Systematic parameterization of polarizable force fields from quantum chemistry data

Lee-Ping Wang ,[†] Jiahao Chen ,[‡] and Troy Van Voorhis [*,‡]

Department of Chemistry, Stanford University, 318 Campus Drive, Stanford, CA 94350, and Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

E-mail: tvan@mit.edu

Abstract

We introduce ForceBalance, a method and free software package for systematic force field optimization with the ability to parameterize a wide variety of functional forms using flexible combinations of reference data. We outline several important challenges in force field development and how they are addressed in ForceBalance, and present an example calculation where these methods are applied to develop a highly accurate polarizable water model.

1 Introduction

Molecular mechanics (MM) using empirical potentials (force fields) is the simulation method of choice for large-scale atomistic systems. Compared to quantum mechanical (QM) calculations, they are computationally far more efficient. Nevertheless, the reliability of MM simulations depends crucially upon the accurate modeling of the essential physical interactions, which in turn is predicated on having accurate parameters.

[*]To whom correspondence should be addressed
[†]Stanford University
[‡]Massachusetts Institute of Technology

Historically, force fields are parameterized by fitting to experimental data. Early examples include interatomic repulsion potentials determined from experimental second virial coefficients,[1] the CFF potential for organic molecules based on experimental geometries and vibrational spectra,[2] and the TIP3P and TIP4P water models[3] which are fitted to the room-temperature density and enthalpy of vaporization of liquid water. More recent examples include the parameterization of protein dihedral potentials to reproduce experimentally observed conformations from NMR measurements.[4–7] On the other hand, the requirement of large amounts of data for a complete parameterization effort has led to the widespread use of QM data in force field development, either directly from potential energies[8,9] and forces,[10–15] or from calculated observables such as vibrational spectra or electrostatic potentials.[16–18] Force fields developed from fitting QM energies and forces have found applications in modeling biomolecular conformation energies,[19] liquid water,[11,20] and materials such as aluminium,[10] iron,[21] and silica;[22] electrostatic potential fitting is commonly applied in general force fields such as GAFF.[23]

The main challenge in force field development is to choose functional forms that are computationally efficient, yet flexible enough to capture the relevant physical interactions in the thermodynamically accessible regions of phase space. An accurate fit of the parameters is also crucial, which necessitates the use of accurate and abundant fitting data from experimental measurements or from QM calculations. Efficient and strictly regularized optimization methods are needed to search the high-dimensional parameter space without overfitting. We have developed an open source software package called ForceBalance[24] that aids in these development and parametrization efforts, which we expect to accelerate the exploration of new force fields and design protocols.

In this article, we briefly review the challenges associated with force field development and parameterization, and how some of these issues are alleviated by using ForceBalance. We then illustrate its use by developing a new polarizable water model with 27 adjustable parameters, including five fluctuating charge sites and an improved functional form for the van der Waals interactions. We validate this new model by comparison with experimental measurements of various properties.

2 Theory

2.1 Objective function

Force field parameterization is essentially an optimization problem in the space of parameters (denoted using k). As described above, the reference data may come from experimental measurements or from QM calculations, and just about any physical quantity can be used in the fitting procedure. To accommodate the diverse choices possible in the fitting procedure, we allow for multiple types of residuals $\mathbf{X}$ to be included in a single objective function χ^2, which is then integrated over the entire configuration space $\mathscr{R}$ of N atoms with some suitable measure $P(\mathbf{r};k)\mathrm{d}\mathbf{r}$ reflecting the thermodynamic ensemble of interest. The integral may be evaluated using any sampling technique, such as molecular dynamics or Metropolis Monte Carlo.

Specializing for now to QM reference calculations of energies and forces, our objective function is defined in Eq. (1a):

$$\chi^2(k;w) = \int_{\mathscr{R}} P(\mathbf{r};k)|\mathbf{X}(\mathbf{r},k;w)|^2 d\mathbf{r}, \tag{1a}$$

$$|\mathbf{X}(\mathbf{r},k;w)|^2 = w\left[\frac{(\Delta E(\mathbf{r},k)-\langle\Delta E\rangle)^2}{\langle E_{\mathrm{QM}}^2\rangle - \langle E_{\mathrm{QM}}\rangle^2}\right] + \frac{1-w}{3N_{\mathrm{atoms}}}\left[\Delta\mathbf{F}(\mathbf{r},k)^{\mathrm{T}}\mathrm{Cov}(\mathbf{F}_{\mathrm{QM}})^{-1}\Delta\mathbf{F}(\mathbf{r},k)\right], \tag{1b}$$

$$\Delta E(\mathbf{r},k) = E_{\mathrm{MM}}(\mathbf{r},k) - E_{\mathrm{QM}}(\mathbf{r}), \tag{1c}$$

$$\Delta\mathbf{F}(\mathbf{r},k) = \mathbf{F}_{\mathrm{MM}}(\mathbf{r},k) - \mathbf{F}_{\mathrm{QM}}(\mathbf{r}), \tag{1d}$$

$$\mathrm{Cov}(\mathbf{F}_{\mathrm{QM}}) = \langle\mathbf{F}_{\mathrm{QM}}\otimes\mathbf{F}_{\mathrm{QM}}\rangle. \tag{1e}$$

where ΔE and $\Delta\mathbf{F}$ represent the energy and force residuals, k is the set of force field parameters, and $\mathrm{Cov}(\mathbf{F}_{\mathrm{QM}})$ is the covariance of the reference QM forces. The energy variance and force covariance introduce appropriate rescalings so that the residuals in the objective function are dimensionless and of unit magnitude. The average energy difference is subtracted out of the energy term, so only relative energies are fitted. The force term is further divided by the number of components so that it has the same scale as the energy term. We introduce an adjustable parameter w which weights

the relative importance of energy residuals ($w = 1$) and force residuals ($w = 0$). In principle w ought to be chosen carefully for each use; for simplicity, we choose $w = \frac{1}{2}$ here, so that both are equally weighted. Information beyond energies and forces are similarly easy to include as squared residuals with appropriate dimensional scaling.

2.2 Nonlinear optimization

We use primarily the L-BFGS algorithm to optimize the objective function, with analytic first derivatives (with respect to parameters) implemented in a modified GROMACS simulation code.[24,25]

While this procedure cannot guarantee a globally optimal fit, we find in practice that the optimization is well-behaved, with the best final parameter values tending to result from a physically motivated initial guess. This is evidenced by numerical studies where we used a force field to produce the reference data and performed an optimization using randomly perturbed parameter values; the optimization reliably converged to the same parameters used to generate the reference data. The stability of our optimized parameters reflect previously reported numerical studies in the literature.[12] In addition, ForceBalance provides some global optimization algorithms such as simulated annealing to handle more problematic situations.

2.3 Self-consistent configurational sampling

The probability distribution $P(r,k)$ in configurational space reflects the thermodynamic ensemble that we work in by providing the appropriate Boltzmann weights to each sampled point r. However, these weights will in general differ between the reference data and that predicted from the force field. In previous work, we showed that a linear combination of the Boltzmann distributions for the force field and the QM reference provides a more consistent result than using either distribution alone.[14] We continue to adopt this approach here. Furthermore, we enforce self-consistency between the sampling simulations, reference QM calculations, and parameter optimizations until self-consistency is reached after a number of generations.[13,26–28] The procedure for doing so is outlined in Figure 1.

