

QUBEKit: Automating the Derivation of Force Field Parameters from Quantum Mechanics

Joshua T. Horton,[†] Alice E. A. Allen,[‡] Leela S. Dodda,[¶] and Daniel J. Cole^{*,†}

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

[‡]*TCM Group, Cavendish Laboratory, 19 JJ Thomson Avenue, Cambridge CB3 0HE,
United Kingdom*

[¶]*Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, United
States*

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

Abstract

Modern molecular mechanics force fields are widely used for modelling the dynamics and interactions of small organic molecules using libraries of transferable force field parameters. For molecules outside the training set, parameters may be missing or inaccurate, and in these cases, it may be preferable to derive molecule-specific parameters. Here we present an intuitive parameter derivation toolkit, QUBEKit (QUantum mechanical BEspoke Kit), which enables the automated generation of system-specific small molecule force field parameters directly from quantum mechanics. QUBEKit is written in python and combines the latest QM parameter derivation methodologies with a novel method for deriving the positions and charges of off-center virtual sites. As a proof of concept, we have re-derived a complete set of parameters for 109 small organic molecules, and assessed the accuracy by comparing computed liquid properties with experiment. QUBEKit gives highly competitive results when compared to standard transferable force fields, with mean unsigned errors of 0.024 g/cm³, 0.79 kcal/mol and 1.17 kcal/mol for the liquid density, heat of vaporization and free energy of hydration respectively. This indicates that the derived parameters are suitable for molecular modelling applications, including computer-aided drug design.

Introduction

Complex biological processes such as protein-ligand binding,^{1,2} enzyme catalysis, and protein folding are often best understood when studied at the atomic scale which has driven an increase in the popularity of molecular mechanics (MM) and computational experiments. The ability of MM to model systems ranging in sizes from thousands to millions of atoms makes it indispensable across a wide range of sciences from biology to materials physics. The key to the general success of MM stems from the force field (FF) and its functional form, which allow the approximate description of the potential energy surface of a system as a simple function of its geometry.³ Ideally, we would like to study any atomic system at the most rigorous level of theory possible using quantum mechanics (QM). The explicit inclusion of electrons in QM is both a strength and a weakness. While QM does offer unprecedented accuracy, it is computationally expensive on the system sizes typically studied in biology, for example. Recent advances in linear-scaling density functional theory (DFT)⁴ and hardware has seen QM calculations on entire proteins become feasible,^{5,6} while QM calculations on small molecules have become routine,³ and offer a plethora of detailed information unobtainable through experiments such as electrostatic potentials and conformation-dependent energy surfaces. The widespread availability of these calculations presents an opportunity to develop new methods^{5,7-9} and tools¹⁰⁻¹⁴ which solely use QM data to derive FF parameters. This is particularly useful in non-standard regions of chemical space, as commonly seen in computer-aided drug design (CADD), which a state-of-the-art FF may struggle to accurately represent using transferable parameters assigned by chemical similarity.¹⁰

Transferable FFs such as GAFF (general AMBER FF),¹⁵ CGENFF (CHARMM general FF)¹⁶ and OPLS-AA¹⁷ are designed to be used in conjunction with their respective highly optimized and benchmarked biological FF counterpart. They are primarily used in simulating drug-like components of systems in CADD and give non-expert users the ability to parametrize highly diverse expanses of chemical space with very little computational cost. The requirement that a FF be transferable stems from two key points, 1) the parametriza-

tion process is a complex and error-prone task that is daunting to the inexperienced user, and 2) an attempt to parametrize all of chemical space would be inconceivable. It is therefore generally assumed that as long as a wide selection of chemical space is covered in the parametrization set then these results can readily be applied to new molecules. Each of the general FFs use libraries composed of thousands of pre-tabulated parameters,¹⁸ intensively fit to experimental and QM data for a set of small molecules that make up their training set. The parametrization goal of these particular FFs focusses on recreating experimental data concerning the condensed phase thermodynamic properties of small organic molecules, such as liquid densities, heats of vaporization and free energies of hydration.¹⁶ This parametrization philosophy follows sound logic as these properties describe the FF's ability to accurately characterize the non-bonded interactions that are also key in protein-ligand binding events.

While a great deal of thought is given to the constituents of the training set and the similarity assignment methodologies, users do not have to explore chemical space for long before they find missing parameters.^{10,19} This has led to the release of several software packages that offer to fill in gaps in their respective FFs for a molecule with missing or untested parameters. These methods often involve QM calculations that are used as reference data in order to fit a specific selection of parameters for the molecule to complete or improve the set. The wide range of methods developed now allows users to potentially re-parametrize entire molecules, however this is rarely done. Instead, these methods are used just to plug holes in the FF and replace as few parameters as possible. While this gap-filling technique has become widely used and often results in accurate parameters, their compatibility with the rest of the parameter set is questionable. This is mainly due to the highly interdependent nature of transferable FF parameters. Thus users of these methods should try to take these interdependencies into account whenever they modify a FF to ensure the overall accuracy of the parameter set is not affected. It should also be noted that efforts like ForceBalance¹⁴ and the Open Force Field Consortium²⁰ aim to expand the areas of chemical space that can be automatically parametrized via well-documented protocols, but ultimately there will always

be gaps.

One way around this would be to derive a system-specific (or “bespoke”) FF for each molecule studied, thereby removing any limitations associated with parameter transferability, and instead adopting a transferable FF derivation methodology akin to the semi-empirical models routinely used for charge derivation. As it is now becoming increasingly common for users to call upon additional software packages to at least optimize their parameter set, it is not unreasonable to suggest extending this to re-derive a new parameter set for the entire molecule. For example, charges and torsion parameters are often targeted for re-derivation and optimization from QM data,²¹ and these methods have thus become widely automated and routine. On the other hand, the remaining FF components (bond stretching, angle bending and Lennard-Jones (L-J) terms) are more difficult to automatically derive because they traditionally involve some form of fitting to experimental data.

However, with several recent developments in parameter derivation methodologies, we now have the means to derive all FF parameters directly from QM with minimal experimental fitting. The first of these is the modified Seminario method,^{7,22} which allows the derivation of bond stretching and angle bending force constants directly from the QM Hessian matrix and optimized geometry. This method has been shown to give highly accurate parameters that are able to reproduce QM vibrational frequencies with an average error of 6.3% for a test set of 70 molecules, which is slightly lower than that achieved by OPLS-AA (7.4%). Secondly, the atoms-in-molecule (AIM) method provides a means to partition the QM molecular electron density amongst the constituent atoms, and hence assign atom-centered partial charges, even for systems comprising many thousands of atoms.^{5,23} Furthermore, the partitioned atomic electron densities can also be used in conjunction with the Tkatchenko-Scheffler (TS) relations²⁴ to calculate all of the L-J parameters for a molecule. This method of using QM-derived non-bonded parameters has been shown to perform well in recreating liquid densities and thermodynamic properties when applied to a test set of 40 organic molecules.⁵ Collectively these methods form the basis of the QUantum mechanical BEspoke (QUBE)

FF.²⁵

Here, we present QUBEKit, a software toolkit designed to help users derive QUBE FF parameters in an intuitive and consistent way that minimizes parameter interdependency. This first iteration combines previously benchmarked QM derivation methods for non-bonded, bond stretching, angle bending and torsion parameters, using the same functional form as the OPLS-AA FF, and is freely available to the community via our Github page (<https://github.com/cole-group/QUBEKit>). Continuing a previous study that benchmarked the accuracy of the atoms-in-molecule non-bonded parameter derivation method, we re-derive parameters for a larger test set comprising over 100 small organic molecules using QUBEKit in a proof-of-concept workflow. We also expand upon the original non-bonded parameter derivation process by adding new fitting parameters that allow the derivation of FF terms for compounds containing bromine, as well as implementing a novel method for the derivation of off-center virtual site positions and charges directly from the QM electron density to model anisotropic electron densities. Combining these techniques we show through the use of the standard FF metrics described that the level of accuracy achievable with a QUBE FF is comparable to that of widely-used general transferable FFs. In this way, we provide the community with a tool for checking and refining parameter sets assigned through chemical similarity, and a starting point for FF improvements through optimization of the derivation protocols.

Theoretical background

FFs are traditionally described using bond-stretching, angle-bending, dihedral rotation, electrostatic and L-J contributions, as exemplified by the OPLS functional form:

$$U = \sum_{\text{Bonds}} \frac{k_r}{2} (r - r_o)^2 + \sum_{\text{Angles}} \frac{k_\theta}{2} (\theta - \theta_o)^2 + \sum_{\text{Pairs}} \frac{q_i q_j}{r_{ij}} + \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{\text{Dihedrals}} \left[\frac{V_1}{2} (1 + \cos(\phi)) + \frac{V_2}{2} (1 - \cos(2\phi)) + \frac{V_3}{2} (1 + \cos(3\phi)) + \frac{V_4}{2} (1 - \cos(4\phi)) \right] \quad (1)$$

The bond-stretching and angle-bending contributions require estimates of the force constants k_r and k_θ respectively as well as reference bond lengths (r_o) and angles (θ_o). The dihedral term is described as a four component cosine series with four corresponding parameters V_1 , V_2 , V_3 , V_4 , where ϕ is the torsion angle of the dihedral being described. The final term accounts for all non-bonded interactions between pairs of atoms separated by a distance r_{ij} . The standard Coulomb potential is used to calculate the interaction between two charges q_i and q_j . Finally the short-range repulsion and longer-range attractive van der Waals interactions are described using the L-J 12-6 potential. Here $A_{ij} = 4\epsilon_{ij}\sigma_{ij}^{12}$ and $B_{ij} = 4\epsilon_{ij}\sigma_{ij}^6$ where the ϵ and σ values of the L-J potential govern the energy well depth and minimum energy separation distance respectively. A complete set of parameters for any molecule described by this FF form requires the derivation of all the parameters of eq 1.

