UNIVERSITY OF CALIFORNIA SAN DIEGO

Improving Force Field Accuracy for Molecular Modeling in Molecular Design

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Liangyue Drew

Committee in charge:

Professor Michael K. Gilson, Chair
Professor J. Andrew McCammon, Co-Chair
Professor Francesco Paesani
Professor Dionicio R. Siegel
Professor Joel Yuen-Zhou

2025

The dissertation of Liangyue Drew is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2025

TABLE OF CONTENTS

Dissertation Approval Page iii

Table of Contents iv

List of Figures vii

List of Tables ix

Acknowledgements x

Vita xi

Abstract of the Dissertation xii

Chapter 1 A Fast, Convenient, Polarizable Electrostatic Model for Molecular Dynamics 1
1.1 Introduction 1
1.2 Methods 5
1.2.1 Optimization of typed polarizabilities 6
1.2.2 RESP-dPol charge model 8
1.2.3 AM1-BCC-dPol charge model 9
1.2.4 Training sets and reference QM data 10
1.2.5 Infrastructure for parameterization 10
1.2.6 Benchmarking of electrostatics models 12
1.3 Results 15
1.3.1 Quality-of-fit to QM ESPs 15
1.3.2 Molecular dipole moments 17
1.3.3 Condensed phase molecular polarizabilities 18
1.3.4 Transferability of AM1-BCC-dPol across molecular conformations 19
1.3.5 Properties of organic liquids from simulations 21
1.4 Discussion 26
1.5 Disclosures 27
1.6 Acknowledgements 27
1.7 Supplementary Information 28
1.7.1 Supporting scripts and data 28
1.7.2 Polarizability parameters 28
1.7.3 MD simulation speed using OpenMM 30
1.7.4 Molecular polarizabilities 31
1.7.5 Chemical structure of molecules used in this work 34
1.7.6 TOC Graphic 40
Bibliography 40

Chapter 2 A Simple, Polarizable, Rigid, 3-point Water Model Using the Direct Polarization Approximation . . . 48
2.1 Introduction . . . 48
2.2 Methods . . . 53
2.2.1 Parameter Optimization . . . 55
2.2.2 Simulation Details . . . 58
2.2.3 Calculation of Physical Properties . . . 59
2.3 Results . . . 62
2.3.1 Initial TIP3P-dPol Model . . . 62
2.3.2 Optimization of dPol Water Models . . . 63
2.3.3 Benchmarks Against Additional Experimental Data . . . 70
2.3.4 Simulation Timing . . . 74
2.4 Discussion . . . 75
2.5 Disclosures . . . 77
2.6 Acknowledgements . . . 77
2.7 Supplementary Information . . . 78
2.7.1 Speed comparison of various polarizable force fields . . . 78
2.7.2 Additional Figures . . . 79
2.7.3 Code Availability . . . 82
2.7.4 TOC Graphic . . . 82
Bibliography . . . 82

Chapter 3 Incorporating the AM1-BCC-dPol Polarizable Model into a Widely Applicable General Force Field . . . 91
3.1 Introduction . . . 91
3.2 Methods . . . 93
3.2.1 Broadened Parameterization of the AM1-BCC-dPol Polarizable Electrostatics model . . . 94
3.2.2 Lennard-Jones Optimization . . . 97
3.2.3 Simulation Details . . . 98
3.3 Results . . . 99
3.3.1 Quality-of-fit to QM ESPs . . . 99
3.3.2 Molecular Dipole Moments . . . 100
3.3.3 Molecular Polarizabilities . . . 101
3.3.4 Physical Properties of Pure Liquids . . . 103
3.4 Discussion . . . 109
3.5 Disclosures . . . 109
3.6 Acknowledgements . . . 109
3.7 Supplementary Information . . . 110
3.7.1 Code availability . . . 110
3.7.2 Additional Figures . . . 110
Bibliography . . . 116

Chapter 4 Improving Force Field Accuracy for Metal-Organic Frameworks 122
4.1 Introduction 122
4.2 Methods 126
4.2.1 Force Fields 126
4.2.2 The Modified Seminario Method 127
4.2.3 Computational Details 129
4.2.4 Calculation of Material Properties 132
4.2.5 Parameter Optimization 134
4.3 Results 136
4.3.1 Force Field Optimization 136
4.3.2 Structural Properties 137
4.3.3 Mechanical Properties 141
4.3.4 Gas Adsorption Isotherm 144
4.4 Discussion 145
4.5 Acknowledgements 145
4.6 Supplementary Information 146
Bibliography 151

LIST OF FIGURES

Figure 1.1: Diagram of Factor-Pol toolkit. 11
Figure 1.2: Comparison of gas-phase molecular dipole moments. 18
Figure 1.3: Comparison of molecular polarizabilities. 19
Figure 1.4: Conformers of alanine dipeptide used to examine the transferability of partial charges. 21
Figure 1.5: Comparison of computed and experimental mass densities for two types of polarizabilities. 23
Figure 1.6: Comparisons of computed and experimental inverse relative dielectric constants (1/D) for two types of polarizabilities. 23
Figure 1.7: Dielectric constant and density calculation test set. 34
Figure 1.8: Polarizability training set. 35
Figure 1.9: Quality-of-fit to QM ESPs data set. 36
Figure 1.10: AM1-BCC-dPol library training set. 37
Figure 1.11: Molecular polarizability data set. 39
Figure 1.12: AM1-BCC-dPol model illustraton figure. 40

Figure 2.1: Contour plots of water properties vs Lennard-Jones parameters ϵ and σ. . . 69
Figure 2.2: Oxygen-oxygen radial distribution function and coordination numbers of several dPol models and other models. 72
Figure 2.3: Calculated temperature-dependent properties using dPol water models. . . 74
Figure 2.4: Distribution of isolated dipole moments in liquid water. 79
Figure 2.5: Convergence of dielectric constants. 79
Figure 2.6: Self-diffusion coefficient as a function of the inverse box size. 80
Figure 2.7: MSD of dPol-opt5 used to compute self-diffusion coefficients using different box sizes. 80
Figure 2.8: RDFs of unnamed water model from Scan 2. 81
Figure 2.9: Polynomial fits to locate possible parameter space. 81
Figure 2.10: dPol water model illustration figure. 82

Figure 3.1: Comparison of 500 gas-phase dipole moments (Debye) computed with the AM1-BCC-dPol mode with QM reference 101
Figure 3.2: Comparison of MM and QM molecular polarizabilities. 103
Figure 3.3: Comparison of calculated and experimental density and heat of vaporization using LJ parameters drawn from RESP2$_{\sigma=0.6}$. 104
Figure 3.4: Comparison of calculated and experimental density and heat of vaporization using the optimized LJ parameters. 105
Figure 3.5: Comparison of calculated and experimental hydration free energy. 107
Figure 3.6: Chemical structures of molecules used in calculation for dipole moments. . 111
Figure 3.7: Labeled comparison of calculated and experimental solvation free energy. . 114
Figure 3.8: Chemical structures of dataset for solvation free energy calculations. 115

Figure 4.1: Metal centers of MOL-5 and MIL-53(Al). 130
Figure 4.2: Metal coordination bonds of MOF-5. 135
Figure 4.3: Stress-Strain relation (C12) computed from linear regression fit with periodic DFT and OPLS4. 136
Figure 4.4: Static visualization of the harmonic motion studied in this work. 140
Figure 4.5: Potential energy curves along the normal mode displacement. 141
Figure 4.6: CO_2 adsorption isotherm of MOF-5 144
Figure 4.7: Convergence tests for the energy cutoffs and K-point grids for periodic DFT calculations in Quantum ESPRESSO. 149
Figure 4.8: MOF-5 input structure is well-equilibrated in a 15 ns MD simulation. . . . 150
Figure 4.9: Loss function as a function of iteration steps for force field optimization. . 150

LIST OF TABLES

Table 1.1: Quality-of-fit of RESP-dPol and AM1-BCC-dPol electrostatic models to baseline QM ESPs. 16
Table 1.2: Average RRMS errors (unitless) of ESPs, relative to reference QM ESPs. . . 20
Table 1.3: Condensed phase mass densities (g/mL) and inverse dielectric constants (1/D). 22
Table 1.4: Element-based polarizabilities ($Å^3$) 28
Table 1.5: Sage LJ-typed polarizabilities ($Å^3$). 29
Table 1.6: MD simulation speed (ns/day) of AM1-BCC-dPol model with OpenMM. . . 30
Table 1.7: Experimental and calculated molecular polarizabilities ($Å^3$) using typed atomic polarizabilities. 31
Table 1.8: Speed comparison of various polarizability models. 33

Table 2.1: Force field parameters of several dPol water model candidates. 63
Table 2.2: Physical properties of water resulted during the optimization and compared to experimental (EXPT) reference and other popular water models at 298K and 1atm. 64
Table 2.3: Speed comparison of various polarizable force fields. 75
Table 2.4: Structure of the present dPol model with those of other polarizable water models. 76

Table 3.1: RRMS (%) of AM1-BCC-dPol and AM1-BCC to baseline QM ESPs. 100
Table 3.2: Atomic polarizabilities ($Å^3$) derived from fitting against V_{diff} 102
Table 3.3: Starting and final optimized LJ parameters. 105
Table 3.4: Liquid properties used as training data. 106
Table 3.5: Hydration free energies of organic molecules from simulations and experiments. 108

Table 4.1: Comparison of MOF-5 elastic tensors between DFT, OPLS4, final force field model (custom OPLS), starting model derived from QM Hessian. 137
Table 4.2: Metal-ligand coordination geometry and relevant force constants derived from QM Hessian. 138
Table 4.3: Comparison of MOF-5 metal-ligand coordination geometry computed from force fields and periodic DFT. 139
Table 4.4: Mechanical properties computed for MOF-5. 142
Table 4.5: Mechanical properties computed for MIL-53(Al). 143
Table 4.6: DFT vibrational frequencies computed for MOF-5 metal center. 146

ACKNOWLEDGEMENTS

Chapter 1, in full, is a reprint of the material as it appears in the Journal of Chemical Theory and Computation, 2024, Wang, Liangyue; Schauperl, Michael; Mobley, David L.; Bayly, Christopher; Gilson, Michael K. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, has been submitted for publication of the material as it may appear in Journal of Chemical Theory and Computation, 2025, Drew, Liangyue W.; Gilson, Michael K. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material; Drew, Liangyue W.; Gilson, Michael K. The dissertation author was the primary investigator and author of this paper.

VITA

2019	Bachelor of Science, South China University of Technology
2020	Master of Science, University of California San Diego
2025	Doctor of Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Improving Force Field Accuracy for Molecular Modeling in Molecular Design

by

Liangyue Drew

Doctor of Philosophy in Chemistry

University of California San Diego, 2025

Professor Michael K. Gilson, Chair
Professor J. Andrew McCammon, Co-Chair

Force fields are widely used computational tools in molecular modeling, providing predictive powers in studying molecular structures, properties, and dynamics. These mathematical functional forms are simplified but effective methods for calculating the interactions between atoms in a molecular system. Traditional additive force fields describe the potential energy of a system by considering bonded interactions and nonbonded interactions. The bonded interactions include bond stretching, angle bending, and torsional rotation, and the nonbonded interactions include van der Waals interactions and electrostatic interactions. Force Fields enable molecular dynamics simulations to study the behaviors of molecules, providing insights into drug design,

materials science, and biochemistry. The development of force fields has progressed to address increasing complexity in molecular systems, incorporating electronic polarization, automatic parameterization, bespoke fitting, and enhanced computational efficiency. In molecular design, force fields allow for the exploration of large chemical space, optimization of molecular conformations, and fast prediction of molecular properties, facilitating the design of novel compounds with desired characteristics. In this dissertation, I aim to develop new methods to improve force field accuracy and advance the balance between computational efficiency and accuracy. I examine several key aspects that influence the accuracy of force fields, focusing on the training data, parameterization methods, and explicit treatment of electronic polarization. First, I show a fast and efficient polarizable electrostatics model for molecular dynamics simulation. Second, I propose a new polarizable water model for improved solvent effects with moderate computational cost. Third, I develop non-bonded force field parameters accompanying the electrostatics model for simulations for physical properties. Furthermore, I explore methods to improve force field accuracy for physical property predictions in metal-organic frameworks, proposing a data-driven parameterization approach to guide the design of functional framework structures. In summary, these advancements aim to improve the accuracy of force fields for molecular modeling that effectively guides molecular design.

Chapter 1

A Fast, Convenient, Polarizable Electrostatic Model for Molecular Dynamics

1.1 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, E_j, leads to an induced dipole as:

$$\mu_{\text{ind,j}} = \alpha_j E_j \tag{1.1}$$

where α_j represents either a scalar polarizability or a polarizability tensor, and in general $E_j = E_{j,\text{ext}} + E_{j,\text{perm}} + E_{j,\text{ind}}$, 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 ($E_{j,\text{ind}}$) 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 Simmonett 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 $E_j \approx E_{j,\text{ext}} + E_{j,\text{perm}}$. 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 of 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[30, 34, 35].

The fluctuating charge (FQ) model describes polarization by allowing charge to flow between atoms in response to the electric field [36–38], so neither inducible dipoles nor added charge centers are required. The modified charges are typically obtained via an electronegativity equalization approach [39, 40], 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[41].

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[42–46]. 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.

1.2 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.

1.2.1 Optimization of typed polarizabilities

Both polarizable electrostatic models developed here (Sections 1.2.2, 1.2.3) 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[47] and parameterized in the SMIRKS Native Open Force Field (SMIRNOFF) format[48]. 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 1.2.4), to reduce concerns about conformational dependencies in our ESP fits[5, 49].

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, $V_{\text{diff,ikl}}$, 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[50] to find typed polarizabilities that minimize the following error function,

$$\chi^2 = \sum_{k=1}^{N_{\text{conf}}} \sum_{l=1}^{6} \sum_{i=1}^{m_k} \left(V_{\text{diff,ikl}} - \sum_{j=1}^{n_k} \frac{\mu_{\text{ind,jkl}} r_{\text{ijk}}}{r_{\text{ij}}^3} \right)^2 \tag{1.2}$$

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, $\mu_{\text{ind, jkl}}$ is the induced dipole on atom j of conformer k with external field direction l, and r_{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 1.2 are given simply by

$$\mu_{\text{ind,jkl}} = \alpha_j E_{\text{ext,jkl}} \tag{1.3}$$

where α_j is the typed polarizability assigned to atom j and $E_{\text{ext,jkl}}$ 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.

1.2.2 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.[43] 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 1.2.5) 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:

$$\chi^2 = \sum_{i=1}^{m} \left(V_{\mathrm{QM,\,i}} - V_{\mathrm{perm,\,i}} - V_{\mathrm{ind,\,i}}\right)^2 + \lambda \left(\sum_{j=1}^{n} q_{\mathrm{j}} - q_{\mathrm{tot}}\right) + a \sum_{j=1}^{n} \left(\sqrt{q_{\mathrm{j}}^2 + b^2} - b\right) \tag{1.4}$$

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:

$$V_{\mathrm{ind,\,i}} = \sum_{j=1}^{n} \frac{\mu_{\mathrm{ind,\,j}} r_{\mathrm{ij}}}{r_{\mathrm{ij}}^3} \tag{1.5a}$$

$$\mu_{\mathrm{ind,\,j}} = \alpha_{\mathrm{j}} \sum_{k \neq j}^{n} f_{\mathrm{jk}} \frac{q_{\mathrm{k}} r_{\mathrm{jk}}}{r_{\mathrm{jk}}^3} \tag{1.5b}$$

A Lagrange multiplier, λ, is used to constrain the sum of atomic charges (q_{j}) to the correct molecular charge, q_{tot} [49, 51]. The last term in Equation 1.4 is the hyperbolic restraint described above.

1.2.3 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 [52] and bond charge correction (BCC) parameters (B_β) predefined based on bond connectivity analysis ($T_{j\beta}$):

$$q_{\mathrm{j}} = q_{\mathrm{j}}^{\mathrm{pre}} + \sum_{\beta=1}^{n_\beta} T_{j\beta} B_\beta \tag{1.6}$$

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:

$$\chi^2 = \sum_{i=1}^{m} \left(V_{\mathrm{QM,\,i}} - V_{\mathrm{pre,\,i}} - \sum_{j=1}^{n} \sum_{\beta=1}^{n\beta} \frac{T_{j\beta} B_\beta}{r_{\mathrm{ij}}} - V_{\mathrm{ind,\,i}} \right)^2 \tag{1.7}$$

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 1.2.4), much as

detailed in Equation 1.2. 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.

1.2.4 Training sets and reference QM data

The polarizability training set consists of 39 small molecules from the RESP2[53] training dataset. An additional set of 80 small molecules from the OpenFF BCC refit study (COH dataset)[54], 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[55, 56] QM level of theory, because prior benchmarks of electronic structure methods for molecular mechanics[57, 58] and polarizable force fields[25, 30, 42] 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[51, 59] 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.

1.2.5 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)[60].

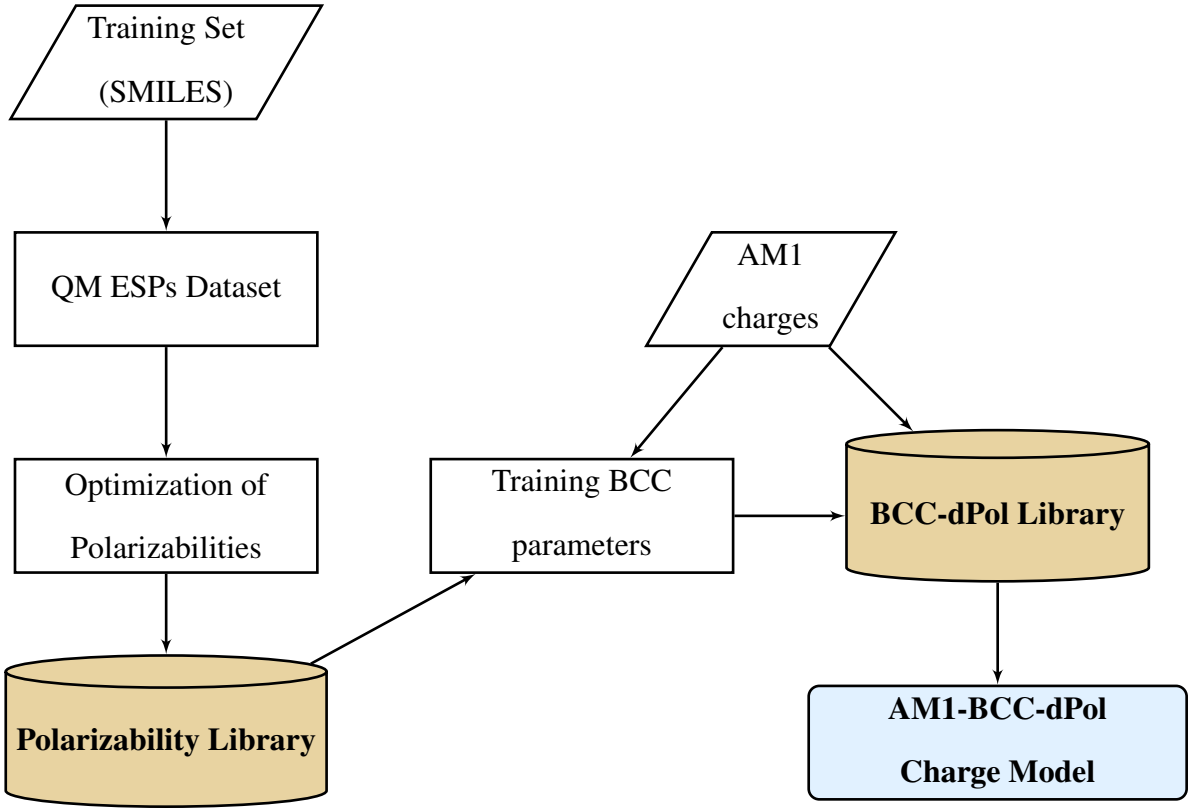


Figure 1.1: Diagram of Factor-Pol toolkit used to carry out QM calculations, derive optimized, typed polarizabilities and polarization-adapted BCCs, and assign RESP-dPol and AM1-BCC-dPol charges and polarizabilities to molecules.

As illustrated in Figure 1.1, 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[61, 62].

1.2.6 Benchmarking of electrostatics models

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 1.8:

$$\mathrm{RRMSE} = \left(\frac{1}{m} \frac{\sum_{i=1}^{m} (V_{\mathrm{Calc,\,i}} - V_{\mathrm{QM,\,i}})^2}{\sum_{i=1}^{m} (V_{\mathrm{QM,\,i}})^2} \right)^{1/2} \tag{1.8}$$

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

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 [51]. Hoewever, 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 calculated as:

$$\mu = \sum_{j=1}^{N} \left(q_j r_j + \mu_{\text{ind, j}} \right) \tag{1.9}$$

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

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[63, 64]:

$$\frac{n^2 - 1}{n^2 + 2} = \frac{4}{3} \pi \rho \alpha_M \tag{1.10}$$

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:

$$\alpha_M = \sum_{j=1}^{N} \alpha_j \tag{1.11}$$

where N is the number of atoms.

Molecular polarizabilities have been used to derive atomic polarizabilities with an additivity method in previous work[65–67]. 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[66, 68].

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[69]. The MPID plugin[61] 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[70] as simulation systems. Scripts to set up simulations are available online[71].

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:

$$\rho = \frac{M}{\langle V \rangle} \tag{1.12}$$

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, 72]:

$$\varepsilon = \varepsilon_\infty + \frac{\langle \mu^2 \rangle - \langle \mu \rangle^2}{3\varepsilon_0 k_B T \langle V \rangle} \tag{1.13}$$

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

$$\varepsilon_\infty = \frac{1}{\varepsilon_0 V} \frac{\mu_{\text{ind}}}{E} + 1 \tag{1.14}$$

where 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:

$$\varepsilon_\infty = \frac{1}{\varepsilon_0 V} \sum_{j=1}^{N} \alpha_j + 1 \tag{1.15}$$

1.3 Results

1.3.1 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 1.1 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[73], 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 1.1: 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

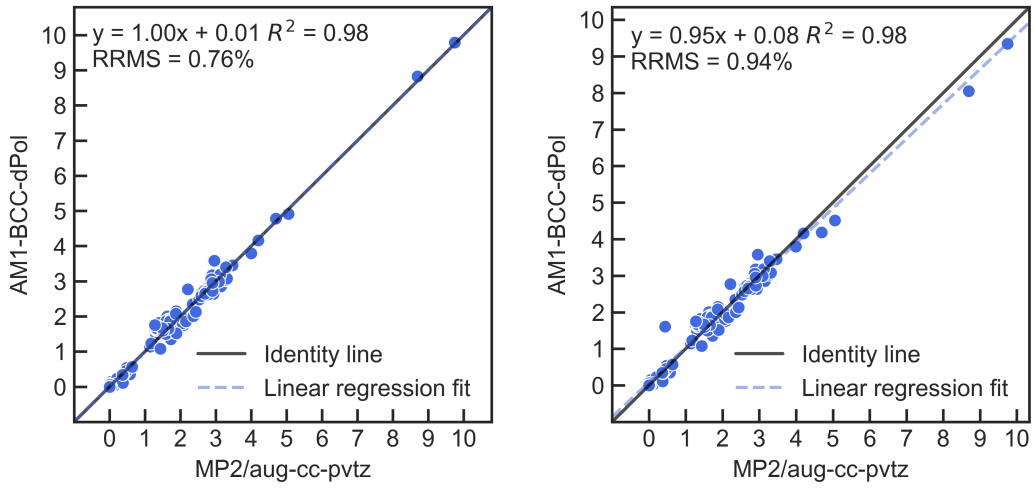
The Sage LJ-based polarizabilities are in principle tuned for the chemical environment of each atom in a molecule, whereas the element-based polarizabilities do not depend on chemical environments. Thus, Sage LJ types might be expected to provide a more accurate description of electronic polarization. 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.

1.3.2 Molecular dipole moments

Previous benchmarking studies on electronic structure methods for gas phase dipole moments[57, 58] 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 1.2, 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

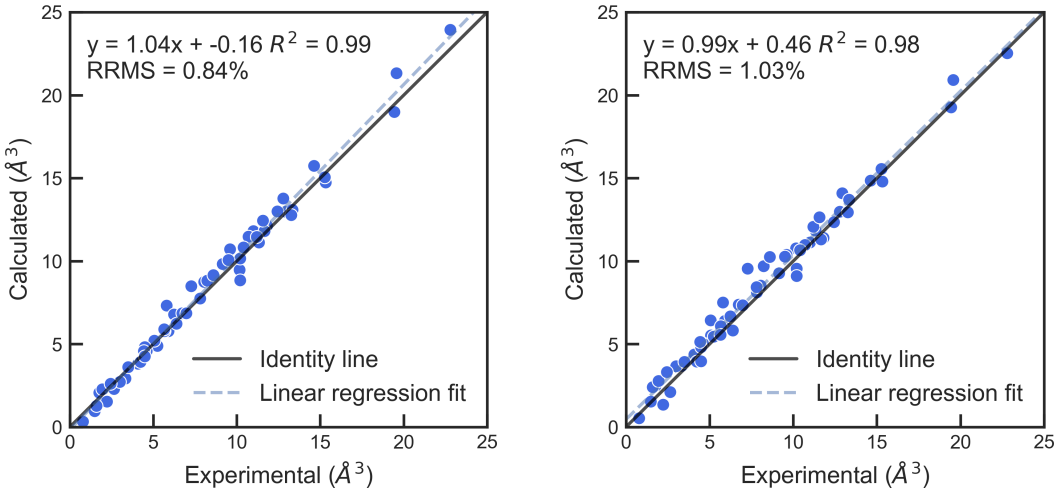
Figure 1.2: Comparison of gas-phase molecular dipole moments (Debye) computed with the AM1-BCC-dPol electrostatics model with QM reference results, for element-based polarizabilities (1.2a) and Sage LJ type-based polarizabilities (1.2b).

1.3.3 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 1.11) agree well with those derived from experimental indices of refraction, as shown in Figure 1.3. 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 1.3: Comparison of molecular polarizabilities computed from our atomic polarizabilities, against experimental reference data, for element-based polarizabilities (1.3a) and Sage LJ type-based polarizabilities (1.3b).

1.3.4 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, 49]. 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 1.4a) and conformer 1 without (Figure 1.4b) an

intramolecular hydrogen bond. Table 1.2 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 1.2: 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.

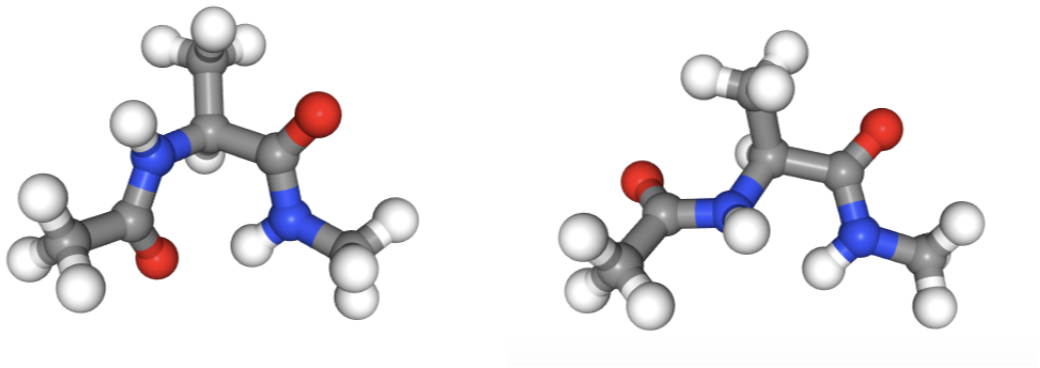
AM1-BCC-dPol	conformer 0	conformer 1
conformer 0	0.82	0.97
conformer 1	0.80	0.75

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 Table1.2, 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 1.4: 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 1.2).

1.3.5 Properties of organic liquids from simulations

We ran simulations of 17 organic liquids (Figure 1.7) 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:

$$U_{\text{elec}} \propto \frac{q_1 q_2}{\varepsilon_0 D} \tag{1.16}$$

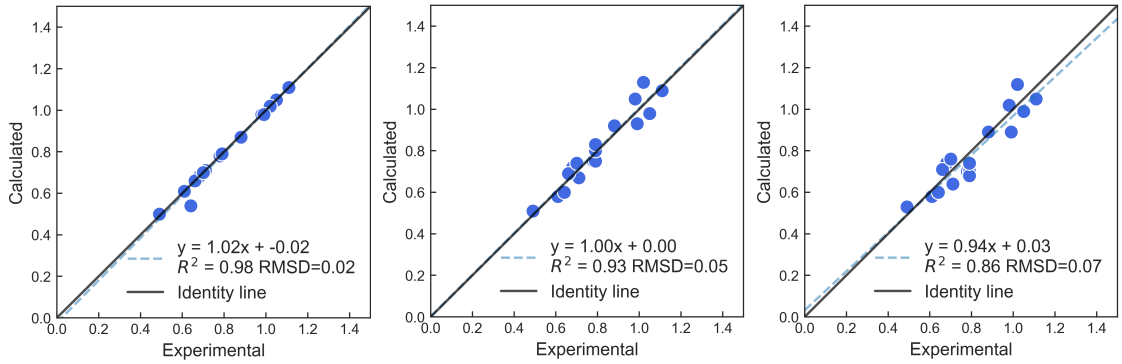
where ε_0 is the permittivity of free space and D is the relative dielectric constant.

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 1.5 and Table 1.3) 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 1.3: 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		**0.05**	**0.07**	**0.02**		**0.05**	**0.06**	**0.33**	

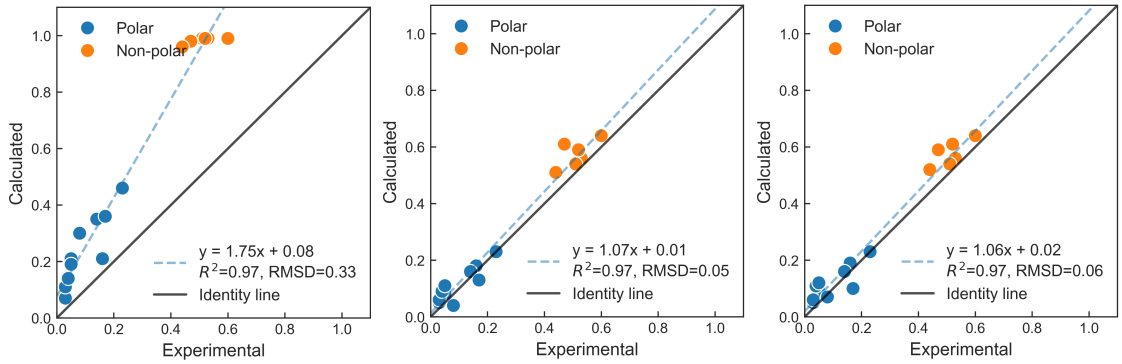


(a) Non-polarizable electrostatic AM1-BCC model

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

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

Figure 1.5: Comparison of computed and experimental mass densities (g/mL) for simulations with the AM1-BCC fixed-charge model (1.5a) and with our AM1-BCC-dPol electrostatic model using element-based polarizabilities (1.5b) and Sage LJ-based polarizabilities (1.5c).



(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 1.6: Comparisons of computed and experimental inverse relative dielectric constants (1/D), for simulations with the AM1-BCC fixed-charge model (1.6a) and with our AM1-BCC-dPol electrostatic model using element-based polarizabilities (1.6b) and Sage LJ-based polarizabilities (1.6c).

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 1.6 and Table 1.3), 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 ~ 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 1.3). 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 1.7). 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 1.3). 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 1.8, 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.

1.4 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[60]. 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.

1.5 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.

1.6 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.

Chapter 1, in full, is a reprint of the material as it appears in the Journal of Chemical

Theory and Computation, 2024, Wang, Liangyue; Schauperl, Michael; Mobley, David L.; Bayly, Christopher; Gilson, Michael K. The dissertation author was the primary investigator and author of this paper.

1.7 Supplementary Information

1.7.1 Supporting scripts and data

Supporting scripts and data to reproduce all data are available at `https://github.com/wwilla7/factorpol-benchmarks-scripts`.

1.7.2 Polarizability parameters

Table 1.4: Element-based polarizabilities ($Å^3$)

Element Type	Polarizability
C	1.642
H	0.163
O	0.642
N	1.042

Table 1.5: Sage LJ-typed polarizabilities (Å^3).

SMIRNOFF Pattern	Polarizability
[#1:1]	0.272
[#1:1]-[#6X4]	0.132
[#1:1]-[#6X4]-[A[1]]	0.161
[#1:1]-[#6X4](-[A[1]])-[A[1]]	0.161
[#1:1]-[#6X4](-[A[1]])(-[A[1]]	0.161
[#1:1]-[#6X4]~[*+1,*+2]	0.132
[#1:1]-[#6X3]	0.198
[#1:1]-[#6X3]~[A[1]]	0.271
[#1:1]-[#6X3](~[A[1]])~[A[1]]	0.271
[#1:1]-[#6X2]	0.372
[#1:1]-[#7]	0.040
[#1:1]-[#8]	0.320
[#6:1]	1.579
[#6X2:1]	1.593
[#6X4:1]	1.593
[#8:1]	1.203
[#8X2H0+0:1]	1.301
[#8X2H1+0:1]	1.301
[#7:1]	1.239
[#1]-[#8X2H2+0:1]-[#1]	1.201
[#1:1]-[#8X2H2+0]-[#1]	0.172

[1] A stands for pattern [#7,#8,#9,#16,#17,#35]

1.7.3 MD simulation speed using OpenMM

Table 1.6: MD simulation speed (ns/day) using OpenMM as MD engine and the MPID plugin for the polarizability models. AM1-BCC-dPol: simulations with AM1-BCC-dpol. AM1-BCC-mPol: matched simulation but with SCF turned on in the MPID plugin. AM1-BCC: simulations with the non-polarizable AM1-BCC model.

SMILE	natoms	AM1-BCC-dPol	AM1-BCC-mPol	AM1-BCC
CO	6	431	351	1287
CC(=O)O	8	404	330	1193
CC	8	384	309	1298
CC=C	9	382	312	1325
CCO	9	381	310	1237
C(CO)O	10	353	297	1210
CC(=O)C	10	371	312	1225
CCC	11	356	288	1228
C1=CC=NC=C1	11	317	262	1229
c1ccccc1	12	387	308	1298
CC(O)C	12	323	260	1310
C1=CC=C(C=C1)N	14	283	239	1339
CCOCC	15	286	239	1233
CNC1=CC=CC=C1	17	288	235	1238
CCCCCC	20	268	224	1151
CCCCCCC	23	231	191	1154
CCCCCCCC	26	224	188	1013
mean	13	334	274	1234

1.7.4 Molecular polarizabilities

Table 1.7: Experimental and calculated molecular polarizabilities ($Å^3$) using typed atomic polarizabilities.

