculations. SOMD interfaces with the OpenMM

toolkit for GPU acceleration, which provides efficient energy evaluation routines whilst allowing flexible definition of force field functional forms via a C++ API.[28] To aid the preparation of QUBE inputs, we make use of the BioSimSpace library, which already facilitates system setup and interoperability between a range of biomolecular simulation packages, such as AMBER, GROMACS, and CHARMM.[29] SOMD has been used within the Sire molecular simulation framework and successfully applied to alchemical free energy studies on a range of drug-like fragments, carbohydrates and host-guest systems.[30–40] The automated setup and processing of alchemical free energy calculations using Sire, BioSimSpace, SOMD and OpenMM has recently been implemented in Cresset's Flare package,[41] and benchmarked against 220 ligands bound to 14 protein targets, with accuracy comparable to previous reports.[3,9,26]

Thus, the SOMD framework provides a promising basis for implementation and future benchmarking of alchemical transformations using the QUBE force field. In this paper, we develop and distribute the file parsers that allow users to run QUBE simulations of protein-ligand complexes in the Sire and OpenMM molecular simulation frameworks. Following validation of the computational approach, we further demonstrate the use of the QUBE force field in alchemical protein-ligand binding free energy simulations using the protein target HIV-1 reverse transcriptase (RT) as a case study.

Computational Methods

Here, we present an overview of the QUBEKit/Sire workflow for molecular modelling (Figure 1), and to demonstrate its use, we have chosen as a case study, two sets of non-nucleoside inhibitors of HIV-1 RT (NNRTIs). HIV-1 RT has been extensively studied both computationally and experimentally, and indeed free energy calculations have been crucial to driving improvements in the potency of NNRTI scaffolds (further background and structural details are provided in the **Supporting Information**).[42–46] Here, we chose 10 compounds from a

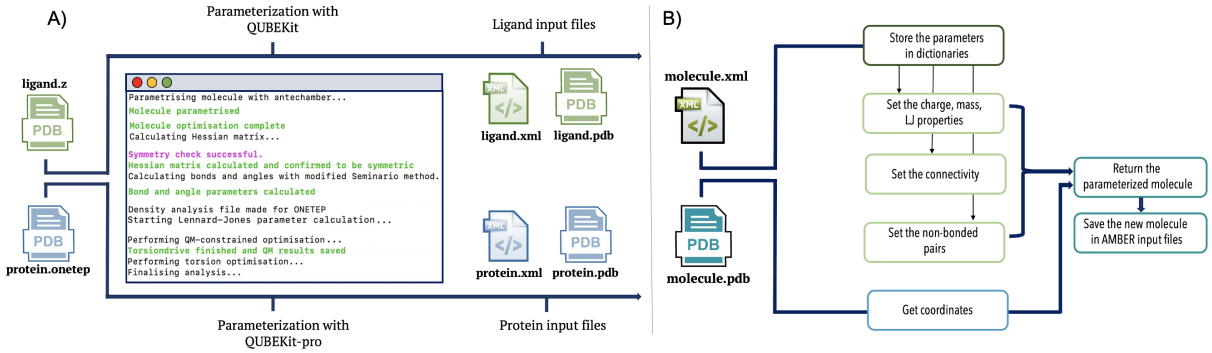

Figure 1: Overview of a) small molecule and protein parameterization using QUBEKit, and b) creation of the molecular systems in Sire for free energy calculations.

series of catechol diethers that incorporate a 7-cyano-2-naphthyl substituent,[43] and 12 compounds comprising indoles, indolizines, and benzofurans from the earlier literature,[42] which we refer to as groups 1 and 2, respectively (full chemical structures are provided in Figure 4 and Table 1). We focused on these two datasets as group 1 exemplifies small side chain transformations, whilst group 2 covers a much wider potency range, including heterocyclic substitutions.