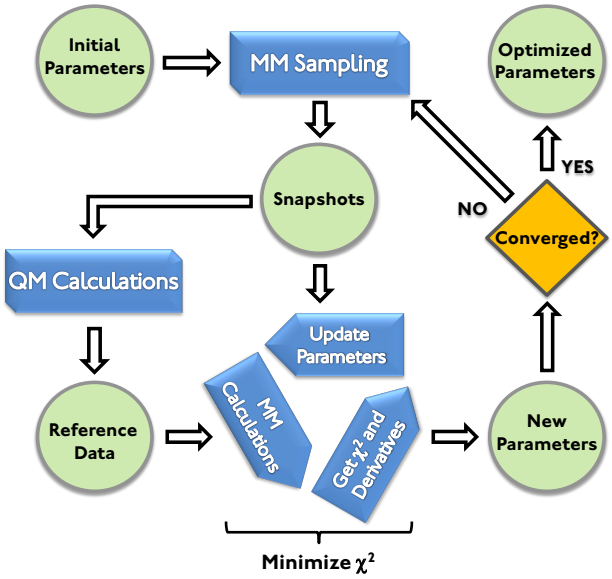


Figure 1: Flowchart for self-consistent force field parameterization showing the flow of reference data (circles) and calculations (boxes): (1, upper left) initial parameters k, (2) sample generation using current parameters k, (3) calculation of corresponding QM reference data, (4) optimization of parameters by minimizing the objective function χ^2, (5) new optimized parameters k. Steps (2) through (5) are repeated until the force field parameters are converged.

2.4 Reweighting of data samples across generations

Since the reference calculations are expensive, we would like to carry over the data from previous generations to aid the optimization process - but it is important to keep in mind that each batch of reference data is sampled using a different force field, and a different thermodynamic ensemble. As the parameters k change between iterations, the measure $P(\mathbf{r};k)\mathrm{d}\mathbf{r}$ reflecting the configurational weights also change correspondingly. Naïvely retaining each reference data sample in the objective function $\chi^2(k)$ would thus bias the final result toward the initial parameters. We addressed this by using the weighted histogram analysis method (WHAM) equations to compute a consistent set of Boltzmann weights for all generations.[29–32] Writing the sampling explicitly as $P(\mathbf{r};k) = \sum_i w_i(k)\delta(\mathbf{r}-\mathbf{r}_i)$, where w_i are the weights from the thermodynamic ensemble, we have at each generation G the self-consistent WHAM weight $P_i(k_j)$ of the i^{th} configuration as

$$P_i(k_G) = \sum_{j=1}^{G} A^{(j)} \frac{w_i(k_G)}{w_i(k_j)}, \tag{2a}$$

$$A^{(j)} = \sum_i P_i(k_j). \tag{2b}$$

where each $A^{(j)}$ is the WHAM weight for the force field at generation j. These are self-consistently determined from these equations beginning from an initial ansatz of equal weights, and the corresponding new weights $P_i(k_G)$ are used in the force field optimization procedure.

Comparing this to the usual application of WHAM—to construct a free energy profile or probability histogram from several simulations that differ by a restraining potential—our current application retains the idea of correcting weights to account for the different measures used to sample different reference data points. To this extent, our present application of WHAM is similar to the multistate Bennett acceptance ratio method for estimating free energy differences.[33]

2.5 Regularization and dimensional rescaling of parameters

Overfitting is a common and onerous problem in force field parameterization. Whenever near- or exact redundancies in the parameter set k are present, optimization algorithms often produce extreme parameter values that are physically counterintuitive or nonsensical. To prevent this, regularization methods are often used to restrain parameters to physically intuitive values.[12,34] A common such method is known as Tikhonov regularization or ridge regression, and involves adding a quadratic penalty to the objective function that restrains parameters to their initial values.

A Bayesian perspective offers a useful framework for choosing the relative scales between the quadratic penalties for each parameter. The quadratic penalty function arises from imposing a Gaussian prior distribution on the force field parameters. As with the objective function in Eq. (1a), the Gaussian widths for each parameter in the prior reflects the intrinsic scale of that parameter, and provides a form of dimensional rescaling that is required to treat parameters with different physical units on the same footing. This is important on two counts. First, parameters in the GROMACS

unit system can vary over six orders of magnitude: bond lengths are on the order of 0.1 while force constants are on the order of 10^5. Second, different parameter types have different inherent variabilities; bond lengths are expected to be correct to within a few percent while atomic partial charges can change sign or fluctuate by several times their initial values.

This Bayesian framework guides, but does not fully automate, our choice of regularization parameters in practical applications. For each parameter, the center of the prior is given by its initial value, and the prior width is the rescaling factor specified at the start of the optimization. For instance, we may choose the prior width for atomic charge parameters to be one elementary charge, and the prior width for a bond length to be 0.01 nm. Note that the prior widths are related to the parameter's natural size, and also incorporate some physical intuition regarding its inherent variability.

3 Application to parametrizing a polarizable water model

To illustrate the power of ForceBalance, we use it to develop a polarizable water model and determine its parameters automatically. The supreme importance of water has spurred intense interest in capturing its extraordinary properties in various theoretical models. The most common point charge models with three or four charge sites have several different published parameter sets depending on the reference data and parameterization strategy, including the TIP3P[3] and SPC/E[35] three-site models and the TIP4P,[3] TIP4P-Ew,[36] TIP4P/Ice[37] and TIP4P/2005[38] four-site models. Other functional forms include a single-site multipole expansion,[39] five-site and six-site models[40,41] and models with explicit three-body interactions.[42]

Recent years have seen increasing interest in the need to treat electronic polarization for more accurate force fields, and thus also for water models;[43–46] moreover, polarizable force fields represent an important frontier for systematic parameterization approaches because of the many-body, nonlinear nature of the dependence of the energies and forces upon the polarization parameters.

3.1 Functional form

Our new water model has five charge sites and a force field of the form

$$E = \sum_{(i,j)\in\text{bonds}} E_{\text{Morse}}(r_{ij}) + \sum_{(i,j,k)\in\text{angles}} E_{\text{U-B}}(\theta_{ijk}, r_{ik}) + E_{\text{QTPIE}}(\{\mathbf{r}_i\}) + \sum_{\text{mol}_i<\text{mol}_j} E_{\text{vdW}}(r_{ij}), \quad (3)$$

where $E_{\text{Morse}}(r_{ij})$ is a Morse potential for the O–H bond vibrations, $E_{\text{U-B}}(\theta_{ijk}, r_{ik})$ is a Urey–Bradley potential for the HOH angle vibration, $E_{\text{QTPIE}}(\{\mathbf{r}\})$ is the electrostatic energy from the QTPIE (charge transfer by polarization current equalization) model,[47–50] and $E_{\text{vdW}}(r_{ij})$ is a pairwise van der Waals interaction. The bond and angle interactions have a total of seven adjustable parameters; the QTPIE interaction has eight parameters and the van der Waals interaction has nine parameters. There are three parameters that determine the position of virtual sites. Overall, ForceBalance treats all 27 adjustable parameters on the same footing and optimizes all of them simultaneously.