Traditionally each term has its own parameter fitting protocol and order that varies between FFs. Our derivation scheme follows this idea, breaking the problem down into individual tasks that have an order of best practice. In particular, we begin by calculating the stretching and bending terms, followed by non-bonded, and finally the dihedral parameters. Next, we shall discuss the motivation behind the derivation and optimization methods combined in QUBEKit.

Bond and Angle Parameters

For each bond and angle in our molecule, we require a force constant and equilibrium value in order to describe the internal energy contribution associated with the vibrational motion. It has been noted that, to describe all of the basic atom type combinations in GAFF, some 20,000 angle parameters would be required.¹⁵ Such large parameter libraries are commonplace with OPLS3 containing 15,236 angle-bending parameters, with a continuing effort to expand this list as new chemistries are encountered.¹⁸ To generate these parameters, general FFs have to use a wide range of reference data combining experiment and QM. QM data actually already play a role in the derivation of the majority of the transferable parameters

in these FFs due to the lack of experimental data available for unique chemical species and the ease of generating accurate QM data on-the-fly. While many of the equilibrium terms are collected from x-ray crystallography and NMR studies of small molecules, some have to be determined from QM predicted minimum energy structures.¹⁵⁻¹⁸ Force constants are then manually fit in an iterative process which aims to recreate the QM vibrational frequencies using an initial guess for the other required parameters as described in the development of CGENFF¹⁶ and AMBER.¹⁵ While this method is effective, it does create interdependencies in the FF parameters as the force constants are dependent on the rest of the original parameter set, meaning that ideally all parameters should be continually updated in a self-consistent fashion until convergence is reached.¹⁶

Instead, we have adopted the modified Seminario method for deriving bond and angle force field parameters. With this method, force constants are calculated directly from QM data and interdependencies between different components of the FF are eliminated, while maintaining low errors in the description of the frequencies of QM normal modes.⁷ The standard Seminario method derives force constants directly from the QM Hessian matrix²² and has been incorporated into specialized FF fitting tools for metal complexes such as the VFFDT plugin,²⁶ or in the MCPB.py²⁷ program which is part of AmberTools. This method estimates force constants by projecting the decomposed forces felt by an atom due to the displacement of a neighboring atom onto their mutual bond vector.²² However, this method results in undesirably stiff force constants due to the double counting of angle bending contributions in larger molecules, as outlined in ref 7. The modified method, however, accounts for an atom's chemical environment and has been shown to recreate QM vibrational frequencies with a low average error of 6.3% across all vibrational modes for a wide range of molecules.⁷ The ability to accurately derive the bonded parameters directly from the QM Hessian matrix without the need for initial parameter guesses clearly simplifies the procedure for non-expert users by removing sources of human error and also speeds up the process making it suitable for automation. We shall also show that the derived force constants

retain a low percentage error in recreating QM vibrational frequencies when combined with the rest of the QUBE FF.

Non-Bonded Parameters

The non-bonded interactions incorporate multiple QM effects, such as electrostatics, induction, dispersion and exchange-repulsion, through effective non-bonded Coulombic and L-J interactions. In fixed point charge models there are many methods to derive partial charges from high-level QM calculations using a mixture of population analysis techniques, but ultimately no unique solution. While *ab initio* calculations yield high-quality charges they are often disregarded as being too computationally expensive and are substituted by a variety of semi-empirical QM based methods. These methods allow the rapid assignment of charges and are heavily parametrized in order to reproduce charges observed at higher levels of theory. For example, GAFF employs Mulliken charges produced from semi-empirical Austin Model 1 (AM1) calculations²⁸ that are then subject to bond charge corrections (BCC) to better recreate experimental hydration free energies.^{29,30} The resulting electrostatic potential is then comparable to that calculated at the HF/6-31G* level which was used to parameterize the AMBER restrained electrostatic potential (RESP) charges.¹⁵ OPLS-AA, on the other hand, uses Cramer-Truhlar CM1A³¹ charges, and recently also included an AM1-BCC inspired localized BCC version of the OPLS-AA/CM1A FF that is available through the LigParGen server.³²⁻³⁴ It should also be noted that as these semi-empirical QM calculations are performed in vacuum they have to be modified to include polarization effects to make them suitable for condensed phase modelling. This is often performed via the inclusion of the BCC mentioned in the case of GAFF and OPLS, and/or in the form of charge scaling factors all of which are only used on neutral molecules.

On the other hand, CGENFF relies heavily on *ab initio* calculations. CGENFF I charges can be first assigned by a similarity search through a library of parametrized fragments or can be derived using MP2/6-31G(d) Merz-Kollman charges.¹⁶ With either starting guess,

the charges are subsequently optimized by fitting to QM-calculated scaled interaction energies at the HF/6-31G(d) level between the molecule and water in a variety of conformations. Again we note the choice of low-level theory, this an artefact from the initial derivation of the CHARMM additive FF, to ensure any new parameters are compatible with the biological CHARMM terms. Importantly this means the overall charge description is compatible between systems that require a mix of transferable and biological FFs.

Computational cost is also kept to a minimum in standard transferable FFs by assigning the L-J parameters from a library of pre-fit parameters. This has become standard practice across transferable FFs, with OPLS3 containing 124 different atom types so far, and many general FFs borrowing terms from their biological counterparts.^{15,18} The L-J potential parameters are often tuned to accurately recreate experimental liquid properties.^{10,16,17,35} While this technique works very well for atoms covered in the original parameterization, more atom types often have to be introduced to account for new chemical environments. During the optimization of the GAMMP/GAFF-LJ* parameters, for example, it was found that for a test set of 430 compounds the 41 standard atom types of GAFF were restricting the maximum achievable accuracy of the FF. The performance was then substantially increased with the addition of 11 new atom types, reducing the average unsigned relative error in the heat of vaporization from 17.9% to 5.9%.¹⁰ Clearly increasing the number of atom types will help increase the overall accuracy of a FF as new exceptions to current atom types arise, logically this implies that system-specific FF parameters have the potential to lead to an overall more accurate FF.

The QUBE FF follows this QM-based philosophy by deriving both L-J parameters and AIM charges from a single ground state QM electron density. The AIM partitioning method divides the total molecular electron density ($n(\mathbf{r})$) into approximately spherical, uniform overlapping atomic densities ($n_i(\mathbf{r})$) via:

$$n_i(\mathbf{r}) = \frac{w_i(\mathbf{r})}{\sum_k w_k(\mathbf{r})} n(\mathbf{r}) \quad (2)$$

The weighting factor $w_i(\mathbf{r})$ is determined by the choice of AIM partitioning method, in our case the density derived electrostatic and chemical charges (DDEC)^{36,37} scheme is employed. This method iteratively optimizes the weighting factor to resemble the spherical average of $n_i(\mathbf{r})$ and the density of a similar reference ion using a mixture of iterative Hirshfeld (IH) and iterative Stockholder atoms (ISA).^{5,36} The charges are then found by integrating the atomic electron density over all space:

$$q_i = z_i - N_i = z_i - \int n_i(\mathbf{r})d^3\mathbf{r} \quad (3)$$

Where N_i is the number of electrons associated with atom i and z_i is the nuclear charge. The electron density, as outlined in Ref 5, is calculated as the direct solution of the inhomogeneous Poisson equation in a medium with a dielectric constant $\epsilon = 4$. It was found that “half-polarizing” the molecule with a low dielectric constant resulted in non-bonded terms that are suitable for condensed phase modelling. Including polarization in this manner allows us to avoid parametrizing any BCC or charge scaling factors as employed by CGENFF, OPLS/CM1A and OPLS/CM5.³⁸

Additionally, the QUBE FF employs the TS method to derive the A_{ij} and B_{ij} terms of the FF in equation 1 by rescaling reference free atom data, proportionally to AIM electron densities.²⁴ The dispersion coefficient B_i is estimated as:

$$B_i = \left(\frac{V_i^{AIM}}{V_i^{free}} \right)^2 B_i^{free} \quad (4)$$

The atomic volume is readily calculated from the same AIM partitioned electron density as used in charge assignment via:

$$V_i^{AIM} = \int r^3 n_i(\mathbf{r})d^3\mathbf{r} \quad (5)$$

The B_i^{free} coefficients are computed using time-dependent density functional theory (TDDFT)

calculations on free atoms in vacuum.³⁹ V_i^{free} is the reference volume of the atom calculated using the MP4(SDQ)/aug-cc-pVQZ method in Gaussian 09⁴⁰ and the chargemol code⁴¹ for each of the elements in our model and these values can be found in Table S1 of the Supporting Information. To ensure that the dispersion and repulsion coefficients result in a minimum in the L-J potential close to the van der Waals radius of the atom, it can be shown that the A_i coefficient can be approximated by:

$$A_i = \frac{1}{2}B_i(2R_i^{AIM})^6 \quad (6)$$

Here we found the AIM effective radius R^{AIM} of each atom by rescaling the reference free atom radius using the TS method:

$$R_i^{AIM} = \left(\frac{V_i^{AIM}}{V_i^{free}} \right)^{1/3} R_i^{free} \quad (7)$$

The only fitting parameters in our model are the eight free atom radii (R_i^{free}), one for each of the elements studied so far (H, C, N, O, F, S, Cl, Br). This version sees the addition of a bromine parameter that was fit in the same spirit as the rest, that is empirically tuning the free atom radii to recreate liquid properties of a selection of bromine-containing molecules (Table S1). A full description of the non-bonded parameter derivation methods can be found in Ref 5.