Ligand Preparation with QUBEKit

QUBEKit[13] interfaces with the Gaussian09[47] and ONETEP[20] QM software packages to perform bond, angle, torsion, charge and Lennard-Jones parameter derivation. BOSS/MCPRO style z-matrices, and the corresponding PDB files, of the 22 NNRTIs were generated using the LigParGen web server.[48] Gaussian09 input files were prepared using QUBEKit. Structural optimizations and Hessian matrix calculations were performed with the ωB97X-D[49] functional and a 6-311++G(d,p) basis set. Harmonic bond stretching and angle bending parameters were derived using the modified Seminario method.[18,19] Non-bonded parameter derivation was performed with the linear-scaling density functional theory code, ONETEP,[20] using previously reported protocols.[13] Density derived electrostatic and chemical (DDEC) electron density partitioning, as implemented in ONETEP, was used to assign atom-centered point charges and atomic volumes.[13,16,50,51] Lennard-Jones parameters were

derived by QUBEKit from the atomic volumes using the Tkatchenko-Scheffler method.[15,17] Torsion parameter fitting on core fragments of the ligand sets followed the general methods used in previous studies,[13,22] though here we employ an interface between QUBEKit and the TorsionDrive package,[52] which improves the quality of both QM and MM torsion scans through its recursive wavefront propagation algorithm. Final force fields were output in xml format. Further descriptions of parameters used in force field derivation may be found in the **Supporting Methods**.

Protein Preparation with QUBEKit and Sire

Simulations for groups 1 and 2 were based on the initial x-ray crystal structures with PDB codes 5ter and 4mfb, respectively.[42,43] The proteins were truncated and underwent parameterization using the QUBE force field (**Supporting Methods**). ONETEP calculations were performed to obtain the ground state electron density, atomic charges and atomic volumes, in the same way as the ligands, as previously described.[22]

To facilitate set-up of QUBE simulations for proteins, the software QUBEKit-pro has been developed and is used here to generate OpenMM xml files from pdb and ONETEP output files. Using QUBEKit as a base has the advantage that most features can be easily applied to the protein, allowing for charge checking and symmetrization, for example. QUBEKit-pro builds the xml by first reading the PDB file to obtain the full topology of the protein or fragment. This provides atomic positions, connectivity, and the amino acid sequence. At this stage, certain groups of atoms are picked out for symmetrization such as hydrogen atoms on the same methyl or amine groups.

With the structure read in, a parametrization step is performed. A stored, general protein xml file contains bond, angle and torsion parameters, which have been generated specifically for compatibility with the QUBE force field[14] and are available for all atoms in standard residues, including NME/ACE caps. This general xml is used to map parameters to the protein in question, using the now stored structure. As described above for the ligands,

atomic charges and volumes are extracted from the ONETEP output file, symmetrized (if required), and used to calculate the Lennard-Jones parameters.[15] QUBEKit is then used to write pdb and xml files for use with OpenMM. Since each atom in the protein is in a unique environment, and therefore has unique charge and Lennard-Jones parameters, each atom in QUBEKit-pro is assigned a unique type.

Molecular systems in Sire can be loaded from AMBER-style input files. The molecule is described primarily with two files: the prmtop (or prm7) file that holds all the parameters and the inpcrd (or rst7) that contains the coordinates of the atoms. Functionality in Sire has been extended to include new features that support parsing of the QUBE pdb and xml input files and that support the OPLS-type potential energy functional form (Figure 1(b)). For the former, an algorithm was implemented that reads the parameters and the coordinates from the xml and pdb input files respectively, and returns formatted AMBER files (prm7/rst7) that contain exactly the same information and can be parsed by SOMD. Geometric combination rules have also been implemented in SOMD to support OPLS-style force fields. Further information and comparison of single point energies between OpenMM and SOMD are provided in the **Supporting Methods**.

Free Energy Simulations

BioSimSpace was used to generate files for free energy calculations of the ligands in bound and unbound states. Free energy simulations were run with SOMD using both QUBE and AMBER/GAFF force field parameterization methods for comparison. Perturbation maps for each HIV-1 RT data set were constructed manually (**Figure S3**). Each perturbation was carried out with two independent simulations, one in each direction. Each bound and unbound simulation was divided into 11 regularly spaced λ windows. The energy of the system was minimized for 1000 cycles by using the steepest descent method and each λ window was run for a total of 4 ns, using a time step of 2 fs. The first 5% of each simulation trajectory was discarded as equilibration. **Table S2** shows that the transformation of **2a** to

2c (and the reverse) in group 2 varies by less than 0.3 kcal/mol with respect to changes in simulation time (in the range 2–4 ns). Free energy changes were calculated from the output using MBAR, and processed by the software FreeEnergyWorkflows[30] to produce the free energy estimates, and associated errors, reported in this manuscript. Further details may be found in the **Supporting Information**.