3.1.1 Intramolecular parameters

The vibrational modes of water can be approximated with simple functional forms owing to the small amplitude of such motions under typical thermodynamic parameters of interest. We chose the Morse potential and Urey–Bradley potential to describe the bond and angle vibrations respectively; these have the well-known functional forms:

$$E_{\text{Morse}}(r_{ij}) = D_{ij}\left[1 - e^{-a_{ij}(r_{ij}-r_{ij}^0)}\right]^2, \quad (4)$$

$$E_{\text{U-B}}(\theta_{ijk}, r_{ik}) = k_{ijk}^{\theta}(\theta_{ijk} - \theta_{ijk}^0)^2 + k_{ik}^{1-3}(r_{ik} - r_{ik}^0)^2. \quad (5)$$

We have found that accounting for some anharmonicity in the vibrations in this way greatly improves the quality of fit for a single water molecule when compared against other alternatives such as harmonic bond-angle cross terms and quartic angle potentials.

3.1.2 Fluctuating charges

QTPIE is a type of fluctuating-charge model[51,52] that has an improved description of charge transfer behavior.[47,48] The charges q_i on each atom i are recomputed for each geometry by minimizing the fluctuating charge energy expression

$$E_{\mathrm{QTPIE}}(\{\mathbf{r}_i\}) = \min_{\{q_i\};Q} \left[\sum_i (\bar{\chi}_i q_i + \frac{1}{2}\eta_i q_i^2) + \sum_{i<j} q_i q_j J_{ij} \right], \tag{6a}$$

$$\bar{\chi}_i = \sum_{i \neq j} \left(\chi_i - \chi_j\right) S_{ij}, \tag{6b}$$

$$J_{ij} = \frac{1}{r_{ij}} \mathrm{erf}\left(\sqrt{\frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j}} r_{ij} \right), \tag{6c}$$

$$S_{ij} = \left[\frac{4\alpha_i \alpha_j}{(\alpha_i + \alpha_j)^2} \right]^{\frac{3}{4}} e^{-\frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j} r_{ij}^2}. \tag{6d}$$

Each fluctuating charge i has three parameters - the electronegativity χ_i, the chemical hardness η_i, and the Gaussian width α_i. This last parameter determines the amount of screening in the charge–charge interactions J_{ij} and also the attenuation S_{ij} of the electronegativity difference between two charge sites. In this study, all pairwise electrostatic interactions were modified to go smoothly to zero at a cutoff distance of 1.2 nm in order to treat periodic systems.

This minimization procedure reduces to solving a linear saddle-point system with the block matrix form

$$\begin{bmatrix} \mathbf{J} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{\chi} \\ \mathbf{0} \end{bmatrix}. \tag{7}$$

where $\mathbf{J}_{ij} = J_{ij}, i \neq j$, $\mathbf{J}_{ii} = \eta_i$ and $\mathbf{K}$ is the charge constraint topology[48] which enforces charge neutrality of individual molecules. The desired charges are the $\mathbf{q}$ part of the solution. The QTPIE model has been implemented into a development version of the GROMACS simulation software which is freely available on the Web.[24]

3.1.3 Virtual sites

Our model contains five fluctuating charge sites with two sites located on hydrogen nuclei, one virtual site on the HOH angle bisector denoted "M" (similar to the TIP4P model), and two out-of-plane virtual sites denoted "L_1" and "L_2". Their positions are determined by:

$$\mathbf{r}_{\mathrm{M}} = \mathbf{r}_{\mathrm{O}} + a_{\mathrm{M}} \left(\mathbf{r}_{\mathrm{OH}_1} + \mathbf{r}_{\mathrm{OH}_2} \right), \tag{8a}$$

$$\mathbf{r}_{\mathrm{L}_1} = \mathbf{r}_{\mathrm{O}} + a_{\mathrm{L}} \left(\mathbf{r}_{\mathrm{OH}_1} + \mathbf{r}_{\mathrm{OH}_2} \right) + c_{\mathrm{L}} \left(\mathbf{r}_{\mathrm{OH}_1} \times \mathbf{r}_{\mathrm{OH}_2} \right), \tag{8b}$$

$$\mathbf{r}_{\mathrm{L}_2} = \mathbf{r}_{\mathrm{O}} + a_{\mathrm{L}} \left(\mathbf{r}_{\mathrm{OH}_1} + \mathbf{r}_{\mathrm{OH}_2} \right) - c_{\mathrm{L}} \left(\mathbf{r}_{\mathrm{OH}_1} \times \mathbf{r}_{\mathrm{OH}_2} \right). \tag{8c}$$

Here $\mathbf{r}_{\mathrm{OH}_1}$ and $\mathbf{r}_{\mathrm{OH}_2}$ denote the displacement vectors from the O atom to the two H atoms; the constants a_{M}, a_{L} and c_{L} are fitting parameters. The out-of-plane sites are needed to describe the nearly isotropic dipole polarizability of the water molecule; without them, the out-of-plane component of the polarizability tensor would be zero. Figure 2 illustrates the charge sites of the water model after all of the fluctuating charge parameters and virtual site positions have been optimized. In the absence of an electric field, the out-of-plane sites are nearly neutral; their main role is to describe the polarizability tensor and not the static charge distribution. The unconventional positions of the out-of-plane sites are determined by the automatic parameterization and are discussed later.

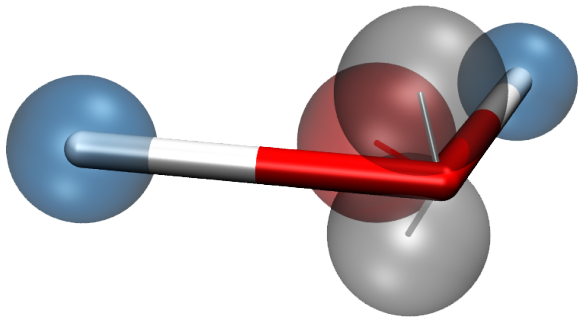


Figure 2: 3D rendering of a model water molecule with the five fluctuating charge sites (spheres) colored to represent their charges in the absence of an electric field. The H atom sites are positive (blue), the in-plane M site is negative (red), and the out-of-plane sites are nearly neutral (gray). The virtual site positions have been fully optimized; note their significant deviation from the "lone pair" positions.

3.1.4 Van der Waals interactions

The van der Waals (vdW) interactions in our water model are described using a new exp–6 functional form with three parameters; we independently parameterize the O–O, H–H and O–H pairwise interactions, so there are a total of nine vdW parameters.