Anisotropy

While atom-centered point charges provide a good representation of the QM electrostatic potential (ESP) if the partitioned atomic electron density is spherical, in many cases this simple representation is inadequate.⁵ This situation occurs when there is significant anisotropy in the underlying electron distribution, and is common in molecules containing nitrogen, sulfur or halogens.⁴² Here, to model electron anisotropy, we employ off-center, “virtual” sites,

which have been shown to be competitive with the use of more computationally expensive higher-order multipole electrostatics.⁴⁵ Virtual sites are commonly used in water models, such as TIP4P,⁴³ and various force fields for modelling lone pairs and σ -holes.⁴⁴ In contrast to those studies and in keeping with our goal of avoiding fitting FF parameters to experiment, in Ref. 5, we proposed a method that relied on the dipole and quadrupole moments of the partitioned atomic electron density, to optimize the charges and locations of virtual sites. However, the method employed did not consistently converge and resulted in a large number of off-center point charges. Modifications were required to correct these issues and improve the usability of the method in an automated high-throughput scenario. Here, we describe a new method for the derivation of virtual site positions and charges directly from the QM electron density. The virtual sites are positioned so as to reproduce as closely as possible the QM ESP of the partitioned atomic electron density. In order to reduce the search space we limit the virtual site positions to those dictated by the symmetry of the atom’s bonding environment. Together these improvements allow us to define virtual sites that improve the electrostatic properties of the simulated molecule in an automated manner.

The QM ESP (Φ_i^{ref}) is calculated from the partitioned atomic electron density ($n_i(\mathbf{r})$). This is advantageous as the method may be applied equally well to both surface and buried atoms. The ESP is taken at a series of points on sets of spheres with radii between 1.4-2.0 times the van der Waals radius of the atom. The error $F(\Phi, \Phi^{ref})$ is given by:

$$F(\Phi, \Phi^{ref}) = \sum_{i=1}^M \frac{|\Phi_i - \Phi_i^{ref}|}{M} \quad (8)$$

where M is the number of sampling points. The MM ESP (Φ_i) is calculated as:

$$\Phi_i = \sum_{j=1}^N \frac{q_j}{4\pi\epsilon_0 r_{ij}} \quad (9)$$

where N is the number of sites on an atom, r_{ij} is the distance from the site to the sampling point and q_j is the charge on site j . An additional threshold parameter (F_{thresh}) was required

to distinguish between atoms that required extra sites and those that did not. Above this threshold the anisotropy method is used, below the threshold, no off-center charges are added. As well as this, extra charges are only added when there is a reduction in error which is controlled by a second parameter (F_{change}).

One Additional Off Center Charge

For atoms with ESP error above the threshold, we begin by attempting to model the anisotropy using a single off-center charge. The vectors for one additional off-center point charge that preserve symmetry are shown in Fig. 1. The vector direction is governed by the number of atoms bonded to the atom exhibiting anisotropy:

1. *One bond.* The atom A (which exhibits anisotropy) has one neighbor, atom B. The vector along which the extra charge is positioned is $\mathbf{r}_1 = \lambda_1 \mathbf{r}_{AB}$, where \mathbf{r}_{AB} is a vector between atom A and atom B and λ_1 is to be determined.
2. *Two bond.* The atom A has two neighbors, atoms B and C. The vector for the extra charge is $\mathbf{r}_1 = \lambda_1 (\mathbf{r}_{AB} + \mathbf{r}_{AC})$, which is along the bisector of the two bond vectors.
3. *Three bond.* The atom A has three neighbors, atoms B, C and D. The vector for the extra charge is $\mathbf{r}_1 = \lambda_1 (\mathbf{r}_{AB} - \mathbf{r}_{AC}) \times (\mathbf{r}_{AD} - \mathbf{r}_{AC})$, which makes an equal angle with all three bond vectors.

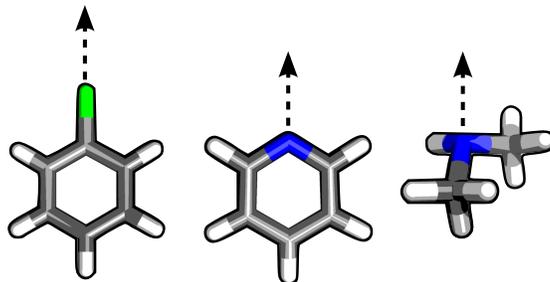


Figure 1: The directions along which off-center point charges are placed for an atom with one, two or three bonds.

After the vector is assigned, the optimal position along the vector and the charge of the off-center point is determined. This is carried out using a grid search of parameters to find the values which best recreate the QM ESP. Assigning a symmetry-derived search direction reduces the number of variables that need to be optimized from four (the x, y, z coordinates and the charge) to two (the distance along the vector and the charge). This simplification is particularly important when multiple off-center point charges are added, as described in the following section. The atom-centered point charge is assigned a value such that the net charge of the atom is unchanged. The method is summarized with a flowchart in Figure S1.

Multiple-Off Center Charges

In Ref. 5, it was often necessary to add more than one off-center point charge to recreate the anisotropy seen in the QM ESP. Therefore, our approach was extended to add multiple charges. Again, the method depends on the number of atoms bonded to the atom exhibiting anisotropy:

1. *One bond.* A second off-center charge is placed along the same vector, $\mathbf{r}_2 = \lambda_2 \mathbf{r}_{\mathbf{AB}}$.
2. *Two bonds.* If two extra point charges are used, the original vector is a line of symmetry. The two charges are then placed in the same plane as the vectors that point from the atom to the neighboring atoms, $\mathbf{r}_{1,2} = \lambda_{\parallel}(\mathbf{r}_{\mathbf{AB}} + \mathbf{r}_{\mathbf{AC}}) \pm \lambda_{\perp}(\mathbf{r}_{\mathbf{AB}} + \mathbf{r}_{\mathbf{AC}}) \times (\mathbf{r}_{\mathbf{AB}} \times \mathbf{r}_{\mathbf{AC}})$, or perpendicular to this plane, $\mathbf{r}_{1,2} = \lambda_{\parallel}(\mathbf{r}_{\mathbf{AB}} + \mathbf{r}_{\mathbf{AC}}) \pm \lambda_{\perp}(\mathbf{r}_{\mathbf{AB}} \times \mathbf{r}_{\mathbf{AC}})$. An example is shown in Fig. 2. A third extra charge can also be added and is placed along the bisector $\mathbf{r}_3 = \lambda_3(\mathbf{r}_{\mathbf{AB}} + \mathbf{r}_{\mathbf{AC}})$.
3. *Three bonds.* A second off-center charge is placed along the same vector, $\mathbf{r}_2 = \lambda_2(\mathbf{r}_{\mathbf{AB}} - \mathbf{r}_{\mathbf{AC}}) \times (\mathbf{r}_{\mathbf{AD}} - \mathbf{r}_{\mathbf{AC}})$. An exception is made for primary amine groups with the second off-center charge placed along the bisector of the NH_2 angle $\mathbf{r}_2 = \lambda_2(\mathbf{r}_{\text{NH}_1} + \mathbf{r}_{\text{NH}_2})$. This is necessary as the regions between the nitrogen and hydrogen atoms exhibit anisotropy in ESP.

A disadvantage of using the partitioned electron density to calculate the QM ESP is that it includes regions that are not accessible during MM simulations, such as between bonds. This is the case for the amine group and results in other regions of the QM ESP not being adequately reproduced. The addition of an off-center site between the nitrogen and hydrogen atoms helps to overcome this issue.

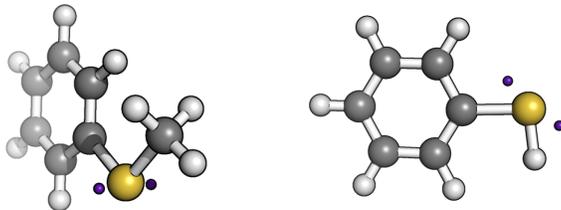


Figure 2: An example of off-center charge placement for the case (left) perpendicular to the plane of the bond vectors or (right) in the plane of the bond vectors.

Torsional Parameters

The final stage in the fitting procedure is the optimization of torsional parameters. Torsional parameters are an important factor controlling the conformational preference of a molecule due to their association with QM stereoelectronic effects, and are therefore often a target for re-optimization.^{8,18,46-51} In this work, we follow the standard procedure of fitting the parameters to minimize the difference between MM and QM constrained one dimensional torsional scans. In particular, we aim to fit the four V_n parameters of the OPLS FF torsion potential shown in equation 1 by automating the scheme outlined in Ref. 47 into QUBEKit with some additional considerations. The steepest descent algorithm is employed to find the torsional parameters that minimize the regularized Boltzmann weighted error function:

$$\Omega = \sqrt{\frac{\sum_{i=1}^n (E_{MM}^i - E_{QM}^i)^2 e^{-E_{QM}^i/k_B T}}{n}} + \lambda \sum_{torsions} \sum_{j=1}^4 |V_j^{ref} - V_j| \quad (10)$$

Where E_{QM} and E_{MM} are the QM and MM optimized energies at each sampled torsional angle, k_B is the Boltzmann constant, T is a temperature weighting factor, n is the number of

sampling points and V_j^{ref} is a reference torsional parameter. Overfitting is often a concern at this point in the fitting process. Here, we introduce a regularization function controlled by a variable parameter λ , which constrains the fitted torsional parameters to be close to the reference values, V_j^{ref} . In this work, V_j^{ref} were taken from the OPLS force field, but could also be set to zero.⁵² It is also important to note that it is not possible to always perfectly recreate the entire QM PES hence users should concentrate on relatively low energy regions as these are most likely to be sampled during room temperature simulations. The weighting temperature T can be adjusted to preferentially weight the low-energy regions of the QM potential energy surface.