SMILES	Experimental	Element-based	Sage LJ-based
C(CCO)CO	10.16	9.48	10.78
CC(C)(C)C	10.20	10.17	9.54
CCO	5.08	4.90	5.52
C1=CC=C(C=C1)[N+](=O)[O-]	12.92	12.99	14.11
CCCCCCCCCCCC	22.80	23.94	22.54
C1CCCCC1	11.00	11.81	11.14
N#N	1.76	2.08	2.48
[H][H]	0.79	0.33	0.54
C(=O)N	4.08	3.81	4.37
C1CCCCCCC1	14.63	15.74	14.85
CNC=O	5.91	5.78	6.41
COC	5.24	4.90	5.45
C=C	4.22	3.94	3.95
CCCCCC	11.80	12.13	11.40
CCCCC=C	11.65	11.81	11.31
N#CCC#N	5.79	7.34	7.52
C1C=CC(=O)O1	7.27	8.50	9.55
CCC#N	6.24	6.78	6.68
COC=O	5.05	5.22	6.43
CC=O	4.59	4.58	5.04
CCCC(=O)OC	11.33	11.12	11.86
C1CCCCCC1	12.79	13.78	12.99
CCCCC(=O)OC	13.34	13.09	13.72
CC#N	4.48	4.82	4.82
C	2.65	2.29	2.12
CC(C)(C)C#N	9.59	10.72	10.39

Continued on next page

Table 1.7 – continued from previous page

SMILES	Experimental	Element-based	Sage LJ-based
CCOCC	10.20	8.84	9.11
CC(=O)N	5.67	5.78	6.09
O	1.49	0.97	1.54
CC(=O)NC	7.82	7.75	8.12
CCC	6.38	6.23	5.83
CC(C)C#N	8.05	8.75	8.53
N	2.22	1.53	1.36
C1CCC=CC1	10.70	11.48	10.98
C1CCCC1	9.15	9.84	9.28
CCCCCCCO	15.33	14.74	14.80
c1ccncc1	9.49	10.07	10.27
CCCCCCO	13.27	12.78	12.95
C1CO1	4.43	4.58	5.13
c1ccc(cc1)O	11.20	11.47	12.08
O=O	1.60	1.28	2.41
[C-]#[O+]	1.95	2.28	2.78
CC	4.48	4.26	3.98
CCCO	6.74	6.87	7.38
CC(C)O	6.97	6.87	7.35
CO	3.32	2.94	3.70
C1CCC(CC1)N	12.44	13.01	12.35
CCCCCCCC(=O)OC	19.44	19.00	19.28
[N-]=[N+]=O	3.00	2.73	3.68
c1ccccc1	10.40	10.83	10.66
C1CC(=O)OC1	8.23	8.83	9.71
C=O	2.45	2.61	3.32
c1ccc(cc1)c2ccccc2	19.57	21.33	20.92
C#C	3.49	3.61	3.93
CCCCCC(=O)OC	15.27	15.06	15.57

Continued on next page

Table 1.7 – continued from previous page

SMILES	Experimental	Element-based	Sage LJ-based
C1CC1	5.64	5.90	5.57
C1COCCO1	8.60	9.16	10.26
C1CCC(CC1)O	11.56	12.45	12.66
CN(C)C=O	7.81	7.75	8.44

Table 1.8: Speed comparison of various polarizability models. Calculations were performed with AM1-BCC-dPol and Sage, using the MPID plugin in OpenMM. Units are ns/day. Systems consist of 256 molecules. Our direct polarization method is denoted as dPol.

Systems	OPT1	OPT2	OPT3	dPol
Ethane	264	228	203	380
Ethanol	273	229	205	376

1.7.5 Chemical structure of molecules used in this work

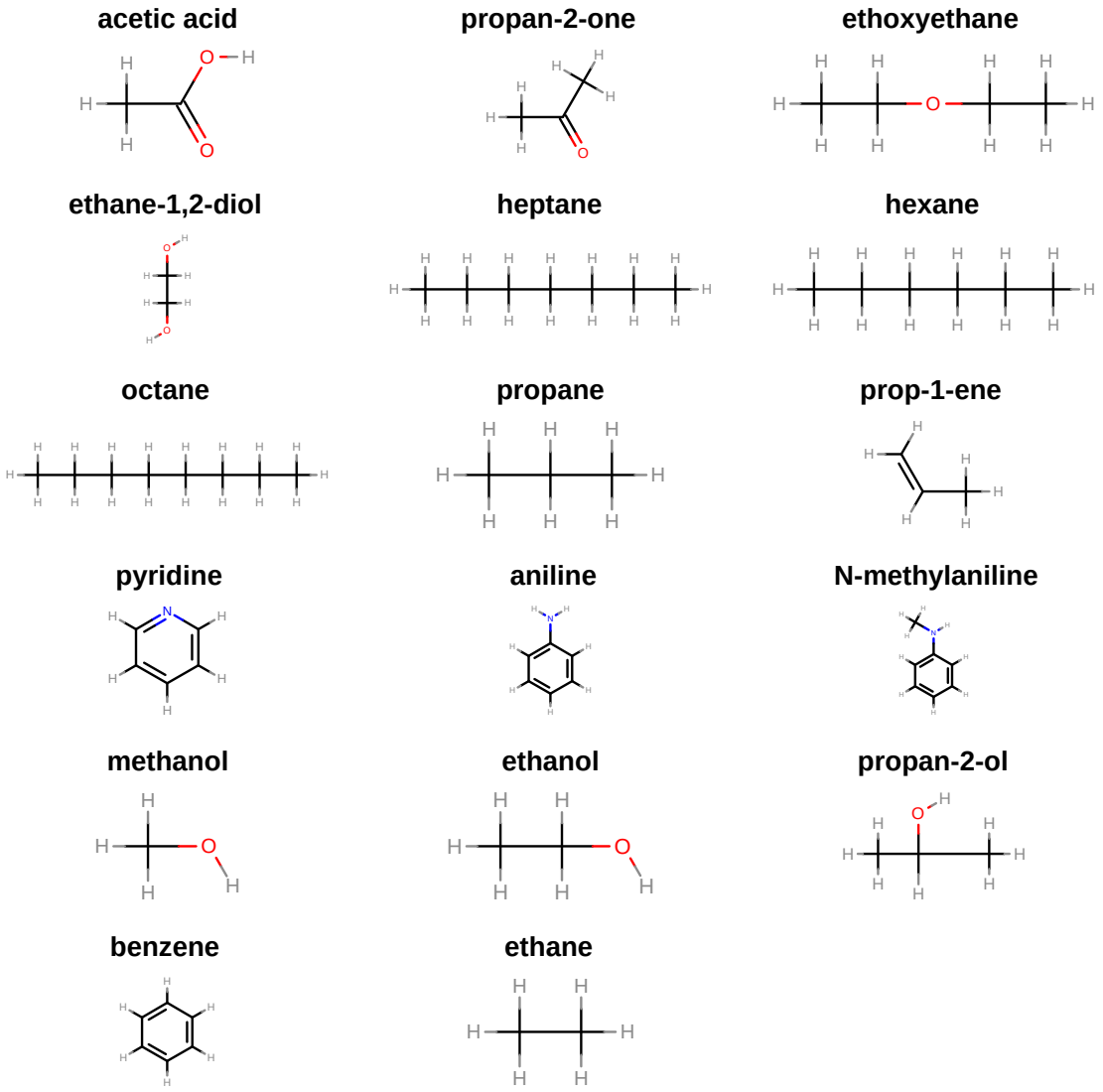


Figure 1.7: Dielectric constant and density calculation test set.

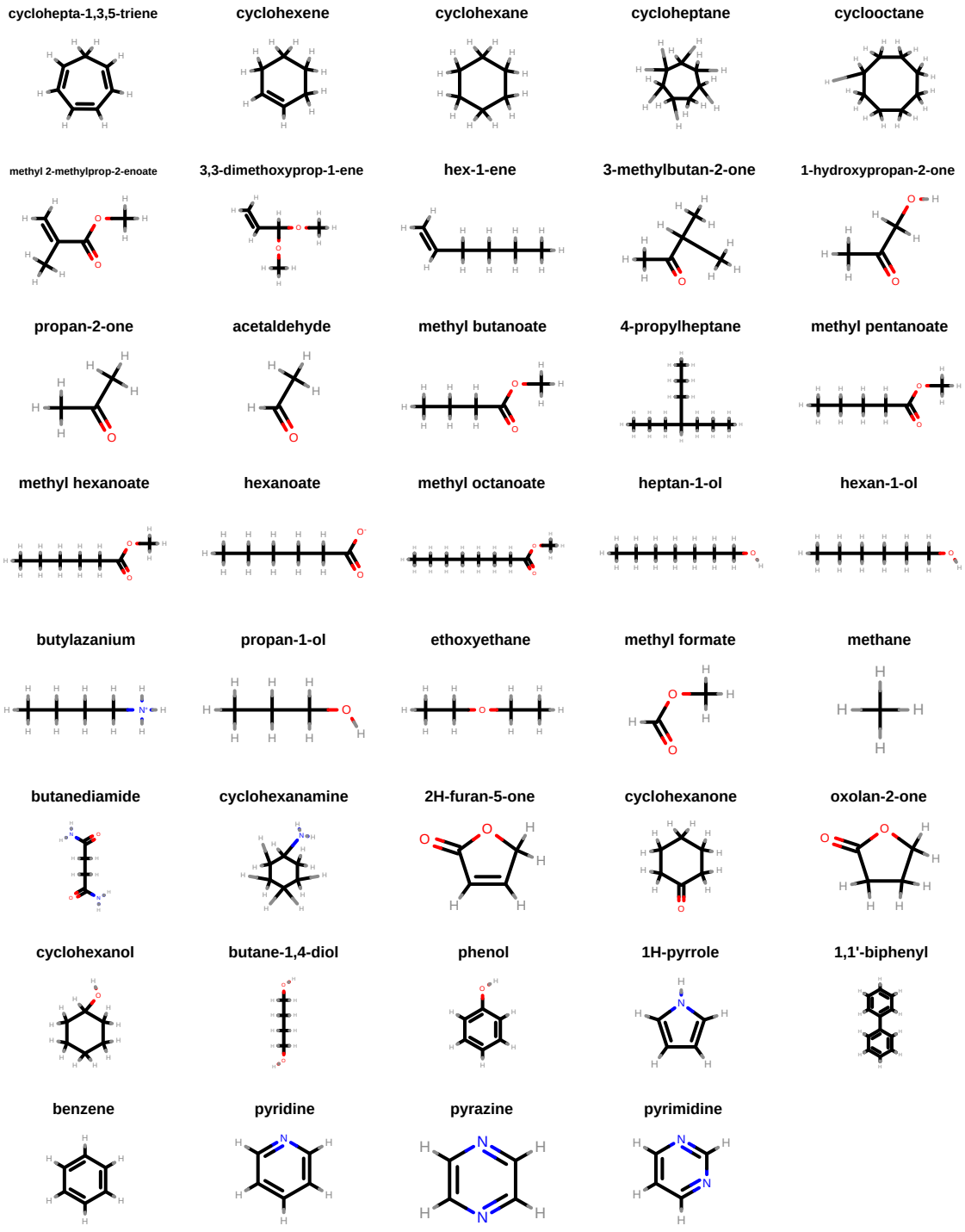


Figure 1.8: Polarizability training set.

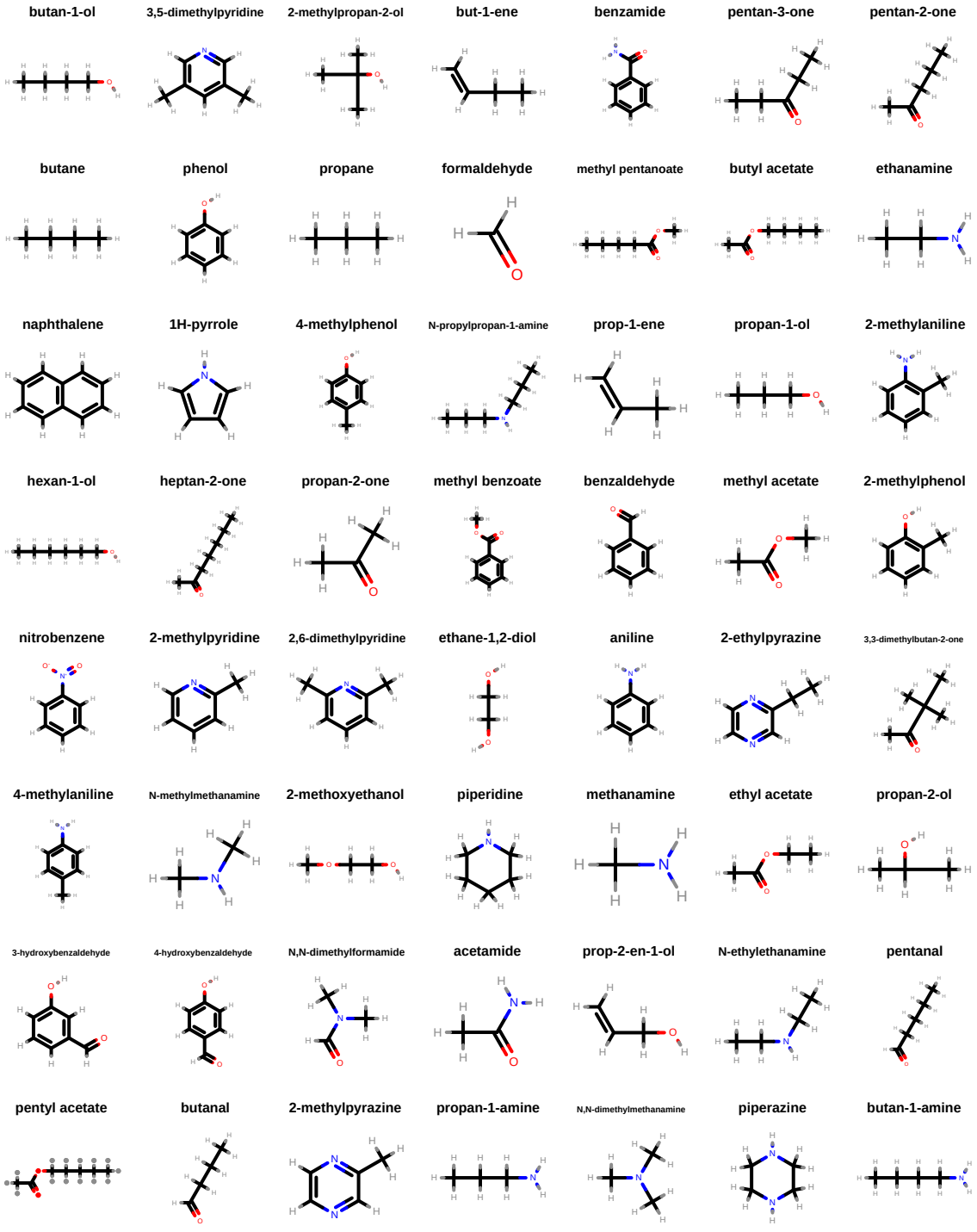


Figure 1.9: Quality-of-fit to QM ESPs data set.

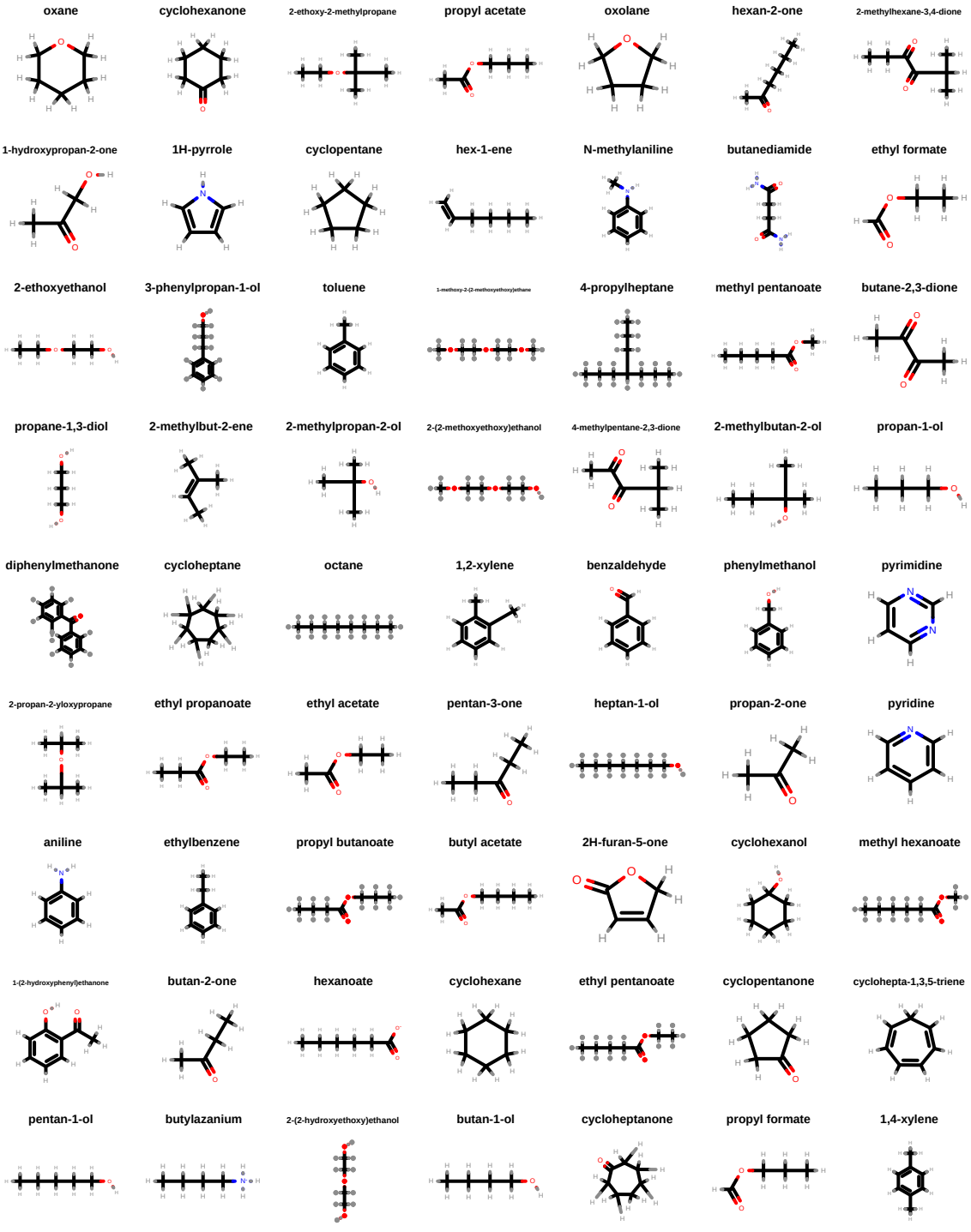


Figure 1.10: AM1-BCC-dPol library training set.

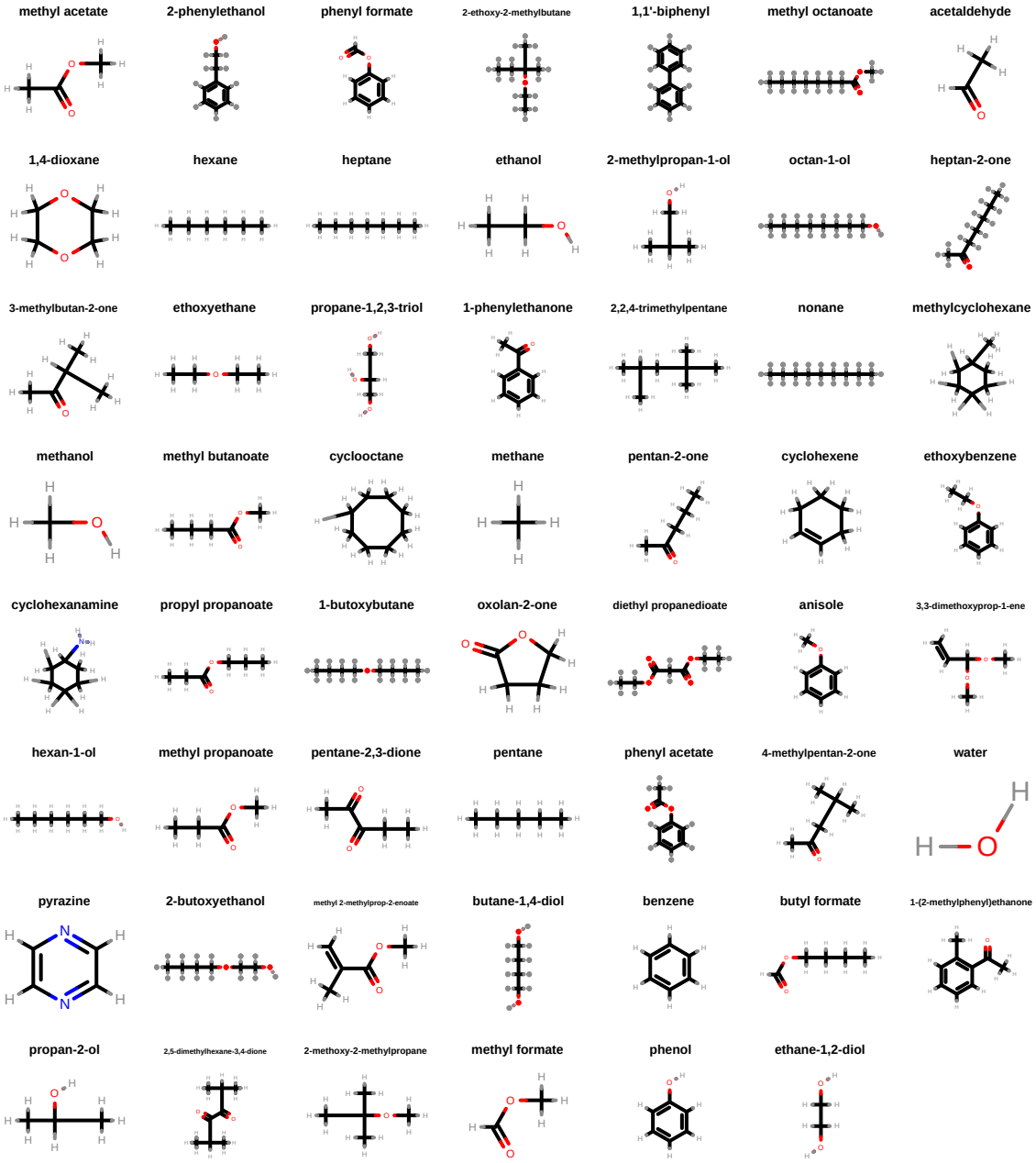


Figure 1.10: AM1-BCC-dPol library training set (Continued).

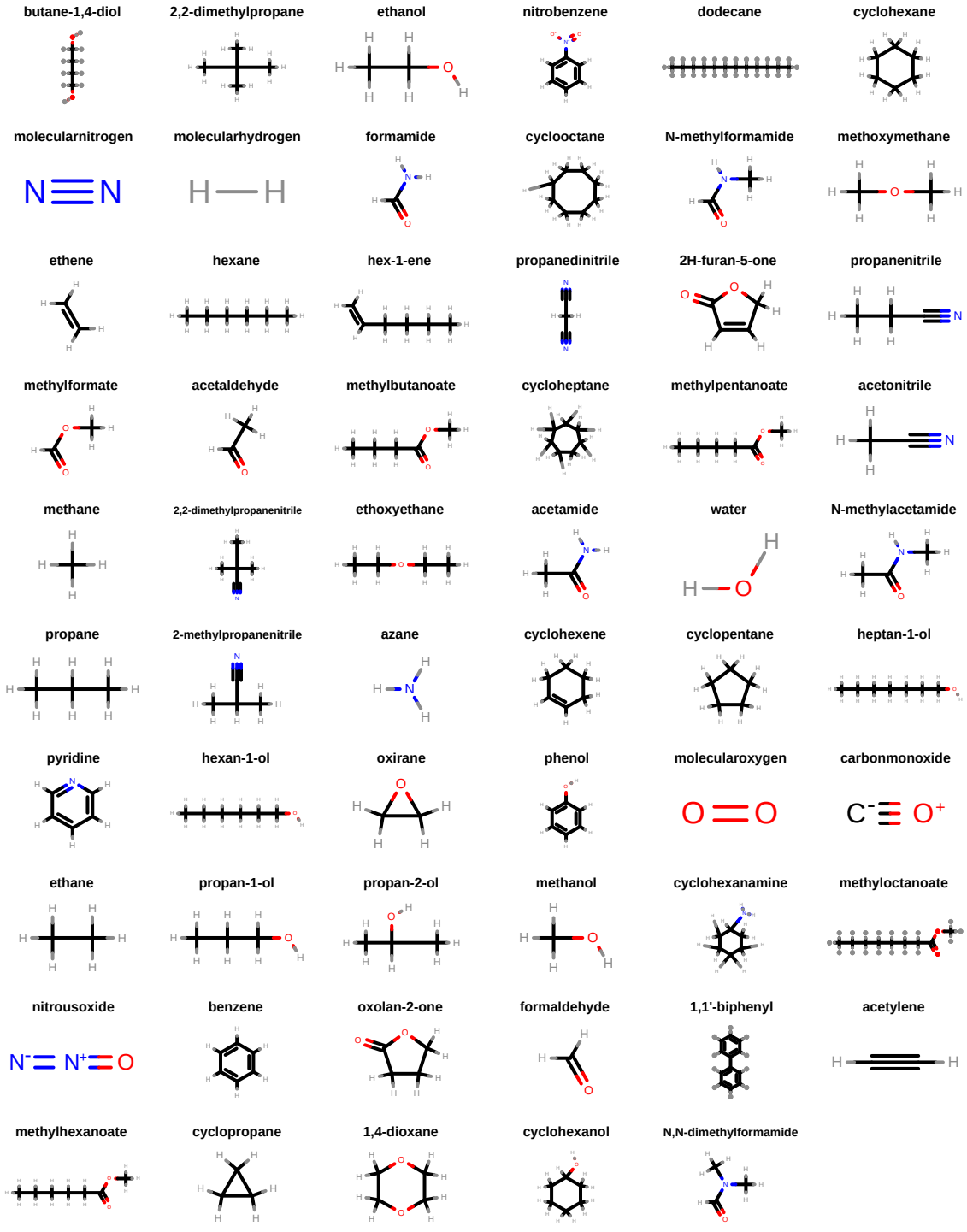


Figure 1.11: Molecular polarizability data set.

1.7.6 TOC Graphic

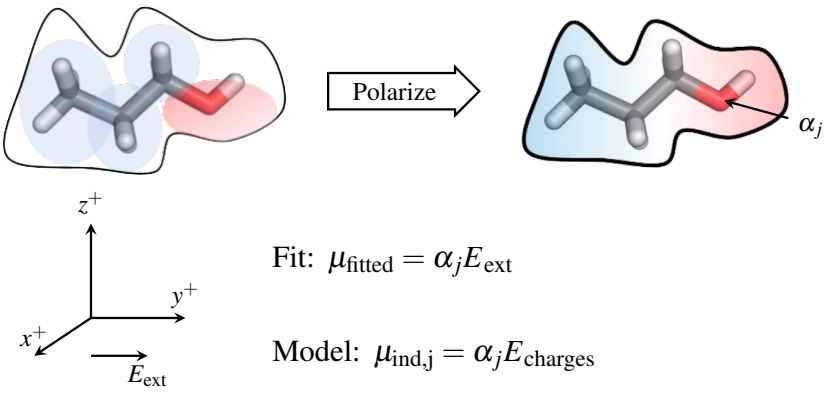


Figure 1.12: AM1-BCC-dPol model illustraton figure.

Bibliography

(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*, Publisher: Wiley, 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*, Publisher: Wiley, 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society, 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: Annual Reviews, 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*, Publisher: American Chemical Society (ACS), 2973–2982.

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

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

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

(21) Bottcher, C. J., *Theory of Electric Polarization, Vol. 1: Dielectrics in Static Fields*, 2nd Revised edition; 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: American Chemical Society (ACS), 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*, Publisher: AIP Publishing, DOI: 10.1063/1.4964866.

(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*, Publisher: Informa UK Limited, 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*, Publisher: American Chemical Society (ACS), 9956–9972.

(30) 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*, Publisher: American Chemical Society (ACS), 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**, ISBN: 9780470845011 Publisher: Wiley, DOI: 10.1002/0470845015.cfa007.

(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*, Publisher: Wiley, 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*, Publisher: AIP Publishing, 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*, Publisher: American Chemical Society (ACS), 5430–5449.

(35) 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.

(36) Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry* **1991**, *95*, Publisher: American Chemical Society (ACS), 3358–3363.

(37) 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*, Publisher: AIP Publishing, 741–754.

(38) 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*, Publisher: Wiley, 1–16.

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

(40) 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*, Publisher: American Chemical Society (ACS), 4315–4320.

(41) 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*, Publisher: American Chemical Society (ACS), 8668–8684.

(42) Litman, J. M.; Liu, C.; Ren, P. Atomic Polarizabilities for Interactive Dipole Induction Models. **2021**, DOI: 10.1021/acs.jcim.1c01307.

(43) 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*, Publisher: Wiley, 1048–1057.

(44) 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*, Publisher: Wiley, 1515–1531.

(45) 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*, Publisher: American Chemical Society (ACS), 621–627.

(46) 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*, Publisher: Wiley, 739–750.

(47) Daylight Chemical Information Systems, I. SMARTS-A Language for Describing Molecular Patterns, Publication Title: 120 Vantis–Suite 550–Aliso Viejo, CA 92656, 2007.

(48) 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*, Publisher: American Chemical Society (ACS), 6076–6092.

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

(50) Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **2020**, *17*, 261–272.

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

(52) 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.

(53) 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*, Publisher: Nature Publishing Group, 1–11.

(54) OpenEye OEAM1BCCELF10Charges, Published: OEAM1BCCELF10Charges - Toolkits – Python.

(55) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, *46*, 618–622.

(56) 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.

(57) 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*, Publisher: American Chemical Society (ACS), 3678–3687.

(58) 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*, Publisher: American Chemical Society (ACS), 249–258.

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

(60) Wang, L. Factor-Pol: A Fast Atom-Centered Typed Isotropic Ready-to-use Polarizable Electrostatic Model. Factor-Pol Model, `https://github.com/wwilla7/factorpol`, [Accessed 6-Jan-2025], 2023.

(61) 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*, Publisher: AIP Publishing, DOI: `10.1063/1.4984113`.

(62) 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`, [Accessed 24-12-2024], 2017.

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

(64) Lorenz, L. Ueber die Refractionsconstante. *Annalen der Physik* **1880**, *247*, Publisher: Wiley, 70–103.

(65) Thole, B. Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics* **1981**, *59*, Publisher: Elsevier BV, 341–350.

(66) Van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole's Model Revisited. *The Journal of Physical Chemistry A* **1998**, *102*, Publisher: American Chemical Society (ACS), 2399–2407.

(67) 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*, Publisher: American Chemical Society (ACS), 3100–3111.

(68) 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*, Publisher: Wiley, 1157–1174.

(69) 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*, ed. by Gentleman, R., Publisher: Public Library of Science (PLoS), e1005659.

(70) 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*, Publisher: Wiley, 2157–2164.

(71) Wang, L. factorpol-benchmarks-scripts: scripts to set up and run benchmarking calculations using factorpol electrostatic model. 2023.

(72) Neumann, M. Dipole moment fluctuation formulas in computer simulations of polar systems. *Molecular Physics* **1983**, *50*, Publisher: Informa UK Limited, 841–858.

(73) 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.

(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*, Publisher: American Chemical Society (ACS), 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*, Publisher: AIP Publishing, DOI: 10.1063/1.4928530.

(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, Version Number: 0.14.5, 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*, Publisher: American Chemical Society (ACS), 3566–3576.

(78) Thompson, M.; Wagner, J.; Gilmer, J. B.; Timalsina, U.; Quach, C. D.; Boothroyd, S.; Mitchell, J. A. OpenFF Interchange, version v0.4.0, 2024.

(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*, Publisher: Royal Society of Chemistry (RSC), 1178–1187.

(80) Horton, J.; Thompson, M.; Wagner, J. OpenFF smirnoff-plugins: Plugins to enable using custom functional forms in SMIRNOFF based force fields, 2022.

(81) Drew, L. W.; Thompson, M. Open Force Field MPID Plugin, 2023.

Chapter 2

A Simple, Polarizable, Rigid, 3-point Water Model Using the Direct Polarization Approximation

2.1 Introduction

Empirical force fields are widely used in molecular simulations to provide mechanistic insight and support molecular design in fields such as materials science and pharmaceutical chemistry. The most widely used force fields estimate the total energy as a sum of what are termed bonded and nonbonded energy contributions. Bonded terms capture the effects of bond stretching, valence angle bending, and torsional rotation, while nonbonded terms capture the remaining interactions among atoms, including, most prominently, van der Waals (vdW) and electrostatic interactions. The former are often modeled in terms of the Lennard-Jones (LJ) potential[1], while the latter are most commonly modeled in terms of the Coulombic interactions of atom-centered fixed charges.

The accuracy of such fixed-charge models is, however, limited by their neglect of the field-

dependent electronic polarization of atoms. The neglect of changes in polarization as a function of atomic coordinates can lead to various nonphysical properties, such as unrealistically low dielectric constants of nonpolar liquids[2, 3], incorrect ionic transport properties[4], and disturbed geometries of active sites of enzymes[5] . Although nonpolarizable force fields can correctly provide the high dielectric constant of liquid water, which is about 80 at room temperature, this is only because the orientational polarizability of individual water molecules makes up for the absence of an explicit treatment of electronic polarizability. This compensation is unlikely to occur within the interior of a folded protein, because the dipoles of a protein contribute little orientational polarizability[6, 7]. As a consequence, it is likely that nonpolarizable force fields overestimate the electrostatic interactions among atoms within proteins. Another concern is that not explicitly accounting for the energy cost of self-polarization of polar molecules when they transfer from gas phase to aqueous solution may lead to errors in computed heats of vaporization, so that training force fields against such phase-change properties will lead to problematic parameters [8, 9]. Moreover, nonpolarizable force fields cannot accurately reproduce the dielectric response of liquids to high-frequency electric fields, whereas induced dipole-based polarizable force fields AMEOBA03[10, 11] and HIPPO[12] can capture the frequency-dependence of the dielectric response [13].

There are currently two main representations of electronic polarization in force fields. One adds, at the center of each atom i, an inducible point-dipole, $\mu_{\text{ind,i}}$ which responds linearly and instantaneously to the local electrical field, E_j[14, 15]; thus,

$$\mu_{\text{ind,i}} = \alpha_i E_i \tag{2.1}$$

Here, α_i, the atom's polarizability, is typically approximated as a scalar so that the induced dipole is in the same direction as the inducing field, though it can, more generally, be a tensor. Potential drawbacks of the inducible dipole representation are that not all simulation codes

support inducible dipoles and that, because the electric field is determined by the point dipoles while the point dipoles are determined by the field, there is a computational cost to determining the self-consistent set of induced dipoles and fields, as recently addressed[16]. Well-developed polarizable force fields using this representation include AMOEBA[11, 17–19] and HIPPO[12] .

The second widely used representation of electronic polarization is the Drude model [20–23] , where each atom is assigned an additional point-charge whose sign is opposite to that of the atomic charge and which is tethered to the atom's center by a Hooke's law spring. This creates a dynamic dipole whose motions are propagated along with those of the atom-centers. Thus, at equilibrium, the time-averaged dipole moment of each atom with its Drude particle depends on the time-averaged local electrical field. The Drude approach is convenient because the motions of the Drude particle are naturally accommodated by standard simulation codes, rather than requiring additional code to handle inducible dipoles. However, it increases the number of particles needing to be simulated, the fluctuation of the Drude dipoles are not entirely physical, and managing them typically requires coupling them to a separate thermostat set to an artificially low temperature and using a relatively short simulation time-step[21]. Nonetheless, this approach, too, has met with considerable success [22, 24].

