A fast, convenient, polarizable electrostatic model for molecular dynamics

Liangyue Wang,[†] Michael Schauperl,[‡] David L. Mobley,[¶] Christopher Bayly,[§] and Michael K. Gilson[*,‖]

[†]*Department of Chemistry and Biochemistry, University of California, San Diego, CA 92093, USA*

[‡]*HotSpot Therapeutics, Inc. Boston, MA 02210, USA*

[¶]*Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697, USA*

[§]*OpenEye Scientific, Cadence Molecular Sciences, Santa Fe, NM 87508, USA*

[‖]*Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, CA 92093, USA*

E-mail: mgilson@ucsd.edu

Abstract

We present an efficient polarizable electrostatic model, utilizing typed, atom-centered, polarizabilities and the fast direct approximation, designed for efficient use in molecular dynamics (MD) simulations. The model provides two convenient approaches to assigning partial charges in the context of the atomic polarizabilities. One is a generalization of RESP, called RESP-dPol, and the other, AM1-BCC-dPol, is an adaptation of the widely used AM1-BCC method. Both are designed to accurately replicate gas-phase QM electrostatic potentials. Benchmarks of this polarizable electrostatic model against gas-phase dipole moments, molecular polarizabilities, bulk liquid densities, and static

dielectric constants of organic liquids, show good agreement with the reference values. Of note, the model yields markedly more accurate dielectric constants of organic liquids, relative to a matched non-polarizable force field. MD simulations with this method, which is currently parameterized for molecules containing elements C, N, O, and H, run about only 3.6-fold slower than fixed charge force fields, while simulations with the self-consistent mutual polarization average 4.5-fold slower. Our results suggest that RESP-dPol and AM1-BCC-dPol afford improved accuracy, relative to fixed charge force fields, and are good starting points for developing general, affordable, and transferable polarizable force fields. The software implementing these approaches has been designed to utilize the force field fitting frameworks developed and maintained by Open Force Field Initiative, setting the stage for further exploration of this approach to polarizable force field development.

1 Introduction

Molecular dynamics (MD) simulations can provide qualitative and quantitative insights into molecular systems without the expense of laboratory experiments. They are widely used to elucidate biomolecular mechanisms and to speed the discovery of new medications. An MD simulation requires a mathematical function called a force field to estimate the forces among atoms of a system, and these forces are used to predict the trajectory of the atoms over time. The design of a force field involves tradeoffs between accuracy, which is essential to generate a conformational ensemble similar to the real world one, and computational speed, which is needed for efficient sampling of thermodynamically and kinetically relevant configurations of the molecular system. In order to achieve a suitable speed-accuracy balance, a typical force field uses a theoretically motivated but ultimately empirical functional form, with an array of adjustable parameters that are tuned against reference data.

The force fields most widely used today approximate the electrostatic interactions of a molecular system in terms of Coulombic interactions among fixed, atom-centered, partial

charges.[1–4] The partial charges of a given molecule are typically chosen to replicate the electrostatic potential (ESP) generated by the molecule in vacuum, as computed by reference quantum mechanical (QM) methods.[1,5–7] Such so-called fixed charge electrostatic models are computationally efficient and have been remarkably useful. However, their accuracy is limited by their neglect of the field-dependent electronic polarization, i.e., of changes in the molecule's electron distribution in space in response to changes in an electric field. Including an explicit treatment of electronic polarizability is expected to improve accuracy, especially in molecular systems with large fields due to ions,[8–15] and in settings where polar molecules move between media of high dielectric constant (e.g., water) and low dielectric constant (e.g., a cell membrane), because changes in dielectric constant lead to changes in the self-polarization of polar molecules.[15–17]

Indeed, the electric field at an atom changes constantly during a simulation because of the movements of other atoms with their partial charges, and these changes induce continual changes in the charge distribution within a molecule and hence of its interatomic forces. These fast charge rearrangements are responsible for the fact that nonpolar organic liquids have static dielectric constants of about 2 and that all organic liquids have high-frequency dielectric constants of about 2.[18–20] Polar liquids have higher static dielectric constants because they have permanent dipole moments which confer orientational polarizability, but this responds more slowly than electronic polarization, so the dielectric constant of a polar liquid falls off with increasing frequency of a time-varying external electric field.[18–20] Electronic polarization also accounts for the fact that polar molecules become less polar when moved from a high-dielectric constant medium to a low-dielectric medium (or vacuum), where the reaction field in the molecule is weaker.[21] A fixed charge force field cannot account for this change, and is thus expected to lose accuracy when used to model processes where the dielectric environment of a molecule changes much. This may also help explain why water molecules treated as polarizable have a greater tendency to occupy nonpolar binding sites than water molecules treated with fixed charges.[22]

We therefore expect that integrating electronic polarizability into a well-parameterized force field can lead to improved accuracy.[16] However, the success of such an advance hinges not only on accuracy but also on how easily the force field can be applied to new molecules and on how well it maintains the efficiency of MD simulations. Pioneering work along these lines dates back to at least the 1970s,[23] and most current polarizable force fields are based on one of three models: induced dipoles, Drude oscillators, or fluctuating charges.

The induced dipole model, which is used in the present study, places a linear, point-polarizability at the center of each atom.[23,24] The field at atom j, $\boldsymbol{E}_j$, leads to an induced dipole as:

$$\boldsymbol{\mu}_{\mathrm{ind,j}} = \alpha_j \boldsymbol{E}_j \tag{1}$$

where α_j represents either a scalar polarizability or a polarizability tensor, and in general $\boldsymbol{E}_j = \boldsymbol{E}_{j,\mathrm{ext}} + \boldsymbol{E}_{j,\mathrm{perm}} + \boldsymbol{E}_{j,\mathrm{ind}}$, where the three contributions are the external field (if present), the field due to the permanent partial atomic charges on other atoms, and the field due to induced dipoles on other atoms. Because the value of each atom's induced dipole depends on the induced dipoles present on other atoms, induced dipoles are usually solved with a self-consistent field (SCF) method, also termed a mutual polarization method. This accounts fully for the contribution of the induced electric field ($\boldsymbol{E}_{j,\mathrm{ind}}$) from the induced dipoles on other atoms. The well-developed AMOEBA force field[25] uses this approach, along with a more detailed multipole representation of the permanent charge distribution. The PHAHST polarizable potentials, designed for material simulations, also utilize the induced dipole model.[26] The complexity of solving for induced dipoles by SCF methods may be reduced by an empirical extrapolation scheme developed by Brooks et al.[27] . An even simpler approach is afforded by the direct approximation of electronic polarization,[28] where the atomic polarizabilities do not "feel" the fields generated by other induced dipoles, but only the fields generated by partial charges and the external field (if any); the approximation is thus $\boldsymbol{E}_j \approx \boldsymbol{E}_{j,\mathrm{ext}} + \boldsymbol{E}_{j,\mathrm{perm}}$. This approach was utilized in the iAMOEBA polarizable water model,[29] which successfully replicates many physical properties of water. However,

the direct approximation has not yet been applied to a wider range of systems. In this work, we extended the usage of the fast direct approximation in various small organic molecules by deriving appropriate force field parameters in the context direct polarization.

The Drude polarizable force field[30] is the polarizable version of the CHARMM family of force fields[2,31,32] . It accounts for polarizability by attaching an auxiliary charged particle to each polarizable atomic center with a harmonic spring.[33] A small charge and mass are transferred from the parent atoms to the Drude particles, allowing them to be treated as dynamical particles that move around their parent atoms. A key advantage of this approach is that the movements of the Drude particles can be accommodated in the usual MD scheme with little modification, rather than requiring the added code needed to handle the inducible dipole model above. However, the added particles make for a somewhat more complex system, a dual-thermostat extended Lagrangian algorithm is used to keep the Drude particles at a very low temperature ($\sim$1K), and a 1-fs time step is employed because the Drude oscillators have high natural frequencies. Following introduction of the Drude polarizable water model 2003[33] , the Drude polarizable force field has successfully been extended to proteins, DNA, lipids, and carbohydrates.[34–36]

The fluctuating charge (FQ) model describes polarization by allowing charge to flow between atoms in response to the electric field,[37–39] so neither inducible dipoles nor added charge centers are required. The modified charges are typically obtained via an electronegativity equalization approach,[40,41] whose solution can be somewhat time-consuming. Because charges can only flow along the direction of bonds, the FQ model does not describe out-of-plane polarization. Nonetheless, the FQ representation can afford accuracy competitive with the induced polarizable representation.[42]

The present study builds on prior advances to propose two polarizable electrostatic models intended to afford a favorable balance of accuracy, ease of use, and computational speed. Both models assign each atom a typed, atom-centered, linear, isotropic polarizability. These empirical atomic polarizabilities are derived from electrostatic potential responses when an

imposed external electric field is present. This method has been widely employed in different flavors of induced dipole model, some examples are included in references.[43–47] In one model, partial charges are assigned with RESP[5]-type fitting to QM ESPs in the context of the typed polarizabilities. In the second model, the AM1-BCC[6,7] method is used, but with a new set of BCCs trained to replicate QM ESPs in the context of our typed polarizabilities. These assignment methods allow facile assignment of polarizabilities and compatible partial charges to a wide range of compounds. In addition, we make consistent use of the direct polarization approximation in order to maximize computational efficiency during MD simulations. The direct approximation also avoids the possibility of "polarization catastrophe" inherent to the SCF approach, a numerical instability where two nearby inducible dipoles mutually polarize each other without limit. The current proof-of-concept models are trained for carbon, hydrogen, oxygen, and nitrogen, but our workflows can be easily applied to the full range of elements in drug-like molecules. We demonstrate the utility and accuracy of these models in calculations on isolated molecules and in simulations of organic liquids. We close with a discussion of next steps to further evaluate these methods and to integrate them into a comprehensive polarizable force field.