In molecules containing multiple flexible dihedral angles, it was found that torsional parameters were best fit in an order that started with rotations that would involve the movement of the fewest number of atoms. For example, a long chain molecule with no repeated dihedral types would be best fit by starting at the ends and working inwards. Larger molecules could also be fragmented during fitting to reduce the computational cost of the fitting procedure. It also should be noted that we do not derive any improper torsion terms in this workflow instead borrowing them from the OPLS-AA FF.

Computational Implementation

QUBEKit has been designed as a python command line toolkit with simple, intuitive commands allowing the user to perform three main tasks: 1) writing QM input files for atoms-in-molecule, Hessian matrix and torsional scan calculations, 2) derivation of MM parameters from the results of the QM calculations, and 3) the output of the parameters in widely-used MM topology and force field files. The only required inputs are the molecule’s structure and some initial reference parameters, which are included in a BOSS style z-matrix, which is freely available via the LigParGen web server (<http://zarbi.chem.yale.edu/ligpargen/>).³²⁻³⁴ The use of a z-matrix at this point makes defining and choosing the dihedral angle to be optimized conceptually straightforward. In this first version of QUBEKit, QM calculations are

currently performed using the Gaussian09⁴⁰ and ONETEP⁴ software packages. QUBEKit interfaces with the BOSS⁵³ molecular simulation package to perform all MM tasks, including torsional fitting and vibrational mode analysis, and incorporates code from the modified Seminario method⁷ to calculate the bond and angle parameters. Future support for additional open source MM and QM software packages is planned. The derived parameters can be written in a variety of MM package formats such as BOSS/MCPRO style (z-matrices, .sb and .par structural and force field files), OpenMM .xml files, and GROMACS .gro and .top files. A full description of the workflow (Figure S3) used in this project with a list of commands can be found on our Github page (<https://github.com/cole-group/QUBEKit>) alongside a tutorial.

Computational Methods

Quantum Mechanical Calculations

All Gaussian09 input files were prepared using QUBEKit, which takes PDB files and the corresponding BOSS/MCPRO style z-matrices generated using the LigParGen web server as input. All optimization routines and frequency calculations used for the bond stretching and angle bending terms were performed with the ω B97X-D⁵⁴ functional using the 6-311++G(d,p) basis set and vibrational scaling factor of 0.957.⁷ Torsional constrained optimizations were performed in Gaussian09⁴⁰ with the same functional and basis set so as to be consistent with the other bonded terms. The torsional scan optimizations were performed in 15° increments from 0° to 360°. The majority of the dihedral parameter fitting was done using no Boltzmann weighting (corresponding to $T=\infty$) and regularization against OPLS reference values was applied with $\lambda = 0.1$. This was only changed in rare cases where it was particularly difficult to recreate the QM energy landscape, in which case $\lambda = 0$ and $T = 2000K$ was used as previously suggested.⁴⁷

Ground-state electron density calculations for non-bonded parameter derivation were performed using the linear-scaling DFT code ONETEP.⁴ Four nonorthogonal generalized

Wannier functions (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 sine (psinc) basis, with a grid size ($0.45a_o$), corresponding to a plane wave cut-off energy of 1020 eV. The PBE exchange-correlation functional was used with PBE OPIUM norm-conserving pseudopotentials.⁵⁵ The calculation was carried out in an implicit solvent using a dielectric of 4 to model induction effects.^{5,56,57} The DDEC module implemented in ONETEP was used to partition the electron density and assign atom-centered point charges and atomic volumes.^{23,58} The charges were assigned with a IH to ISA ratio of 0.02. The ESP error threshold F_{thresh} , was set to 0.9025 kcal/mol. The additional charges are only added if the decrease in ESP error is larger than $F_{change} = 0.0625$ kcal/mol. The locations of the virtual sites were restricted using maximum distance cut-offs chosen by element, as virtual sites near the van der Waals radius can be detrimental. The cut-offs were defined as follows: 0.8Å for N, 1Å for O, S and F, and 1.5Å for Cl and Br.

Pure Liquid Simulations

Pure liquid simulations were performed using OpenMM⁵⁹ with a custom non-bonded potential to describe the mixing rules and 1-4 interactions employed by the OPLS (and QUBE) FF. The required XML files were generated using QUBEKit with extra sites included automatically using the local coordinate site construction function in OpenMM. All extra sites were modelled as individual particles with their own 1-4 interactions, with their only bond being to the parent atom. A plot showing the agreement in single point energies calculated using the correction implemented in OpenMM and the BOSS software can be found in Figure S2. Instructions on how to perform the single point energy check for a new molecule using QUBEKit can be found at the Github (<https://github.com/cole-group/QUBEKit/wiki>) wiki page along with other examples and tutorials.

Simulations were performed in the isothermal-isobaric (NPT) ensemble at 1 atm and comprised 267 molecules in a periodic cubic box. A non-bonded cut-off was used with inter-

molecular interactions being truncated at distances based on molecular size and smoothed over the last 0.5 Å.³³ Long range non-bonded interactions were calculated using the Particle-Mesh-Ewald (PME) method,⁶⁰ with a 0.0005 tolerance error while also applying a long-range correction to the system’s energy. Following minimization of the initial configuration, three nanosecond simulations were run for each molecule using a 1 fs time step. The first nanosecond was treated as equilibration. The liquid and corresponding gas-phase simulations were run at 25°C or the molecule’s boiling point if it was lower. The resulting densities and heats of vaporization were averaged over 2000 data points collected in the production part of the run. The heats of vaporization were computed using equation 8 in Ref. 61. Following their recommended protocol, we employed Langevin dynamics temperature regulation with a collision frequency of 5 ps⁻¹. The pressure was regulated using a Monte Carlo barostat as implemented in OpenMM. Examples of the scripts used for both the liquid and gas simulations along with input files for all molecules in the study can be found in the SI. The uncertainties were found to be less than 0.003 g/cm³ and 0.02 kcal/mol for densities and heats of vapourization respectively. Graphs showing the convergence of the properties with simulation length can also be found in Figures S4 and S5.

Free Energies of Hydration

Free energies of hydration were calculated using GROMACS⁶² due to its ability to include extra sites during alchemical perturbation. All input files were generated using QUBEKit which writes OPLS FF style GROMACS .top and .gro files. The virtual sites were all constructed by hand using the simplest method available for each molecule, with a connection being added between the site and parent to again make the 1-4 interaction lists consistent with OpenMM and BOSS. Each molecule of the test set was annihilated from a cubic box containing approximately 1500 TIP4P water molecules using a two-step approach over 21 λ -windows, first turning of the charges followed by the L-J terms. The solute-solvent non-bonded interactions were switched off via coupling to the λ reaction parameter using soft-core

potentials with settings $\alpha = 0.5$, $p = 1$ and $\sigma = 0.3$.⁶³ The charges were decoupled using λ values of (0.00 0.25 0.50 0.75 1.00) and van der Waals using λ values of (0.00 0.05 0.10 0.20 0.30 0.40 0.50 0.55 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00). The simulations were again run in the NPT ensemble at 1 atm and 25°C. All solvent-solute and solvent-solvent non-bonded interactions were truncated at 10 Å and smoothed over the last 0.5 Å. PME was used with a long-range correction applied to the total energy and pressure. Each λ -window was run using Langevin dynamics and a two femtosecond time step with bonds involving hydrogen constrained using the LINCS algorithm.⁶⁴ The starting configurations at each λ -window were first minimized before being equilibrated twice. The first was a 100 ps run in the canonical ensemble (NVT) followed by a 200 ps run in the NPT ensemble. Finally, the production stage ran for 1 nanosecond and the free energy of hydration was calculated using Bennett’s acceptance ratio as implemented in the GROMACS BAR module.⁶⁵ All uncertainties for the calculations were found to be less than 0.3 kcal/mol.

Results and Discussion

Condensed Phase Properties

A common measure of the quality of FF parameters for use in biomolecular simulations is a comparison of the predicted condensed phase properties of molecules simulated using the FF with experiment. These properties, such as liquid density, the heat of vaporization and free energy of hydration, can be calculated routinely due to low sampling requirements, thus making FF inaccuracies the main contributor to any differences between the computed data and experiment. In this study, we have chosen a benchmark dataset comprising 109 small organic molecules, which are representative of the key functional groups commonly observed in biology and drug design. Importantly most of the molecules used in the set are also part of the training data used during the parametrization of many of the general transferable FFs mentioned, including the OPLS/1.14*CM1A-LBCC FF, which allows for

direct comparison of the FFs. Figure 3 shows the results of the condensed phase property