Results

Validation: Hydration Free Energies

Having set up the computational workflow, and confirmed that single point energies of molecular systems are consistent between SOMD and OpenMM, we turn now to validation of free energies using the QUBE/Sire molecular system. A recent investigation of the reproducibility of (relative and absolute) hydration free energy calculations across contemporary molecular simulation packages (including SOMD and GROMACS) has provided benchmarks and package-specific protocols that serve as a validation test of new free energy procedures.[53] Using the same benchmark test set used in this previous study, we therefore test the agreement between absolute hydration free energies computed using the QUBE force field, as implemented in GROMACS and our new Sire interface. Small molecule input files were prepared using QUBEKit and Sire in the same way as the NNRTI compounds (Computational Methods). GROMACS force field files were generated from the corresponding AMBER input files using BioSimSpace. Free energy protocols were as previously described[53] (**Supporting Methods**) and a tutorial is provided (https://github.com/cole-group/QUBE-SOMD-paper/tree/main/HFE). Figure 2 compares the absolute hydration free energies for nine molecules, computed using SOMD and GROMACS. The mean absolute deviation in the computed hydration free energies is 0.15 kcal/mol which is consistent with the level of reproducibility reported between other packages.[53] Thus we can be confident of our implementation of QUBE in the Sire molecular simulation framework.

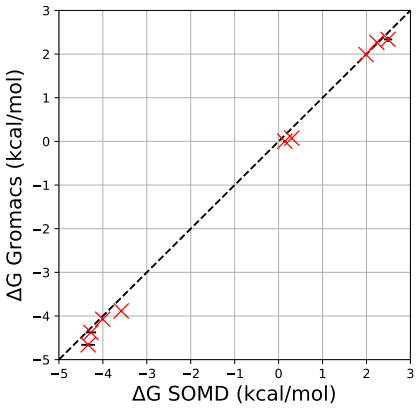

Figure 2: Comparison between absolute hydration free energies computed using the SOMD and GROMACS software packages. Error bars are standard errors in the mean from duplicate runs.

Case Study: Protein-Ligand Binding

To demonstrate the potential for the use of QUBE/SOMD in prospective drug discovery efforts, we analyze the relative alchemical binding free energies of a series of NNRTIs with HIV-1 RT. Crystal structures are available for compounds **1c**,[43] a catechol diether that incorporates a 7-cyano-2-naphthyl substituent, and **2a**,[42] where the 2-napthyl group is replaced by an indolizine. Figure 3 shows overlays of compounds **1c** and **2a** from SOMD simulations using the QUBE force field, with the corresponding crystal structures. In both cases, the cyano group projects out below Trp229 into a solvent-exposed channel, while close aryl-aryl contacts with Tyr188 and Trp229 are maintained. The orientation of residue Tyr181 shows sensitivity to the identity of the ligand, which is consistent with previous simulations[54] and crystallography.[42] In complex with compound **1c**, Tyr181 switches orientation during MD simulations, from a face-to-face interaction with the catechol diether ring of the ligand, to form closer contact with the 2-napthyl group. **Figure S4** reveals that Tyr181 is very flexible, forming both edge-to-face and face-to-face interactions throughout the simulation that more closely resemble the crystal structure. On the other hand, Tyr181 adopts a T-shaped stacking interaction with the central catechol ring of **2a**, which is stable during MD simulations and in agreement with the X-ray crystal structure. A hydrogen bond between one of the