2 Methods

The overall structure of our direct polarizability models is as follows. Each atom has an atom-centered, constant, partial charge, as well as an atom-centered, linear, isotropic point-polarizability. In the direct approximation, the point-polarizabilities feel only by the field generated by the partial charges. That is, they do not feel the other induced dipoles. In accordance with the scaling scheme for short-range intramolecular interactions used in Applequist-like models[23,24] , we apply a scaling factor f_{jk} to exclude 1-2 and 1-3 charge-polarizability interactions ($f_{jk} = 0.0$) and we scale 1-4 interactions by $f_{jk} = 0.5$. These scalings are applied during fitting and also during simulations. Note that 1-2 and 1-3 interatomic

distances vary little across molecular dynamics snapshots, due to the high spring constants of typical bond-stretch and angle-bend terms, so omitting 1-2 and 1-3 charge-polarizability interactions omits a nearly constant term, which may be at least partly adjusted for by suitably chosen partial charges. It is also worth noting that a scaling factor is not needed to avoid polarization catastrophe, because the direct approximation does not have this numerical instability, so f_{jk} could be safely treated as an adjustable parameter to further optimize short-range electrostatics.

2.1 Optimization of typed polarizabilities

Both polarizable electrostatic models developed here (Sections 2.2, 2.3) use typed polarizabilities, and we consider two typing schemes. A first, minimalist, model applies a single value of polarizability to all atoms of a given element (C, H, O, and N). These are termed element-based polarizabilities. The second, more fine-grained, typing scheme applies the same polarizability to all atoms of the same Lennard-Jones (LJ) type, using LJ types from the Open Force Field (OpenFF) Sage force field[4] . These are termed Sage LJ-based polarizabilities. The types are detected with SMARTS patterns[48] and parameterized in the SMIRKS Native Open Force Field (SMIRNOFF) format[49] . Polarizabilities associated with both typing schemes are trained based on a training set of 39 compounds, each in an average of seven conformations (minimum four and maximum 11) differing from all others by at least 0.5 Å root-mean-square deviation (RMSD) (Section 2.4), to reduce concerns about conformational dependencies in our ESP fits.[5,50]

To train the polarizabilities, we computed baseline QM ESPs, $V_{\mathrm{QM,\,ik}}$, for each conformer k in the training set, where i indexes the m_k ESP sampling points around the conformer. For each conformer, we also computed six polarized QM ESPs by imposing uniform external fields of magnitude 0.01 a.u. in the +x, -x, +y, -y, +z, -z directions, and computed their differences relative to the baseline QM ESPs, $V_{\mathrm{diff,ikl}}$, where $l \in [1, \ldots, 6]$ indexes the field directions. Using different field directions allows for averaging over any anisotropy in the

induced polarization. The field strength of 0.01 a.u. was chosen to approximate the external electric field generated by a sodium ion about 4Å away. We used global optimization with the Nelder-Mead method in SciPy[51] to find typed polarizabilities that minimize the following error function,

$$\chi^2 = \sum_{k=1}^{N_{\text{conf}}} \sum_{l=1}^{6} \sum_{i=1}^{m_k} \left(V_{\text{diff,ikl}} - \sum_{j=1}^{n_k} \frac{\boldsymbol{\mu}_{\text{ind,jkl}} \boldsymbol{r}_{\text{ijk}}}{r_{\text{ij}}^3} \right)^2 \tag{2}$$

where N_{conf} is the number of conformations of all molecules in the training set, n_k is the number of atoms in the molecule corresponding to conformation k, $\boldsymbol{\mu}_{\text{ind, jkl}}$ is the induced dipole on atom j of conformer k with external field direction l, and $\boldsymbol{r}_{\text{ijk}}$ is the vector from atom j to ESP point i for conformer k. The second quantity in parentheses is the potential at grid point i generated by the induced dipoles. Because we are using the direct approximation and are fitting polarizabilities to differences in ESP generated by an inducing field, the induced dipoles in Equation 2 are given simply by

$$\boldsymbol{\mu}_{\text{ind,jkl}} = \alpha_j \boldsymbol{E}_{\text{ext,jkl}} \tag{3}$$

where α_j is the typed polarizability assigned to atom j and $\boldsymbol{E}_{\text{ext,jkl}}$ is the external field at atom j for field direction l. Because we are using typed polarizabilities, this polarizability is the same for all instances of the polarizability atom type across all conformers (N_{conf}) of all molecules.

The typed polarizabilities provided by this procedure are independent of the choice of atomic partial charges and therefore allow users to choose among charge-assignment methods – here RESP-dPol and AM1-BCC-dPol. Unlike fitting to molecular dipole moments or molecular polarizabilities, deriving polarizabilities from QM ESPs follows what has arguably been the most successful approach for deriving atomic partial charges (RESP, AM1-BCC). This approach makes physical sense, as it emphasizes the influence of polarization on the strong, short-range intermolecular interactions that are particularly relevant for condensed phase simulations.

2.2 RESP-dPol charge model

The RESP-dPol model assigns partial charges with a RESP-like fitting method in the context of the typed polarizabilities discussed above. It differs from standard RESP because the induced dipoles also contribute to the computed ESPs. Pioneering work by Cieplak et al.[44] used a similar methodology and achieved reasonable agreement with experimental solvation free energies. In RESP-dPol, we adopt the two-stage hyperbolic restraints used in RESP to reduce nonphysical variations of partial charges.[5] The first stage is used to make sure all polar regions are well-fitted, and a weak restraint (a = 0.005 a.u., b = 0.1 a.u.) is used to decrease the overall magnitude of all partial charges. In the second stage, all partial charges in polar regions are fixed and forced symmetry restraints and a strong hyperbolic restraint (a = 0.01 a.u., b = 0.1 a.u.) are used to achieve an optimal description of electrostatics.

We created a toolkit (Section 2.5) that derives RESP-dPol charges for a given molecule in a given conformer by fitting to its baseline QM ESP ($V_{\mathrm{QM,\,i}}$), i.e. its ESP in the absence of an external field, by minimizing the following quantity with a least squares procedure:

$$\chi^2 = \sum_{i=1}^{m} \left(V_{\mathrm{QM,\,i}} - V_{\mathrm{perm,\,i}} - V_{\mathrm{ind,\,i}}\right)^2 + \lambda \left(\sum_{j=1}^{n} q_{\mathrm{j}} - q_{\mathrm{tot}}\right) + a \sum_{j=1}^{n} \left(\sqrt{q_{\mathrm{j}}^2 + b^2} - b\right) \tag{4}$$

where n is the number of atoms, $V_{\mathrm{perm,\,i}} = \sum_{j=1}^{n} \frac{q_{\mathrm{j}}}{r_{\mathrm{ij}}}$ is the Coulomb potential at grid point i generated by the partial charges and $V_{\mathrm{ind,\,i}}$ is the contribution from induced dipoles in the direct approximation:

$$V_{\mathrm{ind,\,i}} = \sum_{j=1}^{n} \frac{\boldsymbol{\mu}_{\mathrm{ind,\,j}} \boldsymbol{r}_{\mathrm{ij}}}{r_{\mathrm{ij}}^3} \tag{5a}$$

$$\boldsymbol{\mu}_{\mathrm{ind,\,j}} = \alpha_{\mathrm{j}} \sum_{k \neq j}^{n} f_{\mathrm{jk}} \frac{q_{\mathrm{k}} \boldsymbol{r}_{\mathrm{jk}}}{r_{\mathrm{jk}}^3} \tag{5b}$$

A Lagrange multiplier, λ, is used to constrain the sum of atomic charges (q_{j}) to the correct molecular charge, q_{tot}.[50,52] The last term in Equation 4 is the hyperbolic restraint described

above.

2.3 AM1-BCC-dPol charge model

The AM1-BCC-dPol model generates partial charges similar in quality to those of RESP-dPol, but without the burden of computing QM ESPs and fitting partial charges to them for each new molecule. Just as in the standard AM1-BCC method[6,7] , AM1-BCC-dPol charges are constructed as a sum of population pre-charges ($q_{\mathrm{j}}^{\mathrm{pre}}$) generated with the fast AM1 method[53] and bond charge correction (BCC) parameters ($\boldsymbol{B}_\beta$) predefined based on bond connectivity analysis ($\boldsymbol{T}_{j\beta}$):

$$q_{\mathrm{j}} = q_{\mathrm{j}}^{\mathrm{pre}} + \sum_{\beta=1}^{n_\beta} \boldsymbol{T}_{j\beta}\boldsymbol{B}_\beta \tag{6}$$

where the summation of the BCC-dPol correction term runs over the total number of bond types (n_β). We use the same BCC types as used in the standard AM1-BCC method.