Despite these and other pioneering advances, polarizable force fields are still not used routinely, presumably due to the difficulty of assigning parameters to diverse molecules, the added computational cost of the simulations, and the relatively modest improvements in simulation accuracy they have so far provided in key applications. We therefore seek an approach that offers a more favorable cost-benefit balance, with the expectation that the research this enables will ultimately lead to polarizable force fields with a decisive accuracy advantage.

In an initial study, we developed AM1-BCC-dPol [3] , a polarizable model of electrostatic interactions which uses atom-centered partial charges and point polarizabilities, and which may be seen as a polarizable version of the widely used AM1-BCC partial charge method[25, 26] . Atoms are assigned pre-optimized point-polarizabilities based simply on element, so that e.g. all

nitrogens get the same polarizability, and partial charges are assigned by the normal AM1-BCC method, except that we optimized a new set of bond-charge corrections (BCCs) to replicate quantum mechanical electrostatic potentials (ESPs) in the context of the point-polarizabilities. Thus, electrostatic parameters are assigned as easily as with the non-polarizable AM1-BCC model. In addition, AM1-BCC-dPol gains computational efficiency by using the direct polarization approximation[27] , where the point polarizabilities respond to the electric field generated by the permanent partial atomic charges but not to the field generated by other induced dipoles. This simplifying approximation avoids the need for a time-consuming, self-consistent field solution of the induced dipoles. For a detailed comparison of the dPol model, the self-consistent field mutual polarization model, Drude oscillators, and the fluctuating charge method, please see our previous work Ref. [3]. The AM1-BCC-dPol electrostatics model is designed as a drop-in substitute for AM1-BCC in the otherwise standard force field functional form that uses the Lennard-Jones approximation of vdW interactions.

The AM1-BCC-dPol model gave a marked improvement in the accuracy of high-frequency dielectric constants of all organic liquids studied and of the static dielectric constants of nonpolar liquids, while affording accuracy of other liquid properties similar to that obtained with an otherwise similar nonpolarizable force field [3]. In addition, simulations with AM1-BCC-dPol ran at least 50% faster[3] than the already efficient OPT1[16] approximation of the full self-consistent treatment of inducible dipoles. However, our initial attempts to integrate AM1-BCC-dPol into a water model did not yield accurate pure water properties. This parallels the situation with many other force fields, where the partial charges (and other parameters) of the water model (e.g. TIP3P[28, 29] , TIP4P[29] , OPC[30] , OPC3[31]) are not derived by the same procedures (e.g. AM1-BCC[25, 26] , RESP[32] , CM4[33]) used for other molecules. Instead, there is typically a dedicated project to optimize charges and other parameters for the water. Note that there is no prior water model that fits our AM1-BCC-dPol paradigm. Perhaps the most closely related one is iAMOEBA[34] , whose success in generating accurate water properties while using

atom-centered inducible dipoles with the direct approximation serves as an inspiration for the present effort. However, iAMOEBA deviates from the AM1-BCC-dPol force field structure, as it uses not only partial charges but also permanent dipole and quadrupole moments, and it does not use the Lennard-Jones potential. In addition, iAMOEBA treats the water model as flexible, and, probably for this reason, is typically run with 0.5 fs time-step to avoid poor energy conservation. However, a time-step of 2 fs may be feasible with the multiple time-step (MTS) algorithm[35].

The present study focuses on the development of a new water model constructed within the AM1-BCC-dPol paradigm. We regard this as an essential next step in the development of a force field designed to be used for biomolecular simulations. We aim for a fast, simple polarizable water model that will integrate well into an AM1-BCC-dPol force field that works well for biopolymers and drug-like molecules. Accordingly, rather than go to e.g., a flexible, four-point model, we develop a rigid, three-point model. Although a four-point model might yield more accurate pure-water properties across a range of conditions[36, 37], it will be slower[38], and four-point water models have not yet yielded better agreement with experiment when used to compute protein-ligand[39] or host-guest binding thermodynamics[40, 41], which are applications of particular interest here.

This paper is organized as follows. In Methods, we describe the optimization pathway to a set of parameters that can reproduce gas-phase and condensed-phase water properties of water under ambient conditions. In Results, we demonstrate the accuracy of the best resulting water model, dPol-opt5, in calculations of important physical properties of pure water that are not included in parameterization. We also compare the speed and performance of the present model with other popular polarizable and non-polarizable water models. In Discussion, we address the importance of the new water model and future work toward a full polarizable force field utilizing the AM1-BCC-dPol approach.

2.2 Methods

We consider water models of the following overall form. The molecule is treated as rigid, and it interacts with other molecules via a single Lennard-Jones site centered on the water oxygen, three constant partial charges centered on the oxygen and hydrogen atoms, and three linear inducible point-dipoles centered on the oxygen and hydrogen atoms. These choices are based on prior studies showing that including additional LJ sites on water hydrogens provides little or no benefit[42], and our own initial studies showing that inducible dipoles only on oxygen led to less satisfactory results. Note that iAMOEBA and AMOEBA also include polarizabilities on all three atoms. The inducible dipoles respond only to the fields generated by the partial charges of other molecules; this is the direct or first-order approximation of polarization[27]. The adjustable parameters of the model are as follows: the H-O bond length (l_{HO}); the H-O-H angle (θ_{HOH}); the partial charge on hydrogen (q_H, where $q_O = -2q_H$); two LJ parameters for oxygen (σ_{LJ} and ε_{LJ}); and the polarizabilities of the oxygen and hydrogen atoms (α_O and α_H). Correspondingly, the potential energy of a collection of water molecules comprising N atoms is given by the sum of the Lennard-Jones energy, the interactions among partial charges, and the direct polarization energy:

$$U_{\mathrm{total}} = U_{\mathrm{LJ}} + U_{\mathrm{ele}}^{\mathrm{perm}} + U_{\mathrm{ele}}^{\mathrm{pol}} \tag{2.2}$$

The Lennard-Jones potential energy is given by

$$U_{\mathrm{LJ}} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.3}$$

where it is understood that i and j are oxygen atoms in different water molecules, r_{ij} is the distance between atoms i and j, ε_{ij} is the depth of the potential well, and σ_{ij} is the distance at which the particle-particle potential energy is zero.

The interaction energy of the partial charges is given by

$$U_{\text{ele}}^{\text{perm}} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{r_{ij}} \tag{2.4}$$

where q_i and q_j are the partial charges of atoms i and j, respectively, and the sum only includes cases where atoms i and j belong to different molecules. Coulomb's constant is omitted for simplicity.

The direct polarization energy is given by:

$$U_{\text{ele}}^{\text{pol}} = -\frac{1}{2} \sum_{j}^{N} \mu_{\text{ind, j}} E_j \tag{2.5}$$

where $\mu_{\text{ind, j}}$, the induced dipole at atom j, is computed as:

$$\mu_{\text{ind, j}} = \sum_{j=1}^{N} \alpha_j E_j \tag{2.6}$$

where α_j is the polarizability of atom j, and E_j is the electric field at atom j due to the electrostatic field generated by the permanent partial charges of other water molecules:

$$E_j = \sum_{i=1, j\neq i}^{N} q_j \frac{r_{ij}}{r_{ij}^3} \tag{2.7}$$

where it is understood that the sum only includes atoms i in molecules other than the molecule containing atom j. It is worth noting that, because induced dipoles do not interact with each other in the direct approximation used here, there is no possibility of a "polarization catastrophe", so a Thole damping function [43] is not required.

2.2.1 Parameter Optimization

Overview

We sought the best parameter set to reproduce a training set of gas-phase and condensed-phase water properties at ambient conditions, detailed in Section 2.2.1. The Nelder-Mead[44] minimization algorithm was used to carry out three optimizations, all initiated from the same water model, which we call TIP3P-dPol. This starting model comprises the QM-derived charges and polarizabilities detailed in Section 2.2.1, the gas-phase experimental geometry, and the LJ parameters of the TIP3P water model[28, 29](Table 2.1). The three optimizations all used the same training set of experimental water properties (Section 2.2.3), but they differed in which parameters were optimized.

In the first optimization, the Lennard-Jones parameters were optimized to reproduce experimental liquid properties detailed below, resulting a model denoted as dPol-opt1. However, as described in Section 2.3.2, this approach consistently led to an excessively high dielectric constant. We next sought to improve the results by optimizing not only the Lennard-Jones parameters but also the OH bond length (l_{OH}) and the H-H bond angle (θ_{HOH}). This second optimization resulted in a model called dPol-opt2. However, dPol-opt2 required a time-step <2 fs to maintain stability during simulations. Finally, the third optimization adjusted all seven parameters, and we extracted the two parameters sets visited during this optimization that gave the lowest values of the objective function; these correspond to models dPol-opt3 and dPol-opt4. Finally, we used a grid search over the two Lennard-Jones parameters to fine-tune the mass density and heat of vaporization, and this led to the best model, dPol-opt5. Following parameterization, we further evaluated the various models against water properties not included in the training set. We also tested the ability of the initial TIP3P-dPol model to replicate experimental data in both the training and test sets. The dPol-opt5 model, now known as the dPol water model, was chosen for use with the AM1-BCC-dPol approach.

The parameterization was carried out with our `dPolfit`[45] Python package, which automates calculation of the physical properties considered here from MD simulations, as well as the value of the objective function (Section 2.2.1). The dPolfit package also allows for optimizing parameters for small organic molecules with the same optimization approaches. The package includes an interface to OpenMM version 7[46] to drive the MD simulations, and it interfaces with Psi4[47] for quantum chemistry calculations, which provides reference data for force field parameterization and evaluation. The SciPy optimize module[48] is the primary optimization interface in the dPolfit package, providing various optimization algorithms such as Nelder-Mead, Newton-CG, BFGS, etc. The Ray distributed framework[49, 50] was implemented to distribute calculations and optimization from single-node to clusters, allowing fast and large-scale parameterization and evaluation. Details of the parameter optimization procedures follow.

Optimization and benchmarking procedures

The free parameters of the water molecule were adjusted using the Nelder-Mead simplex algorithm to optimize agreement with five experimentally determined water properties at 298.15 K and 1 bar: mass density, static dielectric constant, enthalpy of vaporization, coefficient of thermal expansion, and isothermal compressibility. Following optimization, selected models were benchmarked against the following experimental observables outside the training set: self-diffusion coefficient, shear viscosity, O-O radial distribution functions ($g_{OO}(r)$); and values of density, heat of vaporization, dielectric constant, and coefficient of thermal expansion from 249 K to 373 K. Details of the MD simulations and property calculations are provided in Section 2.2.2 and 2.2.3.

The objective function was based on that used in the successful ForceBalance method [51] :

$$L_{\text{tot}} = \sum_{T \in \text{targets}} w_T L_T + 0.01 \sum_{i=1}^{n} (k_i - k_i^0)^2 \tag{2.8}$$

In the first term, T indexes the targeted experimental properties (e.g. water density), w_T is the

weight, which is assigned 0.5 for density, and 0.1 for the remaining properties, and L_T is the absolute relative error of the computed vs experimental value of property T:

$$L_T = \left| \frac{x_{\text{calc}} - x_{\text{expt}}}{x_{\text{expt}}} \right| \tag{2.9}$$

The second term in Equation 2.8 is a regularizer, which weakly restrains the adjustable parameters ($k_i, i = 1, \ldots, n$) to their respective initial values (k_i^0). Note that $n = 2$ for the optimization of model dPol-opt1, n =4 for the optimization of dPol-opt2, and $n = 7$ for the optimization of dPol-opt3 and dPol-opt4.

A grid scan over the two Lennard-Jones parameters was used to minimize the errors in density and heat of vaporization in models resulting from the Nelder-Mead optimization. For each LJ parameter, 50 evenly spaced values were generated, scanning 2500 parameter sets, while the remaining parameters were held fixed. At each grid point, a 1 ns MD simulation was run using a periodic cubic box with 256 water molecules; simulation details are described in Section 2.2.2. Density and heat of vaporization were evaluated based on their relative root mean squared errors (RRMSE) to experiments, and selected models were run for more definitive 50 ns to compute additional properties.

Generation of initial partial charges and atomic polarizabilities

The initial atom-centered partial charges and atomic polarizabilities mentioned in the prior paragraph were derived from quantum mechanical (QM) calculations on a single water molecule, using our previously described procedure[3] . The QM electrostatic potentials (ESP) were computed at the MP2/aug-cc-pVTZ[52, 53] level of theory, using Merz-Singh-Kollman (MSK) style[54, 55] ESP sampling points located outside the van der Waals radii of all atoms. The points were disposed in 10 layers spaced by 0.126 Å, each with a point density of 17/Å^2. First, baseline ESPs were computed in the absence of any external electrical field and were used to

compute the partial charges of water by minimizing the sum of squared error under the constraint that $q_O + 2q_H = 0$:

$$\chi^2 = \sum_{i=1}^{m} \left(V_{\text{QM, ib}} - \sum_{j=1}^{n} \frac{q_{\text{j}}}{r_{\text{ij}}} \right)^2 \tag{2.10}$$

where r_{ij} is the distance between ESP point i and atomic center j, q_j is the ESP derived atomic charge at j, $V_{QM,ib}$ is the baseline QM potential at point i, and m is the number of ESP points. Note that the inducible dipoles play no role in the derivation of these charges because the inducible dipole on a given atom does not respond to charges on other atoms in a 1-2 or 1-3 bonding relationship to it. Then six polarized ESPs were generated with uniform imposed external electric fields ($E_{\text{ext,k}}$) of magnitude 0.01 au in the [+x, -x, +y, -y, +z, -z] directions, where k indexes these directions. The ESP differences between baseline and polarized ESPs, $V_{\text{diff, ik}} = V_{QM,il} - V_{QM,ib}$, were used to derive the polarizabilities (α) of oxygen and hydrogen by minimizing the following sum of squared errors:

$$\chi^2 = \sum_{k=1}^{6} \sum_{i=1}^{m} \left(V_{\text{diff,ik}} - \sum_{j=1}^{n} \frac{\alpha_j E_{\text{ext,jk}} r_{\text{ij}}}{r_{\text{ij}}^3} \right)^2 \tag{2.11}$$

2.2.2 Simulation Details

All MD simulations were performed with OpenMM version 7.7[46, 56] and the MPID (multipole and induced dipole) OpenMM plugin[57–60] with keyword `direct` for direct polarization. The MD time-step was 2 fs, the pressure was maintained at 1 bar using the Monte Carlo barostat, and the simulation temperature was maintained with the Langevin thermostat[61]. The condensed phase simulations used periodic boundary conditions, and long-ranged nonbonded interactions were accounted for with the particle-mesh Ewald method [62] with the short-ranged cutoff set to 9 Å for both electrostatic and LJ interactions. We created an OpenFF SMIRKS Native Open Force Field (SMIRNOFF) plugin, `dPol-Plugin`[63] , to enable the use of the dPol models with OpenFF force fields and the OpenFF software ecosystem.

We used smaller simulations within the iterative parameter optimization loop and then carried out more extensive simulations of the optimized water models. The smaller simulations used a periodic cubic box with 256 water molecules, prepared with PackMol[64]. This was energy-minimized, equilibrated for 8.5 ns, and then run for 7.5 ns at constant temperature and pressure (*NPT*) to generate production data used to calculate the targeted physical properties. Snapshots were saved every 500 ps for post-processing. The larger simulations were similar but used 1000 water molecules and were run for 50 ns at constant pressure and temperature to generate production data. We examined the temperature dependence of water properties by carrying out 28 such simulations at temperatures ranging from 249 K to 373 K. Special procedures used to compute the self-diffusion coefficient and shear viscosity of liquid water are described in Section 2.2.3.

2.2.3 Calculation of Physical Properties

The mean molecular dipole moment of the water molecule in the liquid phase was computed as:

$$\mu_{\mathrm{M,liq}} = \left| \frac{1}{N_{\mathrm{wat}}} \sum_{i=1}^{N_{\mathrm{wat}}} \sum_{j=1}^{3} \left\langle q_j r_{ij} + \mu_{\mathrm{ind,\,ij}} \right\rangle \right| \tag{2.12}$$

where i indexes the N_{wat} water molecules in the simulation, $j \in [1,2,3]$ indexes the atoms in a given water molecule, $\mu_{\mathrm{ind,\,ij}}$ is as given in Equations 2.6-2.7, and angle brackets indicate an average over the simulation snapshots. The gas-phase molecular dipole moment was calculated, without need for simulation data, as:

$$\mu_{\mathrm{M,gas}} = \left| \sum_{j=1}^{3} q_j r_j \right| \tag{2.13}$$

where q_j and r_j are the partial charge and coordinates, respectively, of atom j. The molecular polarizability is simply the sum of all atom-centered isotropic polarizabilities (α_j) at atom j, because, in the direct approximation, only permanent charges on other water molecules and

any external electric field that may be present are considered to induce changes in the water molecule's charge distribution:

$$\alpha_M = \sum_{j=1}^{3} \alpha_j \tag{2.14}$$

The mass density was evaluated as:

$$\rho = \frac{M}{\langle V \rangle} \tag{2.15}$$

where M is the total mass of the system, and V is the volume of the cubic box.

The enthalpy of vaporization was computed as the difference between the condensed-phase (H_l) and gas-phase (H_g) enthalpies:

$$\begin{aligned} \Delta H_{\mathrm{vap}} &= \langle H_\mathrm{g} \rangle - \langle H_\mathrm{l} \rangle \\ &= \langle U_\mathrm{g} \rangle - \langle U_\mathrm{l} \rangle - P(\langle V_\mathrm{g} \rangle - \langle V_\mathrm{l} \rangle) \\ &= \langle U_\mathrm{g} \rangle - \langle U_\mathrm{l} \rangle + RT \end{aligned} \tag{2.16}$$

where $H \equiv U + PV$, U, P, and V are the enthalpy, potential energy, pressure, and volume, respectively, the subscripts g and l denote the gas and liquid phases, and we have used the fact that, at a given temperature, the mean kinetic energy of a water molecule is the same in the condensed and gas phases. The third line of the equation follows from the ideal gas law and the fact that $V_g >> V_l$, where R and T are the gas constant and temperature, respectively. Since the energy cost of polarization – i.e, of distorting the electron cloud from its gas-phase form to its solution-phase form – is implicitly included in the energetics of dPol polarizable model, no empirical self-polarization correction is included.

The static dielectric constant, ε, is computed using the dipole fluctuation approach[65–67] :

$$\varepsilon = \varepsilon_\infty + \frac{\langle \mu^2 \rangle - \langle \mu \rangle^2}{3\varepsilon_0 k_B T \langle V \rangle} \tag{2.17}$$

Here μ is the total dipole moment of the system, ε_∞ is the high-frequency dielectric constant, and k_B is the Boltzmann constant. In the direct approximation used here, the high-frequency dielectric constant is obtained analytically as:

$$\varepsilon_\infty = \frac{1}{\varepsilon_0 V} \sum_{j=1}^{N} \alpha_j + 1 \tag{2.18}$$

Note that, for non-polarizable water models, ε_∞ typically equals 1[29–31] , due to the absence of electronic polarizability; and, for most SCF mutual polarization models[10, 34, 68], ε_∞ is typically estimated with the Clausius-Mossotti equation (e.g., refs [20, 69]). The standard errors in fluctuation properties are calculated using a block averaging method[70] and the 95% confidence intervals are estimated from standard errors of the mean. Good convergence of the dielectric constant was reached by $\sim$35 ns (Figure 2.5).

The coefficient of thermal expansion of the liquid at constant pressure was computed as[71]:

$$\alpha_P = \frac{1}{k_B T} \frac{\langle UV \rangle - \langle U \rangle \langle V \rangle}{\langle V \rangle} \tag{2.19}$$

and the isothermal compressibility was computed as[71]:

$$\kappa_T = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B T \langle V \rangle} \tag{2.20}$$

The self-diffusion coefficient of water obtained from simulations can depend on the size of the simulation box[72]. We estimated the diffusion coefficient in the limit of infinite box size via a protocol similar to that described in Ref.[51]. Thus, box size-dependent diffusion coefficients D_{PBC} (see below) were obtained for cubic water boxes of side length 25, 30, 40, 50, 60, and 90 Å, each with appropriate numbers of water molecules, and the self-diffusion coefficient in the limit

of large box size, D_0, was computed by extrapolating D_{PBC} as the inverse box size goes to zero:

$$D_0 = D_{\mathrm{PBC}} + \frac{2.837297 k_B T}{6\pi\eta L} \tag{2.21}$$

The linear regression fit of the self-diffusion coefficient against the inverse box size L also yields the shear viscosity, η (Equation 2.21)[51, 72]. The values of D_{PBC} were computed from the mean-square displacement (MSD) of water molecules simulated using the microcanonical *NVE* ensemble according to the Einstein relation:

$$D_{PBC} = \frac{1}{6N} \lim_{\mathrm{dt}\to\infty} \frac{\langle |r_{t_0+\mathrm{dt}} - r_{t_0}|^2 \rangle}{\mathrm{dt}} \tag{2.22}$$

where the numerator on the right hand side is the ensemble-averaged MSD of the coordinates r of oxygen atoms from the initial position (t_0) after time dt. N is the total number of oxygen atoms.

The reported values of D_0 and η are ensemble-averages over 100 replicates, where each replicate was initiated from a different snapshot containing position and velocity information drawn from a well-equilibrated NPT trajectory. Each initial snapshot was used as the starting point of a 10 ps MD simulation in the *NVE* ensemble, using the Verlet integrator and a time-step of 1 fs, and saving snapshots every 0.1 ps for post-processing. Sample results are detailed in Figures S3 and S4.

2.3 Results

2.3.1 Initial TIP3P-dPol Model

We applied the AM1-BCC-dPol method of assigning partial charges and atomic polarizabilities to water to construct the TIP3P-dPol model, which we then used as the starting point for all Nelder-Mead optimizations. The TIP3P-dPol model exhibits a gas-phase dipole moment of

1.90 D closely matching the experimental value of 1.86 D. In the condensed phase, the molecular dipole moment fluctuates between 2.2 and 3.1 D, also showing excellent agreement with experimental values. The density of TIP3P-dPol model is 0.997 g/mL, which is closer to experiment at 298K than the density of TIP3P (0.98 g/mL) water. However, the ΔH_{vap} and dielectric constant are not accurately reproduced (Table 2.2).

2.3.2 Optimization of dPol Water Models

Optimization 1

The first optimization, which modified only the two Lennard-Jones parameters, terminated after 50 iterations due to lack of further improvement in the objective function, and generated water model dPol-opt1. This model has a smaller value of σ_O (Table 2.1), which reduced the underestimation of ΔH_{vap} from 18% to 9% but also led to overestimates of the density and dielectric constant (Table 2.2).

Table 2.1: Force field parameters of several dPol water model candidates.

Parameter	TIP3P-dPol	dPol-opt1	dPol-opt2	dPol-opt3	dPol-opt4	dPol-opt5
l_{OH} (Å)	0.9572	0.9572	1.043	1.038	1.038	1.038
θ_{HOH} (deg)	104.52	104.52	110.86	110.8	111.8	111.7
σ_{LJ}^{O} (Å)	3.1508	3.1131	3.2073	3.1897	3.1897	3.2031
ε_{LJ}^{O} (kJ/mol)	0.636	0.6387	0.6651	0.6389	0.8389	0.6357
q_H (e)	0.338	0.338	0.338	0.343	0.357	0.338
α_O (Å^3)	0.997	0.997	0.997	0.899	0.698	0.997
α_H (Å^3)	0.207	0.207	0.207	0.125	0.212	0.207

Optimization 2

Qiu et al.[51], in their work on the TIP3P-FB and TIP3P-ST water models, found that the dielectric constant can be accurately fitted if the geometric parameters are optimized. We therefore

Table 2.2: Physical properties of water resulted during the optimization and compared to experimental (EXPT) reference and other popular water models at 298K and 1atm. Numbers in the parentheses indicate one extra digit in the significant digit, when available.

Property	EXP[61, 73, 74]	dPol-opt5	dPol-opt4	dPol-opt3	dPol-opt2	dPol-opt1	TIP3P[29]	OPC3[31, 75]	AMOEBA[34]	iAMOEBA[34]
ρ (g/mL)	0.997	1.00(6)	0.978	1.00	0.997	1.07	0.982	0.995	1.00	0.997
ΔH_{vap} (kJ/mol)	10.51	10.43	10.43	10.06	10.51	9.55	10.05	10.73	10.48	10.94
ε (unitless)	78.4	80.7	85.5	87.2	90.3	200	94	78.4	81.4	80.7
ε_∞ (unitless)	1.78[76]	1.60	1.46	1.48	1.59	1.63	1	1	1.73[13]	1.76
μ_{g} (D)	1.855	1.89	1.99	1.91	1.92	1.90	2.35	2.43	1.85	1.86
μ_{l} (D)	2.5-3	2.55	2.49	2.66	2.72	2.5	2.35	2.43	2.78	2.78
α (Å^3)	1.44	1.41	1.12	1.15	1.41	1.41	N/A	N/A	1.47	1.82
D_0 (10^{-5}cm$^2 s-1$)	2.3	2.3	2.3	2.7	2.5		6.1	2.3	2.0	2.53
η (mPa s)	0.89	0.82	0.82	0.83	0.88		0.43	0.8	1.08	0.85
α_P (10^{-4}K^{-1})	2.56	7.21	7.8	8.4	8.0	10.9	9.2	4.3	1.9	2.5
κ_T (10^{-6}bar^{-1})	45.3	42.2	44.3	45.4	44.8	47.9	57.4	46	66	41.2
Coordination number	4.4-4.5[77]	4.5	4.6	4.6	4.6		6.2[38]	4.9[38]	5.3	4.4

Note: D_0, η, and coordination numbers were not computed for TIP3P-dPol and dPol-opt1. The coordination numbers of AMOEBA and iAMOBEA models were computed in this work (Figure 2.2).

included l_{OH} and θ_{HOH} as adjustable parameters in the second optimization, which yielded model dPol-opt2. This model has a somewhat longer OH bond-length and HOH bond angle (Table 2.1), leading to a small increment in the dipole moment and stronger hydrogen bonding interactions. Encouragingly, it outperforms TIP3P-dPol and dPol-opt1, especially in relation to the dielectric constant and ΔH_{vap} (Table 2.2). The modest overestimation of the dielectric constant may result from the somewhat large average dipole moment for bulk water monomers (2.72 D). Although this value is within the experimental range of 2.5-3 D, polarizable water models with liquid-phase dipole moments higher than 2.6 D fail to reproduce the dielectric constant, unless also utilizing damping functions or anisotropic polarizabilities.[78–81]. Unfortunately, simulations of dPol-opt2 were unstable when we used a 2 fs time-step. One way to alleviate the unstable issue may be reducing the condensed-phase dipole moments or using a damping function.

Optimization 3

In order to discover a water model that would agree well with experiment while also yielding stable simulations, we widened the parameter search in a third optimization which added the partial charges and polarizabilities as additional adjustable parameters. This optimization of all seven parameters led to two sets of parameters, corresponding to the lowest two values of the objective function encountered during the Nelder-Mead optimization. These parameter sets were both saved and designated dPol-opt3 and dPol-opt4. The two models reproduce the physical properties in the training set with similar overall accuracy (Table 2.2), but the errors are distributed differently: dPol-opt3 yields a more accurate mass density, and dPol-opt4 model is more accurate for ΔH_{vap}. Encouragingly, both models give values of mass density and ΔH_{vap} that are at least as accurate as those from TIP3P, while also giving more accurate static and high-frequency dielectric constants and dipole moments. Given the success of TIP3P, this observation suggested that we were approaching a sufficiently accurate water model. However, the molecular polarizabilities fell from 1.41 $Å^3$, which comes from our QM-trained charges and polarizabilities and is close to the

experimental value of 1.44 Å^3, to $\sim$ 1.1 Å^3, while the magnitude of the partial charges increased somewhat (Table 2.1). The drop in polarizabilities leads to an underestimation of the high-frequency dielectric constant, relative to the experiment (Table 2.2), and the increased polarity leads to a mild overestimation of the gas-phase dipole moments (Table 2.2). For dPol-opt4, there is also a further increase in the value of θ_{HOH}.

To further evaluate dPol-opt3 and dPol-opt4, we computed their O-O radial distribution functions. As shown in Figure 2.2b,c, both gave reasonable results but the second and third peaks are somewhat flattened, especially in the case of dPol-opt4. They also somewhat overestimate the first-shell coordination number, computed as the integral from 0 to $\sim$ 3.5 Å, yielding values of 4.6 and 5.5, compared to the experimental value of 4.4-4.5[77] .

Grid searches over Lennard-Jones parameters

At this point, we had three models, dPol-opt2, dPol-opt3, and dPol-opt4, that gave stable simulations but were highly accurate for only density or only ΔH_{vap}, but not both; and one water model, dPol-opt2, which was accurate for both but which led to unstable simulations. We conjectured that it would be possible to discover parameters that yield stable simulations and accurate results for both of these key observables via focused searches over the Lennard-Jones parameters, which are critical determinants of density and ΔH_{vap} but have a weaker effect on other important properties such as dipole moments and dielectric constants.

Accordingly, we carried out three Lennard-Jones parameter scans as detailed in Section 2.2.1, each with its own fixed geometry, partial charges, and polarizabilities. Scan 1 uses our QM-derived charges (q_H = 0.338) and polarizabilities (α_O = 0.997 Å^3, α_H = 0.207 Å^3), and a water geometry from commonly used water models, $l_{OH} = 1.0$ Å and $\theta_{HOH} = 109.5$ deg. This parameter set gives accurate values of the gas-phase dipole moment and molecular polarizability, 1.88 D and 1.41 Å^3, respectively. Scan 2 used the partial charges, polarizabilities, and geometry of model dPol-opt4. Scan 3 used the partial charges and polarizabilities, of dPol-opt2 but borrowed

the geometry from dPol-opt 4, making a small adjustment of the bond angle to better match the gas-phase dipole moment.

The results of these grid scans are shown in Figure 2.1 as contour plots of density (left hand panels) and heat of vaporization (right hand panels) in the (σ, ε) plane for Scans 1-3 (top, middle, and bottom panels, respectively). It is clear that increasing σ decreases the density and the heat of vaporization, while increasing ε increases the density and heat of vaporization. The red and black lines mark the experimental density and heat of vaporization, respectively. For a water model to correctly capture both properties, its Lennard-Jones parameters must lie at the intersection of these two lines, but parameters far from the scanning plane may be physically unrealistic and therefore may generate other problems.

Scan 1 (Figure 2.1a and 2.1b), which uses the experimental geometry and QM-derived charges and polarizabilities, does not yield an intersection point in or near the scanned range of σ and ε, so we did not pursue it further.

Figure 2.1c and 2.1d show the contour plots for Scan 2, which was intended to correct the slightly underestimated density of dPol-opt4. As shown in the contour plots, experimental density and heat of vaporization can be achieved at ε = 1.006 kJ/mol and σ = 3.161 Å. However, probably as a result of the large ε, this model did not yield an accurate RDF, $g_{OO}(r)$ (Figure 2.8), and therefore was not further tested for other bulk properties.

Scan 3 (Figure 2.1e and 2.1f), was intended to improve the simulation stability of dPol-opt2 while maintaining its favorable properties, discussed above. Because the density and heat of vaporization curves intersect a short distance outside the scan area, we fitted third-order polynomials to them, found the intersection at ε = 0.505 kJ/mol and σ = 3.238 Å, and used these parameters to compute density and heat of vaporization (Figure 2.9). Unexpectedly, this (unnamed) model overestimated both density (1.02 g/L) and heat of vaporization (10.74 kcal/mol), and the Lennard-Jones parameters were also noted to lie outside the range of popular water models, i.e., $\varepsilon \in$ [0.6, 0.9] kJ/mol and $\sigma \in$ [3.0, 3.2] Å (data summarized based on Chaplin's

website[82]). Therefore, it was not pursued further.

Finally, we identified the values of σ and ε within the scan range that yielded the lowest RRMSE of all scans relative to experimental density and heat of vaporization and used 50 ns simulations with this water model, termed dPol-opt5 (Table 2.1), to compute water properties. As listed in Table 2.2, the dPol-opt5 model yielded only 0.9% error in density and 0.86% error in the heat of vaporization. In addition, properties of particular interest for a polarizable model - dielectric constant, high-frequency dielectric constant, dipole moments - are all in excellent agreement with experiment. Further studies of dPol-opt5 and other water models are detailed in the following section.

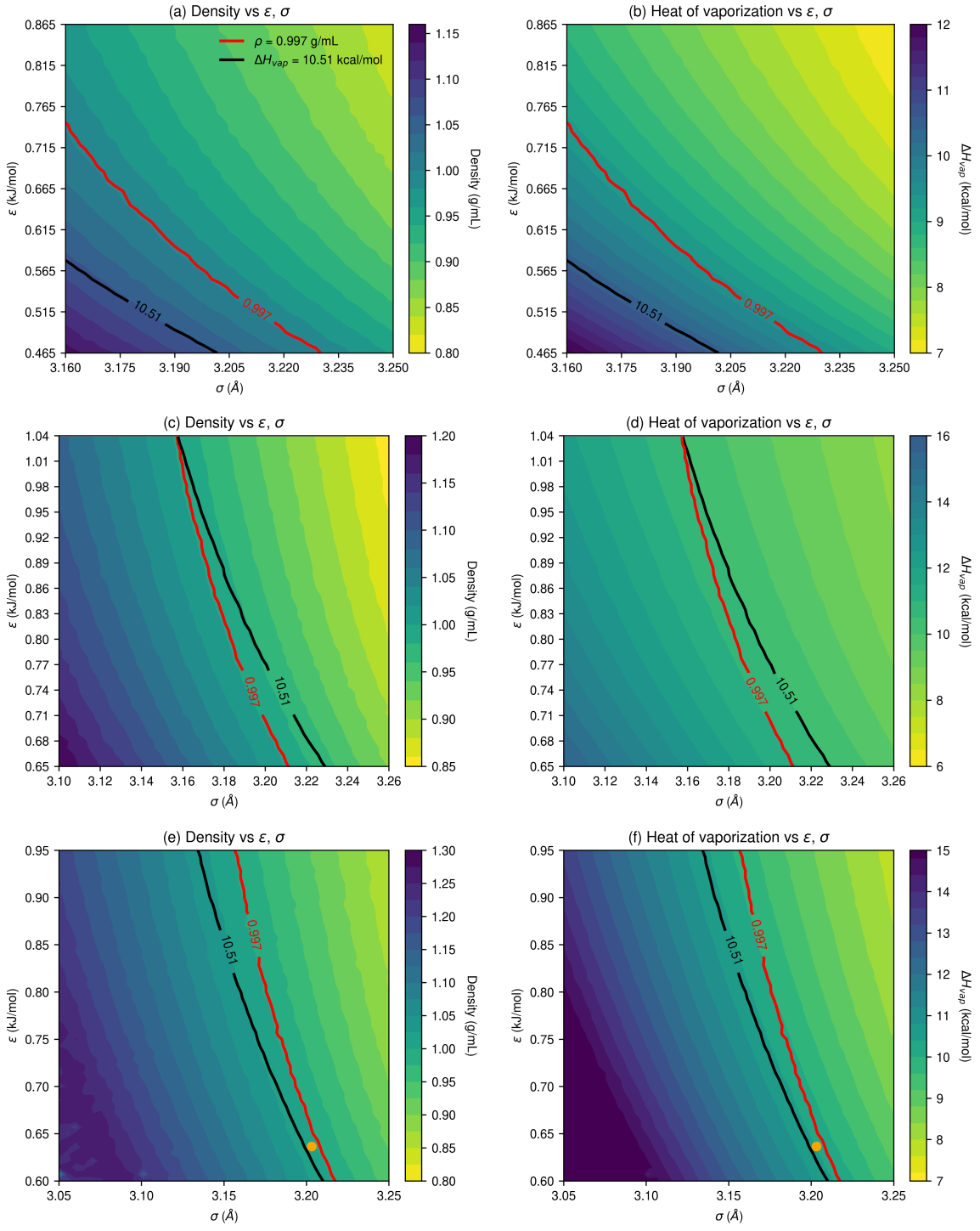


Figure 2.1: Contour plots of water properties vs Lennard-Jones parameters ε and σ. Colored bands represent ranges of the dependent properties: density (1a, 1c, 1e) and heat of vaporization (1b, 1d, 1f). Contour lines connecting points that have the same dependent value. The red lines indicate the dependent values of experimental density, and the black lines indicate the dependent values of experimental heat of vaporization. 1a and 1b are plots for Scan 1, 1c and 1d are for Scan 2, and 1e and 1f are for Scan 3. The orange dot indicates the best resulting model, dPol-opt5.