The exp–6 form models both exchange repulsion and dispersion effects. The repulsion is known to have an approximately exponential form;[53–57] however, the original Buckingham exp–6 function[58] has an unphysical singularity at the origin, which causes severe problems for relatively soft repulsive interactions. We have modified the functional form to eliminate the singularity. The new form is given by E_{vdW} below:

$$E_{\mathrm{vdW}}(r) = \frac{2\varepsilon}{1-\frac{3}{\gamma+3}}\left(\frac{\sigma^6}{\sigma^6+r^6}\right)\left[\frac{3}{\gamma+3}e^{\gamma\left(1-\frac{r}{\sigma}\right)}-1\right], \tag{9a}$$

$$E_{\mathrm{B}}(r) = \frac{\varepsilon}{1-\frac{6}{\gamma}}\left[\frac{6}{\gamma}e^{\gamma\left(1-\frac{r}{\sigma}\right)}-\frac{\sigma^6}{r^6}\right], \tag{9b}$$

where we have also written out the Buckingham interaction E_{B} for comparison. In both interactions σ denotes the minimum energy distance , ε is the well depth, and γ is a dimensionless constant describing the steepness of the repulsion. The Buckingham interaction goes to $-\infty$ at zero separation, but at large values of γ (> 12) there is a substantial barrier that blocks the system from reaching the singularity. However, for softer repulsions the barrier is much lower (Figure 3a), and for small values of γ (< 8) the barrier vanishes completely. Previous attempts to eliminate the singularity have used new functional forms with four or more parameters or piecewise behavior.[58–61] Instead, we have chosen a new vdW empirical potential with just three parameters that nevertheless:

- is analytic and nonsingular,
- tends to an attractive r^{-6} term as $r \to \infty$, describing dipole–dipole dispersion,
- tends to a repulsive exponential term as $r \to 0$ describing exchange repulsion,
- has a well-defined, unique minimum,
- is inexpensive to evaluate, and

- is highly tunable to accommodate interactions between diverse atom types and molecule types.

Furthermore, E_{vdW} is parameterized such that σ, ε, and γ have the same physical interpretations as in the Buckingham potential.

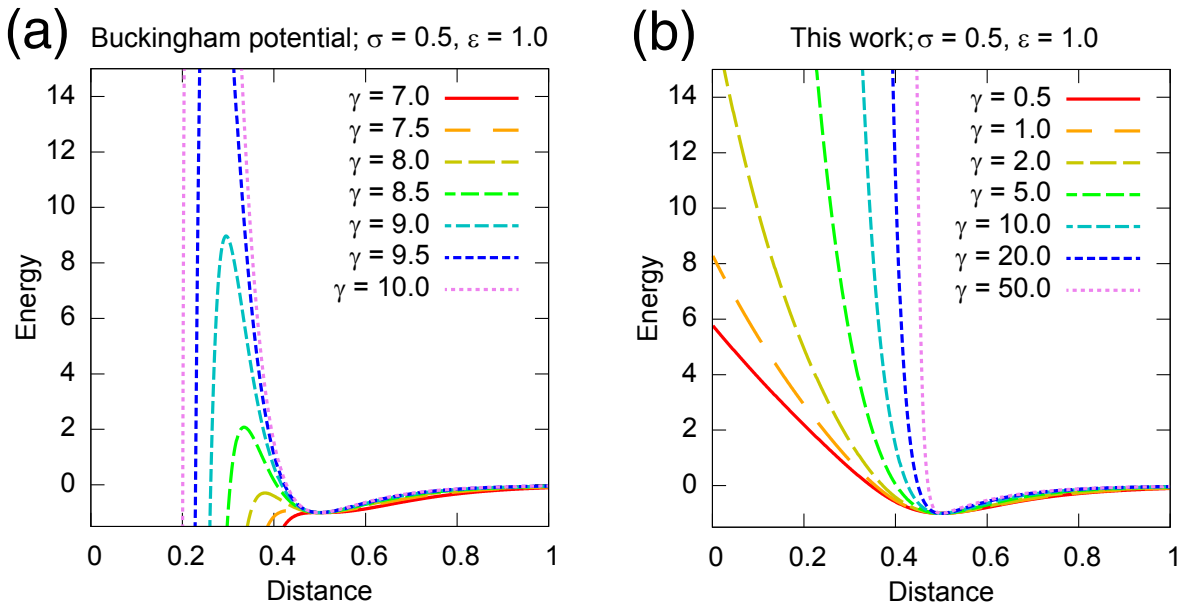


Figure 3: (a) A family of Buckingham potentials with fixed σ and ε parameters with tunable γ. Note that the barrier vanishes for $\gamma < 8$. (b) Our new van der Waals potential.

Our new functional form is guaranteed to have the qualitatively correct behavior for any choice of positive σ, ε, and γ; it has strictly one x-intercept that separates the attractive and repulsive regions, and tends to Ae^{-br} and $-Cr^{-6}$ in the $r < \sigma$ and $r > \sigma$ regions (where A, b, and C are functions of σ, ε, and γ).

Figure 3b shows that the repulsion can be tuned over several orders of magnitude while minimally affecting the attractive region, whereas the Buckingham interaction fails catastrophically for soft repulsions. Furthermore, the new vdW interaction is nearly identical to the Buckingham interaction everywhere except in the region where the Buckingham interaction becomes unphysical. We have implemented the new vdW interaction into our modified copy of GROMACS,[24] and it has proven to be valuable in describing the subtle vdW interactions for hydrogen atoms in hydrogen bonds (see Section 4.2).

3.2 Reference data

The objective function included the following QM reference data:

1. energies and forces for thermally sampled water clusters,
2. dipole and quadrupole moments for an isolated water molecule, and
3. dipole polarizability tensor for an isolated water molecule.

In the objective function, the energy and force residuals were rescaled according to Eq. (1b). The dipole, quadrupole, and polarizability residuals were rescaled to the RMS values of each quantity (1.85 Debye, 2.10 Debye Å, and 1.47 $Å^3$ respectively); all data types were given equal weight. The QM reference data were computed with an RI–MP2[62,63]/aug-cc-pVTZ[64] model chemistry, with additional frozen core and dual–basis[65] approximations to accelerate the calculations. The QM calculations were performed using the Q-Chem quantum chemistry software.[66]

3.3 Parameterization

The water model was automatically parameterized using a self-consistent procedure. The initial guess for force field parameters were derived from a number of sources including the UFF force field,[67] and the SPC/E,[35] TIP4P[3] and TIP5P[40] water models. Where precedent was unavailable (as was the case for fluctuating charge parameters), we guessed the parameters and performed an initial optimization using a simulated annealing algorithm. Table S1 in the Supporting Information lists the starting parameters.

In each generation of the force field parameters, we first obtained 300 samples of water dodecamer configurations by running constant–temperature molecular dynamics and taking snapshots at 10 ps time intervals. A shallow harmonic restraint of 0.5 kJ mol^{-1} nm^{-2} was applied to prevent the trajectories from diverging; this was sufficiently weak to still include samples with one molecule dissociated from the rest of the cluster. We set the prior widths (rescaling factors) by rescaling the parameters within a given type using their geometric mean.

Seventeen self-consistency iterations, using 5100 total cluster configurations, were needed for the parameters to be converged to within 1%; further convergence was not possible due to statistical fluctuations and linear dependency issues. We also performed a multi-cluster fit where we included six sets of clusters with different size (1800 snapshots of 3, 4, 6, 9, 12, and 15-mers); the objective function only decreased by less than 1%, indicating that the force field was robust for different cluster sizes. The final parameter set is given in Table 1.

Table 1: Parameters for the polarizable water model. There are a total of 27 parameters; the average electronegativity has no effect on the interactions.