The training of BCC-dPol parameters is similar to deriving RESP-dPol charges. Both permanent charges and induced dipoles contribute to electrostatic potentials on grid points, and the BCCs are adjusted to minimize deviations from baseline QM ESPs. The contribution to χ^2 from one conformer of one molecule is given by:

$$\chi^2 = \sum_{i=1}^{m} \left(V_{\mathrm{QM,\,i}} - V_{\mathrm{pre,\,i}} - \sum_{j=1}^{n}\sum_{\beta=1}^{n\beta} \frac{\boldsymbol{T}_{j\beta}\boldsymbol{B}_\beta}{r_{\mathrm{ij}}} - V_{\mathrm{ind,\,i}} \right)^2 \tag{7}$$

where $V_{\mathrm{pre,\,i}}$ is the contribution to the computed ESP at site i calculated with AM1 population charges, and $V_{\mathrm{ind,\,i}}$ is the contribution from induced dipoles. The BCCs are fitted by minimizing the sum of χ^2 over all conformers of a training set of 119 molecules (Section 2.4), much as detailed in Equation 2. Once the BCCs have been optimized, the AM1-BCC-dPol method allows both polarizabilities and polarization-consistent partial charges to be assigned to a new molecule as quickly and easily as partial charges alone are assigned with

the traditional AM1-BCC method.

2.4 Training sets and reference QM data

The polarizability training set consists of 39 small molecules from the RESP2[54] training dataset. An additional set of 80 small molecules from the OpenFF BCC refit study (COH dataset)[55] , for a total of 119 molecules, were added to train BCC-dPol parameters (Figure S4).

All molecules used consist only of elements C, H, O, and N, but a diverse range of chemical fragments, including aliphatic carboxylic acids, amides, aldehydes, and aromatic hydrocarbons, are included to ensure the transferability of the resulting parameters. Two charged molecules are included in the training sets to represent electrostatic environments caused by charged proteins or ligands. For molecules with rotatable bonds, multiple conformations are generated and fitted as if they were independent molecules. Baseline and polarized QM ESPs were evaluated at the MP2/aug-cc-pVTZ[56,57] QM level of theory, because prior benchmarks of electronic structure methods for molecular mechanics[58,59] and polarizable force fields[25,35,43] suggest that this QM method affords a favorable balance of efficiency and accuracy. The external electric fields used to generate polarized ESPs have a strength of 0.01 a.u. ESPs were computed at Merz–Singh–Kollman (MSK) style grid points[52,60] located outside the van der Waals radii of all atoms using a spacing of 0.126 Å and a density of 17 points per Å^2 for a total of 10 layers.

2.5 Infrastructure for parameterization

The optimization of polarizabilities and charge models was implemented in an open-source Python package, Fast Atom-Centered Typed Isotropic Ready-to-use Polarizable Electrostatic Model (Factor-Pol)[61] .

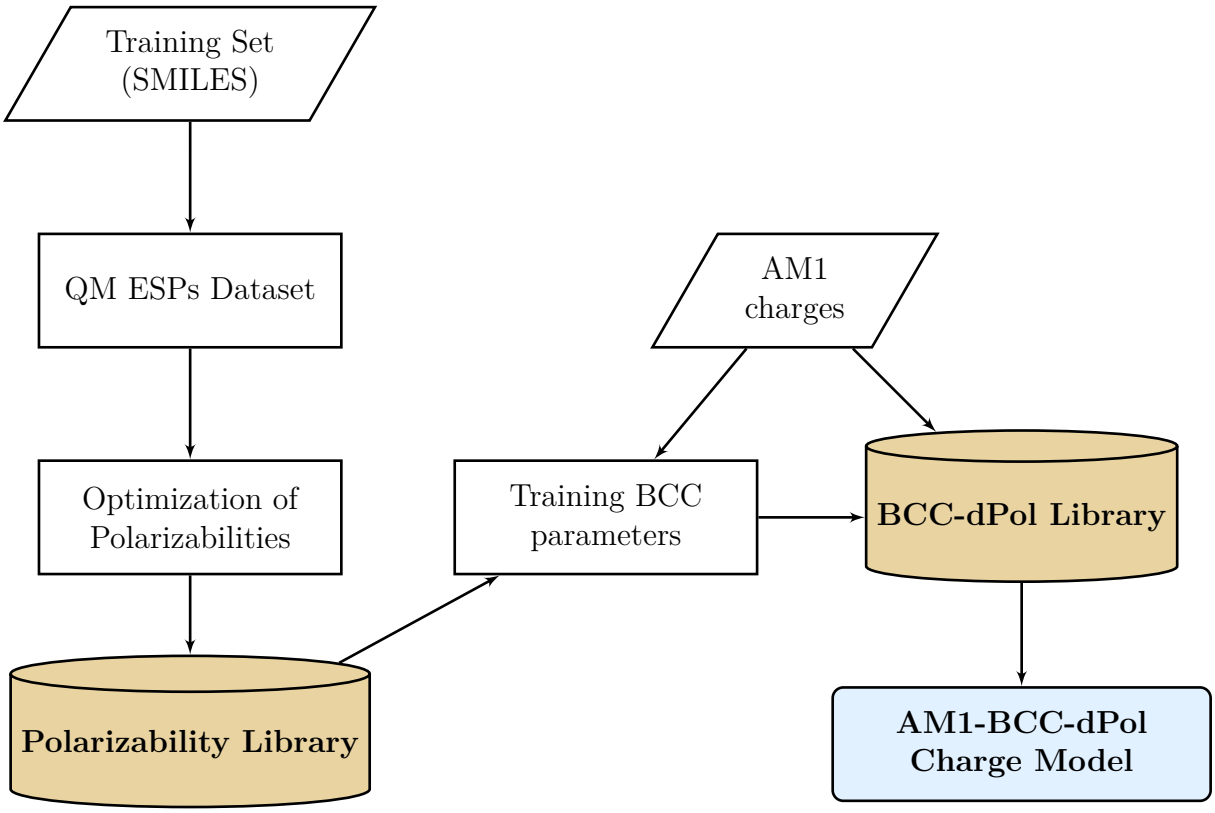


Figure 1: Diagram of Factor-Pol toolkit used to carry out QM calculations, derive optimized, typed polarizabilities and polarization-adapted BCCs, and assign RESP-dPol and AM1-BCC-dPol charges and polarizabilities to molecules.

As illustrated in Figure 1, Factor-Pol combines QM calculations, optimization, and parameterization for current and continued development of the present polarizable electrostatic models. The Factor-Pol package also includes an interface to OpenFF software infrastructure for future development of a full polarizable force field, including tuned LJ parameters, torsion parameters, and a polarizable water model. Factor-Pol also is used to assign RESP-dPol and AM1-BCC-dPol charges and typed polarizabilities to molecules for use in MD simulations with OpenMM and the MPID (multipole and induced dipole) plugin[62,63] .

2.6 Benchmarking of electrostatics models

2.6.1 Accuracy of baseline molecular ESPs

We evaluated the accuracy of both RESP-dPol and AM1-BCC-dPol in terms of relative root mean square errors (RRMSE) to baseline gas-phase QM ESPs on a test dataset (Figure S3).

As done previously,[5] RRMSE was computed as Equation 8:

$$\mathrm{RRMSE} = \left(\frac{1}{m} \frac{\sum_{i=1}^{m} (V_{\mathrm{Calc,\ i}} - V_{\mathrm{QM,\ i}})^2}{\sum_{i=1}^{m} (V_{\mathrm{QM,\ i}})^2} \right)^{1/2} \tag{8}$$

where m is the number of grid points around a molecule and i indexes each grid point.

2.6.2 Gas phase molecular dipole moments

We also evaluated accuracy by comparing the gas-phase molecular dipole moments they yield with corresponding gas-phase QM dipole moments evaluated at the MP2/aug-cc-pVTZ level of theory on the molecules used to train BCC-dPol parameters (Figure S4). It is generally believed that gas-phase partial charges derived from HF/6-31G* QM level of theory are over-polarized for the gas-phase and are therefore fortuitously suitable for use in the condensed phase.[52] However, a fixed, over-polarized electrostatic model is not ideal for use to simulate properties that involve large changes in the electrostatic environment of a molecule, such as the transfer free energies of a molecule from vacuum to water. The RESP-dPol and AM1-BCC-dPol models include explicit polarizability and thus, like other polarizable models, should be able to account for such changes in the electrostatic environment. We therefore would like them to generate accurate, rather than overpolarized, gas-phase molecular dipole moments. The gas-phase permanent molecular dipole moments from our models were cal-

culated as:

$$\mu = \sum_{j=1}^{N} \left(q_j \boldsymbol{r}_j + \mu_{\text{ind, j}} \right) \tag{9}$$

where N is the number of atoms in the molecule, and $\boldsymbol{r}_j$ and $\mu_{\text{ind, j}}$ are, respectively, the coordinates and induced dipole of atom j.