2.3.3 Benchmarks Against Additional Experimental Data

Here, we report further evaluation of models dPol-opt2, dPol-opt3, dPol-opt4, and dPol-opt5, as well as selected prior models, against dynamical, structural, and temperature-dependent experimental observables that were not used in the optimization procedures detailed above.

Structural and dynamical properties

The models' self-diffusion coefficients and shear viscosities were computed from MD simulations and the results are shown in Table 2.2; a sample box-size extrapolation (see Section 2.2.3) is shown in Figure 2.6. All dPol models yield good agreement with the experimental values, avoiding TIP3P's well known overestimation of the self-diffusion coefficient.

The oxygen-oxygen (O-O) radial distribution functions (RDFs) under ambient conditions are plotted in Figure 2.2. Models dPol-opt2 (Figure 2.2 a) and dPol-opt3 (Figure 2.2 b) place the first peak at the right location and give the first-shell coordination numbers in good agreement with experiment (4.4-4.5)[77] . Although the first peak of dPol-opt4 is located correctly, the coordination number is overestimated due to a broad first peak, and g_{OO} is nearly flat beyond the first peak (Figure 2.2 c). Model dPol-opt5 captures all of the experimental features with good fidelity, whereas a number of commonly used water models overestimate the height and coordination number of the first shell [38, 83]. In particular, the first shell g_{OO} of dPol-opt5 is not overstructured, although it has been suggested that overstructuring is difficult to avoid when using the r^{-12} repulsive term of the Lennard-Jones model[84]. The accuracy of g_{OO} for dPol-opt5 suggests that direct polarization can afford good accuracy at short-range. Moreover, the g_{OO} of dPol-opt5 and the direct polarization iAMOEBA model are very similar and more accurate than the mutual polarization AMOEBA model (Figure 2.2d,2.2e,2.2f), further indicating that direct polarization is a viable electrostatic model. The second and third shells of dPol-opt5 are observed at around 4.5 Å and 7 Å, again in close agreement with experiment. We did not examine $g_{OH}(r)$ or $g_{HH}(r)$ because the experimental data are not as well defined as g_{OO}, due to

the scattering of electrons by hydrogen atoms [28, 77]. Overall, our dPol-opt5 model exhibits good water structural properties compared to experiments and commonly used water models.

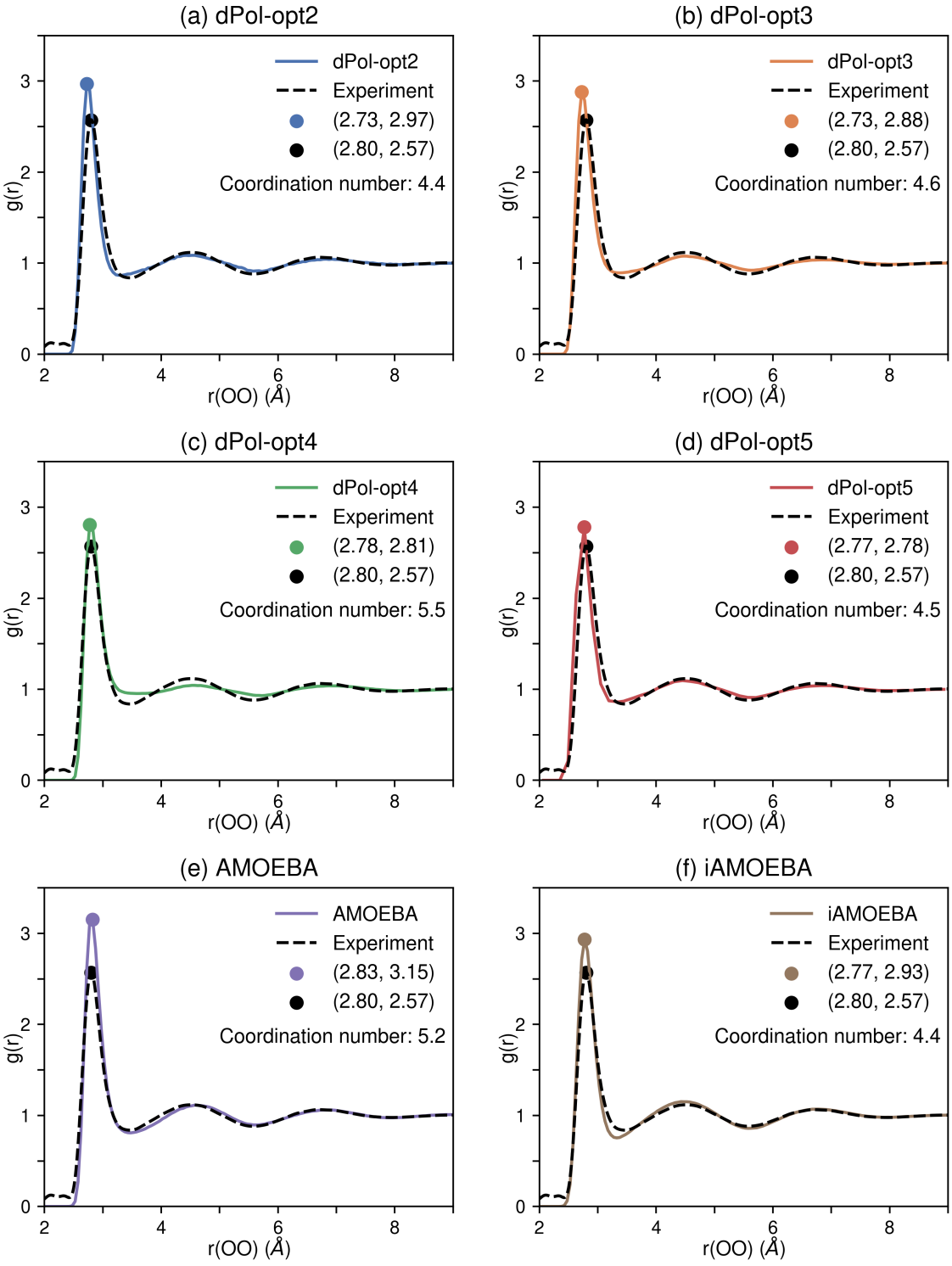


Figure 2.2: Oxygen-oxygen radial distribution function and coordination numbers of dPol-opt2 (a), dPol-opt3 (b), dPol-opt4 (c), and dPol-opt5 (d) and polarizable AMOEBA (e) and iAMOEBA (f) models compared to experimental reference (black). Experimental $g_{OO}(r)$ was obtained from ref[85].

Temperature dependence of water properties

Although our aim is a polarizable water model that works well at the limited range of temperatures most relevant for biomolecular systems, it is informative to examine how well accuracy is maintained as a function of temperature. Accordingly, Figure 2.3 plots the temperature dependence of the heat of vaporization, dielectric constant, density, and coefficient of thermal expansion, computed with models dPol-opt(2,3,4,5) from 250K to 370K, along with the corresponding experimental data.

We find that dPol-opt5 reproduces the temperature dependence of the heat of vaporization well, as it deviates from experiment by less than 0.5 kcal/mol over the full temperature range (Figure 2.3a). The dPol-opt2 and dPol-opt3 models give similar results, but dPol-opt3 underestimates the heat of vaporization across the whole temperature range. The dPol-opt5 model also tracks the temperature dependence of the static dielectric constant well over temperature, with a maximum deviation of 6 from experiment (Figure 2.3b). Models dPol-opt 3 and dPol-opt4 also do reasononably well, but dPol-opt 2 shows larger errors and an incorrect curvature relative to experiment, consistent with the room temperature result in Table 2.2. These deviations might trace to its larger condensed phase dipole moment (Table 2.2) [20, 78–80, 86] and/or an excessively large Kirkwood g-factor[66], which is the ratio of the liquid's dipole moment fluctuation to that which would be obtained in the absence of water-water orientational correlation.

The dPol-opt5 and dPol-opt 2,3 models closely approximate the experimental density at room temperature and capture the negative curvature of the density's temperature, while dPol-opt4 is less accurate even at room temperature (Figure 2.3c). None of the dPol models correctly position the experimental density maximum, but this is not essential for successful application to biomolecular systems near room temperature. The temperature dependence of the coefficient of thermal expansion (Figure 2.3d) tells essentially the same story, since it is directly related to the temperature derivative of density.

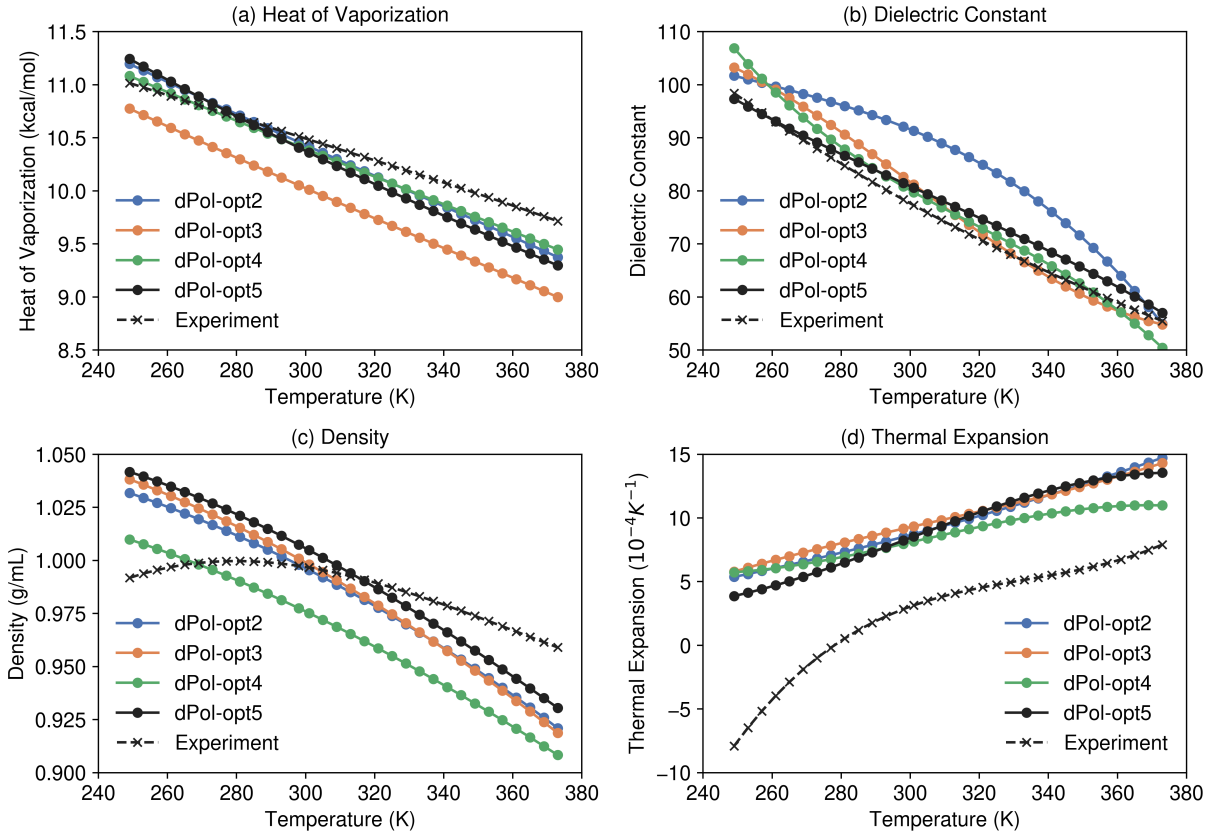


Figure 2.3: Calculated temperature-dependent properties using dPol water models. x-axes are temperature (K), and y-axes are calculated and experimental properties.

2.3.4 Simulation Timing

We compared the simulation speed of the dPol-opt5 model with those of the TIP3P and iAMOEBA models for simulations of 1000 water molecules in a periodic cubic box; see Section 2.2.2 and Section S2.7.1 for details. We used a 2 fs time-step for dPol-opt5 and TIP3P and a 0.5 fs time-step for iAMOEBA, in accordance with the original iAMOEBA paper [34] . We also tested iAMOEBA and AMOEBA with a 2fs time-step using the multiple time step (MTS) algorithm. As shown in Table 2.3, dPol-opt5 runs $\sim$3$\times$ slower than TIP3P with the same time-step, so its speed is comparable to that of the recent fast Drude water model SWM3, which is $\sim$2.6$\times$ slower than TIP3P[87]. The dPol-opt5 (2fs) model is $\sim$4.4$\times$ faster than iAMOEBA

(0.5 fs), largely due to the latter's requirement for a shorter timestep, and it remains faster than both iAMOEBA (2fs/MTS) and AMOEBA (2fs/MTS). The fact that iAMOEBA (2fs) remains slower than dPol (1.2× slower) probably traces to its use of higher-order permanent multipoles.

Table 2.3: Speed comparison of various polarizable force fields. Simulations were performed on NVIDIA GeForce RTX 3090 using OpenMM version 7.7[46, 56], and the tolerance of the mutual induced dipoles of AMOEBA method was set to 1e-5.

	dPol-opt5 (2fs)	iAMOEBA (2 fs with MTS)	iAMOEBA (0.5 fs)	AMOEBA (2fs with MTS)	TIP3P (2fs)
speed (ns/day)	355	285	80	155	1020
slower than TIP3P	2.9	3.6	12.8	6.6	1

2.4 Discussion

We have developed a fast, simple, and accurate water model, dPol-opt5, which is designed to mesh with our AM1-BCC-dPol polarizable electrostatics model [3] . Of note, dPol captures not only the heat of vaporization and density of water at room temperature, but also the self-diffusion coefficient and the high-frequency dielectric constant. Henceforth, the dPol-opt5 model will be referred to as the dPol water model, with the expectation that it will be employed as the default water model within an AM1-BCC-dPol-based polarizable force field.

Table 2.4: Structure of the present dPol model with those of other polarizable water models.

	dPol	iAMEOBA	AMOEBA	SWM4-HLJ	SWM3
Polarization method	Induced dipole	Induced dipole	Induced dipole	Drude oscillator	Drude oscillator
Self-consistent method	No	No	Yes	Yes	Yes
Thole damping	No	Yes	Yes	Yes	Yes
Number of parameters	7	23	26	15	12
Rigid	Yes	No	No	Yes	Yes
Virtual Site	No	No	No	Yes	No
van der Waals	12-6	Buffered 14-7	Buffered 14-7	12-6	12-6
vdW on H	No	No	Yes	Yes	No
Polarizability on Hs	Yes	Yes	Yes	No	No
time-step[a]	2 fs	0.5 fs	1 fs	1 fs	1 fs

a. iAMOEBA and AMOEBA may use a 2 fs time-step with the MTS algorithm.

The dPol water model is structured similarly to TIP3P, in the sense of being a rigid, three-particle model with a single Lennard-Jones site. However, by including atom-centered inducible dipoles, dPol provides a more accurate high-frequency dielectric constant and captures the change in mean dipole moment of a water model going from gas to liquid phase. Although it is, inevitably, slower than TIP3P, dPol model seeks to maximize efficiency not only by keeping the water molecule rigid but also by using the direct polarization approximation, instead of mutual (i.e., self-consistent) polarization.

It is useful to place dPol into the context of other polarizable water models, as done in Table 2.4. Interestingly, dPol has the fewest adjustable parameters; the next fewest is SWM3, with 12, while the most similar other model, iAMOEBA, has 23. In addition, the typical time-step for a polarizable force field is 0.5-1.0 fs, whether due to stability issues related to the self-consistent

calculation of inducible dipoles [88] or to the high frequency mostions of light Drude particles in the Drude approximation. In contrast, dPol by design supports the 2 fs time-step commonly used in biomolecular simulations. In addition, apart from the inducible dipoles, dPol shares a similar functional form with commonly used non-polarizable force fields, like AMBER[89, 90] and GAFF[91], a feature that we anticipate will simplify construction of a comprehensive force field based on this electrostatic model.

As the next step toward a comprehensive polarizable force field based on the direct approximation, we plan to retrain LJ parameters for a wide range of small organic molecules while using AM1-BCC-dPol electrostatics and the dPol water model. The fact that the dPol water model incorporates partial charges and polarizabilities generated from QM calculations by the same basic approach used in formulating the AM1-BCC-dPol electrostatics model suggests that dPol will be highly compatible in this context. In addition, the OpenFF Smirnoff dPol-Plugin makes it possible to integrate direct polarization into the OpenFF force fields parameterization workflow, which paves the path for developing a full direct polarizable force field [92–94]. Thus, we are optimistic that this strategy offers a relatively smooth path to a fast, general, polarizable force field.

2.5 Disclosures

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

2.6 Acknowledgements

We thank the Open Force Field Initiative and the Open Force Field Scientific Advisory Board for help with software infrastructure and scientific feedback. We thank William C. Swope

and Chapin Cavender for helpful discussions. MKG acknowledges funding from NIH grant R01GM132386 to PI Michael Shirts, U. Colorado Boulder. These findings are solely those of the authors and do not necessarily represent the views of the NIH.

Chapter 2, in full, has been submitted for publication of the material as it may appear in Journal of Chemical Theory and Computation, 2025, Drew, Liangyue W.; Gilson, Michael K. The dissertation author was the primary investigator and author of this paper.

2.7 Supplementary Information

2.7.1 Speed comparison of various polarizable force fields

MD simulations were performed on an NVIDIA GeForce RTX 3090 GPU using OpenMM version 7.7[46, 56]. The MPID OpenMM plugin[57–60] was used for the polarizable models. For each simulation, a periodic cubic box with 1000 water molecules was prepared with PackMol[64]. Energy minimization and equilibration were performed on all systems before a 50-ns of production run using *NPT* ensemble. The Langevin integrator was used for all simulations. OpenMM systems for AMOEBA and iAMOEBA were prepared with the force field version built in OpenMM, "amoeba2018.xml" and "iamoeba.xml", respectively. AMOEBA simulations were carried out with a time step of 1 fs as reported in the reference[11]. iAMOEBA simulations were carried out with a time step of 0.5 fs as reported in the reference[34], and a time step of 1 fs and 2 fs (2 fs with multi-timestep algorithm) were carried out to benchmark simulaton timing. The van der Waals and Coulomb cutoff distances for AMOEBA/iAMOEBA simulations were set to 9 Å and 7 Å, respectively. Periodic boundary conditions were applied with the Particle Mesh Ewald method for AMEOBA/iAMOEBA simulations. For dPol simulations, the LJPME method was used to compute long range interactions for both Coulomb and LJ with a cutoff distance of 5.5 Åand tolerance of 1e-4, and the switch distance function was turned off.

2.7.2 Additional Figures

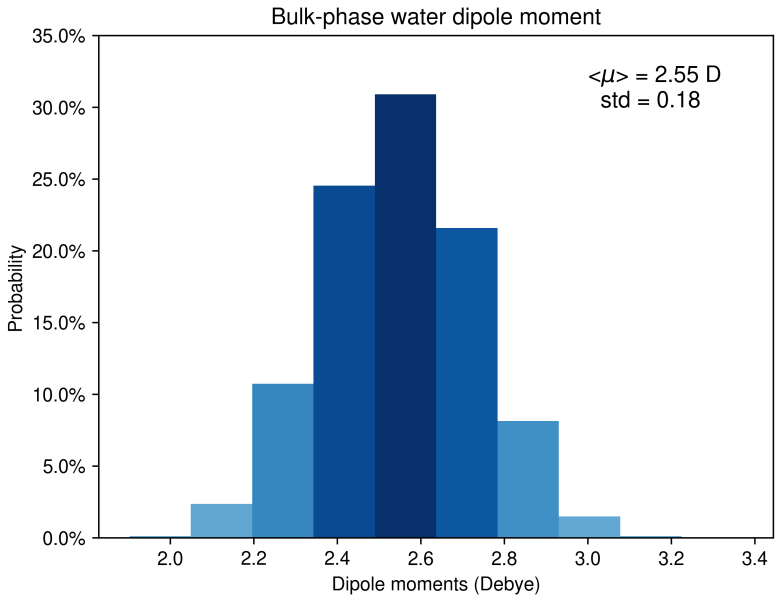


Figure 2.4: Distribution of water molecule instantaneous dipole moments in liquid water for dPol-opt5 at 298K and 1atm. The average dipole moment of a molecule is 2.55 D, with standard deviation 0.18 D.

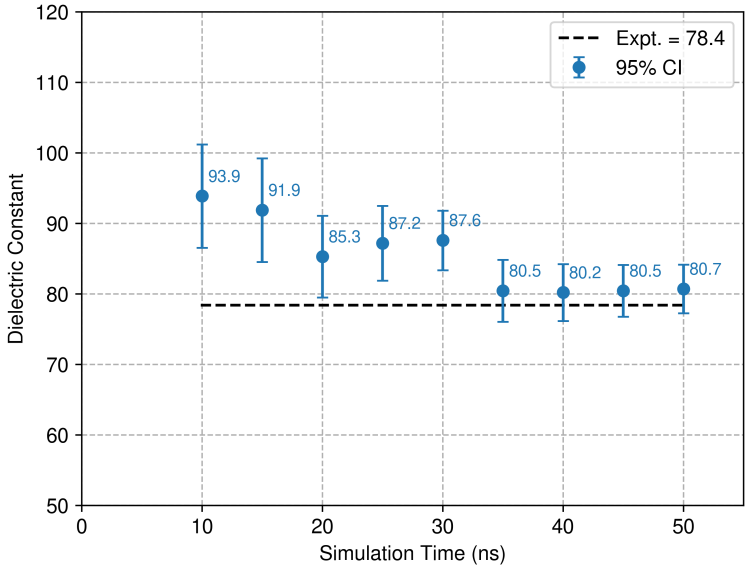


Figure 2.5: Dielectric constants are well-converged in a 50 ns MD simulation. The plotted data are from the dPol-opt5 model.

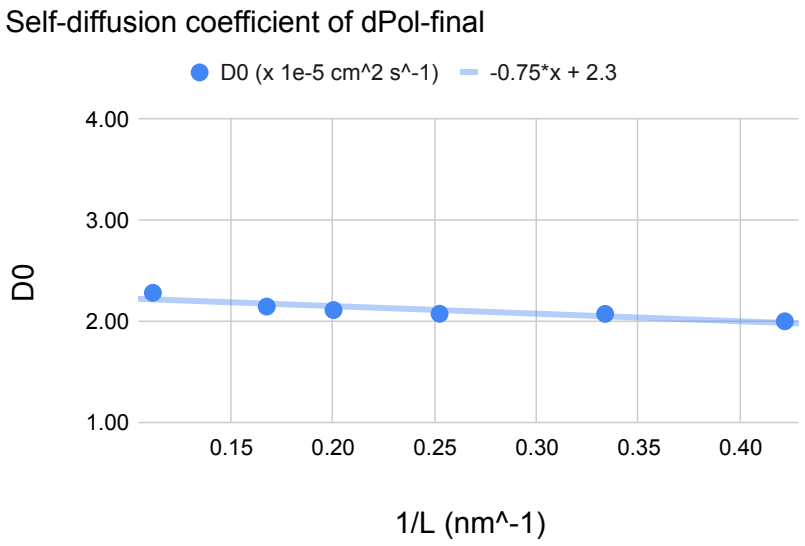


Figure 2.6: Self-diffusion coefficient, D0, (10^{-5}cm$^2 s^{-1}$) of dPol-opt5 model as a function of the inverse box size (1/L), for simulations at 298K and 1 atm

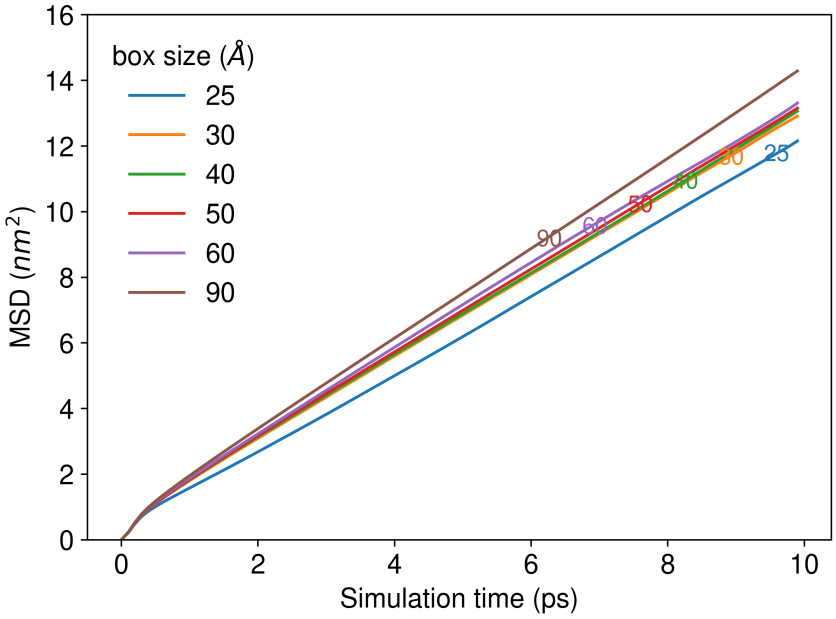


Figure 2.7: Mean-square displacement of dPol-opt5 water molecules was a function of simulation time for various box sizes, at 298 K and 1 atm. MSD was used to compute self-diffusion coefficients using different box sizes.

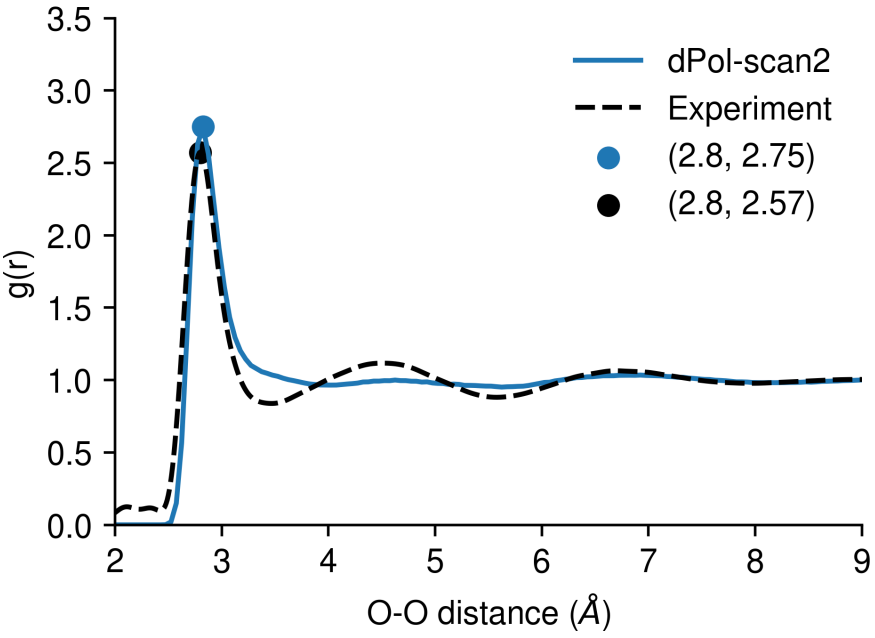


Figure 2.8: Oxygen-oxygen radial distribution function for unnamed water model obtained from the second grid search scan (Scan 2). Although the first peak has the correct position and height, g_{OO} is flat beyond the first peak.

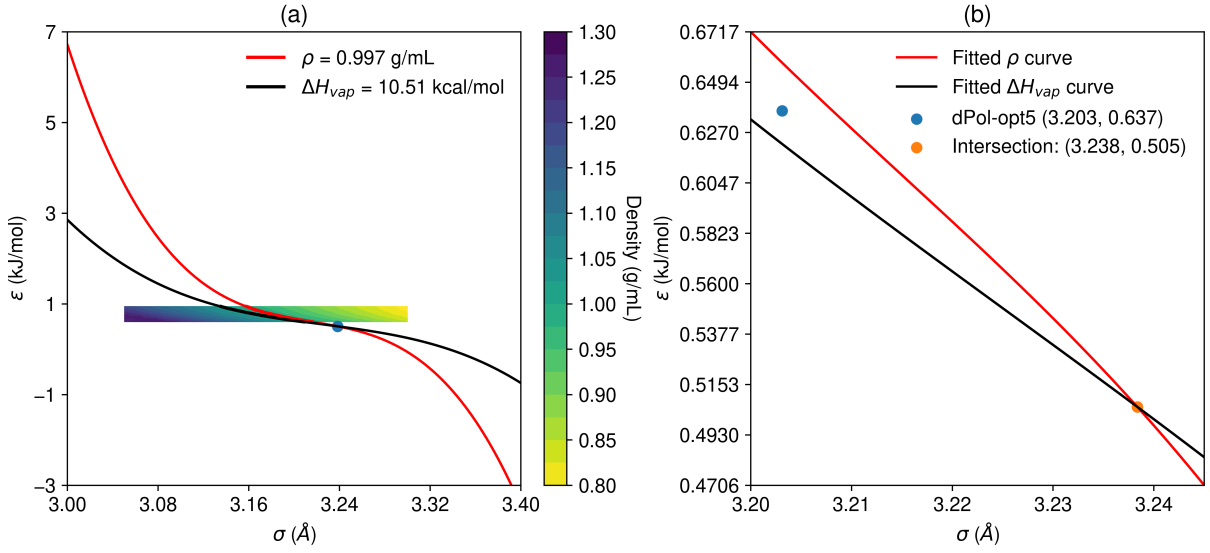


Figure 2.9: Third-order polynomial fits of Scan 3 grid search data to locate the intersection where the experimental density and heat of vaporization may be achieved using the current valence and electrostatics parameters. (a) Fitted curves and, for context, a segment of the scanned grid (contour map). (b) Zoom in near intersection of the fitted curves. Red: line where density matches experiment. Black: line where heat of vaporization matches experiment.

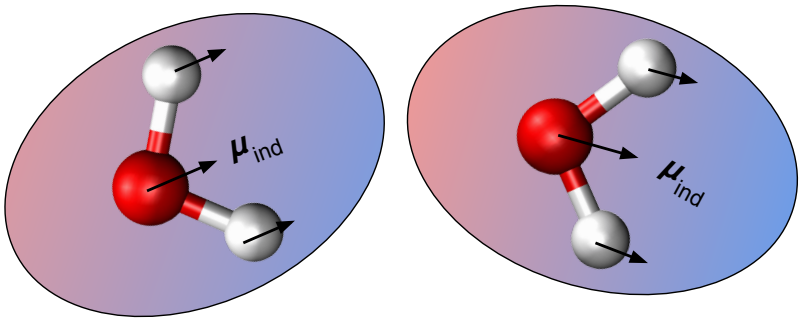


Figure 2.10: dPol water model illustration figure.

2.7.3 Code Availability

`dpol-water`: A repository contains scripts to use the dPol water model.

https://github.com/wwilla7/dpol-water

`dpolfit`: A Python package to develop and use dPol model.

https://github.com/wwilla7/dpolfit

`MPIDOpenMMPlugin`: The plugin to use dPol model with OpenMM.

https://github.com/andysim/MPIDOpenMMPlugin

2.7.4 TOC Graphic

Bibliography

(1) Lennard-Jones, J. E. Cohesion. *Proceedings of the Physical Society* **1931**, *43*, Publisher: IOP Publishing, 461–482.

(2) Leontyev, I.; Stuchebrukhov, A. Accounting for electronic polarization in non-polarizable force fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2613–2626.

(3) Wang, L.; Schauperl, M.; Mobley, D. L.; Bayly, C.; Gilson, M. K. A Fast, Convenient, Polarizable Electrostatic Model for Molecular Dynamics. *Journal of Chemical Theory and Computation* **2024**, *20*, PMID: 38240687, 1293–1305.

(4) Chaban, V. Polarizability versus mobility: atomistic force field for ionic liquids. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16055–16062.

(5) Belyaeva, J.; Zlobin, A.; Maslova, V.; Golovin, A. Modern non-polarizable force fields diverge in modeling the enzyme–substrate complex of a canonical serine protease. *Phys. Chem. Chem. Phys.* **2023**, *25*, 6352–6361.

(6) Gilson, M. K.; Honig, B. H. The dielectric constant of a folded protein. *Biopolymers* **1986**, *25*, 2097–2119.

(7) Simonson, T.; Brooks, C. L. Charge Screening and the Dielectric Constant of Proteins: Insights from Molecular Dynamics. *Journal of the American Chemical Society* **1996**, *118*, 8452–8458.

(8) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *The Journal of Physical Chemistry* **1987**, *91*, 6269–6271.

(9) Swope, W. C.; Horn, H. W.; Rice, J. E. Accounting for Polarization Cost When Using Fixed Charge Force Fields. I. Method for Computing Energy. *The Journal of Physical Chemistry B* **2010**, *114*, PMID: 20540503, 8621–8630.

(10) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *The Journal of Physical Chemistry B* **2003**, *107*, 5933–5947.

(11) 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. J.; 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*, PMID: 20136072, 2549–2564.

(12) Rackers, J. A.; Silva, R. R.; Wang, Z.; Ponder, J. W. Polarizable Water Potential Derived from a Model Electron Density. *Journal of Chemical Theory and Computation* **2021**, *17*, PMID: 34699197, 7056–7084.

(13) Bone, R. A.; Chung, M. K. J.; Ponder, J. W.; Riccardi, D.; Muzny, C.; Sundararaman, R.; Schwarz, K. A new method to calculate broadband dielectric spectra of solvents from molecular dynamics simulations demonstrated with polarizable force fields. *The Journal of Chemical Physics* **2024**, *161*, 064306.

(14) 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.

(15) 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.

(16) 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*, DOI: 10.1063/1.4964866.

(17) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *Journal of Chemical Theory and Computation* **2018**, *14*, PMID: 29438622, 2084–2108.

(18) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *Journal of Chemical Theory and Computation* **2013**, *9*, PMID: 24163642, 4046–4063.

(19) Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *The Journal of Physical Chemistry B* **2015**, *119*, PMID: 25683601, 9423–9437.

(20) 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.

(21) 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.

(22) 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.

(23) 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.

(24) Bedrov, D.; Piquemal, J.-P.; Borodin, O.; MacKerell, A. D. J.; Roux, B.; Schröder, C. Molecular Dynamics Simulations of Ionic Liquids and Electrolytes Using Polarizable Force Fields. *Chemical Reviews* **2019**, *119*, PMID: 31141351, 7940–7995.

(25) 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.

(26) 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.

(27) 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.

(28) Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *Journal of the American Chemical Society* **1981**, *103*, 335–340.

(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(30) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters* **2014**, *5*, PMID: 25400877, 3863–3871.

(31) Izadi, S.; Onufriev, A. V. Accuracy limit of rigid 3-point water models. *The Journal of Chemical Physics* **2016**, *145*, 074501.