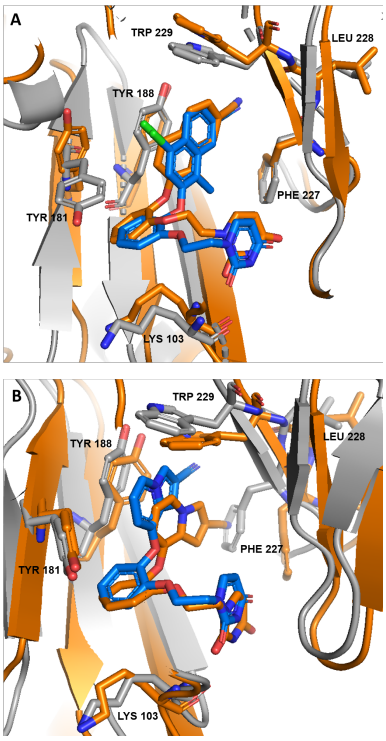

Figure 3: Overlay of **(A)** compound **1c** (orange) with the crystal structure (PDB: 5ter, grey/blue) and **(B)** compound **2a** (orange) and the crystal structure (PDB: 4mfb, grey/blue).

carbonyl oxygens of the uracil group with the backbone of Lys103 remains stable throughout both MD simulations.

Our implementation of the QUBE force field in SOMD now allows us to perform high-throughput alchemical free energy calculations to investigate the effects on binding of small changes to the ligand. A disadvantage of the chosen case study is that enzyme-based activities are unavailable for these compounds. Following previous studies,[54] we provide here EC_{50} data measured using human T-cells infected by wild-type HIV-1 (MT-2 cell assays) for comparison with our computational results. Although it has been demonstrated that cell- and enzyme-based activities are well-correlated for several NNRTI series,[55] we cannot guarantee that experimental data are not affected by, for example, differences in cell permeability between compounds. We therefore focus here on general observations and convergence studies, and refer the reader to previous work where we benchmark the accuracy of QUBE in protein–ligand binding.[21,22]

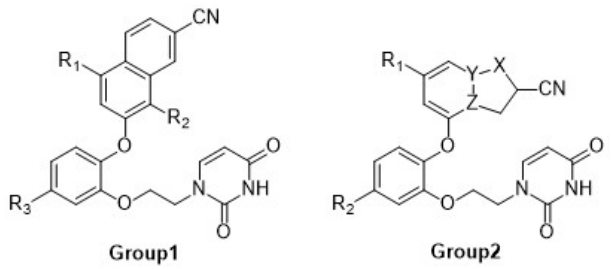

Figure 4: Markush structures of group 1 (2-naphthyl analogs) and group 2 (indole, indolizine and benzofuran analogs).[42,43]

Table 1: Experimental EC_{50} (nM) inhibitory activity in WT HIV-1 RT assays[42,43] and computed relative binding free energies ($\Delta\Delta G$ / kcal/mol) for the NNRTIs studied here. Uncertainties determined from cycle closure are shown in parentheses.[30]

	R_1	R_2	R_3	EC_{50}	$\Delta\Delta G$
1a	F	Cl	H	5.0	0.8 (0.2)
1b	F	Me	H	7.8	1.4 (0.2)
1c	Cl	Me	H	6.2	0.4 (0.3)
1d	Me	Cl	H	5.0	-0.2 (0.3)
1e	Me	Me	H	3.5	0.0
1f	Et	Me	H	6.0	-2.0 (0.8)
1g	Pr	Me	H	21.0	-3.0 (0.8)
1h	iPr	Me	H	16.0	-2.1 (0.8)
1i	Me	F	F	58.0	-3.3 (0.3)
1j	Me	Me	F	1.9	-2.4 (0.5)

	R_1	R_2	X	Y	Z	EC_{50}	$\Delta\Delta G$
2a	H	H	-	N	-	0.38	-0.3 (0.3)
2b	Me	H	-	N	-	0.9	0.1 (0.3)
2c	Me	F	-	N	-	2.0	0.0
2d	F	F	-	N	-	0.4	0.3 (0.1)
2e	H	F	-	N	-	2.7	0.3 (0.4)
2f	H	Cl	-	N	-	5.1	1.0 (0.3)
2g	H	F	-	-	N	17	4.0 (0.4)
2h	H	F	O	-	-	40	0.8 (0.4)
2i	Me	F	O	-	-	260	0.9 (0.4)
2j	H	H	NH	-	-	56	1.6 (0.4)
2k	Me	H	NH	-	-	10	3.2 (0.4)
2m	Cl	H	NH	-	-	340	2.3 (0.4)