Morse	r^0 (nm)	D (kJ mol^{-1})	a (nm^{-1})
r_{OH}	0.10251	183.90	28.134
Angle	θ^0 (deg)	k^θ (kJ mol^{-1} rad^{-2})	
θ_{HOH}	113.73	396.82	
Urey–Bradley	r^0 (nm)	k^{1-3} (kJ mol^{-1} nm^{-2})	
r_{HH}	0.031562	6322.3	
QTPIE	χ (eV)	η (eV)	α (bohr^{-1})
H	-1.1973	18.059	0.35423
M (in-plane)	4.1282	11.209	0.32822
L (out-of-plane)	4.5228	11.596	0.32027
Virtual sites	a	c (nm^{-1})	
M (in-plane)	0.26061	n/a	
L (out-of-plane)	0.10086	2.4066	
van der Waals	σ (nm)	ε (kJ mol^{-1})	γ
r_{HH}	0.99403	0.0005	8.9623
r_{OH}	0.29451	1.6072	5.1196
r_{OO}	0.36174	1.0334	13.256

4 Results and Discussion

4.1 Quality of fit

The final optimized model provides a chemically accurate fit to the reference data. For a total of 5100 snapshots spanning 17 generations, the RMS energy error for the 12-mer is 3.6 kJ/mol. The

RMS force error is 10.3%. In the final generation of the optimization, the RMS energy error and force error change by less than 0.1% from their initial values, thus indicating convergence.

The high quality of fit is reflected across the various physical quantities being fit. First, the quality of the energy fit is evident in the scatter plot of Figure 4. In contrast, the SPC/E point charge model shows a much less satisfactory reproduction of the QM reference data. As SPC/E was parameterized to reproduce experimentally measured properties of water, this demonstrates the potential incompleteness of such models in modeling detailed atomistic interactions. Second, a representative 15-mer geometry is shown in Figure 5 to provide some insight into the quality of the force fit. Both QM reference and MM forces are drawn, but the vectors are often practically coincident. This demonstrates how a force error of 10-11% corresponds to forces that are visually indistinguishable. Third, the dipole and traceless quadrupole moments of the water molecule were also reproduced to within 5% of the reference data. For comparison, the traceless octupole moment, which was not fit, disagrees with the reference data by roughly 15%. Fourth, we found in our multi-cluster study that the RMS energy error scales roughly linearly with cluster size, where each additional water molecule contributes 0.3 kJ/mol to the RMS energy error. The RMS force error has a weaker dependence on the cluster size, ranging from 8.5% for the water trimer to 11.5% for the 15-mer.

After the 16th optimization cycle, the generated force field was fit to 4800 snapshots, but not the final set of 300 snapshots. The small changes in the parameters observed by including the final set of snapshots in the 17th generation indicates that the force field is essentially just as accurate for newly sampled snapshots and snapshots used in the fit. This indicates an absence of overfitting to the QM reference data.

4.2 Optimized model parameters

The final optimized parameters are given in Table 1. While the parameter values should not be interpreted too literally, the final parameter values do contain a few surprises. We now discuss the implications of our fitting results.

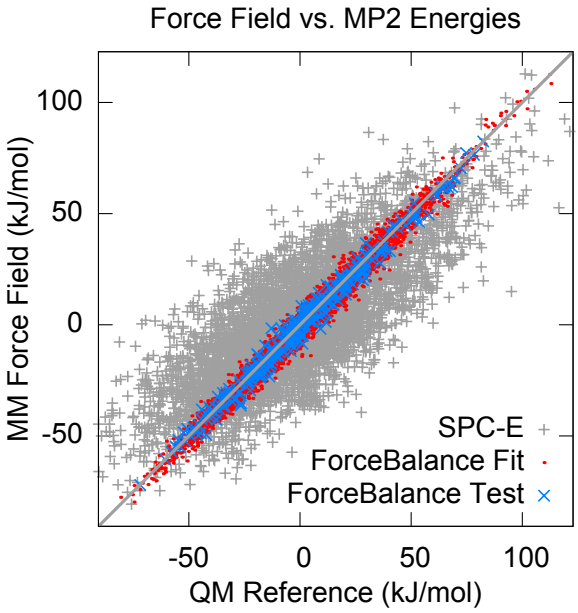


Figure 4: Scatter plots of the relative energies for 5100 configurations of the water 12-mer for SPC/E (grey), the fitted data for our new model (red), and predicted energies for configurations not used in the fit (blue).

The Urey–Bradley interaction has an equilibrium HOH angle of 113 degrees and an equilibrium H–H distance of only 0.039 nm. These do not coincide with the equilibrium geometry of the water monomer. Despite this, the force field provides an equilibrium geometry that closely matches the reference geometry (RMSD = 0.7 pm). This apparent discrepancy can be explained by frustration between the angle term, which favors an increased HOH angle, and the 1–3 term, which favors a shorter H–H distance. At the equilibrium geometry of the water molecule, the Urey–Bradley energy is 44 kJ/mol.

Perhaps our most surprising results are the optimized virtual site positions, which are shown in Figure 2. The position of the oxygen charge site resembles the M site in previous four-site water models. However, the out-of-plane sites (gray spheres in Figure 2) are consistently placed on the same side as the hydrogen atoms. This persists even with initial guesses in the lone pair positions or if the oxygen charge site location was fixed. This is clearly a consequence of the near-isotropy of the dipole polarizability of water, and demonstrates the inadequacy of lone pair models to reproduce the polarization properties of water.

Figure 5: Atomistic forces on a representative water 15-mer from the QM reference data (blue) and the force field (yellow), showing near-coincidence for many atoms. The representative configuration demonstrates a force field with a RMS force error of 11%. The lengths and thicknesses of the force vectors were scaled by an arbitrary global factor to aid visualization.

The QTPIE parameters also demonstrate some interesting trends. As expected, the M and L (in-plane and out-of-plane) charge sites are significantly more electronegative than hydrogen. (There is no charge site on the oxygen atom.) The effective O–H electronegativity difference compares favorably with the QEq parameters:[51] whereas QEq uses an electronegativity difference of 4.2 eV, we find an optimized value of 5.3 eV. As in the QEq model, the absolute electronegativities are not physically significant. We also find that the in-plane site is more electropositive, but also softer, thus resulting in a greater tendency to accumulate negative charge. This reflects slight anistropy of the dipole polarizability, with the out-of-plane polarizability being the smallest component.

The optimized vdW potentials are plotted in Figure 6. The O–O interaction goes to zero at a separation of r_{OO} = 0.316 nm with a well depth of 1.03 kJ/mol; the crossover point agrees well with existing water models,[3,35,36] but the well depth is deeper than expected. The O–H interaction has a crossover point at r_{OH} = 0.24 nm, and the repulsion is significantly softer than the O–O

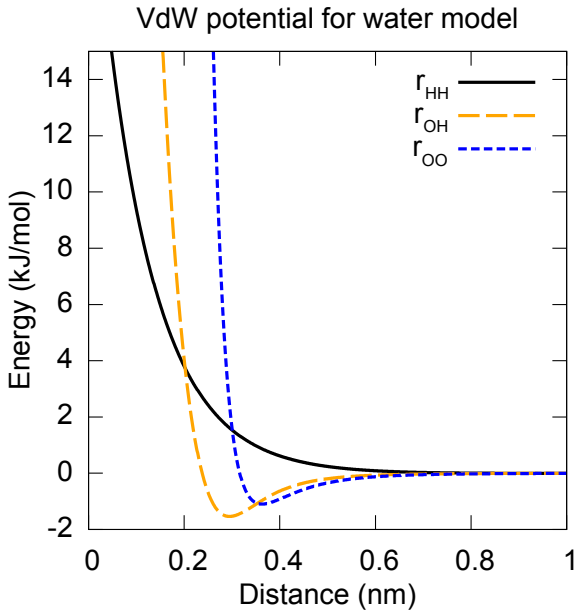


Figure 6: The optimized van der Waals interactions for the water molecule. The O–O potential (blue, dashed) is in close agreement with existing water models, but the well depth is deeper than expected. The O–H potential (orange, dashed) is softly repulsive in the hydrogen bonding region. The H–H potential (black, solid) is entirely repulsive and decays slowly.

interaction (note that the typical O–H distance in a hydrogen bond is 0.07 nm inside the repulsive region.) In stark contrast to the oxygen potentials, the H–H potential is a soft, essentially purely repulsive wall. These observations are consistent with the fact that many water models have no vdW interactions on hydrogens, as a L-J functional form would be too sharply repulsive to provide a good description.