(32) 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.

(33) Olson, R. M.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 4 and Intramolecular Charge Polarization. *Journal of Chemical Theory and Computation* **2007**, *3*, PMID: 26636200, 2046–2054.

(34) 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*, PMID: 23750713, 9956–9972.

(35) Margul, D. T.; Tuckerman, M. E. A Stochastic, Resonance-Free Multiple Time-Step Algorithm for Polarizable Models That Permits Very Large Time Steps. *Journal of Chemical Theory and Computation* **2016**, *12*, PMID: 27054809, 2170–2180.

(36) Vega, C.; Abascal, J. L. F. Simulating water with rigid non-polarizable models: a general perspective. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19663–19688.

(37) Sedano, L. F.; Blazquez, S.; Vega, C. Accuracy limit of non-polarizable four-point water models: TIP4P/2005 vs OPC. Should water models reproduce the experimental dielectric constant? *The Journal of Chemical Physics* **2024**, *161*, 044505.

(38) Kadaoluwa Pathirannahalage, S. P.; Meftahi, N.; Elbourne, A.; Weiss, A. C. G.; McConville, C. F.; Padua, A.; Winkler, D. A.; Costa Gomes, M.; Greaves, T. L.; Le, T. C.; Besford, Q. A.; Christofferson, A. J. Systematic Comparison of the Structural and Dynamic Properties of Commonly Used Water Models for Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **2021**, *61*, PMID: 34406000, 4521–4536.

(39) Huggins, D. J. Comparing the Performance of Different AMBER Protein Forcefields, Partial Charge Assignments, and Water Models for Absolute Binding Free Energy Cal-

culations. *Journal of Chemical Theory and Computation* **2022**, *18*, PMID: 35266690, 2616–2630.

(40) Çınaroğlu, S. S.; Biggin, P. C. Evaluating the Performance of Water Models with Host–Guest Force Fields in Binding Enthalpy Calculations for Cucurbit[7]uril–Guest Systems. *The Journal of Physical Chemistry B* **2021**, *125*, PMID: 33538161, 1558–1567.

(41) Henriksen, N. M.; Gilson, M. K. Evaluating Force Field Performance in Thermodynamic Calculations of Cyclodextrin Host–Guest Binding: Water Models, Partial Charges, and Host Force Field Parameters. *Journal of Chemical Theory and Computation* **2017**, *13*, PMID: 28696692, 4253–4269.

(42) Hagler, A. T.; Huler, E.; Lifson, S. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. *Journal of the American Chemical Society* **1974**, *96*, PMID: 4851860, 5319–5327.

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

(44) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *The Computer Journal* **1965**, *7*, 308–313.

(45) Wang, L. A Python toolkit to create and fit dPol water model against experimental training data, `https://github.com/wwilla7/dpolfit`, [Accessed Jul-09-2024], 2024.

(46) 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*, 1–17.

(47) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: an open-source ab initio electronic structure program. *WIREs Computational Molecular Science* **2012**, *2*, 556–565.

(48) Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **2020**, *17*, 261–272.

(49) Moritz, P.; Nishihara, R.; Wang, S.; Tumanov, A.; Liaw, R.; Liang, E.; Elibol, M.; Yang, Z.; Paul, W.; Jordan, M. I.; Stoica, I. Ray: a distributed framework for emerging AI applications. **2018**, 561–577.

(50) Moritz, P.; Nishihara, R.; Wang, S.; Tumanov, A.; Liaw, R.; Liang, E.; Elibol, M.; Yang, Z.; Paul, W.; Jordan, M. I.; Stoica, I. GitHub - ray-project/ray: Ray is an AI compute engine. Ray consists of a core distributed runtime and a set of AI Libraries for accelerating

ML workloads. — github.com, https://github.com/ray-project/ray, [Accessed 05-03-2025].

(51) Qiu, Y.; Nerenberg, P. S.; Head-Gordon, T.; Wang, L.-P. Systematic Optimization of Water Models Using Liquid/Vapor Surface Tension Data. *The Journal of Physical Chemistry B* **2019**, *123*, PMID: 31314516, 7061–7073.

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

(53) 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.

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

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

(56) 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 7.7.0, https://github.com/openmm/openmm/releases/tag/7.7.0, [Accessed 24-12-2024], 2021.

(57) 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*, DOI: 10.1063/1.4984113.

(58) 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*, DOI: 10.1063/1.4928530.

(59) 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, [Accessed 24-12-2024], 2017.

(60) Simmonett, A.; Wang, L. GitHub - wwilla7/MPIDOpenMMPlugin: OpenMM plugin that implements (an)isotropic polarizable point dipoles and multipoles up to octopoles. — github.com, https://github.com/wwilla7/MPIDOpenMMPlugin, [Accessed 24-12-2024], 2024.

(61) Zhang, Z.; Liu, X.; Yan, K.; Tuckerman, M. E.; Liu, J. Unified Efficient Thermostat Scheme for the Canonical Ensemble with Holonomic or Isokinetic Constraints via Molecular Dynamics. *The Journal of Physical Chemistry A* **2019**, *123*, PMID: 31117592, 6056–6079.

(62) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.

(63) Wang, L.; Thompson, M. OpenFF dPol Plugin for the direct polarization model, `https://github.com/wwilla7/dPol_plugin/tree/mpid`, [Accessed 24-12-2024], 2024.

(64) 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.

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

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

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

(68) Ren, P.; Ponder, J. W. Temperature and Pressure Dependence of the AMOEBA Water Model. *The Journal of Physical Chemistry B* **2004**, *108*, 13427–13437.

(69) Buckingham, A. D.; Sutton, L. E. A theory of the dielectric polarization of polar substances. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1956**, *238*, 235–244.

(70) Flyvbjerg, H.; Petersen, H. G. Error estimates on averages of correlated data. *The Journal of Chemical Physics* **1989**, *91*, 461–466.

(71) Hansen, J.; McDonald, I., *Theory of Simple Liquids*; Academic Press: 1976.

(72) Yeh, I.-C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *The Journal of Physical Chemistry B* **2004**, *108*, 15873–15879.

(73) Wagner, W.; Pruß, A. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *Journal of Physical and Chemical Reference Data* **2002**, *31*, 387–535.

(74) Mills, R. Self-diffusion in normal and heavy water in the range 1-45.deg. *The Journal of Physical Chemistry* **1973**, *77*, 685–688.

(75) Ando, T. Shear viscosity of OPC and OPC3 water models. *The Journal of Chemical Physics* **2023**, *159*, 101102.

(76) Booth, F. The Dielectric Constant of Water and the Saturation Effect. *The Journal of Chemical Physics* **1951**, *19*, 391–394.

(77) Narten, A. H.; Danford, M. D.; Levy, H. A. X-ray diffraction study of liquid water in the temperature range 4–200°C. *Discuss. Faraday Soc.* **1967**, *43*, 97–107.

(78) Halgren, T. A.; Damm, W. Polarizable force fields. *Current Opinion in Structural Biology* **2001**, *11*, 236–242.

(79) Sprik, M. Hydrogen bonding and the static dielectric constant in liquid water. *The Journal of Chemical Physics* **1991**, *95*, 6762–6769.

(80) VALLAURI, M. A. J. P. B. K. H. R. RESEARCH NOTE The use of a point polarizable dipole in intermolecular potentials for water. *Molecular Physics* **1998**, *94*, 873–876.

(81) Millot, J.-C. S. M. T. C. M. C. C. RESEARCH NOTE Static dielectric constant of the polarizable NCC water model. *Molecular Physics* **1998**, *94*, 577–579.

(82) Chaplin, M. Water models — water.lsbu.ac.uk, `https://water.lsbu.ac.uk/water/water_models.html`, [Accessed 14-03-2025].

(83) Camisasca, G.; Pathak, H.; Wikfeldt, K. T.; Pettersson, L. G. M. Radial distribution functions of water: Models vs experiments. *The Journal of Chemical Physics* **2019**, *151*, 044502.

(84) Soper, A. K. Joint structure refinement of x-ray and neutron diffraction data on disordered materials: application to liquid water. *Journal of Physics: Condensed Matter* **2007**, *19*, 335206.

(85) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. *The Journal of Chemical Physics* **2013**, *138*, 074506.

(86) Dang, L. X. The nonadditive intermolecular potential for water revised. *The Journal of Chemical Physics* **1992**, *97*, 2659–2660.

(87) Teng, X.; Yu, W.; MacKerell, A. D. J. Computationally Efficient Polarizable MD Simulations: A Simple Water Model for the Classical Drude Oscillator Polarizable Force Field. *The Journal of Physical Chemistry Letters* **0000**, *0*, PMID: 39841123, 1016–1023.

(88) 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.

(89) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020**, *16*, PMID: 31714766, 528–552.

(90) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11*, PMID: 26574453, 3696–3713.

(91) 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.

(92) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *Journal of Chemical Theory and Computation* **2021**, *17*, PMID: 34551262, 6262–6280.

(93) Boothroyd, S.; Madin, O. C.; Mobley, D. L.; Wang, L.-P.; Chodera, J. D.; Shirts, M. R. Improving Force Field Accuracy by Training against Condensed-Phase Mixture Properties. *Journal of Chemical Theory and Computation* **2022**, *18*, PMID: 35533269, 3577–3592.

(94) 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.

Chapter 3

Incorporating the AM1-BCC-dPol Polarizable Model into a Widely Applicable General Force Field

3.1 Introduction

Molecular dynamics (MD) simulations are widely used to study physical movements of complex biomolecular systems, providing insights into their structural and dynamical properties. Force fields are empirical parameters and functional forms used to estimate the total energy of a physical system. Molecular dynamics simulations predict the system's behavior by solving Newton's equations of motion at short femtosecond-level time-steps. The most widely used force fields typically consist of bonded and non-bonded terms. Bonded terms describe the bond stretching, angle bending, and torsional rotation in terms of force constants corresponding to bond lengths, bond angles, and dihedral angles, respectively. Non-bonded terms capture the interactions between atoms that are not directly captured by bonded terms. These interactions are often modeled using Lennard-Jones potentials for van der Waals forces, and Coulombic potentials

for electrostatic interactions.

In our previous work, we developed AM1-BCC-dPol[1], a polarizable model of electrostatic interactions utilizing the direct polarization approximation[2] . In the AM1-BCC-dPol electrostatics model, partial charges are assigned by performing a fast semi-empirical AM1 calculation[3] , then bond-charge corrections (BCCs) derived in the context of direct polarization (dPol) are applied to the AM1 charges to generate new partial charges. The BCCs used in AM1-BCC-dPol model are a library of typed parameters to refine atomic partial charges, and they are trained to reproduce quantum mechanical electrostatic potentials (QM ESPs) in the context of point-polarizabilities. The AM1-BCC-dPol electrostatic model may be considered as a polarizable version of the widely used AM1-BCC partial charge model[4, 5] . Preliminary results of the AM1-BCC-dPol model in condensed-phase simulations yielded markedly more accurate dielectric constants of organic liquids relative to the non-polarizable AM1-BCC model[1]. Given that we did not optimize a set of compatible Lennard-Jones parameters for AM1-BCC-dPol, we found that the liquid densities computed with the AM1-BCC-dPol model were less accurate compared to results with a matched non-polarizable force field. This aligns with what has been reported by Cerutti et al.[6] and Fennell et al.[7] , where improvements in partial charges alone did not increase the accuracy of simulations.

Non-bonded electrostatic and Lennard-Jones pairwise 12-6 potentials are generally not independent of each other. The partial charges are usually derived from *ab initio* reference data, such as ESPs and dipole moments, while experimental liquid densities and heats of vaporization of small organic liquids[8–10] are widely used to parameterize LJ parameters. These are fitted to calculated liquid properties from molecular dynamics or Monte Carlo (MC) simulations. One limit of this approach is that the values of other predefined parameters, such as partial charges and torsional parameters, can impact the optimized values of LJ parameters. Thus, using alternative charge models or torsional parameters may require reparameterization of LJ parameters.

New atom types are often introduced to improve force field accuracy[11, 12], inevitably

increasing the force field complexity. As a result, a well-developed force field typically includes several atom types for one element. Open Force Field (OpenFF) introduced the direct chemical perception approach to reduce the redundancy in force field parameter types with similar accuracy for liquid properties when compared to the General Amber Force Field[12] (GAFF) [13]. Moreover, Schauperl et al.[14] found that a minimalist set of element-typed LJ parameter types achieved matching force field accuracy with results of the SMIRKS Native Open Force Field (SMIRNOFF). Similarly, our minimalist element-typed polarizabilities showed higher accuracy and transferability than a set of more finely typed polarizabilities based on the SMIRNOFF format[1]. Motivated by these findings, we use the minimalist LJ typing scheme to re-optimize LJ parameters for the AM1-BCC-dPol model, restricting the number of LJ to elements, except for differentiating polar and nonpolar hydrogen.

The present study focuses on extending the coverage of the AM1-BCC-dPol electrostatics model to the full range of elements in drug-like molecules and re-optimizing Lennard-Jones parameters for the AM1-BCC-dPol model. Our previous proof-of-concept models were trained for C, H, O, and N. Here, we extend the coverage of our dPol mode to elements C, H, O, N, P, S, F, Cl, and Br. We benchmark the extended second generation of AM1-BCC-dPol model (AM1-BCC-dPol v2) on the accuracy of reproducing QM baseline ESPs, gas-phase dipole moments, and molecular polarizabilities. We optimize LJ parameters against experimentally measured density and heat of vaporization of pure liquids. The resulting parameters are then tested on calculation for hydration free energy with the recently developed direct polarization dPol water model[15] .

3.2 Methods

The polarizable AM1-BCC-dPol electrostatics model follows the force field structure described previously[1] . Briefly, each atom has an atom-centered constant partial charge and an isotropic scalar polarizability. The induced dipole created at each atom center is evaluated

non-iteratively as the product of its atomic polarizability and the electric field generated by only permanent charges. In accordance with the scaling scheme for short-range electrostatic interactions used in our previous model, we exclude 1-2 and 1-3 charge-polarizability interactions, and we scale 1-4 interactions by a factor of 0.5. The Lennard-Jones 12-6 potential is used to describe the van der Waals interactions. Harmonic potentials are used to describe the potential energies of bond stretching and angle bending, and the energy for the dihedral angle is expressed as a Fourier series. In Methods, we describe the training data to optimize a set of polarizabilities and BCCs with extended coverage in Section 3.2.1; the optimization of Lennard-Jones parameters for the AM1-BCC-dPol electrostatics model in Section 3.2.2; simulation details and calculation of physical properties in Section 3.2.3.

3.2.1 Broadened Parameterization of the AM1-BCC-dPol Polarizable Electrostatics model

Atom-centered polarizabilities and bond-charge-corrections are derived from fitting to quantum mechanical (QM) electrostatic potentials (ESPs). Molecular polarizabilites and molecular dipole moments serve as test data for polarizabilities and partial charges, respectively. We followed the same procedures as described in Ref.[1]. We outline the optimization procedure and the generation of training data in the following.

Dataset and QM Electrostatic Potentials

The QM ESPs were computed for 783 small organic molecules covering the chemical space of C, H, O, N, P, S, F, Cl, and Br. The unpolarized baseline ESPs (V_{base}) were computed without any external electric field. The polarized ESPs (V_{pol}) were computed with uniform imposed external electric fields ($E_{\text{ext,k}}$) of magnitude 0.01 a.u. To mimic the anisotropy effect of induced polarization, these external electric fields were imposed from six directions of $k \in$

[+x, -x, +y, -y, +z, -z]. The unpolarized baseline and polarized QM electrostatic potentials (ESPs) were evaluated at the MP2/aug-cc-pVTZ[16, 17] QM level of theory using Psi4[18] . ESPs were computed at Merz-Singh-Kollman (MSK)[19, 20] style grid points located outside of the van der Waals radii of all atoms. Grid points were disposed in 10 layers with a spacing of 0.126 Å and a density of 17/Å^2.

For molecules with rotatable bonds, multiple conformations are generated and minimized with the B3LYP-D3BJ/DZVP level of QM theory[21–23] . Molecules containing P, S, F, Cl, and Br were curated from two Open Force Field datasets[24, 25] available through QCArchive[26] . A training set of 442 molecules composed of 492 conformations was selected to generate seven sets of QM ESPs for each conformer for optimization of polarizabilities. An additional set of 573 molecules was added to generate unpolarized baseline QM ESPs to train BCC-dPol parameters. The SMILES strings of the dataset are available at the dPol-LJ github repository (`https://github.com/wwilla7/dpol-LJ`).

Typed Atom-centered Polarizabilities

The typed atom-centered polarizabilities were trained against the differences ($V_{\mathrm{diff,k}}$) between polarized QM ESPs and unpolarized QM ESPs. For each imposed external electric field k at grid point i, $V_{\mathrm{diff,ik}} = V_{\mathrm{baseline,i}} - V_{\mathrm{pol,ik}}$. The polarizabilities were derived by minimizing the following sum of squared errors in ESPs:

$$\chi^2 = \sum_{l=1}^{N_{\mathrm{conf}}} \sum_{k=1}^{6} \sum_{i=1}^{m_l} \left(V_{\mathrm{diff,ikl}} - \sum_{j=1}^{n_l} \frac{\mu_{\mathrm{ind,jkl}} r_{\mathrm{ijk}}}{r_{\mathrm{ij}}^3} \right)^2 \tag{3.1}$$

where N_{conf} is the number of conformations of all molecules in the training set, n_l is the number of atoms in the conformer l, i and m_l are the index and total number of ESP sampling points around the conformer l, respectively. The second term in the parenthesis is the potential at grid point i generated by the induced dipole $u_{\mathrm{ind,\,jkl}} = \alpha_j E_{\mathrm{jkl}}$, where α_j is the typed polarizability assigned

to atom j according to its element type. We used the Nelder-Mead[27] method in Scipy[28] to optimize typed polarizabilities for the training set described in Section 3.2.1.

AM1-BCC-dPol electrostatics model

The AM1-BCC-dPol electrostatics model generates high-quality partial charges compatible with the typed atom-centerd polarizabilites. The generation of AM1-BCC-dPol is similar to the standard AM1-BCC method[4, 5], consisting of fast AM1 population charges[3] as precharges (q_j^{pre}) and bond charge corrections (BCC). BCC corrections (B_β) are predefined based on bond connectivity analysis ($T_{j\beta}$):

$$q_j = q_j^{\mathrm{pre}} + \sum_{\beta=1}^{n_\beta} T_{j\beta} B_\beta \tag{3.2}$$

where the summation runs over the total number of bond types (n_β). The BCC types are adopted from the standard AM1-BCC method.

The BCC parameters were trained to reproduce the baseline QM ESPs using a least-squares fitting method. The electrostatic potential at grid point i is computed as:

$$V_{\mathrm{calc,i}} = V_{\mathrm{pre,\,i}} + \sum_{j=1}^{n} \sum_{\beta=1}^{n\beta} \frac{T_{j\beta} B_\beta}{r_{\mathrm{ij}}} + V_{\mathrm{ind,\,i}} \tag{3.3}$$

where $V_{\mathrm{pre,\,i}} = \sum_{j=1}^{n} \frac{q_{\mathrm{j}}^{\mathrm{pre}}}{r_{\mathrm{ij}}}$ is the Coulomb potential at grid point i generated by the AM1 precharge charges, and the $V_{\mathrm{ind,\,i}}$is the contribution from induced dipoles in the direct approximation:

$$V_{\mathrm{ind,\,i}} = \sum_{j=1}^{n} \frac{\mu_{\mathrm{ind,\,j}} r_{\mathrm{ij}}}{r_{\mathrm{ij}}^3} \tag{3.4a}$$

$$\mu_{\mathrm{ind,\,j}} = \alpha_{\mathrm{j}} \sum_{k \neq j}^{n} f_{\mathrm{jk}} \frac{q_{\mathrm{k}} r_{\mathrm{jk}}}{r_{\mathrm{jk}}^3} \tag{3.4b}$$

where q_{k} is the partial charge at atom k, f_{jk} is the scaling factor for short-range electrostatic

interactions. The sum of squared errors (χ^2) of the QM and MM ESPs were optimized as follows:

$$\chi^2 = \sum_{l=1}^{N_{\text{training}}} \sum_{i=1}^{m} (V_{\text{QM, il}} - V_{\text{calc, il}})^2 \tag{3.5}$$

where N_{training} is the number of compounds in the training set. Once the BCC-dPol parameters have been optimized, the AM1-BCC-dPol partial charges are assigned according to Equation 3.2. The `dPolfit` github repository (`https://github.com/wwilla7/dpolfit`) contains Python functions to derive and assign AM1-BCC-dPol partial charges.

3.2.2 Lennard-Jones Optimization

We optimized Lennard-Jones parameters for AM1-BCC-dPol electrostatics model to reproduce experimental observables: densities and heats of vaporization (ΔH_{vap}) of pure organic liquids. We adopted the training set and simple LJ types from Schauperl et al.[29]'s work following the success of the strategy. The `dPolfit` Python program[30] was used to carry out the optimization in the same manner as our previous dPol water model[15] . The free parameters in the LJ optimization consist of five LJ types: C, N, O, polar H, and non-polar H; each LJ type has two parameters: the depth of the LJ potential well (ε) and the distance at which the particle-particle potential energy is zero (σ). A training set was composed of experimental data of densities and heats of vaporization of 15 molecules, covering a variety of functional groups, including alkenes, amides, carboxylic acids, and aromatic hydrocarbons. The Nelder-Mead simplex algorithm[27] was used to optimize the agreement with training data, by minimizing the following objective function:

$$\chi = \sum_{i=1}^{N_{\text{training}}} w_i \left| \frac{x_{\text{calc,i}} - x_{\text{expt,i}}}{x_{\text{expt,i}}} \right| \tag{3.6}$$

where i indexes the N_{training} targeted experimental properties and w_i is the weight assigned for each target, which is 0.5 for density and ΔH_{vap}, respectively. The starting parameters were drawn from the optimized LJ parameters for RESP2$_{\sigma=0.6}$[29] (Table 3.3).

3.2.3 Simulation Details

The `dPolfit` Python package[30] implements all functionalities to apply the AM1-BCC-dPol electrostatics model and LJ optimization. The package calls the `antechamber` program[12, 31] to assign AM1 charges, Psi4 version 1.9[18, 32] for all QM calculations, and OpenMM version 7.7[33, 34] for all MD simulations. The MPID OpenMM plugin[35, 36] with keyword `direct` was used for MD simulations with direct polarization. For condensed-phase simulations, the MD time-step was 2 fs, pressures were maintained at 1 bar using the Monte Carlo barostat, and temperatures were maintained at 298 K using the Langevin thermostat[37] . We used the `dPol-Plugin`[38], a SMIRNOFF plugin to facilitate the interactions with OpenFF force fields[39–41] and software[42–44] . For all liquid systems, the valence terms were assigned from the OpenFF Sage 2.0.0 force field[40] . Cubic liquid boxes composed of 256 molecules were built with the PACKMOL package[45]. For each LJ optimization, liquid boxes were equilibrated for 500 ps and then ran for 1 ns for calculation of densities and heat of vaporization. The `dPolfit` program calls the Ray distributed framework[46, 47] to simultaneously perform MD simulations on multiple compute resources, allowing for fast and large-scale optimization.

The calculation of molecular dipole moments, molecular polarizabilities, mass densities, heats of vaporization, and static dielectric constant follows the methods described in Ref.[1, 15]. The hydration free energies (ΔG_{solv}) were computed alchemically with the dPol model using molecular dynamics simulations, and the workflow was implemented in `dpol-fsolv`[48] Python program. The newly developed dPol water mode was used as the solvent model. Solvated simulation systems were created with the PackMol package[45], consisting of 256 water molecules and one solute molecule. MD simulations were performed separately at alchemical intermediate λ

states, where $\lambda \in [0,1]$, using a spacing of 0.1 for a total of 10 states. The electrostatic component of solvation was computed as the free energy of turning on the solute electrostatics in water, subtracting the free energy of the same transformation in vacuum. The van der Waals component of solvation was computed as the free energy of turning on the Lennard-Jones interaction between the solute and water without electrostatic interactions between them. The Bennett acceptance ration (BAR) method[49] implemented in the `pymbar`[50, 51] program was used to calculate the free energy difference between a number of neighboring intermediate states. Each intermediate state was equilibrated with a 10 ps MD run, and a 3 ns production run over 500 iterations, each consisting of 500 steps using a time-step of 2 fs, in the NPT ensemble. The ΔG_{solv} calculation workflow is also implemented with support of the Ray framework to distribute MD simulations of different intermediate states for fast simultaneous modeling.

3.3 Results

3.3.1 Quality-of-fit to QM ESPs

We first compare the accuracy of reproducing QM ESPs around a given molecule. Our previous study showed that AM1-BCC-dPol trained on a small dataset affords accuracy that is comparable with ESP-based charges, such as RESP-dPol[1] and RESP[52] . In the current work, we trained the AM1-BCC-dPol v2 on a dataset that is an order of magnitude larger than that used in our previous work. As done previously, we evaluate the accuracy of the AM1-BCC-dPol model in terms of relative root-mean-square errors (RRMSE) to baseline gas-phase QM ESPs:

$$\text{RRMSE} = \left(\frac{1}{m} \frac{\sum_{i=1}^{m} (V_{\text{Calc, i}} - V_{\text{QM, i}})^2}{\sum_{i=1}^{m} (V_{\text{QM, i}})^2} \right)^{1/2} \tag{3.7}$$

where i indexes the number of m grid points.

We generated AM1-BCC-dPol charges and QM ESPs for a total of 1096 organic molecules (described in Section 3.2.1), of which 908 were used for training and 157 were used for testing. We also generated AM1-BCC charges using the same conformation with the `antechamber` program for comparison.

Table 3.1: RRMS (%) of AM1-BCC-dPol and AM1-BCC to baseline QM ESPs.

charge model	training set	test set
AM1-BCC-dPol v2	0.29%	0.26%
AM1-BCC	0.34%	0.31%

Table 3.1 shows that AM1-BCC-dPol v2 yields RRMS errors relative to QM ESPs that are lower than the standard AM1-BCC in both the training set and test set. As seen in our previous study, AM1-BCC-dPol partial charges yield marginally lower RRMS errors for the test set than the training set, indicating its transferability and suitability to be used in condensed-phase simulations. The AM1-BCC-dPol charge model is also more transferable across molecular conformations the non-polarizable AM1-BCC, as shown in previous work[1] .

3.3.2 Molecular Dipole Moments

We computed QM molecular dipole moments at the MP2/aug-cc-pVTZ QM level of theory for a total of 500 neutral molecules selected from the BCC training set to benchmark the accuracy of AM1-BCC-dPol v2 model. We chose the higher level MP2 method rather than the commonly used HF/6-31G* level of theory, because the MP2 method provides more truthful dipole moments in gas-phase that are more realistic for simulations in low dielectric environments, while partial charges derived from HF/6-31G* ESPs tend to overpolarize molecules[53, 54].

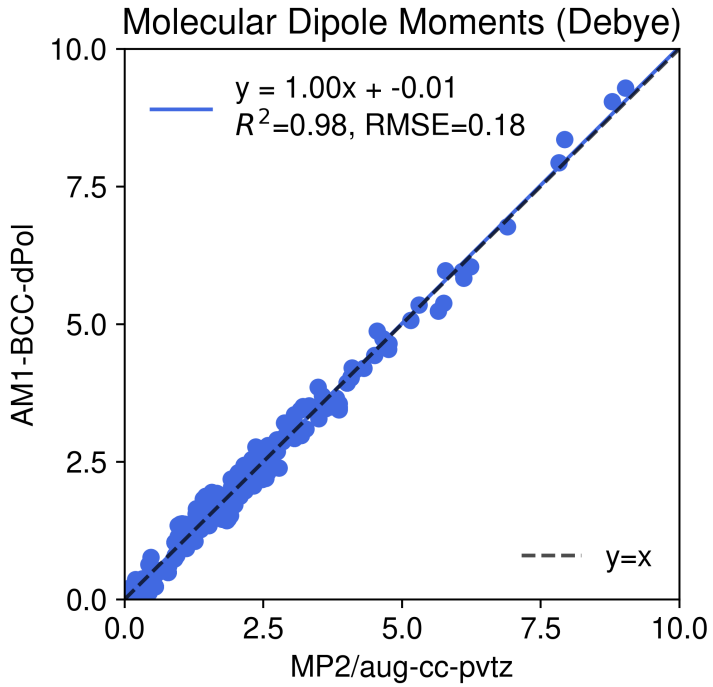


Figure 3.1: Comparison of 500 gas-phase dipole moments (Debye) computed with the AM1-BCC-dPol electrostatic mode with QM reference results computed at MP2/aug-cc-pvtz level of theory.

While not trained on molecular dipole moments, AM1-BCC-dPol v2 predicts dipole moments in excellent agreement with the QM reference, with a regression slope, intercept, and correlation of 1.00, 0.01, and 0.98, respectively. These results highlight that AM1-BCC-dPol v2 generates meaningful partial charges and provides a promising method to generate partial charges for both gas-phase and condensed-phase MD simulations.

3.3.3 Molecular Polarizabilities

We derived the present atomic polarizabilities to best reproduce QM ESPs and capture the intermolecular response of a molecule in electric fields. The resulting polarizability parameters are listed in Table 3.2. Deriving polarizabilities from QM ESPs emphasizes the short-range intermolecular interactions that are more relevant for capturing strong interactions in the condensed

phase. Indeed, Ref.[55] examined two methods of deriving atomic polarizabilities – from *ab initio* ESPs and experimental molecular polarizabilities – both are effective, while the ESP approach is more effective in capturing polarization for nitrogen and oxygen atoms that have long pair electrons.

QM electrostatic potentials provide valuable and convenient reference data to obtain force field parameters; however, parameters derived from QM ESPs do not guarantee accurate predictions of experimental observables, such as molecular polarizabilities. We examined our QM-derived atomic polarizabilities on molecular polarizabilities of 393 molecules curated from the dataset of Ref.[55].

Table 3.2: Atomic polarizabilities ($Å^3$) derived from fitting against differences in baseline and polarized QM ESPs as described in Section 3.2 and Ref.[1].

Index	Element	Polarizability ($Å^3$)
1	H	0.186
2	C	1.435
3	N	1.036
4	O	0.935
5	F	0.150
6	P	2.720
7	S	2.721
8	Cl	2.600
9	Br	5.493

Figure 3.2 shows the exceptional correlation between QM and dPol molecular polarizabilities, with a regression correlation of 0.99 and a root mean square error (RMSE) of 0.49 $Å^3$. Our minimalist element-typed polarizabilities transfer well to molecules with diverse chemical environments. This result indicates that our atom-centered polarizabilities are physically meaningful and accurate.

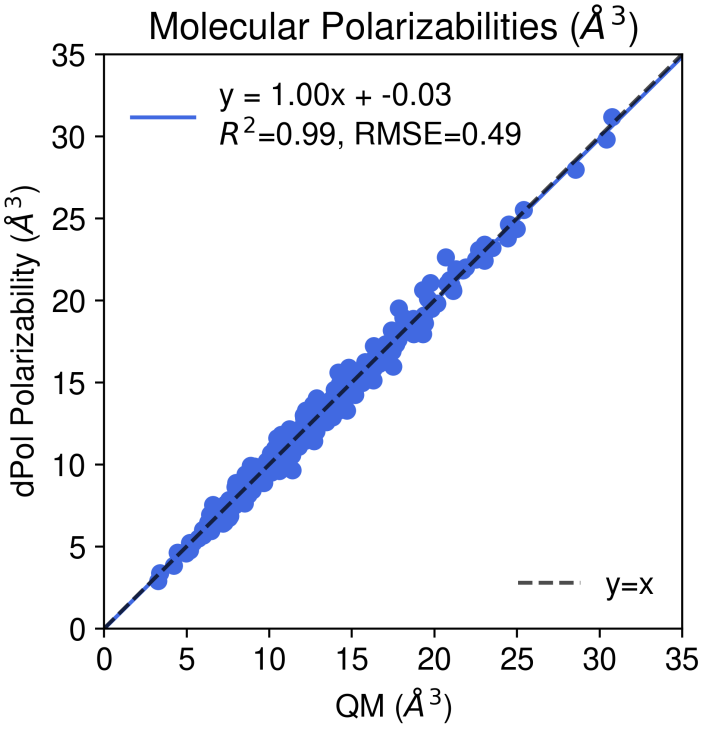


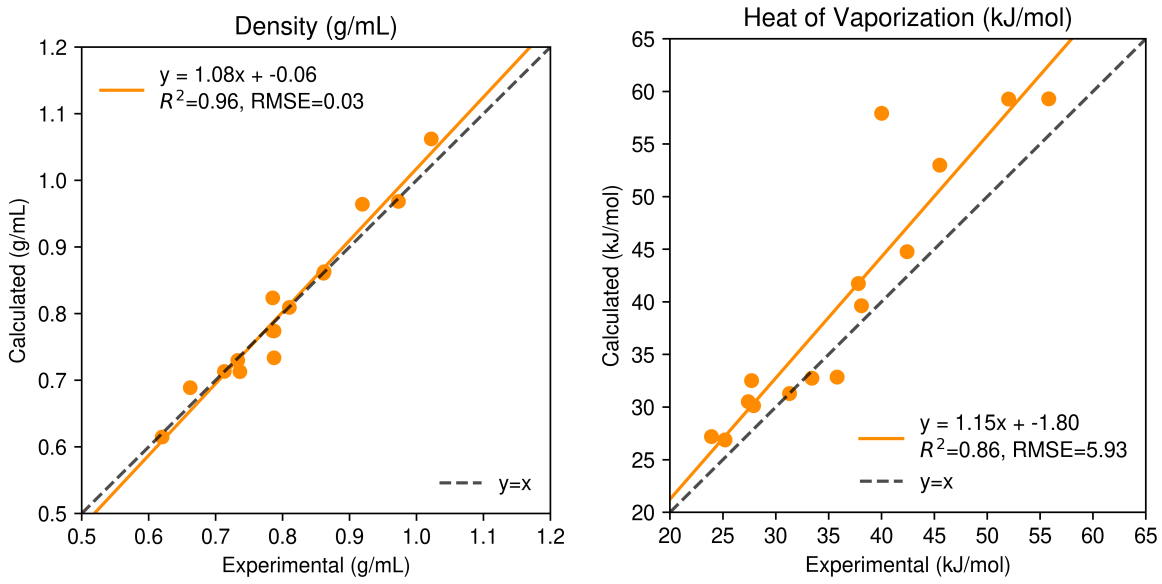
Figure 3.2: Comparison of MM and QM molecular polarizabilities. The MM values were calculated using typed atomic polarizabilities derived using the direct polarization approach. The QM values were obtained from Ref. [55].

3.3.4 Physical Properties of Pure Liquids

The accuracy of non-bonded interactions is determined by both electrostatics and van der Waals interactions. While the first generation of AM1-BCC-dPol was tested on liquid properties without re-optimization of LJ parameters, the liquid densities computed with the polarizable model were somewhat less accurate compared to the non-polarizable model that provided valence and LJ parameters[1]. Here, we re-optimize LJ parameters against densities and heats of vaporization for our AM1-BCC-dPol model to improve the overall force field accuracy.

We chose the minimalist LJ types as the starting point for an efficient systematic optimization of LJ parameters. Figure 3.3 shows the calculated liquid properties vs experimental measurements using AM1-BCC-dPol partial charges and LJ parameters drawn from RESP2$_{\sigma=0.6}$[29] .