2.6.3 Condensed phase molecular polarizability

The polarizability of a molecule, α_M, can be approximated from the measured refractive index, n, of its liquid form via the Lorentz-Lorenz equation[64,65] :

$$\frac{n^2 - 1}{n^2 + 2} = \frac{4}{3} \pi \rho \alpha_M \tag{10}$$

Here ρ is the number density of molecules in the liquid. The present polarizability model treats each atom of a molecule as having an isotropic polarizability α_j that "feels" only permanent charges and any external field that may be present, so the molecular polarizability is simply:

$$\alpha_M = \sum_{j=1}^{N} \alpha_j \tag{11}$$

where N is the number of atoms.

Molecular polarizabilities have been used to derive atomic polarizabilities with an additivity method in previous work.[66–68] Here, instead, we use experimental molecular polarizabilities as an efficient check of the magnitudes of our QM-derived atomic polarizabilities. Reference experimental data were obtained from previous compilations[67,69] .

2.6.4 Condensed phase properties of organic liquids

We used simulations to compute the mass densities and dielectric constants of 17 polar and nonpolar organic liquids using the AM1-BCC-dPol electrostatic model and compared the results to those obtained from matched simulations using the fixed-charge AM1-BCC

electrostatic model. Condensed-phase simulations of these liquids were performed with OpenMM[70] . The MPID plugin[62] was used to enable calculations with AM1-BCC-dPol and typed polarizabilities, using the keyword "direct" to enable direct polarization. For all liquid systems, the valence terms and van der Waals terms were assigned from the OpenFF Sage force field[4] . Cubic liquid boxes composed of 256 molecules were built with Packmol package[71] as simulation systems. Scripts to set up simulations are available online.[72]

Each system was equilibrated with 5 ns simulations in the NVT ensemble, followed by 15 ns NPT production using a time step of 2 fs. MD trajectories for post-processing were saved every 1 ps. For a given MD trajectory, the mass density was evaluated as:

$$\rho = \frac{M}{\langle V \rangle} \tag{12}$$

where M is the total mass of the system, V is the volume of the simulation box, and the angle brackets indicate the average over simulation snapshots.

The dielectric constant of a condensed system is a key factor in determining the strength of electrostatic interactions. Simulations with fixed-charge electrostatic models are notorious for underestimating the dielectric constants of nonpolar liquids, because this derives almost entirely from electronic polarization. In contrast, the dielectric constants of polar liquids derive largely from orientational polarizability, which can be captured reasonably well by a force field that lacks an explicit treatment of electronic polarization. Here, we computed the static dielectric constants of liquids of varying polarity using a dipole fluctuation approach[19,20,73] :

$$\epsilon = \epsilon_\infty + \frac{\langle \mu^2 \rangle - \langle \mu \rangle^2}{3\epsilon_0 k_B T \langle V \rangle} \tag{13}$$

where μ is the total dipole moment of the system, and ϵ_∞ is the high-frequency dielectric constant:

$$\epsilon_\infty = \frac{1}{\epsilon_0 V} \frac{\mu_{\text{ind}}}{\boldsymbol{E}} + 1 \tag{14}$$

where $\boldsymbol{E}$ is the local electric field. Because of the simplicity of the direct polarization

approximation, the high-frequency dielectric constant can be obtained analytically as:

$$\epsilon_\infty = \frac{1}{\epsilon_0 V} \sum_{j=1}^{N} \alpha_j + 1 \tag{15}$$

3 Results

3.1 Quality-of-fit to QM ESPs

The ability to reproduce accurate QM ESPs around a given molecule is essential to reproducing the intermolecular interactions with surrounding molecules for simulations of complex condensed systems. We examined the ability of two polarizability typing schemes, one based simply on elements and the other assigning a separate polarizability to each LJ atom type in the OpenFF Sage force field. Each typing scheme was tested with two charge models, RESP-dPol and AM1-BCC-dPol, both of which are optimized to reproduce baseline gas-phase QM ESPs of training-set molecules in the presence of our typed polarizabilities.

Since RESP-dPol charges are derived by directly fitting to QM ESPs, they afford optimal accuracy on this test, and the comparison with AM1-BCC-dPol provides information on whether the fast AM1-BCC-dPol charge model is suitable for use in condensed-phase simulations. It is worth commenting that the RESP-dPol charge model does not use a training set, and we computed RRMS results for RESP-dPol charges on the training set and test set only for comparison with the AM1-BCC-dPol charge model. Therefore, it is possible for the RESP-dPol to generate lower RRMS errors on the test set than on the training set.

Table 1 shows that AM1-BCC-dPol yields RRMS errors, relative to QM ESPs, that are only marginally higher than those provided by RESP-dPol. This favorable result demonstrates that AM1-BCC-dPol partial charges, which are not fitted to the QM ESPs of each new molecule, are almost as accurate as RESP-dPol charges, which are fitted to each new molecule's QM ESPs. In addition, the AM1-BCC-dPol errors for the test set are very similar

to those for the training set, for both element-based polarizabilities and Sage LJ-based polarizabilities. We note, too, that the RRMS errors of RESP-dPol and AM1-BCC-dPol charge models are similar to those reported for standard RESP charges,[5] indicating that adding polarizabilities does not degrade accuracy. These results suggest that our AM1-BCC-dPol charge model is transferable and is suitable to be used as a fast high-quality alternative to RESP-dPol in condensed-phase MD simulations.

Table 1: Quality-of-fit of RESP-dPol and AM1-BCC-dPol electrostatic models to baseline QM ESPs

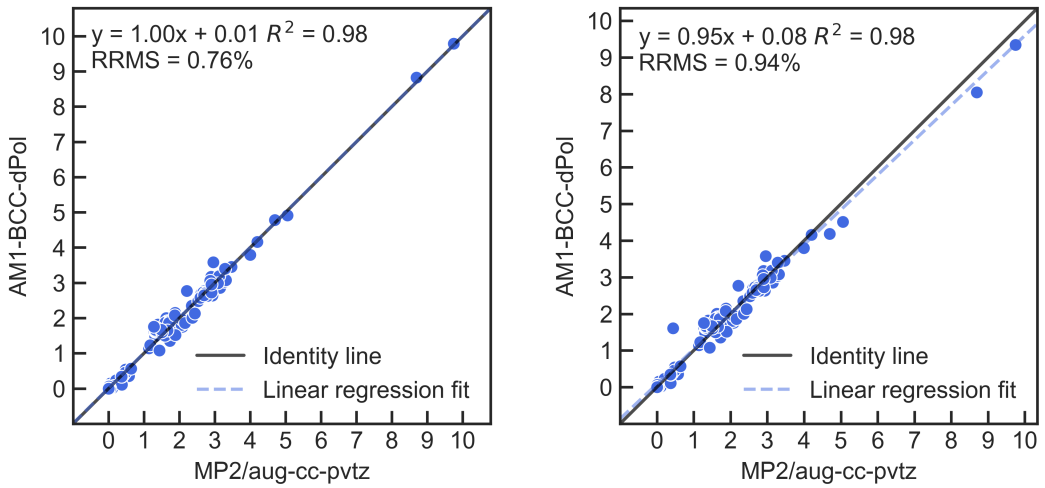
Polarizability Types	**Charge Model**	**RRMS (%)**	
		Training Set	Test Set
Element	RESP-dPol	0.13	0.12
Element	AM1-BCC-dPol	0.18	0.22
Sage LJ	RESP-dPol	0.13	0.13
Sage LJ	AM1-BCC-dPol	0.21	0.21

The Sage LJ-based polarizabilities are in principle tuned for the chemical environment of each atom in a molecule, whereas the element-based polarizabilities do not depend on chemical environments. Thus, Sage LJ types might be expected to provide a more accurate description of electronic polarization. Nonetheless, the RRMS errors, relative to QM ESPs, from the two typing schemes are essentially indistinguishable. In particular, element-based and Sage LJ-based electrostatic models perform almost identically on the test set, although the element-based set has only four polarizabilities whereas the Sage LJ-based set has 14 polarizabilities. We conclude that the minimalist elemental-based polarizability types, paired with the AM1-BCC-dPol charge model, may be sufficient for an accurate polarizable electrostatic model.

3.2 Molecular dipole moments

Previous benchmarking studies on electronic structure methods for gas phase dipole moments[58,59] suggest that HF/6-31G* overpolarizes molecules and that MP2/aug-cc-pVTZ predicts accurate molecular dipole moments. Although the overpolarization that results from fitting partial charges to QM ESPs at the HF/6-31G* level is generally considered as a fortuitous outcome in the RESP and AM1-BCC charge models, as it yields charges suitable for use in condensed phase simulations, it is not ideal for simulations at low dielectric environments, such as in gas phase or inside protein cavities. For the present polarizable model, where polarity adapts to environment, we hope to see gas phase molecular dipole moments that accurately match accurate gas phase QM dipole moments provided by the reliable MP2/aug-cc-pVTZ level of theory.

As shown in Figure 2, molecular dipole moments calculated for 107 neutral molecules with the AM1-BCC-dPol model, using both the element-typed and Sage LJ-typed polarizabilities, agree extremely well with the reference QM results, with relative errors below 1%. Interestingly, the element-based typing scheme gives a slightly lower RRMS error and a linear regression slope slightly closer to unity. Thus, both variants of the AM1-BCC-dPol electrostatics model accurately reproduce reference gas-phase dipole moments, as hoped.