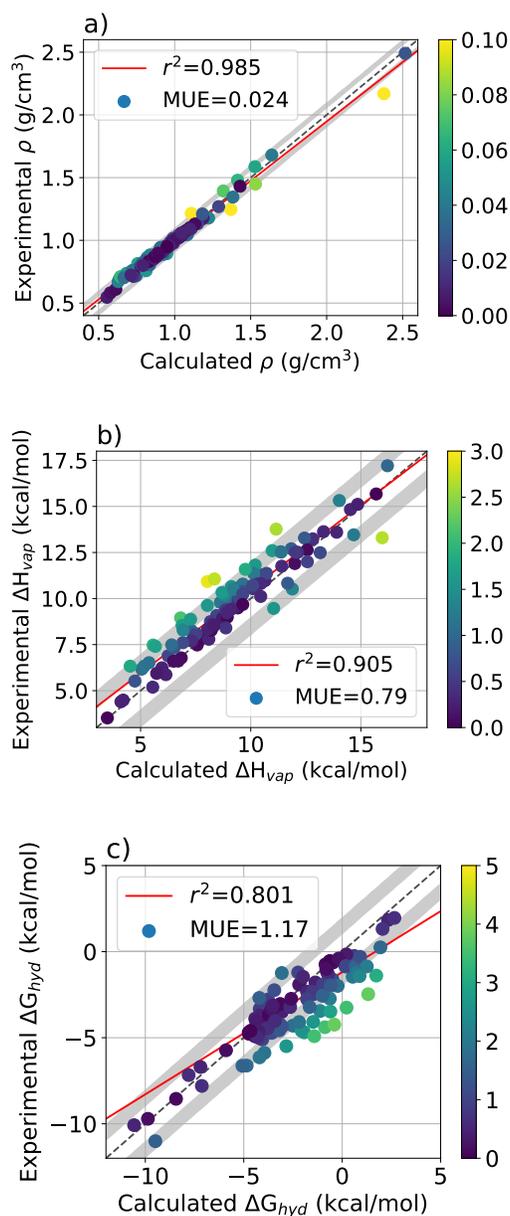


Figure 3: Force field liquid property metrics (a) liquid density, (b) heat of vaporization (c) free energy of hydration. Calculated for the organic molecule test set using QUBE FF parameters. MUE compared to experiment and r^2 correlation are also included.

calculations for the test set where experimental data is available, along with the correlations and mean unsigned errors (MUE), while Table 1 compares the latter with some examples of widely-used transferable FFs.³³ The average errors in the density and heat of vaporization (0.024 g/cm³ and 0.79 kcal/mol, respectively) indicate that QUBE performs extremely well

Table 1: Mean unsigned errors between calculated liquid properties and experiment for various FF parameter sets.

Force field	ρ (g/cm^3)	ΔH_{vap} ($kcal/mol$)	ΔG_{hyd} ($kcal/mol$)
OPLS/1.14*CM1A ³³	0.024	1.40	1.26
GAFF/AM1-BCC ³³	0.039	1.31	0.94
OPLS/CM5 ³³	0.024	1.06	0.94
OPLS/1.14*CM1A-LBCC ³³	0.024	1.40	0.61
DDEC/OPLS ⁵	0.014	0.65	1.03
QUBE (this work)	0.024	0.79	1.17

in the prediction of pure liquid properties, that is despite only using eight fitting parameters in the derivation of non-bonded parameters (the van der Waals radii of the elements H, C, N, O, S, F, Cl, Br used in this study). The general transferable force fields are of similar accuracy, despite being extensively parametrized against data sets similar to these. As we have found previously,⁵ hydration free energies are more difficult to predict (MUE 1.17 kcal/mol). This could be due to limitations in the functional form, particularly the neglect of an explicit polarization term, in describing the transfer of a molecule between low dielectric (vacuum) and high dielectric (water) media. However, the largest outliers in Figure 3(c) are for apolar molecules with a low (less negative) free energy of hydration, for which QUBE under-estimates their solubility. This is particularly problematic for molecules containing benzene rings, and may indicate an imbalance between dispersive and electrostatic contributions to hydration when QUBE is used in combination with a standard transferable water model (TIP4P).

Another potentially problematic group of compounds are aliphatic alcohols as we found the 10 in our test set to have a relatively high MUE (1.27 kcal/mol) in hydration free energy. The poor description of alcohol groups was also previously found to be a trait of the OPLS/CM1A FF.^{33,66} Ref 66 commented on the charges assigned to the head group of 1-octanol by OPLS/CM1A as shown in table 2. They found that scaled CM1A charges were too positive, resulting in the poor prediction of densities and heats of vaporization as shown in table 3. To tackle problematic groups such as these, the OPLS/1.14*CM1A-LBCC

parametrization was developed which adds a systematic bond charge correction to various functional groups and was fit to better reproduce experimental free energies of hydration.³³ In the case of the aliphatic alcohols, the correction transfers a $0.1e^-$ charge to the oxygen of the head group from the neighboring carbon atom as can be seen in table 2. Thus with the same L-J parameters, we see the density, heat of vaporization and free energy of hydration are subsequently improved for 1-octanol as shown in Table 3 along with the values obtained by the QUBE FF. This same BCC was also found to reduce the MUE for the hydration free energy from 1.95 to 0.43 kcal/mol for 32 aliphatic alcohols in the development of the LBCC parameters.³³ Importantly the fitted correction scheme gives roughly the same charge as our AIM partitioning method which demonstrates the successful inclusion of polarization into our charges at the point of derivation rather than via subsequent corrections. We also observe similar σ values between our QUBE FF and OPLS which is reassuring considering OPLS is extensively fit to reproduce liquid properties. While the ϵ values do differ noticeably, it has been found that liquid property predictions can be greatly improved with the systematic tuning of this parameter.⁶¹ However, this would not be compatible with the philosophy of a QM derived FF and future work will instead investigate modifications to the FF functional form.

Finally, it should be noted that there is an increase in the MUE of each of the properties computed using the QUBE FF compared with our original benchmark study (Table 1), which

Table 2: The non-bonded parameters for the head group oxygen in 1-octanol are shown for a variety of FF and charge combinations. The LigParGen server was used to parameterize the OPLS variants, and Antechamber for GAFF with QUBE coming from this work.

Force field	charge	σ	ϵ
OPLS/1.14*CM1A	-0.588	3.120	0.170
OPLS/1.14*CM1A-LBCC	-0.687	3.120	0.170
GAFF/AM1-BCC	-0.598	1.721	0.210
QUBE	-0.673	3.129	0.127

used AIM-derived non-bonded parameters in combination with OPLS bonded parameters (DDEC/OPLS).⁵ This is likely the result of the expanded test set used here as on further inspection of the data concerning only the same molecules that were included in the original benchmark we find the MUEs to be 0.017 g/cm³, 0.59 kcal/mol and 1.08 kcal/mol for the density, heat of vaporization and free energy of hydration respectively, which are very similar to the original values. We therefore, conclude that bonded parameters, while crucial to the conformational preferences of larger molecules, are not too important in the description of the liquid properties of small molecules.

With the inclusion of larger molecules and molecules that contain multiple functional groups, the increase in overall error of the liquid properties is to be expected if we consider the accuracy on a per functional group basis. This effect is exemplified by the case of o-chloroaniline, which has unsigned errors in ΔH_{vap} of 2.61 kcal/mol and in ΔG_{hyd} of 3.49 kcal/mol. By way of comparison, the smaller molecules aniline and chlorobenzene showed unsigned errors in ΔH_{vap} of 1.63 and 1.17 kcal/mol and in ΔG_{hyd} of 2.66 and 1.67 kcal/mol, respectively. This should be kept in mind when applying QUBE (and other force fields) to the study of, for example, absolute protein-ligand binding free energies for larger organic molecules containing multiple functional groups.

Table 3: The liquid properties of 1-octanol predicted using different FF and charge parametrization methods are shown compared with experiment.

Force field	ρ (g/cm ³)	ΔH_{vap} (kcal/mol)	ΔG_{hyd} (kcal/mol)
OPLS/1.14*CM1A	0.807	15.201	-1.26
OPLS/1.14*CM1A-LBCC	0.809	16.038	-3.12
GAFF/AM1-BCC	0.834	20.354	-3.12
QUBE	0.793	16.206	-2.19
Experiment	0.822	17.208	-4.09

Bond, Angle and Dihedral Parameters

As discussed in the previous section, it appears that the bonded parameters have little effect on the accuracy of liquid properties. However, given the importance of torsional parameters in determining conformational preferences of larger molecules, and bond and angle parameters in modelling molecular vibrations, which are important for example in photochemistry applications, we examine the properties of the derived parameters here in more detail.

The first point to note is that by deriving bond and angle parameters directly from the QM Hessian matrix, there is no possibility of missing parameters in the QUBE FF. In contrast, even for this small test set, we found one missing bond parameter and six missing angle parameters using a standard transferable FF. The QUBE predicted values for these terms along with the OPLS atom types are shown in tables S2 and S3. In practice, these parameters would be inferred from similar atom types or re-parameterized by the user, which may introduce inaccuracy. QUBE allows the user to rapidly and automatically derive all necessary parameters with no compromise in accuracy. In this study, the QUBE FF maintains a low mean percentage error in MM vibrational frequencies of 6.5% (MUE of 54 cm^{-1}), which is very similar to the values initially reported reaffirming the wide-scale applicability of the method.⁷

Given the widespread use of transferable bond and angle parameters, it is worth analyzing to what extent these parameters vary in our benchmark test set. Figure 4 plots the range of QUBE bond lengths and force constants for all atoms defined with CT-CT bond types in the test set, and compares them with the OPLS parameters. Further plots like this for all bonds and angles that are present in at least ten of the molecules in the test set can be found in Figures S7-S28. As reported previously,⁷ the modified Seminario method gives bond-stretching force constants that are on average lower than their OPLS counterpart. The QUBE parameters typically span a range of around 0.05 \AA and $100 \text{ kcal/mol/\AA}^2$ for the bond length and force constant respectively, indicating that use of a single average, transferable