4.3 Validation: predicting properties of liquid water

To validate our force field, we compare the predictions made of some properties of liquid water with experimental measurements. This is a particularly severe test of the accuracy and transferability of our intermolecular interactions, as the model was parameterized using data from only gas-phase water clusters.

To compute the liquid water properties, we used a periodic cubic simulation cell with 512 water molecules. Our simulations used a 1.0 fs time step. Dynamics in the NPT ensemble was

achieved using a Nosé-Hoover thermostat[68,69] and a Parrinello-Rahman barostat[70] as implemented in the GROMACS 4.0.7 simulation code.[25] For this study, we performed a total of >20 ns of MD simulation of the water box at 298.15 K and 1 atm.

Table 2: Predicted properties of liquid water and comparisons to experimental measurements; the experimental values are taken from Ref. 71.

Property	**Computed**	**Experiment**
Density (kg m^{-3})	1040 ± 2	1000
Dielectric constant	85 ± 10	78
Dipole moment	2.6	2.3—2.9
Diffusion constant (10^{-5} cm^2) s^{-1}	1.5 ± 0.2	2.3
Temperature of maximum density (°C)	-20 $\pm$ 5	4

Table 2 summarizes some of the properties we have investigated. First, the bulk density disagrees with experiment by 4%. Although a seemingly small error, the high incompressibility of water suggests that pressure fluctuations may not be well-described in our model.

The electrostatic properties of water are also in good agreement with experiment. We calculated a dielectric constant of 85 ± 10, which agrees well with the experimental value of 78.[71] The large error bars result from the slow convergence of fluctuations of the box dipole moment. We also found the dipole moment of the model water molecules to increase from 1.85 D in isolation to 2.63 D in the liquid phase , in good agreement with previous experimental and theoretical assessments of the dipole moment of liquid water molecules.[50,72–74]

Not all the properties we looked at were predicted well. The predicted self-diffusion constant was 1.5 ± 0.2 , in contrast to the experimental value of 2.3.Abnormally low diffusion constants for water molecules have also been observed for other polarizable models of water,[75–79] suggesting that the polarization response is a common cause. Furthermore, our model did not correctly predict the temperature of maximum density for liquid water; the temperature of maximum density was found at -20 °C instead of the experimentally measured 4 °C. This problem could stem from our choice of optimization procedure, which only sampled water molecules from the room temperature canonical ensemble.

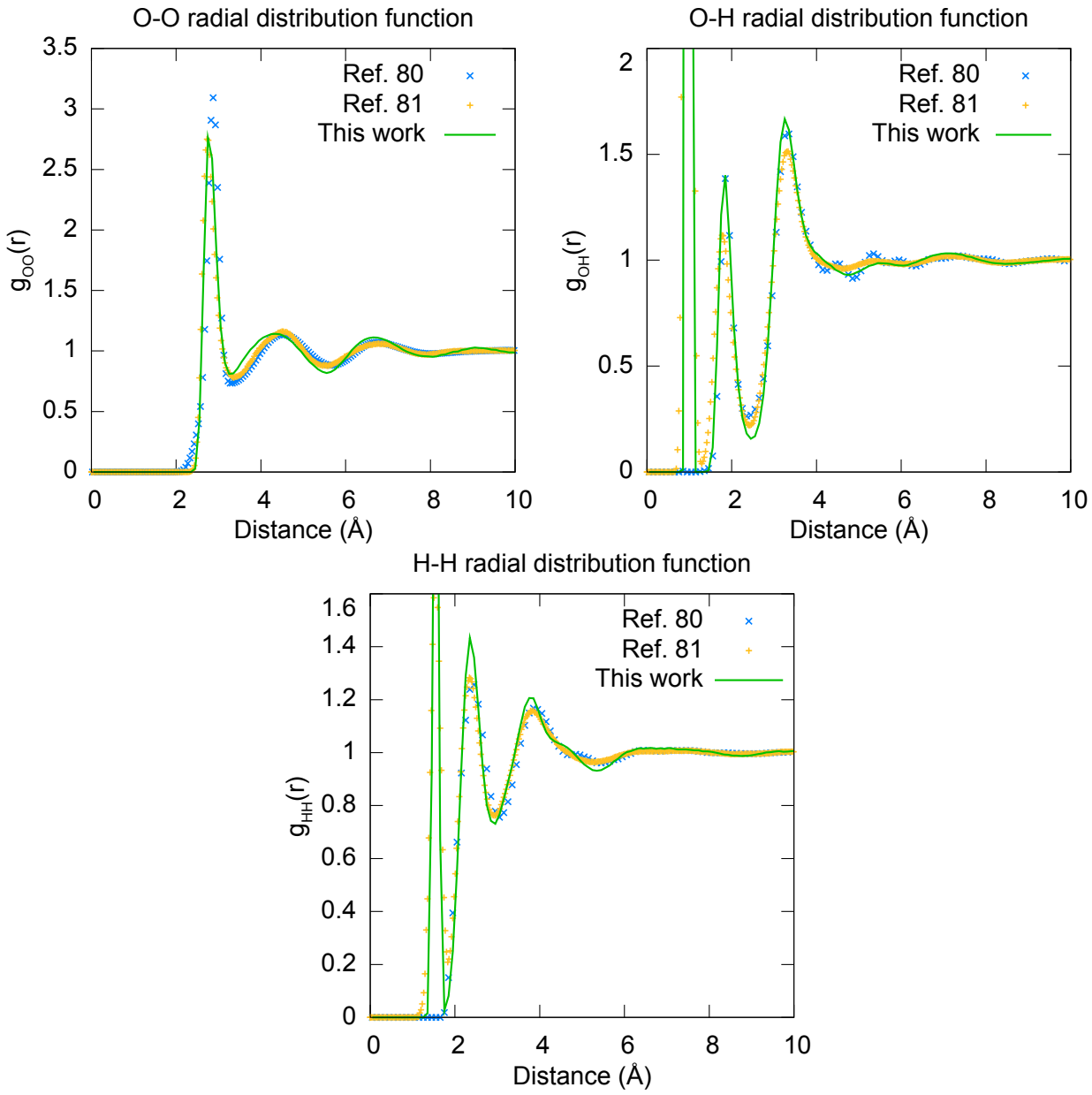


Figure 7: Plots of the (a) O–O, (b) O–H and (c) H–H radial distribution functions as predicted in this work (green), and as measured in Ref. 80 (blue) and Ref. 81 (gold).