Table 1 displays the binding free energies of the 2-napthyl ethers (relative to compound **1e**) computed using the QUBE force field, along with experimental assay data.[43] For comparison, we also report relative binding free energies computed with the AMBER force field, using the same free energy protocols (**Table S4**). Compounds **1a**–**1e** incorporate a series of small methyl and halogen transformations at the R_1 and R_2 positions. With R_2 = Me, the Me and Cl substitutions at R_1 are favored over F (**1e** < **1c** < **1b**) in both computational free energy studies with QUBE and in experimental assays. Experimentally, the two compounds (**1a** and **1d**) with R_2 = Cl are equipotent and less strongly bound than **1e**, while QUBE slightly favors compound **1d**.

It has been hypothesized that bulkier groups at the R_1 position in the 2-napthyl ethers might be accommodated, and confer extra benefit in the common Y181C mutant viral strain of HIV-1 RT.[43] Hence, compounds **1f**–**1h** were added to our benchmark data set. There is evidence of a change in binding mode of these bulkier ligands (**Figure S5**), particularly for **1g** and **1h**. Although we have investigated typical convergence of the free energies over simulation time scales of 2–4 ns (**Table S2**), it may be that significantly longer simulation times and/or enhanced sampling are required here.[56] The two compounds (**1i** and **1j**) with R_3 = F show high potency gains with the QUBE force field. **Figure S6** shows significant movement of the catechol diether group of compound **1j** towards the space between Tyr181 and Lys103, which represents a larger change in conformation than was observed for the other inhibitors. **Table S3** reports a series of repeat runs for several transformations involving compounds **1i** and **1j**. We do not observe very large differences between repeated runs, but do see large hysteresis between forward/reverse transformations. Across the group 1 set, the average hysteresis is 0.76 kcal/mol, which is dominated by transformations involving compounds **1i** and **1j**, as well as **1f** and **1h**, which further supports sampling inadequacies in these cases.

Similarly, Table 1 reports the relative binding free energies of each of the group 2 compounds, along with EC_{50} data from available assays.[42] Compounds **2a** to **2f** represent a series

of indolizines, with methyl and halogen substitutions in the R_1 and R_2 positions. Compounds **2a**, **2b** and **2d** are all sub-nanomolar inhibitors, and are predicted by our QUBE simulations to be essentially equipotent, while the drop in potency of compounds **2e** and **2f** is also reproduced. The isomeric indolizine (**2g**) was included in our test set to analyze the effects of more subtle electrostatic aryl-aryl interactions on binding. QUBE recapitulates the expected drop in potency, though (in agreement with Monte Carlo simulations performed with the OPLS force field[42]) it likely over-estimates the magnitude of this change. Finally, series of benzofurans (**2h** and **2i**) and indoles (**2j**–**2m**) were investigated. The drop in potency (relative to the indolizines) is successfully recovered, which is encouraging for future prospective heterocycle scans. The average hysteresis in forward/reverse transitions is just 0.41 kcal/mol for group 2, indicating satisfactory convergence.

Discussion and Conclusions

We have described an interface between the QUBEKit force field engine, and the Sire molecular simulation framework, for the calculation of alchemical relative binding free energies. As an example application, we have retrospectively analyzed the binding of 22 small molecule inhibitors of HIV-1 reverse transcriptase. With the significant caveat that EC_{50} is not a direct measurement of binding, it appears that the QUBE/SOMD interface is capable of providing a useful guide in design efforts. Where computational and experimental trends in binding are not in agreement, the simulations tend to be characterized by high hysteresis. We propose, in particular, that the present group 1 set may represent an interesting test case for enhanced sampling protocols.