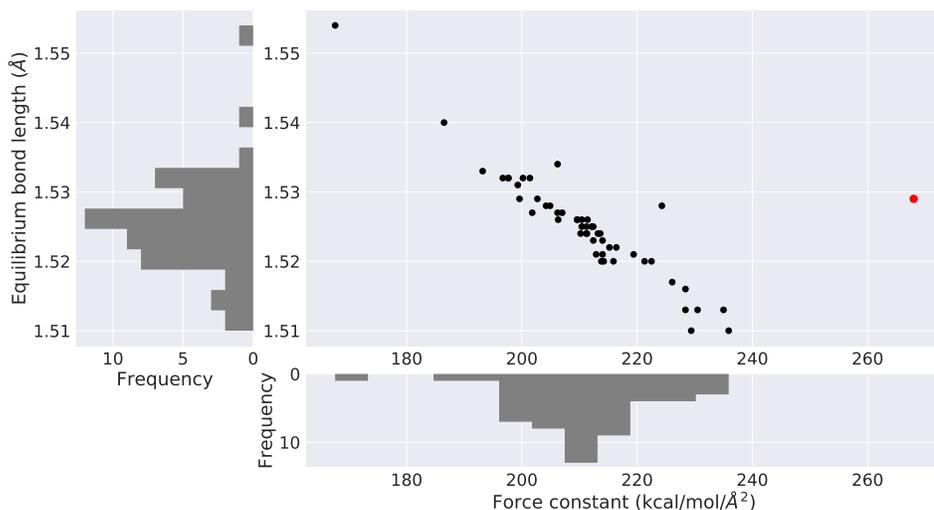


Figure 4: A common bond type is analyzed by comparing the QM predicted equilibrium bond length to the associated derived force constant of each molecule they appear in for the CT-CT bond type. The OPLS parameters are shown in red.

value should not introduce significant error. Interestingly, there is a negative correlation between force constant and equilibrium bond length, supporting the use of bond length to infer force constants in early studies.⁶⁷ These results indicate that it may be possible to derive more explicit algorithms for ‘learning’ force field parameters directly from the molecular geometry. We also envisage QUBE parameters as providing a reasonable starting point for optimization if further fitting to QM potential energy surfaces is desired.¹⁴

Torsional parameters, like the bond and angle parameters, were derived separately for each molecule. Due to the use of virtual sites, we found that parameters were often not transferable between similar molecules, and those that were such as methyl group rotations remained close to the initial OPLS parameters. The overall accuracy of the torsional scan fitting was very good when regularization was used and only a handful of molecules with poor predicted energy surfaces required the setting to be switched off. A sample of torsion fitting data taken directly from the QUBEKit output can be found in the Figures S29-S31 with the overall error and regularization error bias where appropriate. We have also included the dihedral parameters for every molecule in the test set in the SI along with an analysis in

which we have grouped the torsions together based on their OPLS atom types. We envisage these as forming the basis of a community led library to be used to replace or speed up future fitting efforts.

Extra sites

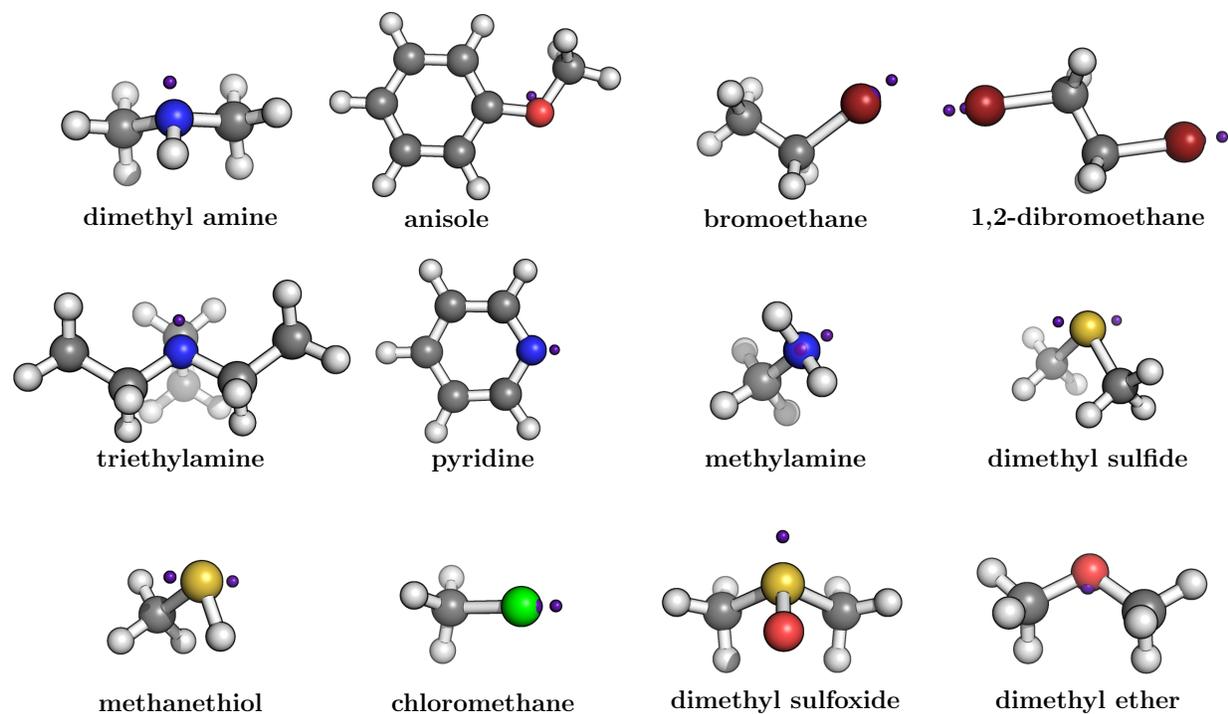


Figure 5: A selection of 12 molecules from the benchmark test set with their extra sites depicted as purple spheres. Charges and positions of the extra sites were derived from the partitioned atomic electron density.

To test the effect of the additional off-center point charges, the liquid properties for the benchmark test set were also calculated in the absence of extra sites. This led to a general worsening of the results with the MUEs becoming 0.023 g/cm³, 0.85 kcal/mol and 1.51 kcal/mol in the density, the heat of vaporization and free energy of hydration respectively (Figure S6). As expected, since it is governed mostly by Lennard-Jones interactions, the error in the density remained approximately constant. However, the decline in accuracy of the other properties indicates that modelling of anisotropy in electron density is required to accurately describe intermolecular interactions. This is consistent with the increasing use of

virtual sites in multiple FFs.^{18,68}

While there is no unique way to derive virtual site parameters, it would seem that deriving the parameters to minimize the error in ESP for an individual atom is effective. Figure 5 shows a selection of molecules from the test set that required virtual sites, the rest of the derived site positions and corresponding charges can also be found in the SI. Here we can see that the derived positions are chemically intuitive, with σ -holes and lone-pairs well-represented. In total 50 of the 109 molecules in the test set required at least one virtual site, and on average a molecule whose functional group ESP error is initially above the chosen threshold requires 2.1 virtual sites. While this is more than is typical in molecular mechanics simulations, the computational cost of virtual sites in an MD simulation is small.⁴⁵ Furthermore, QUBEKit substantially simplifies the process for the user by deriving the virtual site parameters from QM and writing them to simulation-ready input files.

Some molecules with large ESP errors were not assigned off-center virtual sites. Chlorobenzene, for example, was found to have a large ESP error on the Cl atom just below the set threshold of 0.90 kcal/mol. However, the resulting liquid property predictions were not significantly affected, as shown in Table S4. Methanol was another example of a molecule that was not assigned virtual sites despite having an ESP error of 1.50 kcal/mol, which is above the threshold. After performing the grid search it was found that the addition of virtual sites did not substantially reduce the ESP error of the oxygen atom by the required amount F_{change} . This was the case for all aliphatic and aromatic alcohols in the test set which could also contribute to the poor performance of alcohols overall.

Test case: 3-hydroxypropionic acid

While the molecules in the test set represent many of the functional groups often used in CADD, they contain many fewer rotatable dihedral bonds and functional groups than a typical drug-like molecule. Thus following previous work investigating the use of QM derived FF parameters we have used QUBEKit to derive a QUBE FF for 3-hydroxypropionic acid (3-

HA).⁸ The molecule shown in figure 7 incorporates carboxyl and hydroxyl functional groups, has been identified as a potentially useful agent for organic synthesis and is also a surrogate for a typical fragment scaffold. Chen et al. used QM-based fitting techniques to derive the bonded parameters for the molecule from a series of single point energy calculations, with the L-J terms being taken from AMBER and the partial charges assigned according to the CHelpG scheme⁵⁴ in the GAMESS QM software.

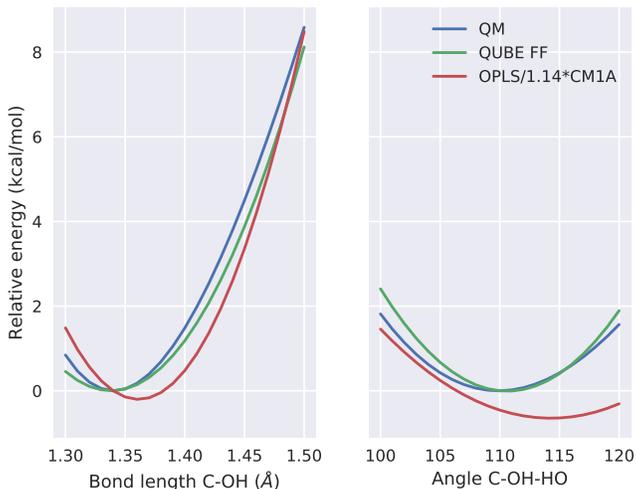


Figure 6: Comparison of the calculated relative single point energies using QM, OPLS and QUBE for C-OH bond-stretching and C-OH-HO angle-bending motions in 3-HA.