Table 3.3 lists the starting and final optimized LJ parameters. The RESP2 is an advanced partial charges model that has its polarity tuned by optimal contributions from gas- and aqueous-phase QM calculations, however, the corresponding optimized LJ parameters yielded overestimated heats of vaporization when used with the explicit polarization AM1-BCC-dPol model. This result aligns with our intuition that using a more accurate electrostatics model without compatible LJ parameters may not improve the accuracy in liquid properties.



(a) Density

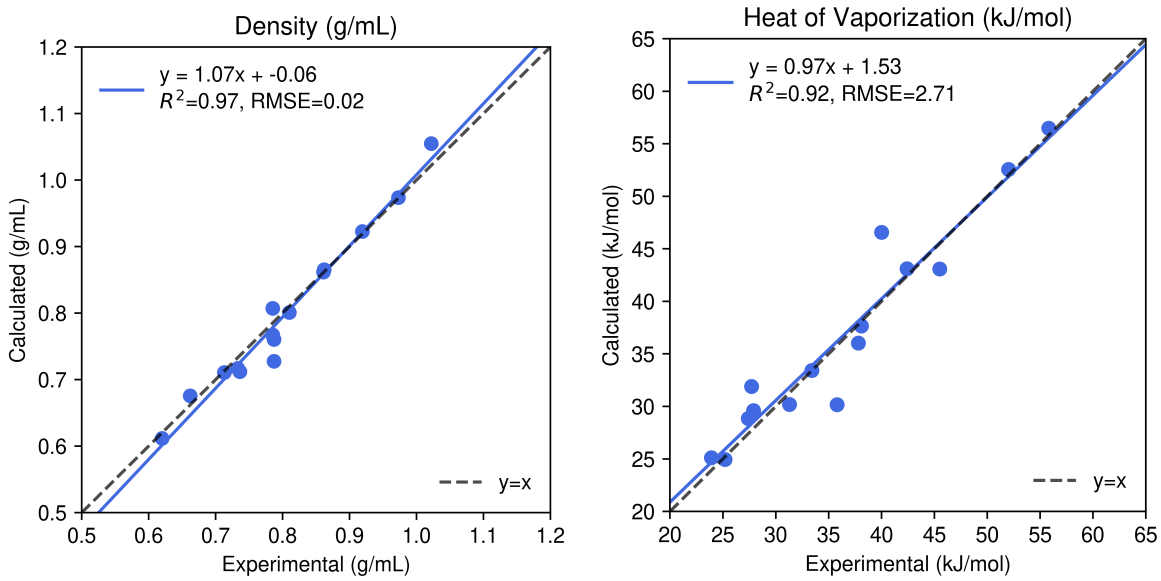
(b) Heat of Vaporization

Figure 3.3: Comparison of calculated and experimental density and heat of vaporization using LJ parameters drawn from RESP2$_{\sigma=0.6}$.

Table 3.3: Starting and final optimized LJ parameters. Starting parameters were drawn from optimized LJ parameters of RESP2$_{\sigma=0.6}$. ε has the unit of kcal/mol, and σ has the unit of Å.

LJ type	Starting values		Optimized values	
	ε	σ	ε	σ
C	0.08400	3.40784	0.08304	3.54894
N	0.20617	3.34313	0.21026	3.40454
O	0.15046	2.96682	0.15005	3.04805
nonpolar H	0.02277	2.56524	0.02230	2.54830
polar H	0.02673	0.96032	0.02714	0.98690

Figure 3.4 shows calculated results using LJ parameters resulting from the optimization. Table 3.4 lists liquid densities and heats of vaporization from experiments and simulations computed with the optimized LJ parameters.



(a) Density

(b) Heat of Vaporization

Figure 3.4: Comparison of calculated and experimental density and heat of vaporization using the optimized LJ parameters.

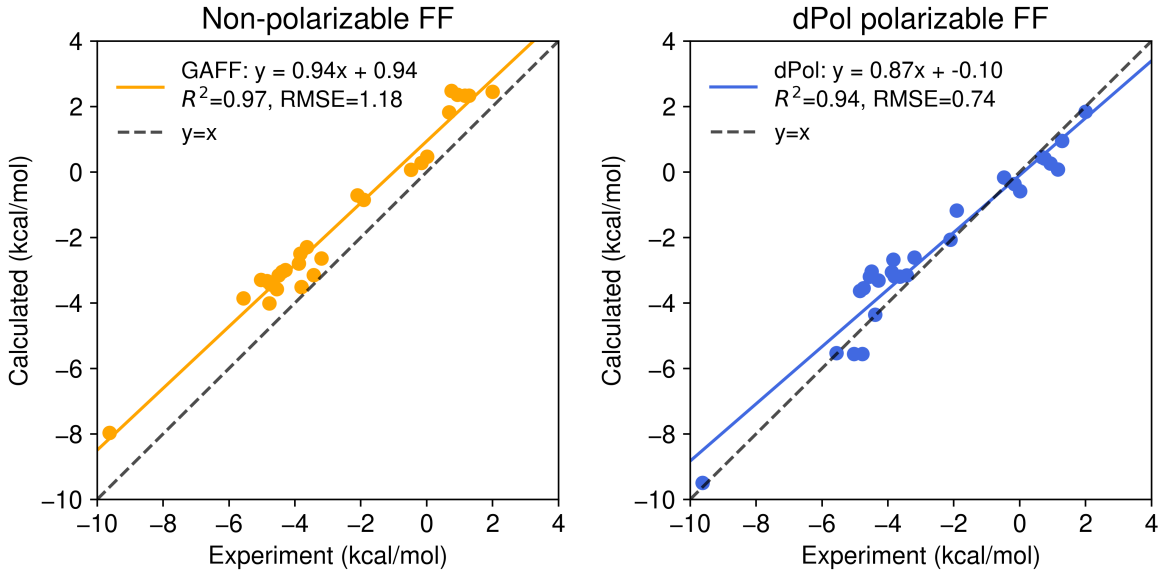
Table 3.4: Density (g/mL) and heat of vaporization (kcal/mol) used as training data. The Expt. columns show reference data obtained from Ref. [29]. The dPol columns show results computed with the final optimized LJ parameters.

Index	Smiles	Density		Heat of Vaporization	
		dPol	Expt.	dPol	Expt.
0	CCOCC	0.71	0.71	6.89	6.55
1	CC1=CC=CC=C1	0.87	0.86	9.00	9.11
2	CO	0.76	0.79	8.61	9.03
3	COCCC	0.71	0.74	7.07	6.67
4	CCCN	0.72	0.73	7.21	8.56
5	C-C#N	0.73	0.79	7.99	7.98
6	CC(C)CC	0.61	0.62	5.96	6.02
7	Cc1ccc(C)cc1	0.86	0.86	10.30	10.13
8	CC(=O)C	0.77	0.79	7.21	7.48
9	c1cocc1	0.97	0.97	7.62	6.62
10	N1(C)CCOCC1	0.92	0.92	11.13	9.56
11	c1ccccc1N	1.05	1.02	13.50	13.34
12	CN	0.68	0.66	6.00	5.71
13	CCCCO	0.80	0.81	12.56	12.43
14	CC(C)O	0.81	0.79	10.30	10.87

LJ optimization with the Nelder-Mead simplex algorithm led to improved accuracy in heats of vaporization while maintaining good agreement with experiments for densities for the 15 molecules in the training set (Figure 3.4). The optimized LJ parameters did not deviate much from the starting parameters. The σ of carbon has the most increased, creating more repulsive interactions between carbons. The optimized LJ parameters were further tested in the calculation of hydration free energies.

We computed ΔG_{solv} for 27 molecules outside of the LJ training data. Calculated ΔG_{solv} were compared with calculated results using the GAFF force field[12] and experimental data, both obtained from the FreeSolv database[56]. Shown in Figure 3.5, the new dPol polarizable model

yielded a lower RMSE of 0.74 kcal/mol in ΔG_{solv} than the RMSE of 1.18 kcal/mol that resulted from the well-developed GAFF and TIP3P models. The non-polarizable GAFF and TIP3P systematically underestimate the hydration free energies. Shown in Table 3.5 and supporting Figures 3.7-3.8 are detailed ΔG_{solv} data, we find that most improvements in dPol model compared to the non-polarizable are polar imidazole and non-polar alkene. The dPol water and non-bonded model capture the changes in electronic polarization in the gas-to-condensed-phase transfer scenario, predicting a more realistic electrostatic component in ΔG_{solv}.



(a) GAFF results obtained from FreeSolv.

(b) dPol model results.

Figure 3.5: Comparison of calculated and experimental hydration free energy. GAFF results are obtained from the FreeSolv database[56].

Table 3.5: Hydration free energies of organic molecules from simulations and experiments computed with dPol model and GAFF. Experiments and GAFF results were obtained from the FreeSolv database.

Index	Smiles	dPol	GAFF	Experimental
0	c1cnc[nH]1	-9.49	-7.97	-9.63
1	C1CNC1	-5.53	-3.86	-5.56
2	C=CCO	-5.56	-3.29	-5.03
3	CCCO	-3.63	-3.33	-4.85
4	c1cc[nH]c1	-5.56	-4.01	-4.78
5	CC(C)O	-3.55	-3.43	-4.74
6	CN	-3.19	-3.58	-4.55
7	CCN	-3.04	-3.16	-4.50
8	CCCN	-4.36	-3.05	-4.39
9	CNC	-3.31	-2.99	-4.29
10	CC#N	-3.06	-2.79	-3.88
11	CCC#N	-2.68	-2.49	-3.84
12	CC(=O)C	-3.18	-3.51	-3.80
13	CCCC#N	-3.19	-2.29	-3.64
14	CCC=O	-3.15	-3.15	-3.43
15	CN(C)C	-2.62	-2.64	-3.20
16	CCOC	-2.06	-0.71	-2.10
17	COC	-1.17	-0.85	-1.91
18	CC#C	-0.17	0.07	-0.48
19	CCC#C	-0.37	0.28	-0.16
20	CCCC#C	-0.59	0.47	0.01
21	CC(=C)C=C	0.45	1.82	0.68
22	C1CC1	0.41	2.48	0.75
23	C=CCC=C	0.26	2.36	0.93
24	C=C(C)C	0.07	2.33	1.16
25	C=C	0.95	2.33	1.28
26	C	1.84	2.45	2.00

3.4 Discussion

We present the second generation AM1-BCC-dPol v2 electrostatics model that extends the chemical spaces covered by our previous proof-of-concept polarizabilities and AM1-BCC-dPol[1] . We also optimized compatible LJ parameters for the AM1-BCC-dPol v2 model for simulations of physical properties. The dPol polarizable model runs $\sim$ 3 times slower than nonpolarizable fixed-charge force fields, as discussed in our previous work[1, 15]. The results here show that the polarizable AM1-BCC-dPol affords higher accuracy than non-polarizable GAFF at a moderate computational cost. We anticipate that the present model will be beneficial for simulations where molecules move between different dielectric environments, such as ligands moving into a protein's interior. We believe that this work provides the foundation for continuous development and testing for applications in biomolecular simulations.

3.5 Disclosures

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

3.6 Acknowledgements

MKG acknowledges funding from NIH grant R01GM132386 to PI Michael Shirts, U. Colorado Boulder. These findings are solely those of the authors and do not necessarily represent the views of the NIH.

Chapter 3, in part, is currently being prepared for submission for publication of the material, Drew, Liangyue W.; Gilson, Michael K. The dissertation author was the primary investigator and author of this paper.

3.7 Supplementary Information

3.7.1 Code availability

`dpol-LJ`: A repository contains scripts to develop and use the dPol Lennard-Jones model.

https://github.com/wwilla7/dpol-LJ

`dpolfit`: A Python package to develop and use dPol model.

https://github.com/wwilla7/dpolfit

`dpol-fsolv`: A repository to set up solvation free energy using dPol polarizable force field.

https://github.com/wwilla7/dpol-fsolv

`MPIDOpenMMPlugin`: The OpenMM plugin to use the dPol model in.MD simulations.

https://github.com/andysim/MPIDOpenMMPlugin

3.7.2 Additional Figures

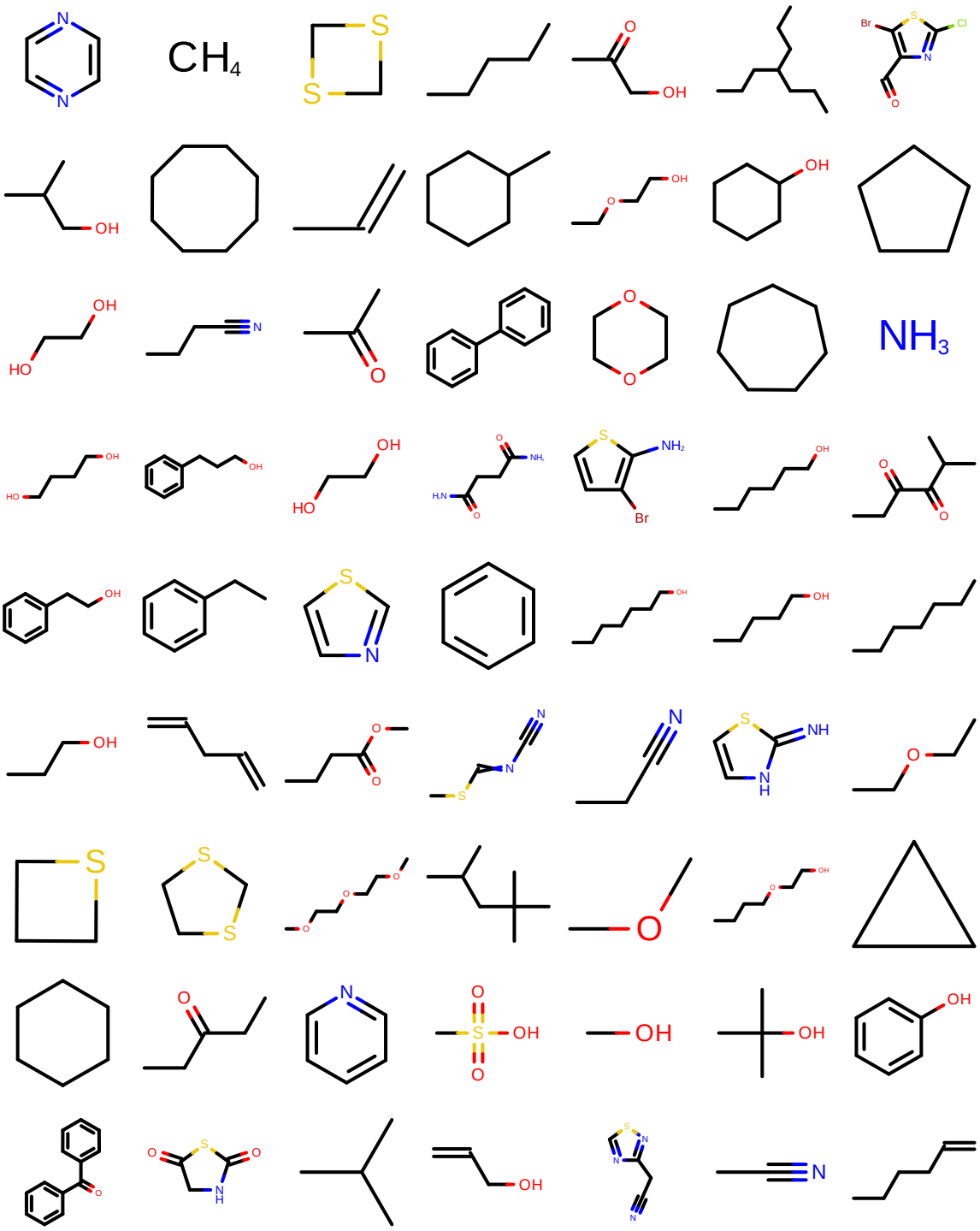


Figure 3.6: Chemical structures of molecules used in calculation for dipole moments.

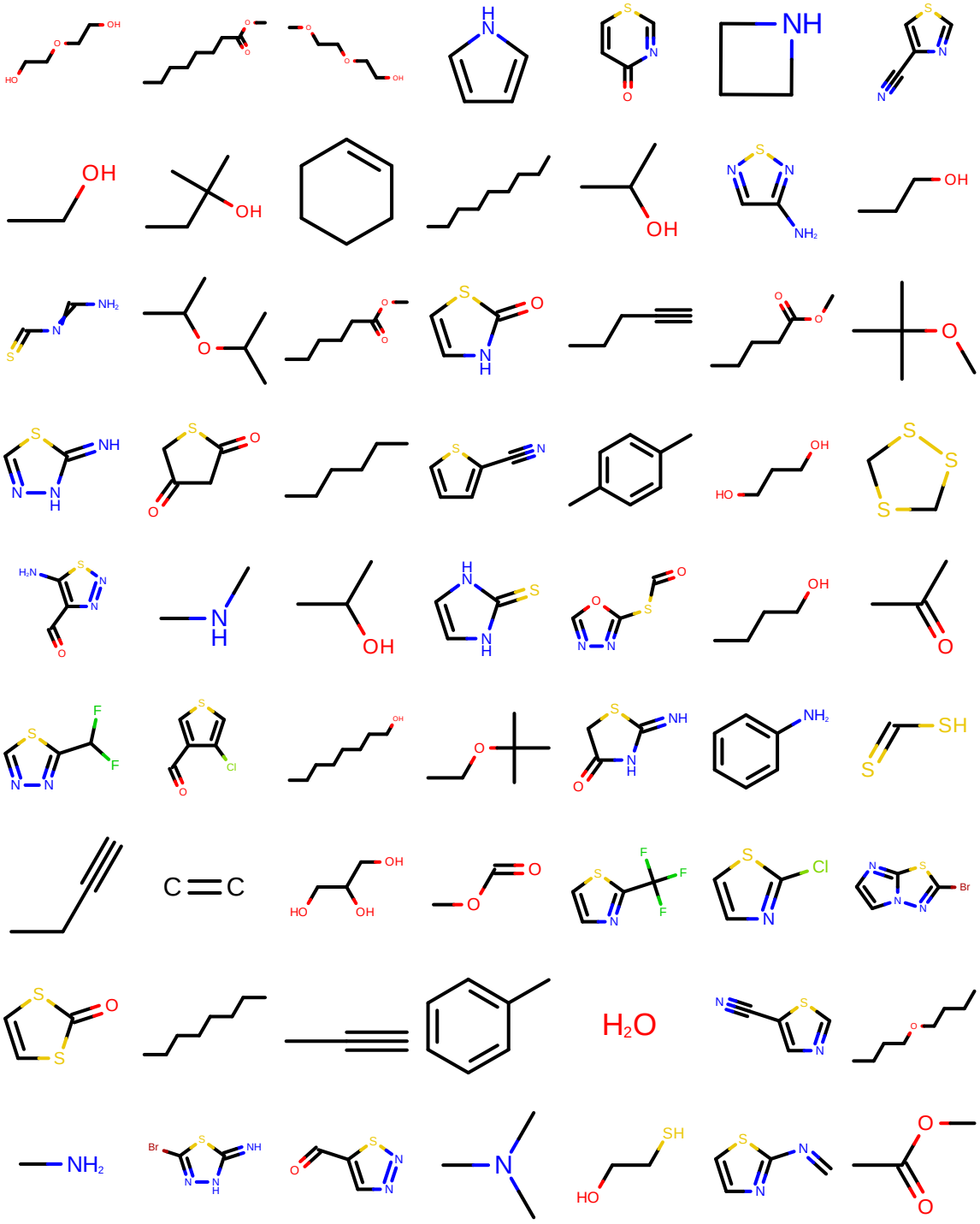


Figure 3.6: Chemical structures of molecules used in calculation for dipole moments (Continued).

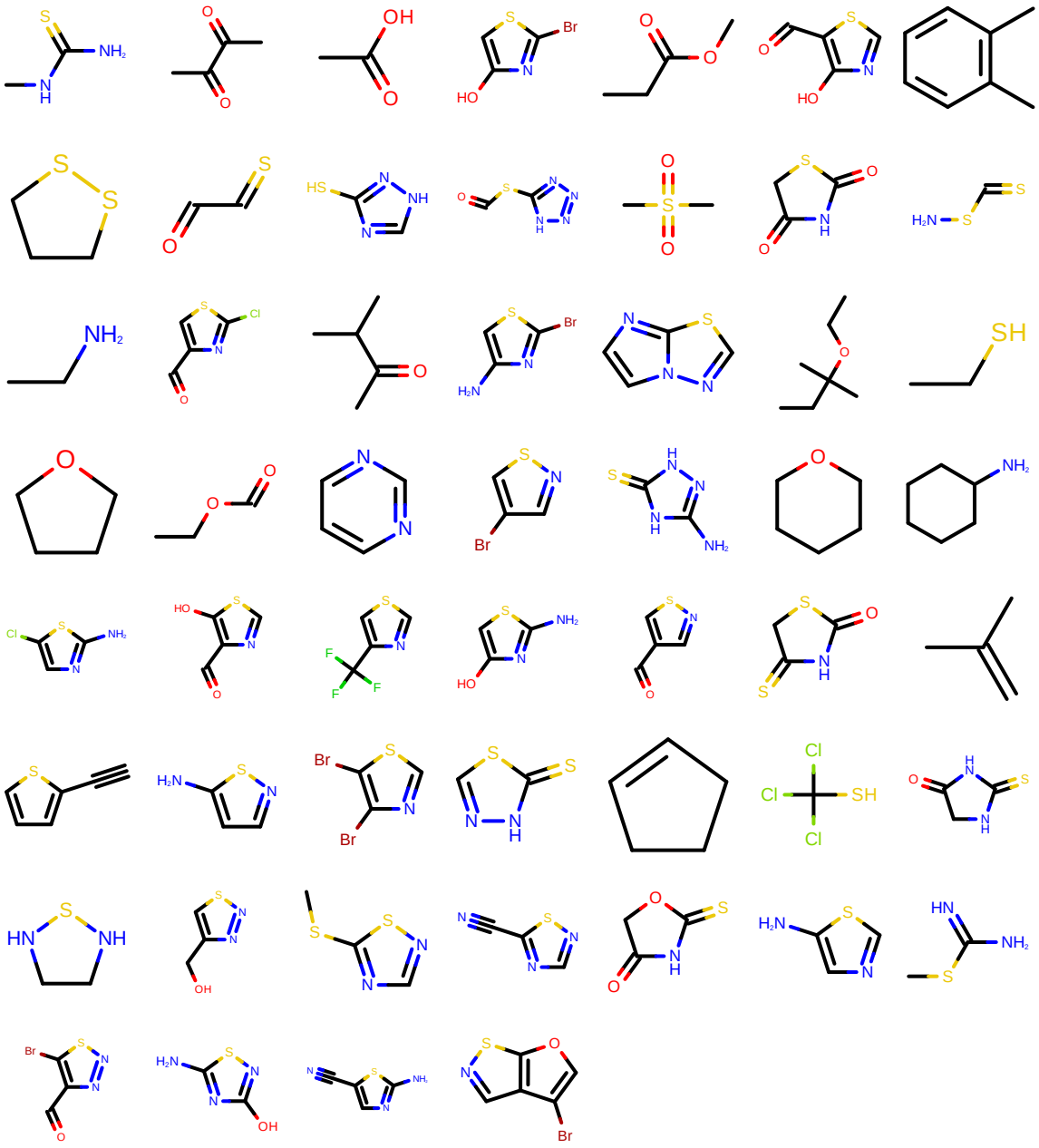


Figure 3.6: Chemical structures of molecules used in calculation for dipole moments (Continued).

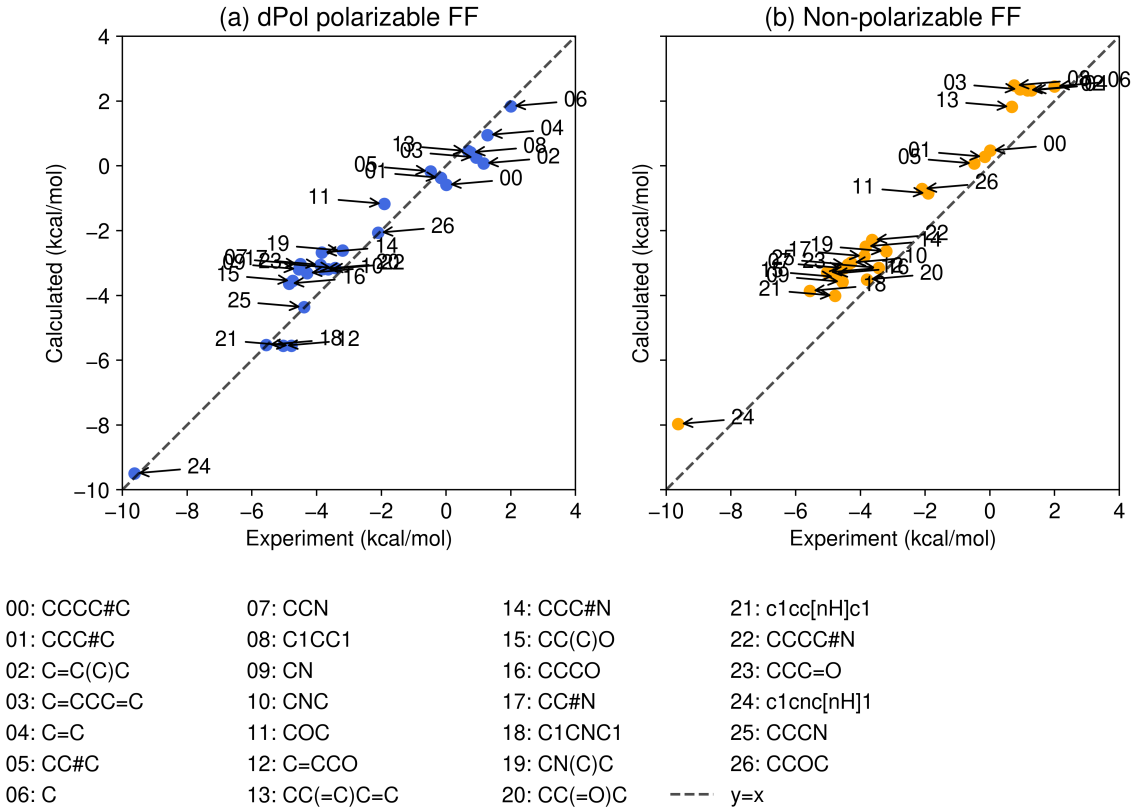


Figure 3.7: Labeled Comparison of calculated and experimental solvation free energy. GAFF results are obtained from the FreeSolv database.

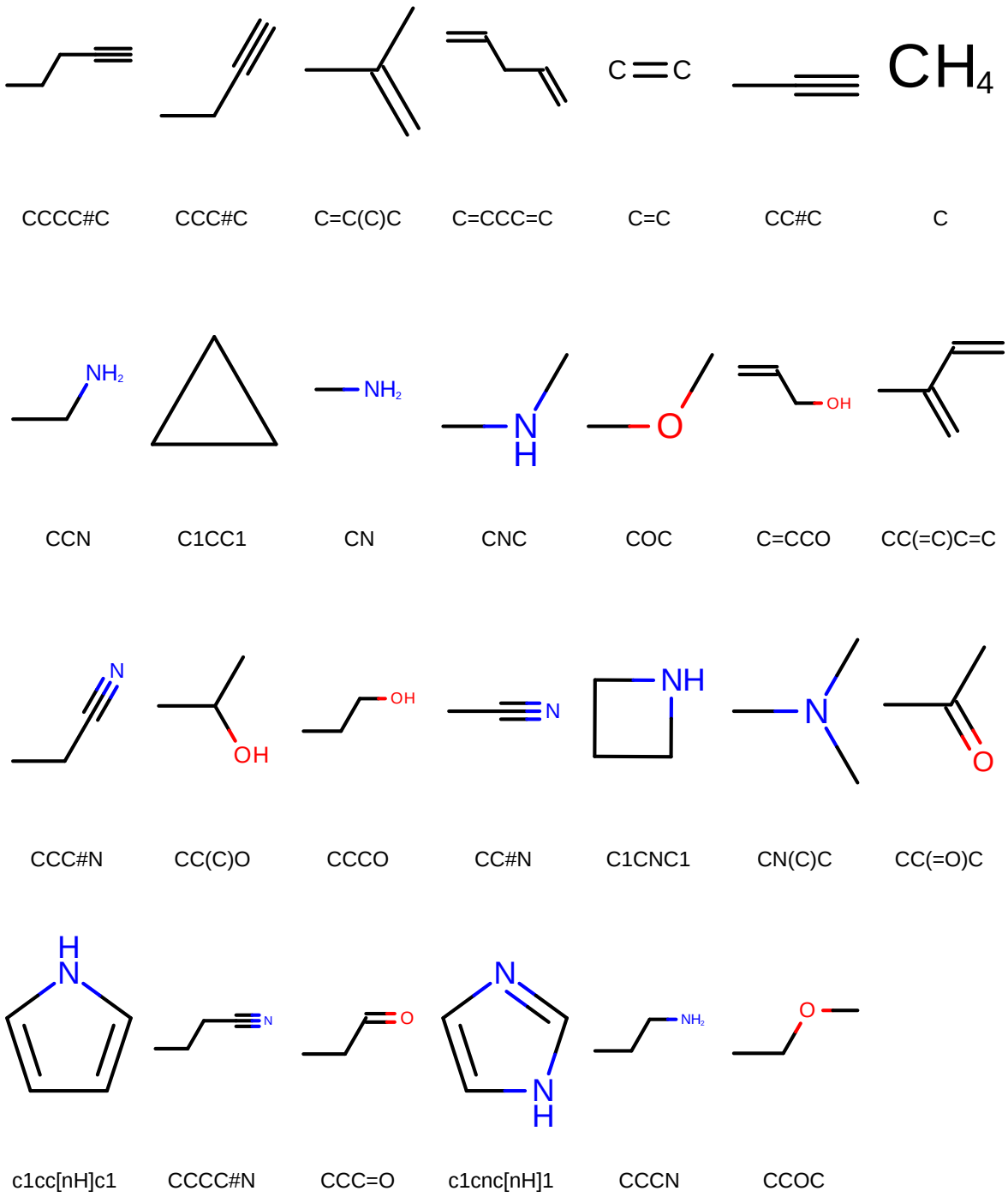


Figure 3.8: Chemical structures of dataset for solvation free energy calculations.

Bibliography

(1) Wang, L.; Schauperl, M.; Mobley, D. L.; Bayly, C.; Gilson, M. K. A Fast, Convenient, Polarizable Electrostatic Model for Molecular Dynamics. *Journal of Chemical Theory and Computation* **2024**, *20*, Publisher: American Chemical Society, 1293–1305.

(2) 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*, Publisher: Informa UK Limited, 181–192.

(3) 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.

(4) 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.

(5) 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.

(6) Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization. *The Journal of Physical Chemistry B* **2013**, *117*, Publisher: American Chemical Society, 2328–2338.

(7) Fennell, C. J.; Wymer, K. L.; Mobley, D. L. A Fixed-Charge Model for Alcohol Polarization in the Condensed Phase, and Its Role in Small Molecule Hydration. *The Journal of Physical Chemistry B* **2014**, *118*, Publisher: American Chemical Society, 6438–6446.

(8) 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*, Publisher: American Chemical Society (ACS), 1657–1666.

(9) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(10) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. Optimized intermolecular potential functions for liquid hydrocarbons. *Journal of the American Chemical Society* **1984**, *106*, Publisher: American Chemical Society, 6638–6646.

(11) Boulanger, E.; Huang, L.; Rupakheti, C.; MacKerell, A. D. J.; Roux, B. Optimized Lennard-Jones Parameters for Druglike Small Molecules. *Journal of Chemical Theory and Computation* **2018**, *14*, 3121–3131.

(12) 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*, Publisher: Wiley, 1157–1174.

(13) 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*, Publisher: American Chemical Society (ACS), 6076–6092.

(14) Schauperl, M.; Kantonen, S. M.; Wang, L.-P.; Gilson, M. K. Data-driven analysis of the number of Lennard–Jones types needed in a force field. *Communications Chemistry* **2020**, *3*, Publisher: Nature Publishing Group, 1–12.

(15) Drew, L. W.; Gilson, M. K. A simple, polarizable, rigid, 3-point water model using the direct polarization approximation, 2025.

(16) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, *46*, 618–622.

(17) 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.

(18) Smith, D. G. A.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; James, A. M.; Lehtola, S.; Misiewicz, J. P.; Scheurer, M.; Shaw, R. A.; Schriber, J. B.; Xie, Y.; Glick, Z. L.; Sirianni, D. A.; O'Brien, J. S.; Waldrop, J. M.; Kumar, A.; Hohenstein, E. G.; Pritchard, B. P.; Brooks, B. R.; Schaefer, H. F.; Sokolov, A. Y.; Patkowski, K.; DePrince, A. E.; Bozkaya, U.; King, R. A.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *The Journal of Chemical Physics* **2020**, *152*, 184108.

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

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

(21) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.

(22) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(23) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

(24) Open Force Field ESP Fragment Conformers.

(25) Open Force Field multi Br ESP Fragments Conformers.

(26) Smith, D. G. A.; Altarawy, D.; Burns, L. A.; Welborn, M.; Naden, L. N.; Ward, L.; Ellis, S.; Pritchard, B. P.; Crawford, T. D. The MolSSI QCArchive project: An open-source platform to compute, organize, and share quantum chemistry data. *WIREs Computational Molecular Science* **2021**, *11*, e1491.

(27) Gao, F.; Han, L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications 2010 51:1* **2010**, *51*, Publisher: Springer, 259–277.

(28) Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **2020**, *17*, 261–272.

(29) 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*, Publisher: Nature Publishing Group, 1–11.

(30) Drew, L. W. dPolfit: A package to derive and use the fast polarizable force field. 2024.

(31) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260.

(32) PSI4: Open-Source Quantum Chemistry, version 1.9.

(33) 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*, ed. by Gentleman, R., Publisher: Public Library of Science (PLoS), e1005659.

(34) 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 7.7.0, 2021.

(35) Simmonett, A. GitHub - andysim/MPIDOpenMMPlugin: OpenMM plugin that implements (an)isotropic polarizable point dipoles and multipoles up to octopoles. — github.com, 2017.

(36) Simmonett, A.; Drew, L. W. MPIDOpenMMPlugin: OpenMM plugin that implements (an)isotropic polarizable point dipoles and multipoles up to octopoles. 2024.

(37) Zhang, Z.; Liu, X.; Yan, K.; Tuckerman, M. E.; Liu, J. Unified Efficient Thermostat Scheme for the Canonical Ensemble with Holonomic or Isokinetic Constraints via Molecular Dynamics. *The Journal of Physical Chemistry A* **2019**, *123*, 6056–6079.

(38) Drew, L. W.; Thompson, M. Open Force Field dPol Plugin for the direct polarization model, 2024.

(39) Wang, L.; Behara, P. K.; Thompson, M. W.; Gokey, T.; Wang, Y.; Wagner, J. R.; Cole, D. J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L. The Open Force Field Initiative: Open Software and Open Science for Molecular Modeling. *The Journal of Physical Chemistry B* **2024**, *128*, Publisher: American Chemical Society, 7043–7067.

(40) 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*, Publisher: American Chemical Society (ACS), 3251–3275.

(41) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *Journal of Chemical Theory and Computation* **2021**, *17*, Publisher: American Chemical Society (ACS), 6262–6280.

(42) 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, Version Number: 0.14.5, 2023.

(43) 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*, Publisher: American Chemical Society (ACS), 3566–3576.

(44) Thompson, M.; Wagner, J.; Gilmer, J. B.; Timalsina, U.; Quach, C. D.; Boothroyd, S.; Mitchell, J. A. OpenFF Interchange, version v0.4.2, 2025.