The computed radial distribution functions of water are plotted in Figure 7, along with two experimentally derived radial distribution functions.[80–82] The plots show that the force field correctly describes the structure of the liquid water at the two-body level, except for some slight overstructuring which is almost certainly due to the lack of nuclear quantum fluctuations. Our results compare favorably with other polarizable models such as AMOEBA[75] and SWM4-DP.[44]

5 Conclusion

Our main goal in this article was to illustrate the utility of systematic optimization methods and their ability to produce high-quality force fields. ForceBalance provides a framework for easily exploring improvements to the force field by changing the optimization strategy, reference data, and functional form. Here we recap four main challenges / sources of error in our strategy and suggest avenues for further improvement.

Reference data: The reference QM data in this study consisted of approximate calculations on finite sized clusters. Our chosen QM method (MP2 / aug-cc-pVTZ) is an incomplete treatment of electron correlation and may suffer from basis set truncation errors, whereas the most accurate benchmark calculations for water cluster binding energies employ the CCSD(T) method with complete basis extrapolation.[83,84] Our choice of QM method was a compromise between the requirements for high accuracy, comprehensive configurational sampling, and large cluster sizes. In the future we plan to improve our model by incorporating more types of data into our objective function, such as QM calculations with periodic boundary conditions or experimental data.

Functional form: Our choice of functional form was adequate for fitting the MP2/aTZ energies and forces with chemical accuracy; the QTPIE interaction was a good description of the intermolecular electrostatics, and we introduced a new van der Waals interaction with a well-behaved repulsive wall. However, there still remains a small amount of energy and force error, and our electrostatic treatment does not describe intermolecular charge transfer. ForceBalance provides a framework for rapidly exploring the space of functional forms, and in the future we plan to explore functional forms of varying complexity and different treatments of intermolecular charge transfer.

Sampling: We showed that the 27 parameters in our model converged after 17 cycles of self-consistent optimization. The WHAM-based convergence accelerator was instrumental in achieving convergence because it allows us to use reference data from previous generations without biasing the final result. Despite our reference data being entirely based on small clusters, our resulting force field performed quite accurately for several experimentally measured properties of liquid water. However, our model failed to find the temperature of maximum density of liquid water; we

propose that sampling from different tempratures may produce a model that is better adapted to temperature variations.

Regularization: Overfitted models are characterized by unphysical parameter values and poor predictive power. In ForceBalance the physically acceptable range of parameter values is quantified using a Bayesian prior. In the present study we find that the converged parameters are physically interpretable (albeit not too literally), and the force field is able to accurately predict energies and forces outside of the training set. We observed that several force field parameters fell outside our initial expectations, but the differences are not so great that they are patently unphysical. These borderline cases are the most interesting, because they suggest that the true physical interaction differs from our intuitive expectations, and in the future may guide us toward designing better functional forms.

Classical approximation: Quantum nuclear effects have a profound impact on the behavior of water, giving rise to phenomena such as deuterium fractionation and the kinetic isotope effect, and also affecting the condensed phase properties. It is generally acknowledged that including quantum nuclear effects in a simulation leads to reduced structure,[85,86] reduced phase transition temperatures[87] and increased diffusion constants[88,89] compared to a classical simulation; there is also strong evidence for a competing stabilizing effect from intramolecular zero-point vibrations.[90] In the present study, our use of classical molecular dynamics is a major approximation that does not treat the quantum nuclear effects. This may have contributed to the observed discrepancies with experimental measurements. Empirical force fields implicitly include these effects by fitting to the experimental data directly - but some major issues remain, such as the unphysically large temperature gap between the freezing point and density maximum in almost all models.[91] The combination of accurate *ab initio*-based parameterization of a polarizable force field, a sound treatment of quantum nuclear effects, and improved algorithms for more rapid simulations[92] may be sufficient to afford quantitative agreement with experiment comparable to empirical models; this is an exciting topic for future study.

The water model presented in this work is not a bid for the most accurate polarizable water

model - it is quite accurate for some, but not all properties of liquid water. More importantly, the procedure outlined in this work can be easily applied to other systems - this holds great promise for researchers who wish to perform MM simulations but lack the force field for their molecules of interest. In the future, we hope that systematic parameterization methods like ForceBalance will contribute to the molecular simulation community by improving the accuracy of MM simulations and aiding in our search for the essential physical interactions that govern the dynamics of molecules.

Acknowledgement

This work was funded by ENI S.p.A. as part of the Solar Frontiers Research Program. We thank Prof. Vijay Pande (Stanford) and Prof. Todd J. Martínez (Stanford) for helpful discussions.

References

(1) Westheimer, F. H.; Mayer, J. E. *J. Chem. Phys.* **1946**, *14*, 733–738.

(2) Lifson, S.; Warshel, A. *J. Chem. Phys.* **1968**, *49*, 5116–5129.

(3) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(4) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.

(5) Li, D.-W.; Brueschweiler, R. *J. Chem. Theory Comput.* **2011**, *7*, 1773–1782.

(6) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–1958.

(7) Nerenberg, P. S.; Head-Gordon, T. *J. Chem. Theory Comput.* **2011**, *7*, 1220–1230.

(8) Brommer, P.; Gaehler, F. *Model. Simul. Mater. Sc.* **2007**, *15*, 295–304.

(9) Toth, G. *J. Phys.-Condens. Mat.* **2007**, *19*, 335222.

(10) Ercolessi, F.; Adams, J. B. *Europhys. Lett.* **1994**, *26*, 583–588.

(11) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896–10913.

(12) Youngs, T. G. A.; Del Popolo, M. G.; Kohanoff, J. *J. Phys. Chem. B* **2006**, *110*, 5697–5707.

(13) Akin-Ojo, O.; Song, Y.; Wang, F. *J. Chem. Phys.* **2008**, *129*, 064108.

(14) Wang, L.-P.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*, 231101.

(15) Sparta, M.; Hansen, M. B.; Matito, E.; Toffoli, D.; Christiansen, O. *J. Chem. Theory Comput.* **2010**, *6*, 3162–3175.

(16) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(17) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. *J. Comput. Chem.* **1995**, *16*, 1357–1377.

(18) Jakalian, A.; Jack, D. B.; Bayly, C. I. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(19) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908–5920.

(20) Akin-Ojo, O.; Wang, F. *J. Comput. Chem.* **2011**, *32*, 453–462.

(21) Laio, A.; Bernard, S.; Chiarotti, G. L.; Scandolo, S.; Tosatti, E. *Science* **2000**, *287*, 1027–1030.

(22) Umeno, Y.; Kitamura, T.; Date, K.; Hayashi, M.; Iwasaki, T. *Comput. Mater. Sci.* **2002**, *25*, 447–456.

(23) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(24) Wang, L.-P. ForceBalance: Systematic force field optimization. `https://simtk.org/home/forcebalance`, Date Accessed: November 14, 2012.

(25) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(26) Ischtwan, J.; Collins, M. A. *J. Chem. Phys.* **1994**, *100*, 8080–8088.

(27) Paramore, S.; Cheng, L. W.; Berne, B. J. *J. Chem. Theory Comput.* **2008**, *4*, 1698–1708.