(a) Element-based (b) Sage LJ-based

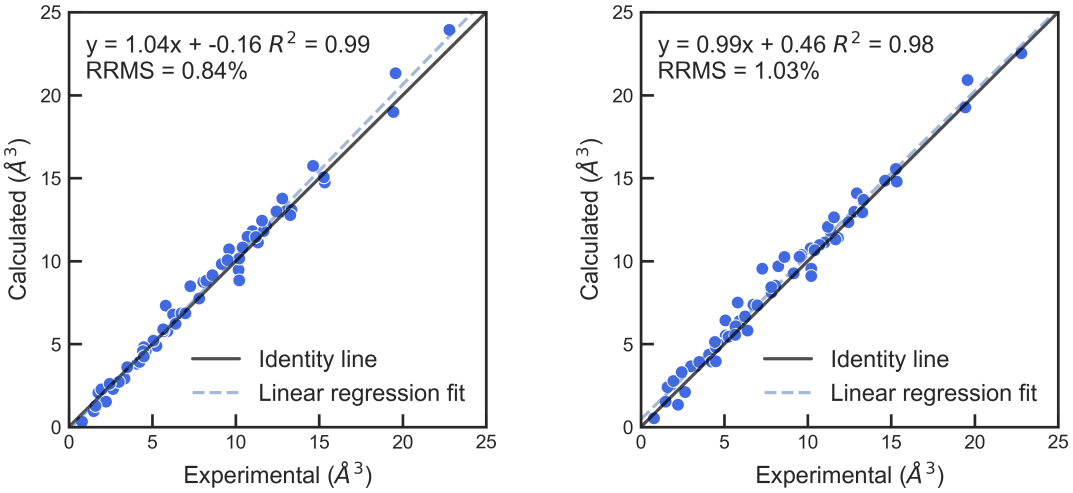
Figure 2: Comparison of gas-phase molecular dipole moments (Debye) computed with the AM1-BCC-dPol electrostatics model with QM reference results, for element-based polarizabilities (2a) and Sage LJ type-based polarizabilities (2b).

3.3 Condensed phase molecular polarizabilities

The present atomic polarizabilities were derived to best fit QM ESPs and thus capture the intermolecular interactions of a system in a dynamic electrostatic environment. However, there is no guarantee that these polarizabilities will also yield accurate molecular polarizabilities. Comparing with these experimental observables thus offers an independent check of the robustness and transferability of our model.

We find that molecular polarizabilities computed from both element-based and Sage LJ-typed atomic polarizabilities (Equation 11) agree well with those derived from experimental indices of refraction, as shown in Figure 3. It is worth noting that, although the element-typed polarizabilities include four types, the minimalist element-typed polarizabilities transfer well to molecules with diverse chemical environments that are not included in the training data, such as nitriles and alkynes. The chemical structures of molecules with available experimental measurements are presented in Figure S5. These results suggest that our atomic polarizabilities based on induced gas-phase QM ESPs retain their broader

physical meaning and can be valid for use in molecular simulations.



(a) Element typed polarizabilities (b) Sage LJ-typed polarizabilities

Figure 3: Comparison of molecular polarizabilities computed from our atomic polarizabilities, against experimental reference data, for element-based polarizabilities (3a) and Sage LJ type-based polarizabilities (3b).

3.4 Transferability of AM1-BCC-dPol across molecular conformations

For a fixed charge model, the partial charges that best replicate QM ESPs for one conformation of a molecule are not optimal for other conformations, and this problem is usually referred to as conformational dependency[5,50] . Thus, losses in accuracy are inevitable when fixed charge models are used in simulations where the molecular conformation varies. We conjectured that the polarizable AM1-BCC-dPol and RESP-dPol model would better preserve accuracy across multiple conformations of a molecule compared to the non-polarizable AM1-BCC model, because of the ability of the inducible dipoles to respond to changes in field due to changes in conformation. As an illustrative test of this idea, we examined the transferability of the AM1-BCC-dPol and RESP-dPol electrostatic models across two conformations of alanine dipeptide, conformer 0 with an intramolecular hydrogen bond (Figure 4a) and conformer 1 without (Figure 4b) an intramolecular hydrogen bond. Table 2 shows the

average unitless RRMS errors of gas phase ESPs computed with conformer 0 and evaluated on conformers 0 and 1, and vice version, for the non-polarizable AM1-BCC model, and the polarizable models AM1-BCC-dPol and RESP-dPol.

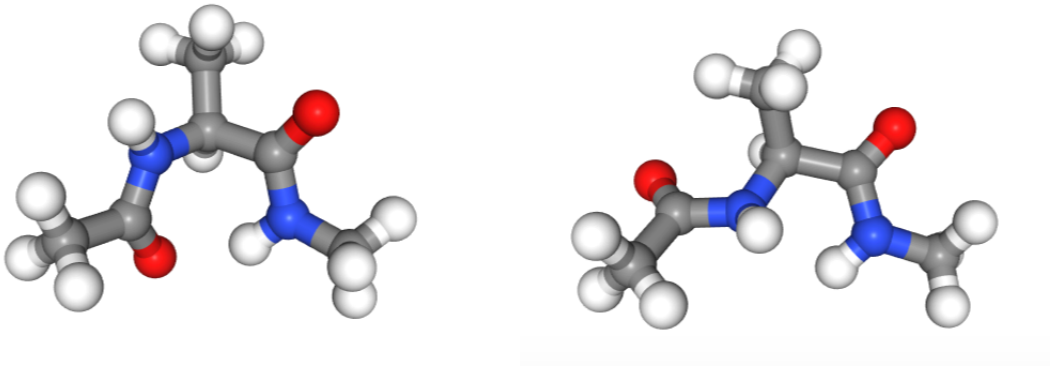
Table 2: Average RRMS errors (unitless) of ESPs, relative to reference QM ESPs, computed using partial charges derived from one conformer (column 1) and tested for a second conformer (columns 2 and 3), for electrostatic models AM1-BCC, AM1-BCC-dPol, and RESP.

AM1-BCC-dPol	conformer 0	conformer 1
conformer 0	0.82	0.97
conformer 1	0.80	0.75

RESP-dpol	conformer 0	conformer 1
conformer 0	0.63	1.19
conformer 1	0.93	0.75

AM1-BCC	conformer 0	conformer 1
conformer 0	0.95	1.28
conformer 1	0.87	1.13

As shown in Table2, AM1-BCC-dPol provides consistently low errors ($< 1\%$) across both the training and test conformers. RESP-dPol provides slightly greater accuracy overall when trained and tested on the same conformation (diagonal elements of each subtable), but less accuracy when trained on one conformer and tested against the conformer (off-diagonal elements). These increased errors presumably reflect overtraining. The nonpolarizable AM1-BCC model is less accurate than AM1-BCC-dPol in all four cases and less accurate than RESP-dPol in all but one case. Overall, these results suggest that our AM1-BCC-dPol model is more transferable and accurate than AM1-BCC, at least in part because its inducible dipoles respond appropriately to changing electrostatic environments. The simplicity, transferability, and accuracy afforded by AM1-BCC-dPol make it an attractive charge model for use in MD simulations where conformational charges play an important role.



(a) Conformer 0 (b) Conformer 1

Figure 4: The two conformers of alanine dipeptide used to examine the transferability of charges trained for one conformation and tested on a second conformation (see Table 2).

3.5 Properties of organic liquids from simulations

We ran simulations of 17 organic liquids (Figure S1) at constant pressure to compare the ability of force fields using various electrostatics models to provide accurate mass densities and dielectric constants. We tested the fixed charge AM1-BCC model, AM1-BCC-dPol with element-based polarizability types, and AM1-BCC-dPol with Sage LJ-based polarizability types. We drew all other force field parameters (Lennard-Jones and valence terms) from the OpenFF Sage force field. As previously done,[74] we focused on the reciprocals (D^{-1}) of the dielectric constant (D). This approach is motivated by the fact that electrostatic energies are proportional not to the dielectric constant but to its reciprocal:

$$U_{\text{elec}} \propto \frac{q_1 q_2}{\epsilon_0 D} \tag{16}$$

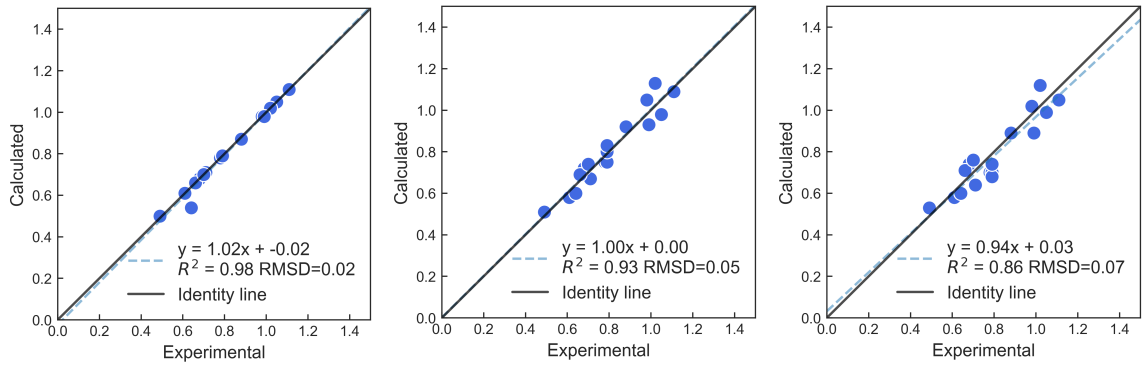
where ϵ_0 is the permittivity of free space and D is the relative dielectric constant.