Importantly, our bespoke parameter derivation methodologies offer the potential for substantial improvements in accuracy. By deriving as many force fields parameters as practical directly from QM, rather than fitting to experiment, the process of force field design becomes an exercise in accurately mapping QM data onto physically-motivated parameters

and functional forms.[13,57,58] In developing the QUBE non-bonded parameters for protein-ligand complexes such as these, only seven parameters have been directly fit to experiment (the van der Waals radii of seven elements), with the remainder being derived from QM. Thus, future parameter and functional form updates should be relatively straightforward to automate. Overall, this first generation interface between QUBE and SOMD offers a robust, adaptable platform, with access to GPU-accelerated dynamics, that will substantially improve our ability to validate and apply bespoke QM-derived force fields in computer-aided drug design.

Data and Software Availability

Sire (https://github.com/michellab/Sire), BioSimSpace (https://github.com/michellab/BioSimSpace) and QUBEKit (https://github.com/qubekit/QUBEKit) are freely available at Github. Data and tutorials accompanying this paper are freely available at: https://github.com/cole-group/QUBE-SOMD-paper.

Conflict of Interest Disclosure

J.M. is a current member of the Scientific Advisory Board of Cresset.

Acknowledgement

This work has been performed using the Cambridge Tier-2 system operated by the University of Cambridge Research Computing Service (http://www.hpc.cam.ac.uk) funded by EPSRC Tier-2 capital grant EP/P020259/1, as well as the Rocket High Performance Computing service at Newcastle University. DJC acknowledges financial support from a UKRI Future Leaders Fellowship (MR/T019654/1) and the EPSRC (EP/R010153/1).

Supporting Information Available

Background to HIV-1 RT, supporting methods, validation of single point energies, free energy mappings, supporting MD snapshots, convergence data, and AMBER free energies (PDF).

References

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

(2) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937.

(3) Song, L. F.; Merz, K. M. Evolution of Alchemical Free Energy Methods in Drug Discovery. *J. Chem. Inf. Model.* **2020**,

(4) Mey, A. S. J. S.; Allen, B.; Bruce Macdonald, H. E.; Chodera, J. D.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations. *Living J. Comp. Mol. Sci* **2020**,

(5) Schindler, C. E. M. et al. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *J. Chem. Inf. Model.* **2020**, *60*, 5457–5474.

(6) 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*, 3696–3713.

(7) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 3499–3509.

(8) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr., A. CHARMM General Force Field (CGenFF): A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comp. Chem.* **2010**, *31*, 671–690.

(9) 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.

(10) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.

(11) Grimme, S. A general quantum mechanically derived force field (QMDFF) for molecules and condensed phase simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497–4514.

(12) Rackers, J. A.; Wang, Q.; Liu, C.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. An optimized charge penetration model for use with the AMOEBA force field. *Phys. Chem. Chem. Phys.* **2017**, *19*, 276–291.

(13) 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.

(14) 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.

(15) 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.

(16) Manz, T. A.; Sholl, D. S. Improved atoms-in-molecule charge partitioning functional for simultaneously reproducing the electrostatic potential and chemical states in periodic and nonperiodic materials. *J. Chem. Theory Comput.* **2012**, *8*, 2844–2867.

(17) 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*, 6–9.

(18) Seminario, J. M. Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.* **1996**, *60*, 1271–1277.

(19) 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.* **2018**, *14*, 274–281.

(20) Prentice, J. C. et al. The ONETEP linear-scaling density functional theory program. *J. Chem. Phys.* **2020**, *152*, 174111.

(21) Cole, D. J.; Cabeza De Vaca, I.; Jorgensen, W. L. Computation of protein-ligand binding free energies using quantum mechanical bespoke force fields. *MedChemComm* **2019**, *10*, 1116–1120.

(22) Horton, J. T.; Allen, A. E.; Cole, D. J. Modelling flexible protein-ligand binding in p38α MAP kinase using the QUBE force field. *ChemComm* **2020**, *56*, 932–935.

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

(24) Manz, T.; Chen, T.; Cole, D. J.; Limas, N. G.; Fiszbein, B. New scaling relations to compute atom-in-material polarizabilities and dispersion coefficients: part 1. Theory and accuracy. *RSC Adv.* **2019**, *9*, 19297–19324.

(25) Visscher, K. M.; Geerke, D. P. Deriving Force-Field Parameters from First Principles Using a Polarizable and Higher Order Dispersion Model. *J. Chem. Theory Comput.* **2019**, *15*, 1875–1883.