Here we compare the QUBE FF to OPLS for 3-HA, giving a performance overview of the derivation versus transferable techniques. Figure 6 compares the two force fields with QM single point calculations. Since we compute the bond and angle force constants in a one-off calculation directly from the QM Hessian matrix, with no iterative fitting, it is not obvious how accurate they will be in reproducing QM conformational energetics when combined with the rest of the QUBE FF parameters. However, Figure 6 reveals that the QUBE FF reproduces extremely well, not only the QM minimum energy conformations, but also describes small changes in these same bond lengths and angles. This is also well replicated across all calculated vibrational modes for the molecule with an average percentage error of 6.7% compared to the QM vibrational frequencies.

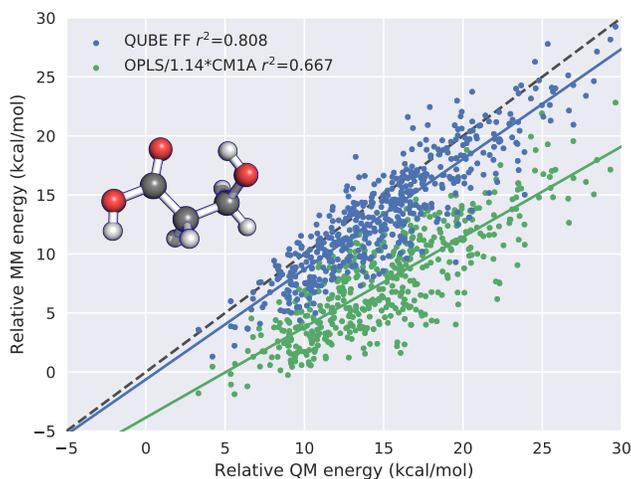


Figure 7: Comparison between the relative QM and MM energies using the QUBE FF and OPLS for 500 conformations extracted from a MD simulation of 3-HA which is shown as an inset.

Finally, with the goal of evaluating the ability of QUBE to recreate intramolecular energetics including torsional rotations, a liquid simulation of 3-HA solvated in a box containing 1000 TIP4P water molecules was performed. We then extracted 500 conformations from the simulation and computed the relative single point energies of each snapshot of the 3-HA molecule using OPLS, QUBE and QM (with the same functional and basis sets as used for the parameter derivation). Figure 7 shows the correlation between the relative MM and QM energies for 3-HA. Compared to OPLS, the correlation is improved from 0.667 to 0.808. Thus, despite the high flexibility of 3-HA, QUBE is not only able to reproduce the minimum energy structure, but also sample physically reasonable structures in liquid simulations, which is encouraging for future use in computer-aided drug design.

Conclusions

With the spread of low-cost computing and access to automated software,^{10–12,14} it is becoming increasingly common for users to perform parameter set optimization prior to running molecular mechanics simulations. However, this optimization is typically limited to the charge and torsional parameters, for which well-established protocols for fitting to QM data

exist. In this paper, we present the QUBEKit software for automated derivation of virtually all force field parameters required to model the dynamics of small organic molecules. Namely, we bring together methods for charge and Lennard-Jones parameter derivation from atoms-in-molecule analysis of the QM electron density, and a method for deriving bond and angle force constants directly from the QM Hessian matrix. We also include a novel method for off-center virtual site derivation along with an implementation of a standard one-dimensional torsion fitting scheme.

Overall, we achieve mean unsigned errors of 0.024 g/cm³, 0.79 kcal/mol and 1.17 kcal/mol in the prediction of liquid densities, heats of vaporization and free energies of hydration for a benchmark set of 109 molecules, compared to experiment. Hence, we achieve comparable accuracy to standard, transferable force fields. Importantly, however, to describe all molecules in the benchmark data set, we have *only fit 8* parameters to experimental data (the van der Waals radii of eight elements in vacuum). This reduction in empiricism has two key advantages. Firstly, it has the potential to substantially simplify the FF fitting process, since the parameters come directly from QM and do not rely on the collection of experimental fitting data, which is time-consuming for small molecules, and is rarely done for larger molecules. Secondly, the ease of FF design presents the opportunity to derive new protocols and even move beyond the standard functional form of the FF. Opportunities for FF improvement include i) update of the atoms-in-molecule partitioning scheme,⁶⁹⁻⁷³ ii) the introduction of more rigorous descriptions of van der Waals interactions,⁷⁴⁻⁷⁶ iii) inclusion of explicit polarization, and iv) a more accurate functional form for the short-range repulsion.^{74,77} Such efforts would typically require significant re-parameterization of the FF libraries, but could potentially require only a few lines of code in QUBEKit.

One example of the iterative improvement of FF design protocols, is the addition in this paper of a new method for off-center virtual site parameter derivation for the modelling of anisotropic electron density. Compared to our previous method,⁵ the parameter derivation process is faster and more user-friendly. By deriving the virtual site charges and positions

from the molecular symmetry and partitioned atomic electron density, we do not require any experimental data for fitting. Furthermore, since the bond, angle and Lennard-Jones parameter derivation methods are independent of the charge derivation, we can trivially add extra sites without substantially altering the force field. Notably, the mean unsigned error in the free energies of hydration of our benchmark set increases to 1.51 kcal/mol if virtual sites are not included. QUBEKit writes the virtual site positions in OpenMM .xml file format for ease-of-use.

In contrast to our previous work,⁵ we have supplemented the atoms-in-molecule non-bonded parameters with molecule-specific bonded parameters derived from the QM Hessian matrix and torsional scans. In agreement with our previous study,⁷ we showed that the so-called modified Seminario method is able to reproduce QM normal mode vibrational frequencies to high accuracy (6.5% here). Closer examination of bond and angle force field parameters for widely used atom types reveals that bond and angle parameters are reasonably transferable between closely-related molecules. Such analyses of more complex molecules could be used to identify problems with standard force fields where bonded parameters may require re-fitting or the inclusion of more atom types. In addition, we have shown that for the molecule 3-HA, QM relative energies of an ensemble of structures are modelled well with the QUBE FF when combined with torsional fitting. It should be noted that torsional fitting is the major computational expense in QUBE (since it requires a constrained QM optimization at each torsion angle), and methods to reduce this expense are under investigation. In this regard, we have provided a library of all of the torsional parameters derived in this study, and these could be used as initial estimates for approximate dynamics, or as an initial guess in the fitting process. Improper torsional parameters are not derived in this study, and we have used those from the OPLS FF here.

We have provided with this paper the QUBEKit software toolkit, tutorials and data sets (<https://github.com/cole-group/QUBEKit>). This first version utilizes the BOSS molecular mechanics software,⁵³ and Gaussian09⁴⁰ and ONETEP⁴ QM packages for parameter deriva-

tion. Simulation-ready files are output in OpenMM, Gromacs or BOSS format. Future work will widen the choice of available software, particularly for parameter derivation. In parallel, we are also releasing the QUBE protein force field,²⁵ which employs the same parameter derivation techniques, alongside a torsion library for the twenty naturally-occurring amino acids. Together these tools will allow users to derive compatible and accurate QUBE force fields for both proteins and small molecules for use in computer-aided drug design applications.

Acknowledgement

This research made use of the Rocket High Performance Computing service at Newcastle University. The authors acknowledge financial support from EPSRC grant EP/R010153/1 (DJC) and the EPSRC Centre for Doctoral Training in Computational Methods for Materials Science under grant EP/L015552/1 (AEAA).

Supporting Information Available

QUBEKit can be downloaded and installed from our Github page (<https://github.com/cole-group/QUBEKit>). Free atom reference data, anisotropy workflow, single point BOSS and OpenMM energy comparison, QUBEKit workflow, N-(1-methylethyl)-2-propanamine and N,N-dimethylaniline liquid density and heat of vaporization property convergence, liquid density and thermodynamic property predictions excluding extra sites, QUBE predicted values for missing bond-stretching and angle-bending terms, QUBE predicted bond type analysis, QUBE predicted angle type analysis, some example dihedral fitting plots and chlorobenzene liquid property data. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- (1) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical free energy calculations for drug discovery. *J. Chem. Phys.* **2012**, *137*, 230901.
- (2) Jorgensen, W. L. The many roles of computation in drug discovery. *Science (New York, N.Y.)* **2004**, *303*, 1813–8.
- (3) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035–4061.
- (4) 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*.
- (5) 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.
- (6) Cole, D. J.; Hine, N. D. M. Applications of large-scale density functional theory in biology. *J. Phys. Condens. Matter* **2016**, *28*, 393001.
- (7) 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.
- (8) Chen, S.; Yi, S.; Gao, W.; Zuo, C.; Hu, Z. Force field development for organic molecules: Modifying dihedral and 1-n pair interaction parameters. *J. Comput. Chem.* **2015**, *36*, 376–384.
- (9) Grimme, S. A general quantum mechanically derived force field (QMDF) for molecules and condensed phase simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497–4514.