(45) 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*, Publisher: Wiley, 2157–2164.

(46) Moritz, P.; Nishihara, R.; Wang, S.; Tumanov, A.; Liaw, R.; Liang, E.; Elibol, M.; Yang, Z.; Paul, W.; Jordan, M. I.; Stoica, I. Ray: a distributed framework for emerging AI applications. *Proceedings of the 13th USENIX Conference on Operating Systems Design and Implementation* **2018**, 561–577.

(47) Moritz, P.; Nishihara, R.; Wang, S.; Tumanov, A.; Liaw, R.; Liang, E.; Elibol, M.; Yang, Z.; Paul, W.; Jordan, M. I.; Stoica, I. Ray is an AI compute engine. Ray consists of a core distributed runtime and a set of AI Libraries for accelerating ML workloads.

(48) Drew, L. W. A repository to set up solvation free energy using dPol polarizable force field, 2025.

(49) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Physical Review Letters* **2003**, *91*, Publisher: American Physical Society, 140601.

(50) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129*, 124105.

(51) Shirts, M.; Beauchamp, K.; Naden, L.; Chodera, J.; jaimergp; Martiniani, S.; Stern, C.; Henry, M.; Thompson, M.; Fass, J.; Gowers, R.; McGibbon, R. T.; Dice, B.; Jones, C.; Dotson, D. L.; York, F.; Pulido, I.; Burgin, T. choderalab/pymbar: 4.0.3 Support for Python 3.12 and Jax 0.3.25+, version 4.0.3, 2024.

(52) 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*, Publisher: American Chemical Society, 10269–10280.

(53) 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*, Publisher: American Chemical Society (ACS), 249–258.

(54) 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*, Publisher: American Chemical Society (ACS), 3678–3687.

(55) Litman, J. M.; Liu, C.; Ren, P. Atomic Polarizabilities for Interactive Dipole Induction Models. **2021**, DOI: 10.1021/acs.jcim.1c01307.

(56) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 711–720.

Chapter 4

Improving Force Field Accuracy for Metal-Organic Frameworks

4.1 Introduction

Metal-organic frameworks (MOF) are a fascinating class of materials with unique properties and a wide range of potential applications. The metal ions connect with organic molecules to form strong, symmetrical, and directional building blocks. These building blocks, termed secondary building units (SBUs), construct a tunable, highly ordered, and highly porous crystalline structure with desirable properties. The highly ordered and porous crystalline structures make it easy to characterize and understand the structure-property relationships. The tunable pore size and shape provide a high surface area for gas adsorption and catalysis. By introducing different metal ions or/and organic linkers into the MOF crystalline structure, scientists can design new materials with tailored properties.

Various methods have been used to study MOF structures. Experimental studies such as synthesis, X-ray diffraction, gas sorption analysis, mechanical testing, etc., are crucial for determining and understanding the structural properties and mechanisms of MOFs. Computational

studies such as quantum mechanics (QM), molecular mechanics (MM), and machine learning (ML) have been applied to model MOF materials to accelerate the innovation and prediction of their properties and applications. Density Functional Theory (DFT) provides valuable information about the secondary building units to understand how the metal ions and organic linkers arrange themselves. Periodic DFT offers the ability to simulate periodic, repeating crystal structures and predict important structural properties for understanding MOFs' mechanical properties. Molecular dynamics (MD) and Monte Carlo (MC) simulations complement DFT studies by capturing materials' dynamic behavior and flexibility over time. MD simulations are useful in understanding the vibrational modes and flexible gate-opening breathing behaviors. Compared to DFT methods, MD approaches are computationally less demanding with appropriate force field parameterization. With well-designed experimental or simulated data to build predictive models, ML approaches are powerful to guide suitable structure-property designs with target characteristics.

Force field accuracy is essential in reliable and meaningful MD simulations. The majority of force fields use mathematical approximations to describe the potential energy surface, and they may be decomposed into bonded interactions, including bond stretching, angle bending, and torsional rotation, and non-bonded interactions, which describe van der Waals interactions via Lennard-Jones (LJ) terms and electrostatics via Coulombic terms. In particular, MOFs pose unique challenges due to the charged complex metal-organic interactions, and gas-MOF interactions in gas adsorption. While force fields such as OPLS[1–5], AMBER[6], and CHARMM[7–10] are widely used for biomolecules and small organic molecules, they struggle to accurately describe the metal-ligand bonding characteristics due to the lack of describing delocalized electrons and metal coordination geometries. Special parameterization for metal-ligand centers is crucial for obtaining reliable MD simulation results. Specialized force fields tailored for MOFs, such as UFF[11], Quick-FF[12], and MOF-FF[13], are developed with a variety of functional forms and unique parameter sets. Several molecular dynamics studies for gas adsorption calculations

showed good agreement between experiments and simulations using specialized MOF force fields[14, 15].

Typically, nonbonded models are used to describe metals, where the metal ions are represented as van der Waals spheres with formal charges, and the ligand-metal interactions are described through Coulomb electrostatics and LJ terms. Additionally, harmonic stretch and bend potentials can be applied between the metal ion and coordinating organic ligands or surrounding residues to preserve the metal-ligand coordination geometry. Such restraints are termed zero-order bond restraints in OPLS3e[3]. The zero-order bond restraints model all metal-ligand coordination with same force constants: a bond stretch force constant of 300 kcal $mol^{-1}Å^{-2}$ and a bend force constant of 60 kcal $mol^{-1}Å^{-2}$[16]. Formal charges on the metal ions and partial charges on ligands are used to describe electrostatic interactions. While both nonbonded and the zero-order bond models can describe the metal-ligand coordination geometry with reasonable accuracy, they often show unexpected behavior when used in complex chemical environments. The reason for this is that metals can have many different types of ligands and metal-coordination geometries[17] . Force fields parameterized for a specific metal-ligand coordination may not transfer well to other metal-coordination geometries. Beyond the zero-order bond restraints model, the covalent bond model[18] is also useful in metal-organic simulations with a guaranteed preservation of metal-ligand coordination geometry. The bonded model treats the metal with its ligands as covalent bonds, and metal-ligand interactions are modeled with bonded terms and nonbonded terms. The partial charge on metal is generated with the same charge model as ligands, as opposed to an integer value in the nonbonded and zero-order bond models.

Force field parameters are typically parameterized by fitting the force field to reproduce experimental data or high-quality quantum chemical calculations. In general, the bond stretching and angle bending are fitted to vibrational frequencies and energy minimum geometries. In OPLS3[2], the bond and angle force constants are derived from second derivatives and minimum geometries obtained by DFT calculations. Because valence terms are insensitive to the other

force field terms, they can be determined without nonbonded terms. Force constants of valence terms can also be derived from QM Hessian matrices. One of the most widely used methods is the Seminario method[19]. In the Seminario method, the partial QM Hessian matrix components are projected onto relevant bonds and angles to derive force field harmonic force constants. This method has been successfully adopted to derive force fields for metalorganics, such as zeolitic imidazolate frameworks (ZIFs)[20–22]. Additionally, a modified Seminario method introduced by Allen et al.[23] has shown improved accuracy in reproducing QM normal mode frequencies compared to the original Seminario method. A recent work utilizing the modified Seminario method has shown the applicability of this method in metal-containing supramolecular structures[24].

Reliable MD simulations are invaluable in revealing the dynamic behaviors of MOF materials. These materials feature dynamic crystalline structures that can undergo reversible structural deformations upon gas adsorption or be induced by temperature and pressure changes. Understanding the elastic behavior and structural stability of MOFs is fundamental to predicting adsorption isotherms for chemical engineering applications of gas capture, separation, and purification. Mechanical properties such as Young's modulus, shear modulus, bulk modulus, and Poisson's ratio are key quantities in describing the elastic properties of MOFs. Moreover, the complex relationship between stress and strain can be quantified by a fourth-rank elastic tensor, which at most consists of 21 independent components following the symmetry of stress and strain tensors. These mathematical entities are useful data for training a force field to predict characteristic properties of MOF materials.

In this work, we present a proof-of-concept force field model based on the OPLS4[4] force field with special parameterization for MOF simulations, termed custom OPLS. The metal-ligand interactions are described with a bonded representation, the coordination geometry is obtained from periodic DFT calculations, and the metal-center force constants are determined by metal-core cut out from periodic structures and fine-tuned against elastic tensors. The force field is

validated on QM minimized geometry, normal modes, and mechanical properties. Moreover, the applicability of OPLS4 and custom OPLS4 for gas adsorption is tested on adsorption performance of MOF-5 for carbon dioxide.

4.2 Methods

We maintained the functional form of the OPLS force field in parameterization for metal-organic frameworks. In Methods, we describe the force field structure and the covalently bonded model for MOFs; methods to use the modified Seminario method to derive force field parameters for "covalently bonded" metal-ligand centers in Section 4.2.2; computational details of DFT calculations and molecular dynamics simulations for MOF in Section 4.2.3; methods to compute mechanical properties and gas adsorption isotherm in Section 4.2.4; force field parameter optimization against metal-ligand coordination geometry and elastic tensors in Section 4.2.5.

4.2.1 Force Fields

The functional form of the OPLS force field is given by[2]:

$$\begin{aligned} U(r^N) = & \sum_{i=1}^{N_{bonds}} k_i(l_i - l_{i,\mathrm{eq}})^2 + \sum_{i=1}^{N_{angle}} k_i(\theta_i - \theta_{i,\mathrm{eq}})^2 \\ & + \sum_{i=1}^{N_{dihedrals}} \left[\frac{V_1}{2}(1+cos\varphi) + \frac{V_2}{2}(1-cos2\varphi) + \frac{V_3}{2}(1+cos3\varphi) + \frac{V_4}{2}(1-cos4\varphi) \right] \\ & + \sum_{i=1}^{N} \sum_{j=i+1}^{N} (4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}) \end{aligned} \quad (4.1)$$

Bond stretch and angle bending energies are represented by a harmonic potential where the subscript eq denotes the equilibrium bond length l and angle θ. The torsional energies are represented by a truncated Fourier series that sums over all dihedral angles. In OPLS4[4], the metal-ligand interactions are often described using zero-order bond restraints. The OPLS zero-

order bond restraints contain a fixed stretch and bend force constant of 300 kcal/mol and 60 kcal/mol, respectively[3]. The metal-ligand geometry is restrained to the input coordination geometry.

Alternatively, the metal-ligand interactions can be represented by covalent bonds. Force constants for metal-ligand bonds and angles are parameterized separately to reproduce reference data. Non-integral partial charges are computed for metals and coordinating residues. A metal center is cut out from the periodic system for force field parameterization. The experimental crystal structures and periodic density functional theory (DFT) calculations are used to derive force field parameters and benchmark mechanical properties.

4.2.2 The Modified Seminario Method

The modified Seminario method was developed by Allen et al. in 2018 based on the original Seminario method[19] to parameterize force constants for harmonic potentials from the quantum mechanically computed Hessian matrix of the molecule. In this section, we briefly describe the two methods.

The reaction force σF due to a small displacement σx in a molecular system of N atoms can be described according to Hooke's Law:

$$\sigma F = -[k]\sigma x \tag{4.2}$$

where [k] is the 3N$\times$N force constant matrix, Hessian, of the molecule. The Hessian matrix is a symmetrical square matrix containing the second-order partial derivatives of the potential energy with respect to the atomic coordinates. Consider a system of two atoms A and B, the force felt by

atom A due to displacement of atom B is given by $\sigma F_A = -[k_{AB}]\sigma x_B$, and $[k_{AB}]$ is given as:

$$[k_{AB}] = \begin{bmatrix} \frac{\partial^2 E}{\partial x_A \partial x_B} & \frac{\partial^2 E}{\partial x_A \partial y_B} & \frac{\partial^2 E}{\partial x_A \partial z_B} \\ \frac{\partial^2 E}{\partial y_A \partial x_B} & \frac{\partial^2 E}{\partial y_A \partial y_B} & \frac{\partial^2 E}{\partial y_A \partial z_B} \\ \frac{\partial^2 E}{\partial z_A \partial x_B} & \frac{\partial^2 E}{\partial z_A \partial y_B} & \frac{\partial^2 E}{\partial z_A \partial z_B} \end{bmatrix} \tag{4.3}$$

The eigenvectors v_i and eigenvalues λ_i of k_{AB} are the directions of the displacements of the normal modes and the force constants corresponding to the displacements, respectively. To compute force constants in the direction of the bond (k_l), each eigenvector is projected onto the direction of the bond vector, $\hat{u}^{AB}$:

$$k_l = \sum_{i=1}^{3} \lambda_i |\hat{u}^{AB} \cdot \hat{v}_i^{AB}| \tag{4.4}$$

We used the conventional definition of bonded atoms specified by the connectivity.

Consider a system of two bonds AB and CB, which form an angle ∠ABC. The angle force constant (k_θ) is computed as the projection of each eigenvector onto the direction perpendicular to bond AB and CB. The unit vector perpendicular to the plane ABC is given by:

$$\hat{u} = \frac{\hat{u}^{AB} \times \hat{u}^{CB}}{|\hat{u}^{AB} \times \hat{u}^{CB}|} \tag{4.5}$$

where $\hat{u}$ is the unit vector perpendicular to the plane ABC, and $\hat{u}^{AB}$ and $\hat{u}^{CB}$ are the unit vectors of bond AB and CB. The unit vectors perpendicular to the bonds AB and CB on the plane ABC are $\hat{u}^{PA} = \hat{u} \times \hat{u}^{AB}$ and $\hat{u}^{PC} = \hat{u} \times \hat{u}^{CB}$, respectively. The bond force constants projected onto the direction of unit vectors are calculated as:

$$k_{PA} = \sum_{i=1}^{3} \lambda_i^{AB} |\hat{u}^{PA} \cdot \hat{v}_i^{AB}| \tag{4.6}$$

and:

$$k_{PC} = \sum_{i=1}^{3} \lambda_i^{CB} |\hat{u}^{PC} \cdot \hat{v}_i^{CB}| \tag{4.7}$$

The angle force constant is calculated as the equivalent spring constant of two springs connected in series:

$$\frac{1}{k_\theta} = \frac{1}{R_{AB}^2 k_{PA}} + \frac{1}{R_{CB}^2 k_{PC}} \tag{4.8}$$

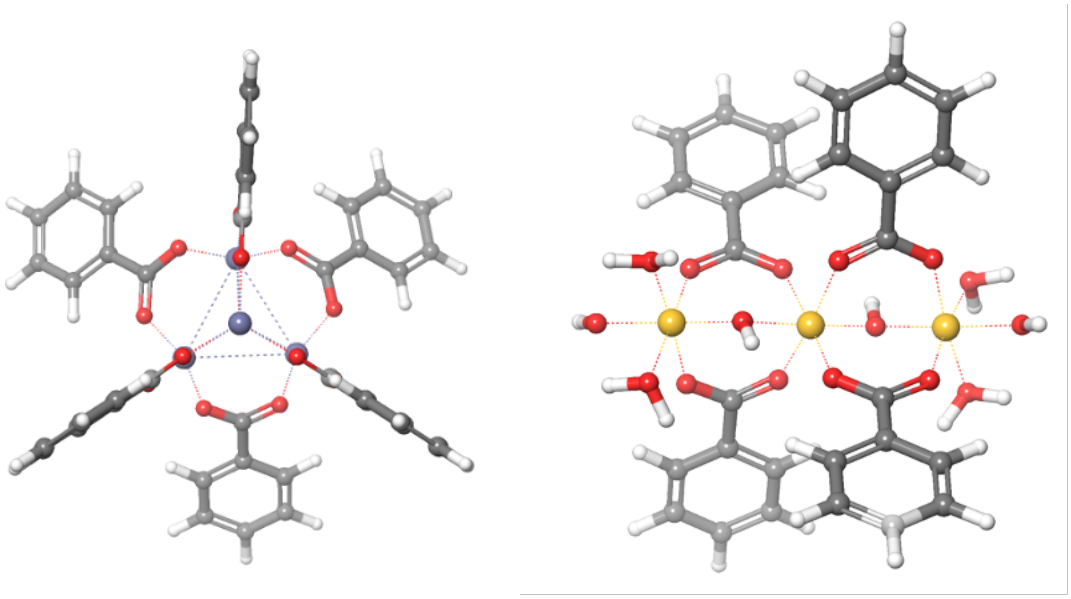
In the original Seminario method, the change of energy associated with the displacement of atom A along the direction of $\hat{u}^{PA}$ is only considered for the change in ∠ABC. However, for larger systems where atom A is also involved in other angles, the displacement along the direction of $\hat{u}^{PA}$ may also contribute to neighboring angles. Allen et al.[23] found that the mean of the additional contribution from all neighboring angles results in the most accurate force constants that give the most reasonable agreement with the QM vibration frequency spectra. In the modified Seminario method, for a central atom B, the angle force constant of ∠ABC is given by[23]:

$$\frac{1}{k_\theta} = \frac{1 + \frac{\sum_{i=1}^{N} |\hat{u}^{PA} \cdot \hat{u}^{PA_i}|}{N-1}}{R_{AB}^2 k_{PA}} + \frac{1 + \frac{\sum_{i=1}^{M} |\hat{u}^{PC} \cdot \hat{u}^{PC_i}|}{N-1}}{R_{CB}^2 k_{PC}} \tag{4.9}$$

where N and M are the number of angles that have atom B as the central atom. In this work, we apply the modified Seminario method to derive bond and angle force constants for metal-organic ligand interactions.

4.2.3 Computational Details

We considered two MOFs in the training set to study rigid MOF-5 and flexible MIL-53(Al). The crystal structures of MOF-5 (CSD identifier: SAHYIK)[25, 26] and MIL-53(Al) (CSD identifier: SABVUN02)[27] were obtained from the Cambridge Structural Database (CSD). Periodic structures were created using the Build Cell program of the Schrödinger Suite 2024-3[28] . Periodic systems were optimized with Periodic DFT using Schrödinger Quantum ESPRESSO



(a) MOL-5 metal center (b) MIL-53(Al) metal center

Figure 4.1: Metal centers of MOL-5 and MIL-53(Al).

Interface[29] . Metal cores and coordinating ligands were truncated from the optimized periodic structures to perform geometry optimization, *ab initio* Hessian, and vibrational frequencies using Jaguar DFT functions[30]. The MOF-5 metal center consists of a Zn_4O center and six repeating ligand 1, 4-benzenedicarboxylate (BDC) ligands (Figure 4.1a). The MIL-53(Al) metal center consists of three Al(OH) chains and four repeating BDC ligands (Figure 4.1b).

DFT Calculations

The conventional unit cell of 76 atoms of MIL-53(Al) was used for periodic DFT calculations. A primitive unit cell consisting of 106 atoms was used for framework MOF-5 rather than the conventional unit cell of 424 atoms due to the increased computational cost. In all calculations, the Perdew-Burke-Ernzerhof (PBE) ultrasoft pseudopotentials, generalized gradient approximation (GGA) of the PBE functional, and DFT-D3 dispersion correction were used. The self-consistent convergence tests were carried out for plane-wave cutoff energies of 40 Ry to 80 Ry with a charge density multiplier value of 5.0. Brillouin zone partition tests were performed to

properly simplify calculations. For MOF-5, a plane-wave cutoff energy of 40 Ry, and k-point sampling of 1x1x1 including Γ point were selected. For MIL-53(Al), a plane-wave cutoff energy of 40 Ry, and k-point sampling using grid plane distance of 0.05 (1/Å) including Γ point were used. In all calculations, atomic positions and cell volume were all first fully relaxed with an energy convergence threshold of 1×10^{-6} Ry and the option "optimize FFT grid". DFT property calculations for elastic constants were performed on the fully relaxed structures. The relaxed atom positions, lattice parameters, and elastic constants are used as reference for force field training.

Metal centers were relaxed with the ωB97x-D3 long-range corrected hybrid density functionals with dispersion corrections and the LACV3P+* basis set. The vibrational frequencies and Hessian matrices were computed with the same level of DFT theory. Vibrational frequencies and Hessian matrices were used to derive force constants based on the modified Seminario method[23] . We applied a scaling factor of 0.985 to vibrational frequencies according to the Database of Frequency Scale Factors for Electronic Model Chemistries Version 5[31].

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were performed with the Desmond program of the Schrödinger Suite 2024-3[28]. Periodic supercell structures were created for MOF-5 (2 x 2 x 2) and MIL-53(Al) (6 x 4 x 4) for condensed-phase MD simulations with periodic boundary conditions. The supercell systems were relaxed using the default relaxation protocol in the Schrödinger Suite. The relaxation protocol includes a series of Brownian minimizations and short, low temperatures, small time-step MD simulations in either *NVT* or *NPT* ensemble. The fully relaxed MOF-5 structure was equilibrated in *NPT* ensemble. The well-equilibrated system was run for 10 ns at 1 bar and 15 K using a time-step of 2 fs and an anisotropic MTK Barostat. The fully relaxed MIL-53(Al) structure was equilibrated at 15K, 300K, and 375K, to observe the temperature-induced open-pore and narrow-pore breathing behavior[32, 33]. The well-equilibrated snapshot was extracted for property calculations. A custom-built version of the

Desmond program was used for MD simulations with customized metal-ligand atom types.

4.2.4 Calculation of Material Properties

Mechanical Properties

The Elastic Constants Panel implemented in Schrödinger Suite 2024-3 was used to set up and perform all elastic constant calculations for MOFs. The stress-based method was chosen to set up uniaxial and biaxial tension simulations, and the stresses were derived from the strained simulation and were used in a linear fit to obtain the elastic tensors. For both types of tensions, the number of increments was set as four, the size of increments was set as 0.0025, and the number of simulation runs per tension was eight. For uniaxial tension, the independent increments and decrements in lattice parameters a, b, and c were taken to set up strain simulations. For biaxial tension, increments and decrements in each pair of a, b, and c were taken to set up strain simulations. The total number of simulations using the configuration described above results in 49 simulations to obtain the full elastic tensors. For each simulation, a 1 ns Brownian minimization using time step of and time step of 1.0 fs was performed at 10 K.

The elastic tensor C is a fourth-order tensor describing the stress-strain (σ-ε) relation:

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} \tag{4.10}$$

where i,j,k,l are Cartesian indices, taking values x, y, z. Both σ and ε are symmetric tensors, and

the stress-strain tensor can be transformed in Voigt notation:

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{12} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{13} & C_{23} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{14} & C_{24} & C_{34} & C_{44} & C_{45} & C_{46} \\ C_{15} & C_{25} & C_{35} & C_{45} & C_{55} & C_{56} \\ C_{16} & C_{26} & C_{36} & C_{46} & C_{56} & C_{66} \end{bmatrix} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ 2\varepsilon_{23} \\ 2\varepsilon_{13} \\ 2\varepsilon_{12} \end{bmatrix} \tag{4.11}$$

MOF-5 has cubic symmetry, which leads to a simplified elastic tensor of three independent constants: C11, C12, and C44. The MIL-53(Al) has an orthorhombic crystal structure, which has nine independent elastic constants: C11, C22, C33, C44, C55, C66, C12, C13, and C23.

From the elastic tensors, mechanical properties were derived following the formalism of the Materials Project[34, 35]. In this work, we focused on elastic constants including Young's modulus, bulk modulus, shear modulus, Poisson's ratio, and universal anisotropy.

Gas Adsorption Isotherm

The Penetrant Loading Calculation Panel implemented in Schrodinger Suite 2024-3[28] was used to perform simulations of the loading of the MOF-5 periodic system by carbon dioxide (CO_2) at a series of vapor pressures, and the penetrant weight percentages versus vapor pressures curve was plotted to obtain the gas adsorption isotherm. The protocol for simulating loading of small gas adsorption capacity is described in Ref [36] .

For each adsorption isotherm, the vapor pressures of the penetrant ranged from 0 bar to 45 bar, with an increment of 3 bar. The loading capacity of the MOFs is determined with cycles of 2-stage molecular simulations. The first stage is a grand canonical Monte Carlo (GCMC) simulation in the grand canonical (μVT) ensemble to insert the gas molecule to the vacancies in the system under appropriate chemical potential (μ). The chemical potential μ is the cost in energy to add a particle to a N particle system, may be divided into ideal and excess parts,

$\mu = \mu_{\text{ideal}} + \mu_{\text{excess}}$ [37]. The ideal part in grand canonical ensemble is given as[38]:

$$\mu = k_B T \ln(N\lambda_{\text{th}^3}/V) \tag{4.12}$$

where k_{B} is the Boltzmann constant, T is the temperature, and λ_{th} is the thermal wavelength. The excess chemical potential is the difference between the chemical potential of the system and that of an ideal gas at the same conditions. In this work, we used an empirical excess chemical potential of -2.0 kcal/mol for CO_2 using OPLS4. Additionally, an excess chemical potential of -2.5 kcal/mol was empirically chosen to use with the custom OPLS force field. In the future work, we will rigorously define the excess chemical potential of CO_2 Widom insertion method[39, 40].

Gas adsorption measurements are reported in terms of absolute, excess, and net adsorption [41] . The adsorption amount obtained by GCMC simulation is the absolute adsorbed amount, and the adsorption isotherm obtained with experiments is the excess adsorbed amount. We compared the simulation and experimental data following the relation[42]:

$$m_{\text{ab}} = m_{\text{excess}} + \rho_{\text{gas}} V_{\text{free}} \tag{4.13}$$

where m_{ab} is the absolute adsorbed mass, m_{excess} is the excess adsorbed mass, ρ_{gas} is the ideal gas density at each vapor pressure, and V_{free} is the volume of the free MOF materials.

4.2.5 Parameter Optimization

The structure-property relationship is crucial for materials design and prediction of material behaviors, and also provides useful information for force field development. In this section, we discuss a workflow to optimize a custom OPLS force field against elastic tensors.

The metal coordination bonds were created between organic ligands and metals within the coordination distance of the metal ions (Figure 4.2).

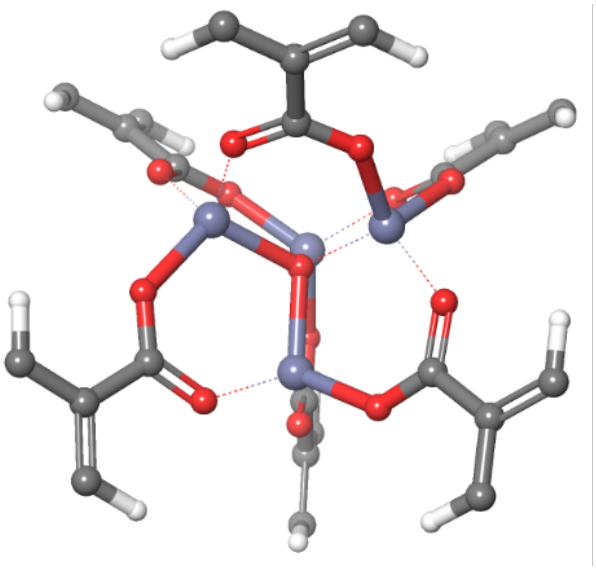


Figure 4.2: Metal coordination bonds of MOF-5.

Partial charges were assigned to the metal and ligands using the CM1A-BCC charge model[43, 44] to create a neutral system. The metal center coordination geometry was maintained according to the gas-phase DFT reference, and the bond and angle force constants were parameterized using the modified Seminario method. The average of fitted force constants over the same atom type was used as the initial force constant per type for optimization against the stress-strain relationship.

With the force constants fixed, we optimized the coordination geometry (bond length r, angle θ) against periodic DFT structures, and stress tensor σ:

$$\chi^2 = \sum_{i=1}^{N} \frac{(r_{\mathrm{FF,i}} - r_{\mathrm{DFT,i}})^2}{r_{\mathrm{DFT,i}}^2} + \sum_{i=1}^{M} \frac{(\theta_{\mathrm{FF,i}} - \theta_{\mathrm{DFT,i}})^2}{\theta_{\mathrm{DFT,i}}^2} + \sum_{i=1}^{5} \frac{(\sigma_{\mathrm{FF,i}} - \sigma_{\mathrm{DFT,i}})^2}{\sigma_{\mathrm{DFT,i}}^2} \tag{4.14}$$

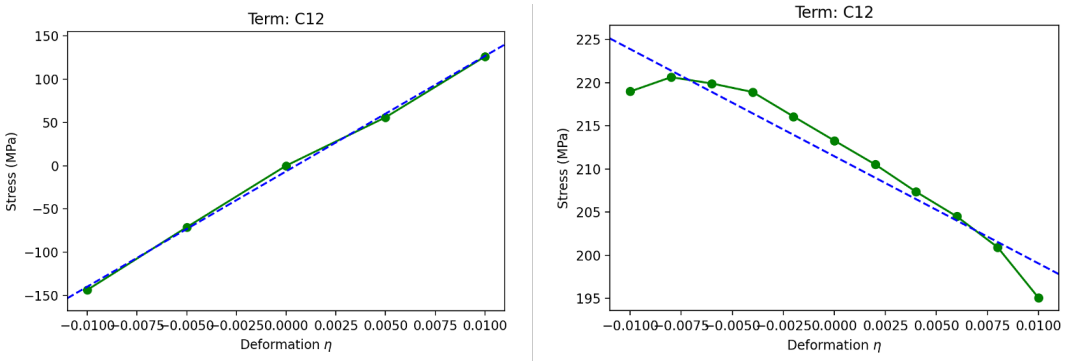
where the subscript FF and DFT denote values generated by the force field and DFT, respectively. The N and M are the number of bonds and angles in a system. The third term sums the relative squared errors over five stresses computed for deformation of ± 0.01, ± 0.005, and no deformation. The objective functions were minimized in cycles of system equilibration and elastic tensor calculations, and the final optimized custom force field was validated in calculations of

mechanical properties.

4.3 Results

4.3.1 Force Field Optimization

We compute elastic constants on MOF-5 with DFT and OPLS4 as the baseline results, shown in Table 4.1. The starting structures of OPLS4 are well-equilibrated as described in Section 4.2.3. OPLS4 cannot reproduce the symmetry of the MOF-5 elastic tensor, indicated by vastly different C11, C22, and C33 values. The OPLS4 misrepresents the stress-strain relation of C12, predicting positive stress without deformation (Figure 4.3).



(a) Periodic DFT (b) OPLS4

Figure 4.3: Stress-Strain relation (C12) computed from linear regression fit with periodic DFT and OPLS4.

The negative C12 value (-0.60) indicates that the stress-strain relationship of uniaxial deformations is not physically meaningful.

Table 4.1: Comparison of MOF-5 elastic tensors between DFT, OPLS4, final force field model (custom OPLS), starting model derived from QM Hessian. Units of elastic constants are GPa.

Elastic Constant	DFT	OPLS4	Custom OPLS	Hessian Fitted
C11	27.80	10.40	34.20	22.70
C22	27.80	115.80	34.90	22.20
C33	27.80	6.60	33.20	20.20
C12	10.60	-0.60	5.80	1.80
C44	3.60	1.20	0.60	0.80
RMSE	n/a	37.90	5.00	6.30

To address the discrepancy between Periodic DFT and OPLS4 in MOFs (Figure 4.3), we used MOF-5 as a proof-of-concept case to optimize custom OPLS against the y axis stresses generated by uniaxial deformations along lattice parameter *a*, which were used in linear regression fit to derive the C12 elastic constant term, as described in Section 4.2.5. Throughout the optimization, we evaluated the relative-root-mean-squared-error (RRMSE) to periodic DFT and selected the model with the lowest RRMSE as the final custom OPLS. The loss function (RRMSE) was minimized from 1.37% to 0.09% (Figure 4.9). The RMSE of calculated elastic tensors to periodic DFT results reduced from 37.9 GPa of OPLS4 to 5.0 GPa of custom OPLS (Table 4.1). The symmetry of the elastic constant was well reproduced in the custom OPLS model. To further validate the final custom OPLS model, we evaluated structural properties and material properties in the following sections.

4.3.2 Structural Properties

MOF-5 Metal Center Geometry

The DFT fully relaxed equilibrium geometry of the MOF-5 metal center is shown in Figure 4.2. Each Zinc (Zn) atom is covalently bonded to two oxygen (O) atoms, and all Zn-O bonds have the same equilibrium distance. The angles formed by the two carboxylate O atoms

and Zn atoms are the same, but the angles formed by one of the carboxylate O atoms, one Zn atom, and the metal center O atom are slightly different, shown in Table 4.2. While the force constants fitted from the Hessian show negligible differences, and they only represent a relatively small energy scale, it is still important to separate O-Zn-O angles into two bending types to maintain the accurate equilibrium metal-ligand coordination geometry.

Table 4.2: Metal-ligand coordination geometry and relevant force constants derived from QM Hessian. All Zn atoms are typed as Zn, O1 and O2 denote carboxylate oxygens, and Oc denotes the metal center oxygen. Bond length and angle use units of Åand degree. Bond stretch force constants are in the unit of kcal/mol/Å/Å. Angle stretch force constants are in the unit of kcal/mol/rad/rad. All force constants are converted to be consistent with the OPLS force field functional form convention.

Symbols	Hessian fitted force constants	Force Field equilibrium geometry	Periodic DFT equilibrium geometry
Zn-Oc	74.75	2.00	1.97
Zn-O1	74.13	1.99	1.97
Zn-O2	74.13	1.99	1.97
O2-Zn-Oc	40.19	110.06	111.49
O1-Zn-Oc	46.95	110.06	111.49
O2-Zn-O2	36.67	108.87	107.38
O1-Zn-O1	36.68	108.87	107.38
O1-Zn-O2	36.68	108.87	107.38
Zn-Oc-Zn	45.05	109.47	109.47
C-O1-Zn	33.88	132.81	130.61
C-O2-Zn	33.87	132.81	130.61

Starting with the initial geometry obtained from the crystal structure, the metal core atomic positions were minimized using the OPLS4 and custom OPLS force field, and compared with the geometry optimized with periodic DFT. We chose the periodic DFT optimized geometry as the reference because it provides reasonable elastic tensors. Table 4.3 shows the relevant metal-ligand coordination geometry of MOF-5, and the RRMSE of custom OPLS is 0.23%, significantly

lower than the error of 0.75% resulting from OPLS4. The bonded model utilized in custom OPLS4 preserves the metal-ligand coordination geometry more effectively than the zero-order bond model. The default zero-order bond parameters was used in the OPLS4 calculation, and an optimization without such restraints is not possible to maintain the tetrahedron geometry of the metal core.

Table 4.3: Relevant MOF-5 metal-ligand coordination geometry computed from OPLS4, custom OPLS, and periodic DFT. RRMSE of force field results relative to DFT results.