(26) Wang, L. et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.

(27) Sire Molecular Simulation Framework. 2019; `http://siremol.org`.

(28) 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*, 1–17.

(29) Hedges, L.; Mey, A.; Laughton, C.; Gervasio, F.; Mulholland, A.; Woods, C.; Michel, J. BioSimSpace: An interoperable Python framework for biomolecular simulation. *J. Open Source Softw.* **2019**, *4*, 1831.

(30) Mey, A. S.; Jiménez, J. J.; Michel, J. Impact of domain knowledge on blinded predictions of binding energies by alchemical free energy calculations. *J. Comput. Aided Mol. Des.* **2018**, *32*, 199–210.

(31) Bosisio, S.; Mey, A. S.; Michel, J. Blinded predictions of host-guest standard free energies of binding in the SAMPL5 challenge. *J. Comput. Aided Mol. Des.* **2017**, *31*, 61–70.

(32) Granadino-Roldán, J. M.; Mey, A. S.; Pérez González, J. J.; Bosisio, S.; Rubio-Martinez, J.; Michel, J. Effect of set up protocols on the accuracy of alchemical free energy calculation over a set of ACK1 inhibitors. *PLoS ONE* **2019**, *14*, 1–20.

(33) De Simone, A.; Georgiou, C.; Ioannidis, H.; Gupta, J., A. A. Juárez-Jiménez; Doughty-Shenton, D.; Blackburn, E. A.; Wear, M. A.; Richards, J. P.; Barlow, P. N.; Carragher, N.; Walkinshaw, M. D.; Hulme, A. N.; Michel, J. A computationally designed binding mode flip leads to a novel class of potent tri-vector cyclophilin inhibitors. *Chem. Sci.* **2019**, *10*, 542–547.

(34) Papadourakis, M.; Bosisio, S.; Michel, J. Blinded predictions of standard binding free energies: lessons learned from the SAMPL6 challenge. *J. Comput. Aided Mol. Des.* **2018**, *32*, 1047–1058.

(35) Rodil, A.; Bosisio, S.; Ayoup, M. S.; Quinn, L.; Cordes, D. B.; Slawin, A. M.; Murphy, C. D.; Michel, J.; O'Hagan, D. Metabolism and hydrophilicity of the polarised 'Janus face' all-*cis* tetrafluorocyclohexyl ring, a candidate motif for drug discovery. *Chem. Sci.* **2018**, *9*, 3023–3028.

(36) Georgiou, C.; McNae, I.; Wear, M.; Ioannidis, H.; Michel, J.; Walkinshaw, M. Pushing the Limits of Detection of Weak Binding Using Fragment-Based Drug Discovery: Identification of New Cyclophilin Binders. *J. Mol. Biol.* **2017**, *429*, 2556–2570.

(37) Bosisio, S.; Mey, A. S.; Michel, J. Blinded predictions of distribution coefficients in the SAMPL5 challenge. *J. Comput. Aided Mol. Des.* **2016**, *30*, 1101–1114.

(38) Mey, A. S.; Juárez-Jiménez, J.; Hennessy, A.; Michel, J. Blinded predictions of binding modes and energies of HSP90- ligands for the 2015 D3R grand challenge. *Bioorg. Med. Chem.* **2016**, *24*, 4890–4899.

(39) Calabrò, G.; Woods, C. J.; Powlesland, F.; Mey, A. S.; Mulholland, A. J.; Michel, J. Elucidation of Nonadditive Effects in Protein-Ligand Binding Energies: Thrombin as a Case Study. *J. Phys. Chem. B* **2016**, *120*, 5340–5350.

(40) Mishra, S. K.; Calabró, G.; Loeffler, H. H.; Michel, J.; Koča, J. Evaluation of Selected

Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3333–3345.

(41) Kuhn, M.; Firth-clark, S.; Tosco, P.; Mey, A. S. J. S.; Mackey, M.; Michel, J. Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *J. Chem. Inf. Model.* **2020**, *60*, 3120–3130.