All three electrostatics models (AM1-BCC, AM1-BCC-dPol with element-based polarizability types, and AM1-BCC-dPol with Sage LJ types) give good agreement with experimental mass densities (Figure 5 and Table 3) with RMS errors of 0.03, 0.05, and 0.07 g/mL respectively. Compared to results from the non-polarizable Sage force field, liquid densities

computed with polarizable electrostatic models are somewhat less accurate, but this is as expected given that the Sage FF was fitted against liquid state data using the non-polarizable AM1-BCC electrostatics model. We anticipate that retraining the LJ and torsional terms against liquid state data in the context of AM1-BCC-dPol will lead to liquid densities as accurate as those obtained with the non-polarizable AM1-BCC model.

Table 3: Condensed phase mass densities (g/mL) and inverse dielectric constants (1/D) of organic liquids from simulations and experiment computed with the following electrostatics models: element-typed AM1-BCC-dPol (E-dPol), Sage LJ-typed AM1-BCC-dPol (S-dPol), and AM1-BCC (noPol).

		Density				Inverse Dielectric Constant			
Liquid	Temp (K)	E-dPol	S-dPol	noPol	Expt.	E-dPol	S-dPol	noPol	Expt.
acetic acid	293.15	0.98	0.99	1.05	1.05	0.18	0.19	0.21	0.16
propan-2-one	298.15	0.75	0.70	0.78	0.78	0.08	0.19	0.21	0.05
ether	293.15	0.67	0.64	0.71	0.71	0.23	0.47	0.46	0.23
ethane-1,2-diol	293.15	1.09	1.05	1.11	1.11	0.05	0.08	0.07	0.03
heptane	293.15	0.72	0.74	0.68	0.68	0.55	0.55	0.99	0.52
hexane	293.15	0.69	0.71	0.66	0.66	0.56	0.56	0.99	0.53
octane	293.15	0.74	0.76	0.70	0.70	0.54	0.54	0.99	0.51
propane	293.15	0.51	0.53	0.50	0.49	0.64	0.64	0.99	0.60
prop-1-ene	220.15	0.58	0.58	0.61	0.61	0.61	0.59	0.98	0.47
pyridine	293.15	1.05	1.02	0.98	0.98	0.04	0.07	0.30	0.08
aniline	293.15	1.13	1.12	1.02	1.02	0.16	0.25	0.35	0.14
methylaniline	300.15	0.93	0.89	0.98	0.99	0.13	0.10	0.36	0.17
methanol	293.15	0.75	0.70	0.79	0.79	0.06	0.10	0.11	0.03
ethanol	293.15	0.80	0.68	0.79	0.79	0.09	0.19	0.14	0.04
2-propanol	293.15	0.83	0.74	0.79	0.79	0.11	0.18	0.19	0.05
benzene	293.15	0.92	0.89	0.87	0.88	0.51	0.52	0.96	0.44
ethane	95.15	0.60	0.60	0.54	0.64	0.59	0.61	0.99	0.52
RMSD vs. Expt		**0.02**	**0.05**	**0.07**		**0.05**	**0.06**	**0.33**	

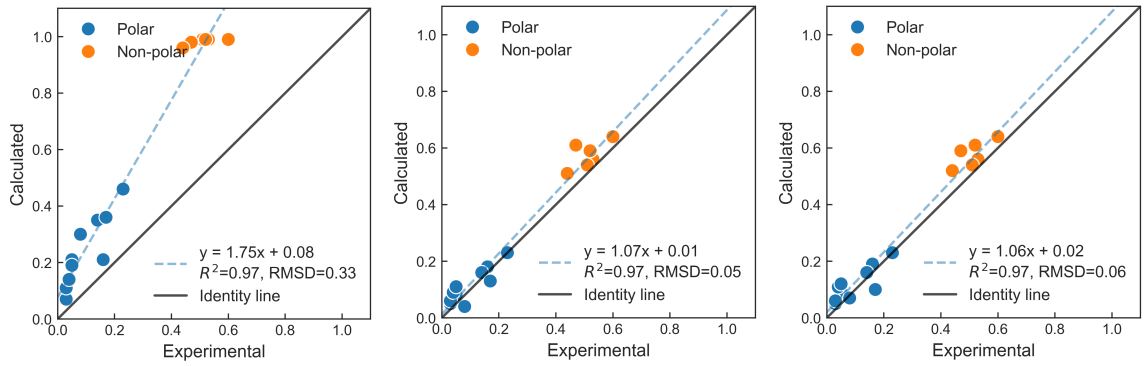


(a) Non-polarizable electrostatic AM1-BCC model

(b) Polarizable AM1-BCC-dPol model with element-based polarizabilities

(c) Polarizable AM1-BCC-dPol model with Sage LJ-based polarizabilities

Figure 5: Comparison of computed and experimental mass densities (g/mL) for simulations with the AM1-BCC fixed-charge model (5a) and with our AM1-BCC-dPol electrostatic model using element-based polarizabilities (5b) and Sage LJ-based polarizabilities (5c).



(a) Non-polarizable electrostatics model AM1-BCC.

(b) Polarizable AM1-BCC-dPol model with element-based polarizabilities

(c) Polarizable AM1-BCC-dPol model with Sage LJ-based polarizabilities.

Figure 6: Comparisons of computed and experimental inverse relative dielectric constants (1/D), for simulations with the AM1-BCC fixed-charge model (6a) and with our AM1-BCC-dPol electrostatic model using element-based polarizabilities (6b) and Sage LJ-based polarizabilities (6c).

We also find that going from the nonpolarizable AM1-BCC model to the AM1-BCC-dPol models dramatically improves the accuracy of the dielectric constants computed for organic liquids, (Figure 6 and Table 3), with root nean square deviation (RMSD) of the inverse dielectric constant (1/D) improving from 0.33 to 0.05 and the slope going from 1.75 to 1.07, in the case of element-based polarizabilities, and similar results for Sage LJ-based

polarizabilities. Although improvements are seen across the range of dielectric constants, nonpolar liquids show the greatest improvement. Precisely because they are nonpolar, the orientational polarizability of a nonpolar liquid is negligible, so omitting electronic polarizability causes them to have dielectric constants of about one; i.e., near zero polarizability. This has the problematic consequence that electrostatic interactions in a nonpolar liquid modeled with a nonpolarizable force field will be overestimated about 2-fold. Including electronic polarizability via the AM1-BCC-dPol model makes these liquids polarizable and thus corrects the dielectric constants to $\sim$ 2. These results suggest that AM1-BCC-dPol will more faithfully model electrostatic interactions in nonpolar media and the lipid membranes of cells.

Interestingly, the dielectric constants of a small range of alcohols are significantly underestimated in the AM1-BCC-dPol simulations (Table 3). Although this underestimation might trace to inadequate electronic polarizability, this seems unlikely, because AM1-BCC-dPol gives good agreement with experimental molecular polarizabilities for these compounds (Table S4). In addition, the dielectric constants of these highly polar liquids are dominated by orientational polarization, so that adding electronic polarization does not have a dramatic effect. This is apparent by comparing AM1-BCC-dPol with the nonpolarizable but otherwise similar Sage force field (Table 3). Note that previous simulations with the nonpolarizable GAFF/AM1-BCC FF similarly underestimate the dielectric constants of these liquids (methanol 20.0, ethanol 11.4, 2-propanol 10.5).[74] Although all of these underestimates might trace to too-weak partial charges, they may also result from an overly low Kirkwood g-factor[20] . The g-factor is the ratio of the liquid's dipole moment fluctuations to those which would be obtained if there were no correlations among the molecules in the liquid. Thus, a higher g-factor gives a higher dielectric constant, other things being equal. Importantly, the g-factor depends on not just the electrostatic model but also on other force field parameters, such as Lennard-Jones parameters. Thus, correcting these dielectric constants may require a more holistic refitting of the force field.

We chose the direct approximation of polarization primarily for the sake of computational speed relative to a full SCF treatment of inducible dipoles. We compared the average speed of 15 ns NPT production simulations using a standard fixed-charge AM1-BCC model, our direct approximation AM1-BCC-dPol, and a full self-consistent mutual polarization model using the same parameters as AM1-BCC-dPol and also executed with the MPID plugin in OpenMM, using keyword "direct" and "mutual" for direct and mutual polarization, respectively. All calculations were run on the same hardware. Simulations with direct polarization average 3.6-fold slower than fixed charge simulations, while simulations with the full mutual polarization calculations, with convergence set to 10^{-5}, average 4.5-fold slower. We also compared the speed of the direct approximation with the available OpenMM implementation of the empirical extrapolation scheme for efficient treatment of induced dipoles, namely the OPTn methods[27,75] , for two sample systems, ethane and ethanol. As detailed in Table S5, the direct method is about twice as fast as OPT3 and 50% faster than OPT1. (Runs with OPT0, which is theoretically equivalent to the direct polarization, failed for currently unknown reasons, and hence are not included here.)