- (10) Boulanger, E.; Huang, L.; Rupakheti, C.; MacKerell, A. D.; Roux, B. Optimized Lennard-Jones Parameters for Druglike Small Molecules. *J. Chem. Theory Comput.* **2018**, *14*, 3121–3131.
- (11) Prampolini, G.; Campetella, M.; De Mitri, N.; Livotto, P. R.; Cacelli, I. Systematic and Automated Development of Quantum Mechanically Derived Force Fields: The Challenging Case of Halogenated Hydrocarbons. *J. Chem. Theory Comput.* **2016**, *12*, 5525–5540.
- (12) Zahariev, F.; De Silva, N.; Gordon, M. S.; Windus, T. L.; Dick-Perez, M. ParFit: a Python-based object-oriented program for fitting molecular mechanics parameters to ab-initio data. *J. Chem. Inf. Model.* **2017**, *57*, 391–396.
- (13) Betz, R. M.; Walker, R. C. Paramfit: Automated optimization of force field parameters for molecular dynamics simulations. *J. Comput. Chem.* **2015**, *36*, 79–87.
- (14) Wang, L. P.; Martinez, T. J.; Pande, V. S. Building force fields: An automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891.
- (15) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (16) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (17) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OLPS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

- (18) Harder, E. et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
- (19) Rizzo, R. C.; Jorgensen, W. L. OPLS All-Atom Model for Amines: Resolution of the Amine Hydration Problem. *J. Am. Chem. Soc.* **1999**, *121*, 4827–4836.
- (20) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Open Force Field Consortium: Escaping atom types using direct chemical perception with SMIRNOFF v0.1. **2018**, bioRxiv:10.1101/286542.
- (21) Riquelme, M.; Lara, A.; Mobley, D. L.; Verstraelen, T.; Matamala, A. R.; Vöhringer-Martinez, E. Hydration Free Energies in the FreeSolv Database Calculated with Polarized Iterative Hirshfeld Charges. *J. Chem. Inf. Model.* **2018**, *58*, 1779–1797.
- (22) Seminario, J. M. Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.* **1996**, *60*, 1271–1277.
- (23) Lee, L. P.; Cole, D. J.; Skylaris, C. K.; Jorgensen, W. L.; Payne, M. C. Polarized protein-specific charges from atoms-in-molecule electron density partitioning. *J. Chem. Theory Comput.* **2013**, *9*, 2981–2991.
- (24) 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.
- (25) Allen, A. E.; Robertson, M. J.; Payne, M. C.; Cole, D. J. Introducing QUBE: Quantum Mechanical Bespoke Force Fields for Protein Simulations. In preparation (**2018**).
- (26) Zheng, S.; Tang, Q.; He, J.; Du, S.; Xu, S.; Wang, C.; Xu, Y.; Lin, F. VFFDT: A New Software for Preparing AMBER Force Field Parameters for Metal-Containing Molecular Systems. *J. Chem. Inf. Model.* **2016**, *56*, 811–818.

- (27) Li, P.; Merz, K. M. MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56*, 599–604.
- (28) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (29) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (30) 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*, 1623–1641.
- (31) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class IV charge models: A new semiempirical approach in quantum chemistry. *J. Comput. Aided Mol. Des.* **1995**, *9*, 87–110.
- (32) 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.
- (33) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. Localized Bond Charge Corrected CM1A Charges for Condensed-Phase Simulations Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121*, 3864–3870.
- (34) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6665–70.

- (35) Damm, W.; Frontera, A.; Rives, J. T.; Jorgensen, W. L. OPLS All-Atom Force Field for Carbohydrates. *J. Comp. Chem.* **1997**, *18*, 1955–1970.
- (36) 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.
- (37) 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.
- (38) Dodda, L. S.; Vilseck, J. Z.; Cutrona, K. J.; Jorgensen, W. L. Evaluation of CM5 Charges for Nonaqueous Condensed-Phase Modeling. *J. Chem. Theory Comput.* **2015**, *11*, 4273–4282.
- (39) Chu, X.; Dalgarno, A. Linear response time-dependent density functional theory for van der Waals coefficients. *J. Chem. Phys.* **2004**, *121*, 4083–4088.
- (40) Frisch, M. J. et al. Gaussian09 Revision E.01. Gaussian Inc. Wallingford CT 2009.
- (41) Manz, T. A.; Gabaldon Limas, N. Chagemol program for performing DDEC analysis. Available from <http://ddec.sourceforge.net> (accessed September 2014).
- (42) Kramer, C.; Spinn, A.; Liedl, K. R. Charge Anisotropy: Where Atomic Multipoles Matter Most. *J. Chem. Theory Comput.* **2014**, *10*, 4488–4496.
- (43) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (44) Yan, X. C.; Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. Improved Description of Sulfur Charge Anisotropy in OPLS Force Fields: Model Development and Parameterization. *J. Phys. Chem. B* **2017**, *121*, 6626–6636.

- (45) Unke, O. T.; Devereux, M.; Meuwly, M. Minimal distributed charges: Multipolar quality at the cost of point charge electrostatics. *J. Chem. Phys.* **2017**, *147*, 161712.
- (46) Dubay, K. H.; Hall, M. L.; Hughes, T. F.; Wu, C.; Reichman, D. R.; Friesner, R. A. Accurate force field development for modeling conjugated polymers. *J. Chem. Theory Comput.* **2012**, *8*, 4556–4569.
- (47) 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.
- (48) Wildman, J.; Repiščák, P.; Paterson, M. J.; Galbraith, I. General Force-Field Parametrization Scheme for Molecular Dynamics Simulations of Conjugated Materials in Solution. *J. Chem. Theory Comput.* **2016**, *12*, 3813–3824.
- (49) Wang, R.; Ozhgibesov, M.; Hirao, H. Partial hessian fitting for determining force constant parameters in molecular mechanics. *J. Comput. Chem.* **2016**, 2349–2359.
- (50) Dahlgren, M. K.; Schyman, P.; Tirado-Rives, J.; Jorgensen, W. L. Characterization of biaryl torsional energetics and its treatment in OPLS all-atom force fields. *J. Chem. Inf. Model.* **2013**, *53*, 1191–1199.
- (51) Zgarbová, M.; Rosnik, A. M.; Luque, F. J.; Curutchet, C.; Jurečka, P. Transferability and additivity of dihedral parameters in polarizable and nonpolarizable empirical force fields. *J. Comput. Chem.* **2015**, *36*, 1874–1884.
- (52) Vanommeslaeghe, K.; Yang, M.; Mackerell, A. D. Robustness in the fitting of molecular mechanics parameters. *J. Comput. Chem.* **2015**, *36*, 1083–1101.
- (53) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26*, 1689–1700.

- (54) 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.
- (55) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (56) Womack, J. C.; Anton, L.; Dziedzic, J.; Hasnip, P. J.; Probert, M. I.; Skylaris, C. K. DL-MG: A Parallel Multigrid Poisson and Poisson-Boltzmann Solver for Electronic Structure Calculations in Vacuum and Solution. *J. Chem. Theory Comput.* **2018**, *14*, 1412–1432.
- (57) Dziedzic, J.; Helal, H. H.; Skylaris, C. K.; Mostofi, A. A.; Payne, M. C. Minimal parameter implicit solvent model for ab initio electronic-structure calculations. *Europhys. Lett.* **2011**, *95*.
- (58) Lee, L. P.; Gabaldon Limas, N.; 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.
- (59) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Simmonett, A. C.; Harrigan, M. P.; Brooks, B. R.; Pande, V. S. OpenMM 7 : Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.
- (60) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method A smooth particle mesh Ewald method. *J. Chem Phys.* **1995**, 8577–8593.
- (61) Wang, J.; Hou, T. Application of molecular dynamics simulations in molecular property prediction. 1. Density and heat of vaporization. *J. Chem. Theory Comput.* **2011**, *7*, 2151–2165.

- (62) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.
- (63) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- (64) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Johannes, G. E. M. F. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (65) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (66) Caleman, C.; Van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; Van Der Spoel, D. Force field benchmark of organic liquids: Density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *J. Chem. Theory Comput.* **2012**, *8*, 61–74.
- (67) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- (68) Doherty, B.; Zhong, X.; Acevedo, O. Virtual Site OPLS Force Field for Imidazolium-Based Ionic Liquids. *J. Phys. Chem. B* **2018**, *122*, 2962–2974.
- (69) Manz, T. A.; Limas, N. G. Introducing DDEC6 atomic population analysis: Part 1. Charge partitioning theory and methodology. *RSC Advances* **2016**, *6*, 47771–47801.
- (70) Limas, N. G.; Manz, T. A. Introducing DDEC6 atomic population analysis: Part 2. Computed results for a wide range of periodic and nonperiodic materials. *RSC Advances* **2016**, *6*, 45727–45747.

- (71) Manz, T. A. Introducing DDEC6 atomic population analysis: Part 3. Comprehensive method to compute bond orders. *RSC Advances* **2017**, *7*, 45552–45581.
- (72) Limas, N. G.; Manz, T. A. Introducing DDEC6 atomic population analysis: Part 4. Efficient parallel computation of net atomic charges, atomic spin moments, bond orders, and more. *RSC Advances* **2018**, *8*, 2678–2707.
- (73) Manz, T.; Chen, T.; Cole, D. J.; Limas, N. G.; Fiszbein, B. Introducing DDEC6 Atomic Population Analysis: Part 5. New Method to Compute Polarizabilities and Dispersion Coefficients. **2018**, 10.26434/chemrxiv.7205693.v1.
- (74) Stone, A. J. Intermolecular Potentials. *J. Science* **2008**, *321*, 787–788.
- (75) Walters, E. T.; Mohebifar, M.; Johnson, E. R.; Rowley, C. N. Evaluating the London Dispersion Coefficients of Protein Force Fields Using the Exchange-Hole Dipole Moment Model. *J. Phys. Chem. B* **2018**, *122*, 6690–6701.
- (76) Misquitta, A. J.; Stone, A. J. ISA-Pol: Distributed polarizabilities and dispersion models from a basis-space implementation of the iterated stockholder atoms procedure. **2018**, arxiv:1806.06737.
- (77) Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. Beyond Born-Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields. *J. Chem. Theory Comput.* **2016**, *12*, 3851–3870.

Graphical TOC Entry