(42) Lee, W.-G.; Gallardo-Macias, R.; M. Frey, K.; A. Spasov, K.; Bollini, M.; S. Anderson, K.; L. Jorgensen, W. Picomolar Inhibitors of HIV Reverse Transcriptase Featuring Bicyclic Replacement of a Cyanovinylphenyl Group. *J. Am. Chem. Soc.* **2013**, *135*, 16705–16713.

(43) Lee, W.-G.; H. Chan, A.; A. Spasov, K.; S. Anderson, K.; L. Jorgensen, W. Design, Conformation, and Crystallography of 2-Naphthyl Phenyl Ethers as Potent Anti-HIV Agents. *ACS Med. Chem. Lett.* **2016**, *7*, 1156–1160.

(44) Thakur, V. V.; Kim, J. T.; Hamilton, A. D.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. Optimization of pyrimidinyl- and triazinyl-amines as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem.* **2006**, *16*, 5664–5667.

(45) Ruiz-Caro, J.; Basavapathruni, A.; Kim, J. T.; Bailey, C. M.; Wang, L.; Anderson, K. S.; Hamilton, A. D.; Jorgensen, W. L. Optimization of diarylamines as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem.* **2006**, *16*, 668–671.

(46) Jorgensen, W. L. Computer-aided discovery of anti-HIV agents. *Bioorg. Med. Chem.* **2016**, *24*, 4768–4778.

(47) Frisch, M. J. et al. Gaussian09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

(48) Dodda, L. S.; De Vaca, I. C.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server:

An automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017**, *45*, W331–W336.

(49) Chai, J. D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(50) Manz, T. A.; Sholl, D. S. Chemically meaningful atomic charges that reproduce the electrostatic potential in periodic and nonperiodic materials. *J. Chem. Theory Comput.* **2010**, *6*, 2455–2468.

(51) Lee, L. P.; Limas, N. G.; Cole, D. J.; Payne, M. C.; Skylaris, C. K.; Manz, T. A. Expanding the scope of density derived electrostatic and chemical charge partitioning to thousands of atoms. *J. Chem. Theory Comput.* **2014**, *10*, 5377–5390.

(52) Qiu, Y.; Smith, D. G.; Stern, C. D.; Feng, M.; Jang, H.; Wang, L. P. Driving torsion scans with wavefront propagation. *J. Chem. Phys.* **2020**, *152*, 244116.

(53) Loeffler, H. H.; Bosisio, S.; Duarte Ramos Matos, G.; Suh, D.; Roux, B.; Mobley, D. L.; Michel, J. Reproducibility of Free Energy Calculations across Different Molecular Simulation Software Packages. *J. Chem. Theory Comput.* **2018**, *14*, 5567–5582.

(54) Vilseck, J. Z.; Armacost, K. A.; Hayes, R. L.; Goh, G. B.; Brooks, C. L. Predicting Binding Free Energies in a Large Combinatorial Chemical Space Using Multisite λ Dynamics. *J. Phys. Chem. Lett.* **2018**, *9*, 3328–3332.

(55) Rizzo, R. C.; Udier-Blagović, M.; Wang, D.-P.; Watkins, E. K.; Kroeger Smith, M. B.; Smith, R. H.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of Activity for Nonnucleoside Inhibitors with HIV-1 Reverse Transcriptase Based on Monte Carlo Simulations. *J. Med. Chem.* **2002**, *45*, 2970–2987.

(56) Lim, N. M.; Wang, L.; Abel, R.; Mobley, D. L. Sensitivity in Binding Free Energies Due to Protein Reorganization. *J. Chem. Theory Comput.* **2016**, *12*, 4620–4631.

(57) Kantonen, S. M.; Muddana, H. S.; Schauperl, M.; Henriksen, N. M.; Wang, L. P.; Gilson, M. K. Data-Driven Mapping of Gas-Phase Quantum Calculations to General Force Field Lennard-Jones Parameters. *J. Chem. Theory Comput.* **2020**, *16*, 1115–1127.

(58) Spicher, S.; Grimme, S. Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angew. Chem. Int. Ed.* **2020**, *59*, 15665–15673.

Graphical TOC Entry