Note that the MPID plugin used here carries along a full, anisotropic tensor representation of polarizability and a set of permanent multiples. Thus, we implement our simpler model zeroing the off-diagonal tensor elements and the permanent multipoles, but carrying along these more detailed terms necessarily slows the calculations. We are currently exploring how much further speedup can be attained with a more streamlined implementation. It is worth noting that the OpenMM implementation of iAMOEBA model achieved a speedup of simulations by a factor of 1.5 to 6 over full mutual polarization in the AMOEBA model, depending on the choices of convergence parameter in the SCF iterations.[29] Therefore, we expect that greater speed of our direct polarization model will be achievable by further software engineering of the OpenMM polarizability plugin.

4 Discussion

We have presented a novel polarizable electrostatic model that utilizes the direct polarization approximation, typed atomic polarizabilities, and a fast AM1-BCC-dPol charge assignment method. A RESP-inspired model, called RESP-dPol, is also available. Our overall approach is designed to capture the key consequences of electronic polarizability in simulations while remaining convenient to use and computationally tractable. The present results indicate that the polarizable AM1-BCC-dPol model provides a realistic representation of electrostatics at moderate computational cost. Its success may trace in part to the fact that, like some of the most widely used fixed-charge electrostatic models, our model is tuned to replicate gas-phase QM ESPs. Calculations of liquid dielectric constants indicate that the use of a polarizable electrostatic model is crucial to replicate these important experimental observables, for both non-polar and polar liquids.

We anticipate that the present approach will enhance the accuracy of simulations of systems where molecules move between regions with very different dielectric constants, such as water and a lipid membrane, or with very different ambient electric fields, such as from water to a protein pocket with many ionized side-chains. These are settings in which large changes in electronic polarization occur, making it important to account for them in detail. While the current polarizable electrostatic model can be used with valence and van der Waals terms from the existing OpenFF Sage force field in MD simulations, the accuracy of such simulations will be improved by adjusting the LJ and torsional parameters against reference experimental and QM data in the context of the polarizable force field. In the meantime, a compatible and fast polarizable water model is essential to achieve accurate simulations of molecular recognition. The Factor-Pol software infrastructure that implements both training and use of these models is shared open source at GitHub repository Factor-Pol.[61] It can be easily used in the OpenFF software ecosystem, i.e., with OpenFF Toolkit,[76] OpenFF Evaluator,[77] and OpenFF Interchange[78] to enable seamless use with the valence

and Lennard-Jones parameters of the OpenFF Sage force field in OpenMM. We have also implemented a version of the Smirnoff-plugin[79,80] , named MPID-plugin[81] , to introduce electronic polarization to the OpenFF force field fitting software stack. This integration sets the stage for planned future work aimed at generating more comprehensive force fields that fully integrate the polarizable electrostatics models. Currently, we are developing a polarizable water model that utilize direct polarization, atom-centered polarizabilities and partial charges.

In summary, we have presented a polarizable electrostatic model that is fast, due to its use of the direct polarization approximation, and both general and convenient to apply, due to the use of typed polarizabilities and the AM1-BCC-dPol method of assigning consistent partial charges. We believe that the present work provides the foundation of a generally parameterized polarizable force field that can be used in a wide range of applications in biomolecular simulations.

5 Acknowledgements

MKG acknowledges funding from the National Institute of General Medical Sciences (R01GM061300). These findings are solely of the authors and do not necessarily represent the views of the NIH.

6 Disclosures

MKG has an equity interest in and is a cofounder and scientific advisor of VeraChem LLC. He is also on the Scientific Advisor Boards of Denovicon.

DLM serves on the scientific advisory boards of OpenEye Scientific Software, Cadence Molecular Sciences and Anagenex, and is an Open Science Fellow with Psivant Sciences.

Supporting Information Available

The Supporting Information is available free of charge online at.

Table S1 and S2, typed polarizability parameters; Table S3, MD simulation speed (ns/day) using OpenMM; Table S4, benchmarking results for molecular polarizabilities; Table S5, speed comparison of various polarizability models; Figure S1-S5, chemical structures of molecules used in this work.

References

(1) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005**, *26*, 1668–1688.

(2) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **1983**, *4*, 187–217.

(3) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* **1988**, *110*, 1657–1666.

(4) Boothroyd, S.; Behara, P. K.; Madin, O. C.; Hahn, D. F.; Jang, H.; Gapsys, V.; Wagner, J. R.; Horton, J. T.; Dotson, D. L.; Thompson, M. W.; Maat, J.; Gokey, T.; Wang, L.-P.; Cole, D. J.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Shirts, M. R.; Mobley, D. L. Development and Benchmarking of Open Force Field 2.0.0: The Sage Small Molecule Force Field. *Journal of Chemical Theory and Computation* **2023**, *19*, 3251–3275.

(5) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993**, *97*, 10269–10280.

(6) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational Chemistry* **2000**, *21*, 132–146.

(7) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry* **2002**, *23*, 1623–1641.

(8) Jorgensen, W. L.; McDonald, N. A.; Selmi, M.; Rablen, P. R. Importance of Polarization for Dipolar Solutes in Low-Dielectric Media: 1, 2-Dichloroethane and Water in Cyclohexane. *Journal of the American Chemical Society* **1995**, *117*, 11809–11810.

(9) Jorgensen, W. L.; Jensen, K. P.; Alexandrova, A. N. Polarization Effects for Hydrogen-Bonded Complexes of Substituted Phenols with Water and Chloride Ion. *Journal of Chemical Theory and Computation* **2007**, *3*, 1987–1992.

(10) Schyman, P.; Jorgensen, W. L. Exploring Adsorption of Water and Ions on Carbon Surfaces Using a Polarizable Force Field. *The Journal of Physical Chemistry Letters* **2013**, *4*, 468–474.

(11) Zhang, J.; Yang, W.; Piquemal, J.-P.; Ren, P. Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. *Journal of Chemical Theory and Computation* **2012**, *8*, 1314–1324.

(12) Prajapati, J. D.; Mele, C.; Aksoyoglu, M. A.; Winterhalter, M.; Kleinekathöfer, U. Computational Modeling of Ion Transport in Bulk and through a Nanopore Using the Drude Polarizable Force Field. *Journal of Chemical Information and Modeling* **2020**, *60*, 3188–3203.

(13) Li, H.; Ngo, V.; Da Silva, M. C.; Salahub, D. R.; Callahan, K.; Roux, B.; Noskov, S. Y. Representation of Ion–Protein Interactions Using the Drude Polarizable Force-Field. *The Journal of Physical Chemistry B* **2015**, *119*, 9401–9416.

(14) Savelyev, A.; MacKerell, A. D. Balancing the Interactions of Ions, Water, and DNA in the Drude Polarizable Force Field. *The Journal of Physical Chemistry B* **2014**, *118*, 6742–6757.

(15) Grossfield, A.; Ren, P.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *Journal of the American Chemical Society* **2003**, *125*, 15671–15682.

(16) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual Review of Biophysics* **2019**, *48*, 371–394.

(17) Peng, X.; Zhang, Y.; Chu, H.; Li, Y.; Zhang, D.; Cao, L.; Li, G. Accurate Evaluation of Ion Conductivity of the Gramicidin A Channel Using a Polarizable Force Field without Any Corrections. *Journal of Chemical Theory and Computation* **2016**, *12*, 2973–2982.

(18) Onsager, L. Electric Moments of Molecules in Liquids. *Journal of the American Chemical Society* **1936**, *58*, 1486–1493.

(19) Kirkwood, J. G. On the Theory of Dielectric Polarization. *The Journal of Chemical Physics* **1936**, *4*, 592–601.

(20) Kirkwood, J. G. The Dielectric Polarization of Polar Liquids. *The Journal of Chemical Physics* **1939**, *7*, 911–919.

(21) Bottcher, C. J. *Theory of Electric Polarization, Vol. 1: Dielectrics in Static Fields*, 2nd ed.; Elsevier Science: Amsterdam, New York, 1973.

(22) Chung, M. K. J.; Miller, R. J.; Novak, B.; Wang, Z.; Ponder, J. W. Accurate Host–Guest Binding Free Energies Using the AMOEBA Polarizable Force Field. *Journal of Chemical Information and Modeling* **2023**, *63*, 2769–2782.

(23) Applequist, J.; Carl, J. R.; Fung, K.-K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *Journal of the American Chemical Society* **1972**, *94*, 2952–2960.

(24) Applequist, J. Atom charge transfer in molecular polarizabilities: application of the Olson-Sundberg model to aliphatic and aromatic hydrocarbons. *The Journal of Physical Chemistry* **1993**, *97*, 6016–6023.

(25) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B* **2010**, *114*, 2549–2564.

(26) Hogan, A.; Space, B. Next-Generation Accurate, Transferable, and Polarizable Potentials for Material Simulations. *Journal of Chemical Theory and Computation* **2020**, *16*, 7632–7644.

(27) Simmonett, A. C.; Pickard, F. C.; Ponder, J. W.; Brooks, B. R. An empirical extrapolation scheme for efficient treatment of induced dipoles. *The Journal of Chemical Physics* **2016**, *145*.