Symbols	OPLS4	Custom OPLS	Periodic DFT
Zn-Oc	2.05	1.98	1.97
Zn-O1	2.04	1.98	1.97
Zn-O2	2.04	1.98	1.97
Zn-Zn	3.35	3.26	3.21
O2-Zn-Oc	109.00	109.80	111.49
O1-Zn-Oc	109.00	109.80	111.49
O2-Zn-O2	109.90	106.80	107.38
O1-Zn-O1	109.90	109.40	107.38
O1-Zn-O2	109.90	109.50	107.38
Zn-Oc-Zn	109.50	109.30	109.47
C-O1-Zn	135.10	130.80	130.61
C-O2-Zn	135.10	130.80	130.61
RRMSE (%)	0.75	0.23	n/a

Characteristic Normal Modes of MOF-5

The strength, flexibility, and stiffness of a material are all influenced by the vibrational motions of its building blocks. For MOFs, understanding the vibrational characteristics of the secondary building blocks is fundamental to creating reliable force field models to simulate the mechanical response to external forces and predict mechanical properties. To validate our custom OPLS model, we studied several characteristic normal modes of the MOF-5 metal cores (Figure 4.4), and compared the potential energy curves of the harmonic motion along the

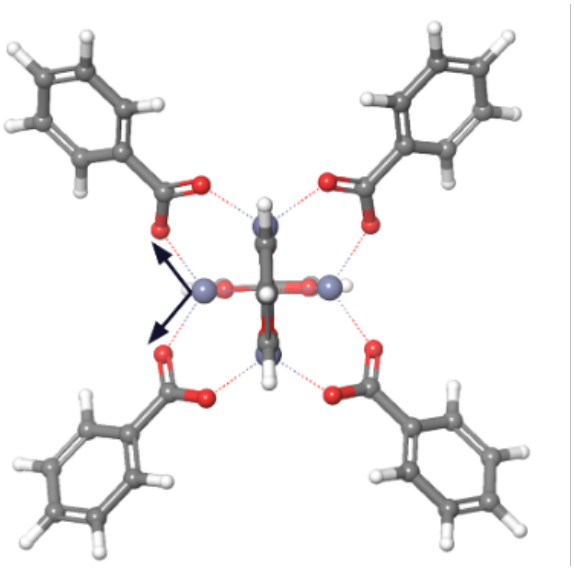


Figure 4.4: Static visualization of the harmonic motion studied in this work. The angle O1-Zn-O2 indicated by the black arrow in the figure exhibits a bending motion.

normal mode displacements. The angle studied is associated with a QM normal mode with a vibrational frequency of 52.09 cm^{-1}, which is the 5th from the lowest frequencies (Table 4.6). Accurately modeling this low-frequency motion is crucial to successfully modeling the flexibility and elasticity of MOF-5. Similar analysis on the metal center also applies to other MOF structures. By solving the eigenvalue problem for the QM Hessian matrix, we can reproduce the QM potential energy surface as a simple harmonic oscillator. The eigenvalues (λ) of the Hessian are the force constants acting on all atoms, and the eigenvectors (v) are the corresponding displacements. The DFT potential energies are given by Hooke's Law:

$$E_{\mathrm{DFT}} = \frac{1}{2}\lambda v^2 \tag{4.15}$$

Figure 4.5 shows that our custom OPLS can recapitulate the QM harmonic motions more closely than OPLS4, which indicates a more accurate description of the restoring forces when atoms are displaced from equilibrium.

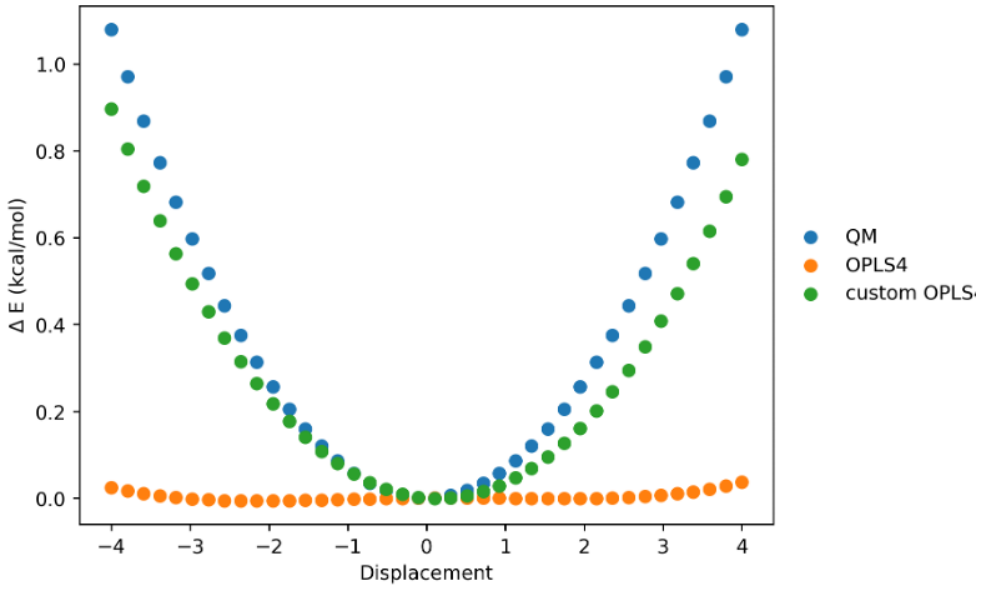


Figure 4.5: Potential energy curves along the normal mode displacement.

4.3.3 Mechanical Properties

The full elastic tensors and elastic constants were computed using well-equilibrated supercell structures and methods described in Section 4.2.4. In this section, we compare the calculation results of periodic DFT, OPLS4, and custom OPLS. For cubic crystal material MOF-5, the 6 x 6 elastic tensor has three independent components due to symmetry: C11, C12, and C44. It is worth noting that the Elastic Constant Panel sets up simulations to derive the full 6 x 6 elastic tensor, and a reliable force field should be able to predict the symmetries of stress and strain tensors. Table 4.1 presents the partial elastic tensor components of MOF-5. Compared to DFT reference, the custom OPLS results have significantly lower RMSE values than baseline OPLS4 results. For the custom OPLS force field, the RMSE for the elastic tensor C11, C22, C33, C44, and C12 is 6.3 GPa, whereas the RMSE for the OPLS4 force field is 37.9 GPa. The high RMSE value of OPLS4 is largely due to the missing symmetry of the elastic tensor. This result is not surprising because the OPLS4 metal-ligand coordination geometry is not well preserved, and the characteristic angular normal mode is not well captured. While the custom OPLS was trained on

metal-ligand coordination geometry and elastic tensor C12, the resulting force field is transferable to elastic tensors C11 and C44.

Next we tested the accuracy of Custom OPLS in calculation for mechanical properties. We derived five mechanical properties from the computed elastic tensors, shown in Table 4.4. Overall, the Custom OPLS results are more accurate than OPLS4, with the RMSE values of 3.8 and 12.0 GPa, respectively. While the custom OPLS results mostly agree with DFT reference, the bulk modulus has a notable difference. This result may indicate that the bulk modulus could be related not only to the metal-core force fields, but also to other aspects of the force fields. While the RMSE of mechanical properties computed using Custom OPLS is significantly lower than OPLS4, it is higher than the Hessian fitted model, with values of 8.0, 12.0, and 3.8 GPa, respectively. This may indicate overfitting to the C12 term, which leads to the worsening universal elastic anisotropy. This result implies that a weak regularization term to restrain the adjustable parameters to their respective Hessian fitted values may improve the optimization.

Table 4.4: Mechanical properties computed for MOF-5. Units of mechanical properties are GPa.

Mechanical Properties	DFT	OPLS4	Custom OPLS	Hessian Fitted
Young's modulus	7.00	13.70	9.80	7.80
Bulk modulus	16.20	8.40	15.10	8.30
Shear modulus	2.40	5.60	3.50	2.90
Universal elastic anisotropy	7.50	31.90	25.00	10.00
Isotropic Poisson ratio	0.40	0.20	0.40	0.30
RMSE	n/a	12.00	8.00	3.80

The MIL-53(Al) mechanical properties were performed for the large pore equilibrated conformation obtained at 15K, and compared with DFT calculations performed with the fully relaxed cell using the large pore configuration obtained from the crystal structure, shown in Table 4.5. Unlike the rigid frameworks MOF-5, MIL-53(Al) is characterized by a "wine rack"

structure, which exhibits a "breathing effect" where the pore size and shape can change reversibly in response to external stimuli, such as temperatures and pressures[18] . The structural transition is associated with a unit cell volume change up to 40% [45, 46]. Predicting the equilibrium between the large-pore and narrow-pore of MIL-53(Al) is not a trivial task for molecular dynamics[46, 47].

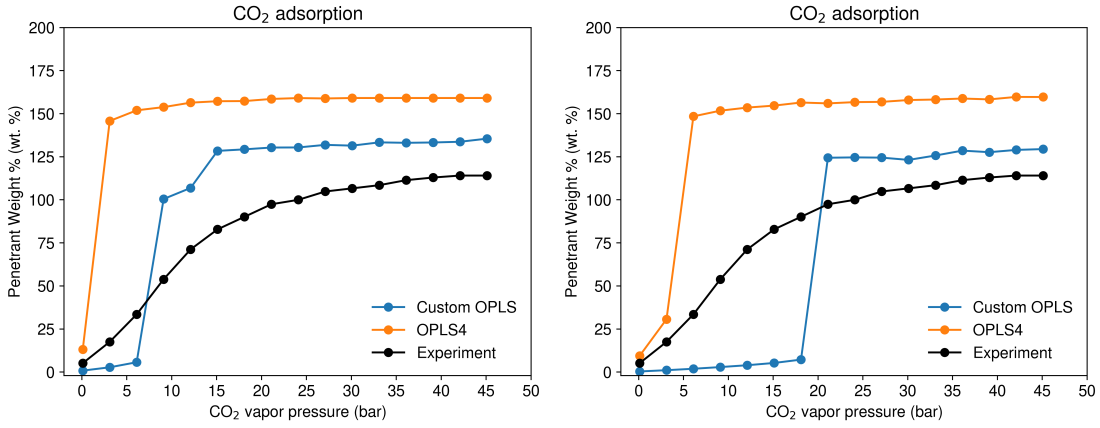
Table 4.5: Mechanical properties computed for MIL-53(Al).

(GPa)	DFT	OPLS4	Custom OPLS4
Young's modulus	22.90	33.64	25.57
Bulk modulus	20.60	24.86	17.86
Shear modulus	8.71	13.20	10.14
Universal elastic anisotropy	33.26	66.67	70.24
Isotropic Poisson ratio	0.31	0.27	0.26
RMSE	n/a	15.94	16.64

Neither OPLS4 nor Custom OPLS can reproduce a reasonable universal elastic anisotropy for MIL-53(Al). The universal elastic anisotropy measures how much a material's elastic properties differ in different directions. The high anisotropy of MIL-53(Al) is directly linked with its "wine rack" structure [48, 49] . For breathing MOFs, the stress-strain linearity becomes nonlinear when experiencing structural transitions[50]. We suspect the mechanical stress exerted on to MIL-53(Al) during the elastic constant calculation reached the threshold that the given "large-pore" starring structure cannot stand, and led to the unexpected high anisotropy of elastic properties. Nevertheless, using OPLS4 and Custom OPLS, we were able to recapitulate the narrow pore to large pore transition from 300K to 375K at 1 bar, which shows good agreement with the experimental reference[32] . The current results provide useful data for future development in force field for flexible MOFs from mechanical properties.

4.3.4 Gas Adsorption Isotherm

The porous nature of MOF-5 provides a massive surface area for gas molecules to attach, and this process is reversible by changing pressure and temperature. This feature makes MOF-5 a promising candidate for CO_2 separation and storage. We tested the applicability of using OPLS4 and custom OPLS to simulate CO_2 uptake of MOF-5, and compared the results with experimental reference[51] . The purpose of this study is to identify issues of gas adsorption with OPLS4 and evaluate the new custom OPLS for MOF-5. The ambient condition CO_2 uptakes at a range of vapor pressures are plotted as penetrant weight percent versus vapor pressures as shown in Figure 4.6.



(a) $\mu_{\text{excess}} = -2$ kcal/mol. (b) $\mu_{\text{excess}} = -2.5$ kcal/mol.

Figure 4.6: CO_2 adsorption isotherm of MOF-5. Excess chemical potentials used for CO_2 at all vapor pressures are: (a) -2 kcal/mol and (b) -2.5 kcal/mol.

With the recommended excess chemical potential (μ_{excess}) of -2 kcal/mol, both OPLS4 and custom OPLS overestimate the CO_2 uptake, while custom OPLS shows slightly improved performance than OPLS4. To test the sensitivity of excess chemical potential in gas adsorption calculations, $\mu_{\text{excess}} = -2.5$ kcal/mol was empirically chosen for custom OPLS4 (Figure 4.6b). While CO_2 loading results were still poor, it is evident that Custom OPLS4 can predict more physically meaningful CO_2 uptake than OPLS4. Future work is designed to rigorously define

the excess chemical potentials as a function of vapor pressures for CO_2 using OPLS force fields, and the effects of different vapor pressures on excess chemical potential need to be carefully accounted for in GCMC simulations.

4.4 Discussion

We have developed a workflow to derive force fields for metal-organic frameworks from periodic DFT calculations and mechanical property calculations. We present a covalently bonded representation for molecular dynamics simulations in the context of OPLS force fields. The current proof-of-concept model is developed and validated on the rigid MOF-5 and the flexible MIL-53(Al), but our workflows can be easily applied to other MOF structures. We incorporate elastic constants calculations into the force field optimization, which leads to more accurate mechanical properties compared to OPLS4 and force fields derived from QM Hessian. We have also tested the validity of the bonded model by showing that the force field predicted structural properties and mechanical properties are in good agreement with the DFT reference. The predicted gas adsorption properties of the custom OPLS are more accurate than OPLS4 within current calculation protocols. We anticipate that this workflow can be automated to systematically derive force field parameters for metal-organic frameworks.

4.5 Acknowledgements

I thank Schrödinger, Inc., for financial support and resources. I thank Steven Dajonawicz for helpful comments.

This chapter contains work that I completed during a summer internship at Schrödinger, Inc. under the direction of Steven Dajnowicz. Liangyue Drew is the sole author of this chapter.

4.6 Supplementary Information

Table 4.6: DFT vibrational frequencies computed for MOF-5 metal center. Frequency has in the unit of cm^{-1}

Index	Frequency	Index	Frequency	Index	Frequency
0	48.11	87	639.46	174	1,194.24
1	48.15	88	639.47	175	1,194.27
2	48.29	89	639.51	176	1,194.28
3	51.97	90	639.6	177	1,215.29
4	52.06	91	639.63	178	1,215.38
5	52.09	92	639.64	179	1,215.77
6	54.04	93	699.44	180	1,215.87
7	61.06	94	699.46	181	1,215.92
8	61.12	95	700.71	182	1,217.63
9	65.39	96	700.73	183	1,345.68
10	65.43	97	700.75	184	1,345.70
11	65.46	98	705.52	185	1,345.73
12	66.32	99	714.66	186	1,346.03
13	66.35	100	714.66	187	1,346.06
14	66.42	101	714.67	188	1,346.07
15	78.19	102	715.02	189	1,357.65
16	78.28	103	715.02	190	1,357.66
17	78.32	104	715.04	191	1,357.68
18	78.41	105	744.3	192	1,357.80
19	78.41	106	744.31	193	1,357.81
20	81.09	107	744.32	194	1,357.82
21	81.11	108	747.13	195	1,482.59
22	81.27	109	747.14	196	1,482.65
23	84.59	110	747.14	197	1,485.71
24	107.64	111	851.54	198	1,485.76
25	107.65	112	851.56	199	1,485.80
26	107.66	113	851.56	200	1,502.23
27	116.07	114	852.41	201	1,502.26
28	116.09	115	852.44	202	1,502.31

Continued on next page

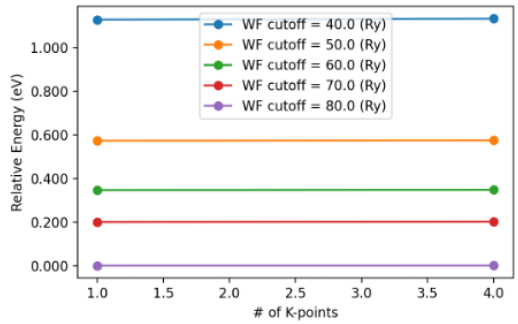
Table 4.6 – continued from previous page

Index	Frequency	Index	Frequency	Index	Frequency
29	116.13	116	852.46	203	1,503.27
30	120.71	117	864.99	204	1,503.32
31	120.87	118	865.01	205	1,503.36
32	120.88	119	866.48	206	1,506.12
33	139.87	120	866.49	207	1,554.22
34	139.97	121	866.51	208	1,554.26
35	150.55	122	870.33	209	1,554.49
36	150.58	123	886.91	210	1,554.53
37	150.64	124	887.54	211	1,554.55
38	156.58	125	887.54	212	1,557.32
39	156.71	126	887.55	213	1,622.31
40	156.73	127	887.82	214	1,622.31
41	162.52	128	887.82	215	1,622.34
42	194.6	129	986.43	216	1,662.89
43	194.64	130	986.43	217	1,662.90
44	194.65	131	986.45	218	1,662.90
45	203.02	132	986.86	219	1,684.28
46	203.03	133	986.86	220	1,684.29
47	203.04	134	986.86	221	1,684.31
48	237.11	135	1,025.37	222	1,694.95
49	237.16	136	1,025.72	223	1,694.95
50	237.17	137	1,025.72	224	1,694.96
51	247.76	138	1,025.77	225	1,694.96
52	247.83	139	1,025.87	226	1,694.98
53	261.72	140	1,025.87	227	1,695.19
54	317.44	141	1,031.97	228	1,708.41
55	317.44	142	1,032.05	229	1,708.43
56	317.46	143	1,032.06	230	1,708.44
57	320.51	144	1,032.14	231	3,190.43
58	320.52	145	1,032.20	232	3,190.43
59	320.55	146	1,032.36	233	3,190.43
60	426.18	147	1,037.02	234	3,190.43
61	426.26	148	1,037.03	235	3,190.44

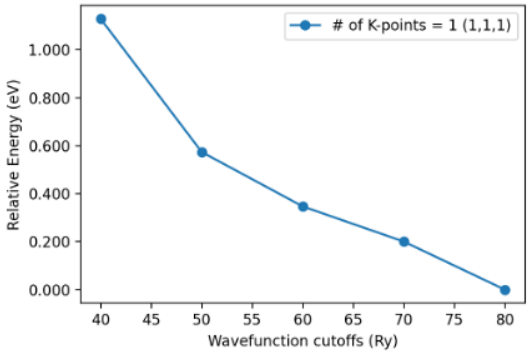
Continued on next page

Table 4.6 – continued from previous page

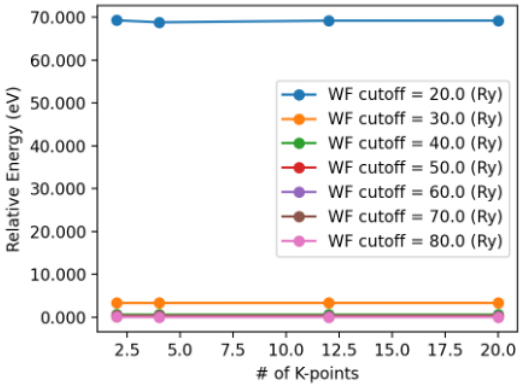
Index	Frequency	Index	Frequency	Index	Frequency
62	426.26	149	1,037.03	236	3,190.44
63	426.29	150	1,037.13	237	3,203.92
64	426.32	151	1,037.13	238	3,203.92
65	426.34	152	1,037.14	239	3,203.93
66	457.49	153	1,065.89	240	3,203.95
67	457.5	154	1,065.91	241	3,203.96
68	457.51	155	1,065.98	242	3,203.96
69	458.29	156	1,066.00	243	3,214.88
70	458.3	157	1,066.01	244	3,214.89
71	458.31	158	1,066.37	245	3,214.89
72	494.32	159	1,117.55	246	3,214.90
73	494.44	160	1,117.56	247	3,214.92
74	494.44	161	1,117.67	248	3,215.00
75	499.48	162	1,117.73	249	3,232.37
76	500.53	163	1,117.83	250	3,232.38
77	500.53	164	1,117.84	251	3,232.41
78	515.76	165	1,191.28	252	3,232.45
79	515.98	166	1,191.31	253	3,232.47
80	515.98	167	1,191.57	254	3,232.48
81	608.95	168	1,191.61	255	3,233.99
82	608.97	169	1,191.61	256	3,234.01
83	608.98	170	1,192.29	257	3,234.02
84	614.31	171	1,194.17	258	3,234.03
85	614.33	172	1,194.17	259	3,234.05
86	614.34	173	1,194.23	260	3,234.12



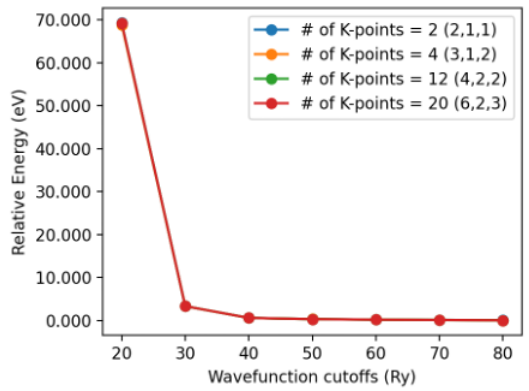
(a) MOF5: Convergence with the wavefunction cutoffs.



(b) MOF5: Convergence with the wavefunction cutoff for a given k-point grid.



(c) MIL-53(Al): Convergence with the wavefunction cutoffs.



(d) MIL-53(Al): Convergence with the wavefunction cutoff for a given k-point grid.

Figure 4.7: Convergence tests for the energy cutoffs and K-point grids for periodic DFT calculations in Quantum ESPRESSO.

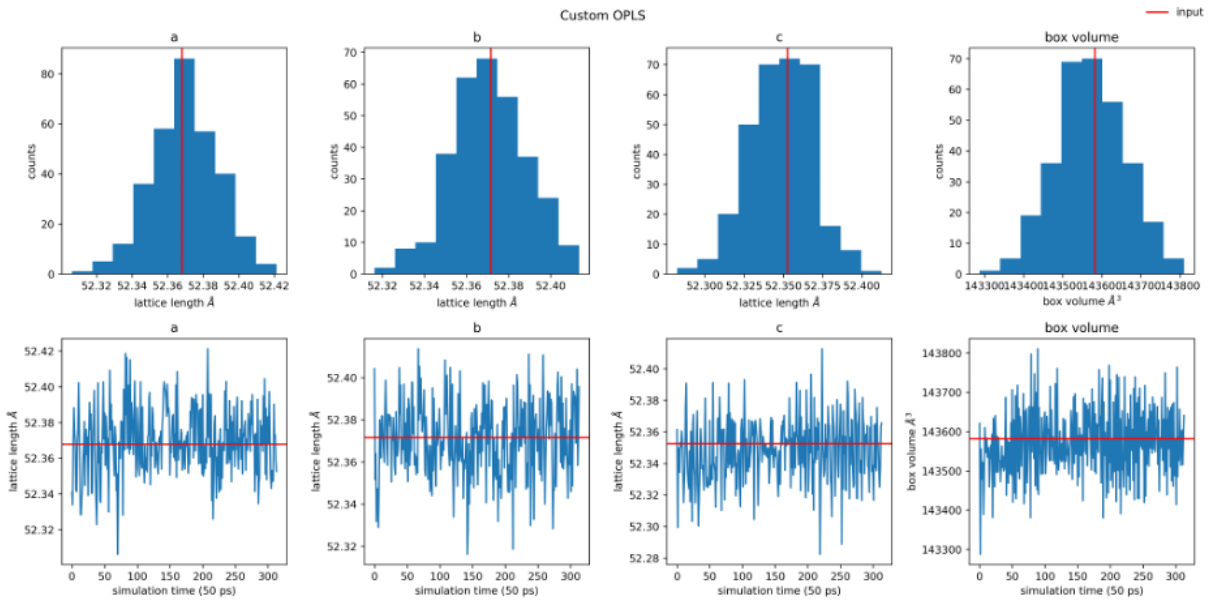


Figure 4.8: MOF-5 input structure is well-equilibrated in a 15 ns MD simulation. The subplots are for lattice parameters (a, b, c) and simulation box volume.

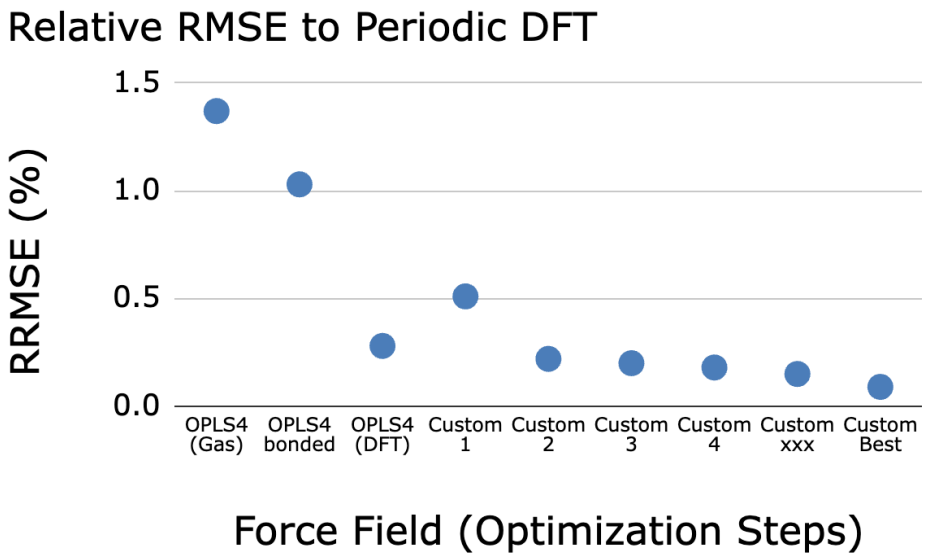


Figure 4.9: Loss function as a function of iteration steps for force field optimization. Some intermediate iterations were not plotted.

Bibliography

(1) 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*, Publisher: American Chemical Society (ACS), 1657–1666.

(2) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation* **2016**, *12*, 281–296.

(3) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *Journal of Chemical Theory and Computation* **2019**, *15*, 1863–1874.

(4) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *Journal of Chemical Theory and Computation* **2021**, *17*, 4291–4300.

(5) Damm, W.; Dajnowicz, S.; Ghoreishi, D.; Yu, Y.; Ganeshan, K.; Madin, O.; Rudshteyn, B.; Hu, R.; Wu, M.; Shang, Y., et al. OPLS5: Addition of Polarizability and Improved Treatment of Metals. **2024**, DOI: `https://doi.org/10.26434/chemrxiv-2024-2svr7`.

(6) 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*, Publisher: Wiley, 1668–1688.

(7) 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**, ISBN: 9780470845011 Publisher: Wiley, DOI: `10.1002/0470845015.cfa007`.

(8) 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*, Publisher: Wiley, 187–217.

(9) 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. *Journal of Computational Chemistry* **2009**, *31*, Publisher: Wiley, 671–690.

(10) 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*, Publisher: Wiley, 1545–1614.

(11) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society* **1992**, *114*, Number: 25, 10024–10035.

(12) Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T.; Schmid, R.; Waroquier, M.; Van Speybroeck, V. QuickFF: A program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input. *Journal of Computational Chemistry* **2015**, *36*, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.23877, 1015–1027.

(13) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. MOF-FF – A flexible first-principles derived force field for metal-organic frameworks. *physica status solidi (b)* **2013**, *250*, 1128–1141.

(14) Dokur, D.; Keskin, S. Effects of Force Field Selection on the Computational Ranking of MOFs for CO_2 Separations. *Industrial & Engineering Chemistry Research* **2018**, *57*, Number: 6, 2298–2309.

(15) Altintas, C.; Erucar, I.; Keskin, S. High-Throughput Computational Screening of the Metal Organic Framework Database for CH_4 /H_2 Separations. *ACS Applied Materials & Interfaces* **2018**, *10*, Number: 4, 3668–3679.

(16) Dajnowicz, S.; Ghoreishi, D.; Modugula, K.; Damm, W.; Harder, E. D.; Abel, R.; Wang, L.; Yu, H. S. Advancing Free-Energy Calculations of Metalloenzymes in Drug Discovery via Implementation of LFMM Potentials. *Journal of Chemical Theory and Computation* **2020**, *16*, Number: 11, 6926–6937.

(17) Hu, L.; Ryde, U. Comparison of Methods to Obtain Force-Field Parameters for Metal Sites. *Journal of Chemical Theory and Computation* **2011**, *7*, Publisher: American Chemical Society, 2452–2463.

(18) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. J. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *Journal of Chemical Theory and Computation* **2010**, *6*, 2935–2947.

(19) Seminario, J. M. Calculation of intramolecular force fields from second-derivative tensors. *International Journal of Quantum Chemistry* **1996**, *60*, 1271–1277.

(20) Krokidas, P.; Castier, M.; Moncho, S.; Brothers, E.; Economou, I. G. Molecular Simulation Studies of the Diffusion of Methane, Ethane, Propane, and Propylene in ZIF-8. *The Journal of Physical Chemistry C* **2015**, *119*, 27028–27037.

(21) Krokidas, P.; Castier, M.; Moncho, S.; Sredojevic, D. N.; Brothers, E. N.; Kwon, H. T.; Jeong, H.-K.; Lee, J. S.; Economou, I. G. ZIF-67 Framework: A Promising New Candidate for Propylene/Propane Separation. Experimental Data and Molecular Simulations. *The Journal of Physical Chemistry C* **2016**, *120*, 8116–8124.

(22) Verploegh, R. J.; Kulkarni, A.; Boulfelfel, S. E.; Haydak, J. C.; Tang, D.; Sholl, D. S. Screening Diffusion of Small Molecules in Flexible Zeolitic Imidazolate Frameworks Using a DFT-Parameterized Force Field. *The Journal of Physical Chemistry C* **2019**, *123*, 9153–9167.

(23) Allen, A. E. A.; Payne, M. C.; Cole, D. J. Harmonic Force Constants for Molecular Mechanics Force Fields via Hessian Matrix Projection. *Journal of Chemical Theory and Computation* **2018**, *14*, 274–281.

(24) Piskorz, T. K.; Lee, B.; Zhan, S.; Duarte, F. Metallicious: Automated Force-Field Parameterization of Covalently Bound Metals for Supramolecular Structures. *Journal of Chemical Theory and Computation* **2024**, *20*, Publisher: American Chemical Society, 9060–9071.

(25) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **1999**, *402*, Number: 6759, 276–279.

(26) Eddaoudi, M.; Li, H.; Reineke, T.; Fehr, M.; Kelley, D.; Groy, T. L.; Yaghi, O. Design and synthesis of metal-carboxylate frameworks with permanent microporosity. *Topics in Catalysis* **1999**, *9*, Number: 1, 105–111.

(27) Ling, Y.; Sun, T.; Guo, L.; Si, X.; Jiang, Y.; Zhang, Q.; Chen, Z.; Terasaki, O.; Ma, Y. Atomic-level structural responsiveness to environmental conditions from 3D electron diffraction. *Nature Communications* **2022**, *13*, Publisher: Nature Publishing Group, 6625.

(28) Schrödinger Release Notes - Release 2024-3, Schrödinger `https://www.schrodinger.com/life-science/download/release-notes/release-2024-3/` (accessed 04/19/2025).

(29) Quantum ESPRESSO Interface | Schrödinger Materials Science, Schrödinger `https://www.schrodinger.com/platform/products/ms-quantum-espresso-gui/` (accessed 04/19/2025).

(30) Jaguar | Schrödinger Materials Science, Schrödinger https://www.schrodinger.com/platform/products/ms-jaguar/ (accessed 04/19/2025).

(31) S. Kanchanakungwankul; J. L. Bao; J. Zheng; I. M. Alecu; B. J. Lynch; Y. Zhao; D. G. Truhlar Database of Frequency Scale Factors for Electronic Model Chemistries, 2021.

(32) Loiseau, T.; Serre, C.; Huguenard, C.; Fink, G.; Taulelle, F.; Henry, M.; Bataille, T.; Férey, G. A rationale for the large breathing of the porous aluminum terephthalate (MIL-53) upon hydration. *Chemistry (Weinheim an Der Bergstrasse, Germany)* **2004**, *10*, 1373–1382.

(33) Isaeva, V. I.; Kustov, L. M. The application of metal-organic frameworks in catalysis (Review). *Petroleum Chemistry* **2010**, *50*, 167–180.

(34) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*, 011002.

(35) Elastic Constants | Materials Project Documentation https://docs.materialsproject.org/methodology/materials-methodology/elasticity (accessed 04/22/2025).

(36) Sanders, J. M.; Misra, M.; Mustard, T. J.; Giesen, D. J.; Zhang, T.; Shelley, J.; Halls, M. D. Characterizing moisture uptake and plasticization effects of water on amorphous amylose starch models using molecular dynamics methods. *Carbohydrate Polymers* **2021**, *252*, 117161.

(37) Attard, P. A grand canonical simulation technique for dense and confined fluids with application to a Lennard-Jones fluid. *The Journal of Chemical Physics* **1997**, *107*, 3230–3238.

(38) Blundell, S. J.; Blundell, K. M. In *Concepts in Thermal Physics*; Oxford University Press: 2009.

(39) Widom, B. Some Topics in the Theory of Fluids. *The Journal of Chemical Physics* **1963**, *39*, 2808–2812.

(40) Widom, B. Potential-distribution theory and the statistical mechanics of fluids. *The Journal of Physical Chemistry* **1982**, *86*, 869–872.

(41) Brandani, S.; Mangano, E.; Sarkisov, L. Net, excess and absolute adsorption and adsorption of helium. *Adsorption* **2016**, *22*, 261–276.

(42) Yang, X.; Arami-Niya, A.; Lyu, J.; Guo, X. Net, Excess, and Absolute Adsorption of N2, CH4, and CO2 on Metal–Organic Frameworks of ZIF-8, MIL-101(Cr), and UiO-66 at 282–361 K and up to 12 MPa. *Journal of Chemical & Engineering Data* **2021**, *66*, Publisher: American Chemical Society, 404–414.

(43) 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.

(44) 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.

(45) Férey, G.; Serre, C.; Devic, T.; Maurin, G.; Jobic, H.; Llewellyn, P. L.; De Weireld, G.; Vimont, A.; Daturi, M.; Chang, J.-S. Why hybrid porous solids capture greenhouse gases? *Chem. Soc. Rev.* **2011**, *40*, 550–562.

(46) Ma, Q.; Yang, Q.; Ghoufi, A.; Férey, G.; Zhong, C.; Maurin, G. Guest dependent pressure behavior of the flexible MIL-53(Cr): A computational exploration. *Dalton Trans.* **2012**, *41*, 3915–3919.

(47) Rogge, S.; Vanduyfhuys, L.; Ghysels, A.; Waroquier, M.; Verstraelen, T.; Maurin, G.; Van Speybroeck, V. A Comparison of Barostats for the Mechanical Characterization of Metal–Organic Frameworks. *Journal of Chemical Theory and Computation* **2015**, *11*, Publisher: American Chemical Society, 5583–5597.

(48) Ortiz, A. U.; Boutin, A.; Fuchs, A. H.; Coudert, F.-X. Metal–organic frameworks with wine-rack motif: What determines their flexibility and elastic properties? *The Journal of Chemical Physics* **2013**, *138*, Publisher: American Institute of Physics, 174703.

(49) Ortiz, A. U.; Boutin, A.; Fuchs, A. H.; Coudert, F.-X. Anisotropic Elastic Properties of Flexible Metal-Organic Frameworks: How Soft are Soft Porous Crystals? *Physical Review Letters* **2012**, *109*, Publisher: American Physical Society, 195502.

(50) Neimark, A. V.; Coudert, F.-X.; Boutin, A.; Fuchs, A. H. Stress-Based Model for the Breathing of Metal-Organic Frameworks. *The Journal of Physical Chemistry Letters* **2010**, *1*, Publisher: American Chemical Society, 445–449.

(51) Millward, A. R.; Yaghi, O. M. Metal-Organic Frameworks with Exceptionally High Capacity for Storage of Carbon Dioxide at Room Temperature. *Journal of the American Chemical Society* **2005**, *127*, Publisher: American Chemical Society, 17998–17999.