(28) Kong, L. T.; Denniston, C.; Muser, M. H.; Qi, Y. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10195–10203.

(29) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1988**, *61*, 2635–2638.

(30) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.

(31) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339–1350.

(32) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(33) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.

(34) Liu, P.; Shi, Q.; Daume, H.; Voth, G. A. *J. Chem. Phys.* **2008**, *129*, 214114.

(35) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(36) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(37) Abascal, J. L. F.; Sanz, E.; Fernandez, R. G.; Vega, C. *J. Chem. Phys.* **2005**, *122*, 234511.

(38) Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2005**, *123*, 234505.

(39) Te, J. A.; Ichiye, T. *J. Chem. Phys.* **2010**, *132*, 114511.

(40) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, PII [S0021–9606(00)50820–4].

(41) Nada, H.; van der Eerden, J. P. J. M. *J. Chem. Phys.* **2003**, *118*, 7401–7413.

(42) Kumar, R.; Skinner, J. L. *J. Phys. Chem. B* **2008**, *112*, 8311–8318.

(43) Ren, P. Y.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(44) Lamoureux, G.; MacKerell, A. D.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185–5197.

(45) Fanourgakis, G. S.; Xantheas, S. S. *J. Chem. Phys.* **2008**, *128*, 074506.

(46) Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. *J. Chem. Phys.* **2010**, *132*, 014309.

(47) Chen, J.; Martinez, T. J. *Chem. Phys. Lett.* **2007**, *438*, 315–320.

(48) Chen, J.; Hundertmark, D.; Martinez, T. J. *J. Chem. Phys.* **2008**, *129*, 214113.

(49) Chen, J.; Martinez, T. J. *Chem. Phys. Lett.* **2008**, *463*, 288–288.

(50) Chen, J.; Martinez, T. J. *J. Chem. Phys.* **2009**, *131*, 044114.

(51) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358–3363.

(52) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

(53) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325–340.

(54) Ahlrichs, R.; Hoffmannostenhof, M.; Hoffmannostenhof, T.; Morgan, J. D. *Phys. Rev. A* **1981**, *23*, 2106–2117.

(55) Hurst, G. J. B.; Fowler, P. W.; Stone, A. J.; Buckingham, A. D. *Int. J. Quantum Chem.* **1986**, *29*, 1223–1239.

(56) Wheatley, R. J.; Price, S. L. *Mol. Phys.* **1990**, *69*, 507–533.

(57) Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **1994**, *94*, 1723–1765.

(58) Buckingham, R. A. *Planet. Space Sci.* **1961**, *3*, 205–216.

(59) Carra, S.; Konowalow, D. D. *Nuovo Cimento* **1964**, *34*, 205–&.

(60) Barker, J. A.; Pompe, A. *Aust. J. Chem.* **1968**, *21*, 1683–&.

(61) Ahlrichs, R.; Penco, R.; Scoles, G. *Chem. Phys.* **1977**, *19*, 119–130.

(62) Weigend, F.; Haser, M. *Theor. Chem. Acc.* **1997**, *97*, 331–340.

(63) Weigend, F.; Kohn, A.; Hattig, C. *J. Chem. Phys.* **2002**, *116*, 3175–3183.

(64) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(65) Wolinski, K.; Pulay, P. *J. Chem. Phys.* **2003**, *118*, 9497–9503.

(66) Shao, Y. et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

(67) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(68) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(69) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(70) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(71) CRC Handbook of Chemistry and Physics, 91st edition. 2010-2011; `http://www.hbcpnetbase.com/`, Date Accessed: February 8, 2011.

(72) Sprik, M. *J. Chem. Phys.* **1991**, *95*, 6762–6769.

(73) Odutola, J. A.; Dyke, T. R. *J. Chem. Phys.* **1980**, *72*, 5062–5070.

(74) Lee, H. M.; Suh, S. B.; Lee, J. Y.; Tarakeshwar, P.; Kim, K. S. *J. Chem. Phys.* **2000**, *112*, PII [S0021–9606(00)30322–1].

(75) Ren, P. Y.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

(76) Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R. H.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999**, *103*, 4730–4737.

(77) Liem, S. Y.; Popelier, P. L. A.; Leslie, M. *Int. J. Quantum Chem.* **2004**, *99*, 685–694.

(78) Fanourgakis, G. S.; Xantheas, S. S. *J. Phys. Chem. A* **2006**, *110*, 4100–4106.

(79) Rick, S. W. *J. Chem. Phys.* **2001**, *114*, 2276–2283.

(80) Soper, A. K.; Phillips, M. G. *Chem. Phys.* **1986**, *107*, 47–60.

(81) Soper, A. K.; Ricci, M. A. *Phys. Rev. Lett.* **2000**, *84*, 2881–2884.

(82) Soper, A. K. *Chem. Phys.* **2000**, *258*, 121–137.

(83) Bates, D. M.; Tschumper, G. S. *J. Phys. Chem. A* **2009**, *113*, 3555–3559.

(84) Yoo, S.; Apra, E.; Zeng, X. C.; Xantheas, S. S. *J. Phys. Chem. Lett.* **2010**, *1*, 3122–3127.

(85) Wallqvist, A.; Berne, B. J. *Chem. Phys. Lett.* **1985**, *117*, 214–219.

(86) Kuharski, R. A.; Rossky, P. J. *J. Chem. Phys.* **1985**, *82*, 5164–5177.

(87) McBride, C.; Noya, E. G.; Aragones, J. L.; Conde, M. M.; Vega, C. *Phys. Chem. Chem. Phys.* **2012**, *14*, 10140–10146.

(88) Lobaugh, J.; Voth, G. A. *J. Chem. Phys.* **1997**, *106*, 2400–2410.

(89) Miller, T. F.; Manolopoulos, D. E. *J. Chem. Phys.* **2005**, *123*, 154504.

(90) Habershon, S.; Markland, T. E.; Manolopoulos, D. E. *J. Chem. Phys.* **2009**, *131*, 024501.

(91) Fernandez, R. G.; Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2006**, *124*, 144506.

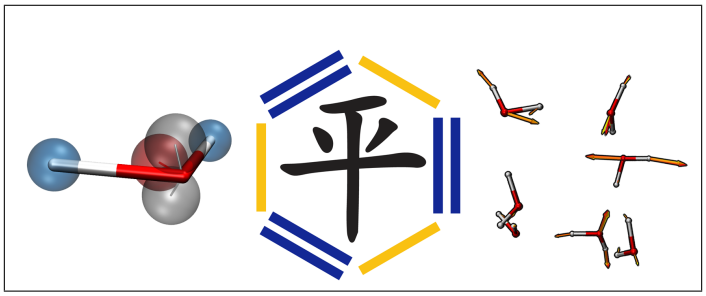
(92) Fanourgakis, G. S.; Markland, T. E.; Manolopoulos, D. E. *J. Chem. Phys.* **2009**, *131*, 094102.

Supporting Information Available

Starting parameters for the optimization are provided in Table S1.

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

Graphical TOC Entry



ForceBalance is a method and free software package for automatic force field parameterization. Left: The polarizable water model that is optimized in this work. Middle: The ForceBalance logo; the Chinese character means “balance”. Right: A plot of the QM (yellow) and MM (red) atomistic forces on a cluster of water molecules, illustrating the quality of the fit.