(28) Straatsma, T. P.; McCammon, J. A. Molecular Dynamics Simulations with Interaction Potentials Including Polarization Development of a Noniterative Method and Application to Water. *Molecular Simulation* **1990**, *5*, 181–192.

(29) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. *The Journal of Physical Chemistry B* **2013**, *117*, 9956–9972.

(30) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chemical Reviews* **2016**, *116*, 4983–5013.

(31) MacKerell, A. D.; Brooks, B.; Brooks, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parameterization. *Encyclopedia of Computational Chemistry* **1998**,

(32) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009**, *30*, 1545–1614.

(33) Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *The Journal of Chemical Physics* **2003**, *119*, 5185–5197.

(34) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *Journal of Chemical Theory and Computation* **2013**, *9*, 5430–5449.

(35) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell Jr, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chemical Reviews* **2016**, *116*, 4983–5013.

(36) Lin, F.-Y.; Huang, J.; Pandey, P.; Rupakheti, C.; Li, J.; Roux, B.; MacKerell Jr, A. D. Further Optimization and Validation of the Classical Drude Polarizable Protein Force Field. *Journal of Chemical Theory and Computation* **2020**, *16*, 3221–3239.

(37) Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry* **1991**, *95*, 3358–3363.

(38) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. Parametrizing a polarizable force field from ab initio data. I. The fluctuating point charge model. *The Journal of Chemical Physics* **1999**, *110*, 741–754.

(39) Patel, S.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *Journal of Computational Chemistry* **2003**, *25*, 1–16.

(40) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.

(41) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity-equalization method for the calculation of atomic charges in molecules. *Journal of the American Chemical Society* **1986**, *108*, 4315–4320.

(42) Li, A.; Voronin, A.; Fenley, A. T.; Gilson, M. K. Evaluation of Representations and Response Models for Polarizable Force Fields. *The Journal of Physical Chemistry B* **2016**, *120*, 8668–8684.

(43) Litman, J. M.; Liu, C.; Ren, P. Atomic polarizabilities for interactive dipole induction models. *Journal of Chemical Information and Modeling* **2021**, *62*, 79–87.

(44) Cieplak, P.; Caldwell, J.; Kollman, P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *Journal of Computational Chemistry* **2001**, *22*, 1048–1057.

(45) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *Journal of Computational Chemistry* **2002**, *23*, 1515–1531.

(46) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from ab Initio Quantum Chemistry. *The Journal of Physical Chemistry A* **2003**, *108*, 621–627.

(47) Antila, H. S.; Salonen, E. On combining Thole's induced point dipole model with fixed charge distributions in molecular mechanics force fields. *Journal of Computational Chemistry* **2015**, *36*, 739–750.

(48) Daylight Chemical Information Systems, I. SMARTS-A Language for Describing Molecular Patterns. `https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html`, 2007; [Accessed 11-Dec-2023].

(49) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *Journal of Chemical Theory and Computation* **2018**, *14*, 6076–6092.

(50) Reynolds, C. A.; Essex, J. W.; Richards, W. G. Atomic charges for variable molecular conformations. *Journal of the American Chemical Society* **1992**, *114*, 9075–9079.

(51) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, .; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.;

Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **2020**, *17*, 261–272.

(52) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry* **1990**, *11*, 431–439.

(53) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class IV charge models: A new semiempirical approach in quantum chemistry. *Journal of Computer-Aided Molecular Design* **1995**, *9*, 87–110.

(54) Schauperl, M.; Nerenberg, P. S.; Jang, H.; Wang, L.-P.; Bayly, C. I.; Mobley, D. L.; Gilson, M. K. Non-bonded force field model with advanced restrained electrostatic potential charges (RESP2). *Communications Chemistry* **2020**, *3*, 44.

(55) Boothroyd, S. OpenFF BCC Refit Study COH. `https://github.com/openforcefield/qca-dataset-submission/tree/master/submissions/2020-10-30-OpenFF-BCC-Refit-Study-COH`, 2020; [Accessed 15-Jun-2023].

(56) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, 618–622.

(57) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.

(58) Hickey, A. L.; Rowley, C. N. Benchmarking Quantum Chemical Methods for the Calculation of Molecular Dipole Moments and Polarizabilities. *The Journal of Physical Chemistry A* **2014**, *118*, 3678–3687.

(59) Zhou, A.; Schauperl, M.; Nerenberg, P. S. Benchmarking Electronic Structure Methods for Accurate Fixed-Charge Electrostatic Models. *Journal of Chemical Information and Modeling* **2019**, *60*, 249–258.

(60) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry* **1984**, *5*, 129–145.

(61) Wang, L. GitHub - wwilla7/Factor-Pol — github.com. `https://github.com/wwilla7/factorpol`, 2023; [Accessed 28-Jan-2023].

(62) Huang, J.; Simmonett, A. C.; Pickard, F. C.; MacKerell, A. D.; Brooks, B. R. Mapping the Drude polarizable force field onto a multipole and induced dipole model. *The Journal of Chemical Physics* **2017**, *147*.

(63) Simmonett, A. GitHub - andysim/MPIDOpenMMPlugin: OpenMM plugin that implements (an)isotropic polarizable point dipoles and multipoles up to octopoles. — github.com. `https://github.com/andysim/MPIDOpenMMPlugin`, 2017; [Accessed 11-12-2023].

(64) Lorentz, H. A. Ueber die Beziehung zwischen der Fortpflanzungsgeschwindigkeit des Lichtes und der Körperdichte. *Annalen der Physik* **1880**, *245*, 641–665.

(65) Lorenz, L. Ueber die Refractionsconstante. *Annalen der Physik* **1880**, *247*, 70–103.

(66) Thole, B. Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics* **1981**, *59*, 341–350.

(67) van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole's Model Revisited. *The Journal of Physical Chemistry A* **1998**, *102*, 2399–2407.

(68) Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations II: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies. *The Journal of Physical Chemistry B* **2011**, *115*, 3100–3111.

(69) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(70) 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 Computational Biology* **2017**, *13*, e1005659.

(71) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009**, *30*, 2157–2164.

(72) Wang, L. GitHub - wwilla7/factorpol-benchmarks-scripts — github.com. `https://github.com/wwilla7/factorpol-benchmarks-scripts`, 2023; [Accessed 17-Oct-2023].

(73) Neumann, M. Dipole moment fluctuation formulas in computer simulations of polar systems. *Molecular Physics* **1983**, *50*, 841–858.

(74) Beauchamp, K. A.; Behr, J. M.; Rustenburg, A. S.; Bayly, C. I.; Kroenlein, K.; Chodera, J. D. Toward Automated Benchmarking of Atomistic Force Fields: Neat Liquid Densities and Static Dielectric Constants from the ThermoML Data Archive. *The Journal of Physical Chemistry B* **2015**, *119*, 12912–12920.

(75) Simmonett, A. C.; Pickard, F. C.; Shao, Y.; Cheatham, T. E.; Brooks, B. R. Efficient treatment of induced dipoles. *The Journal of Chemical Physics* **2015**, *143*.

(76) Wagner, J.; Thompson, M.; Mobley, D. L.; Chodera, J.; Bannan, C.; Rizzi, A.; trevorgokey; Dotson, D. L.; Mitchell, J. A.; jaimergp; Camila; Behara, P.; Bayly, C.; Horton, J.; Wang, L.; Pulido, I.; Lim, V.; Sasmal, S.; SimonBoothroyd; Dalke, A.; Smith, D.; Horton, J.; Wang, L.-P.; Gowers, R.; Zhao, Z.; Davel, C.; Zhao, Y. openforcefield/openff-toolkit: 0.14.5 Minor feature release. 2023; `https://doi.org/10.5281/zenodo.10103216`, [Accessed 29-12-2023].

(77) Boothroyd, S.; Wang, L.-P.; Mobley, D. L.; Chodera, J. D.; Shirts, M. R. Open Force Field Evaluator: An Automated, Efficient, and Scalable Framework for the Estimation of Physical Properties from Molecular Simulation. *Journal of Chemical Theory and Computation* **2022**, *18*, 3566–3576.

(78) Thompson, M.; Wagner, J.; Gilmer, J. B.; Timalsina, U.; Quach, C. D.; Boothroyd, S.; Mitchell, J. A. OpenFF Interchange. 2022; `https://github.com/openforcefield/openff-interchange`, [Accessed 29-12-2023].

(79) Horton, J. T.; Boothroyd, S.; Behara, P. K.; Mobley, D. L.; Cole, D. J. A transferable double exponential potential for condensed phase simulations of small molecules. *Digital Discovery* **2023**, *2*, 1178–1187.

(80) Horton, J.; Thompson, M.; Wagner, J. OpenFF smirnoff-plugins: Plugins to enable using custom functional forms in SMIRNOFF based force fields. `https://github.com/openforcefield/smirnoff-plugins`, 2022; [Accessed 14-12-2023].

(81) Wang, L.; Thompson, M. OpenFF MPID Plugin. `https://github.com/openforcefield/MPID_plugin`, 2023; [Accessed 14-12-2023].

TOC Graphic