THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Automated Optimization of Water−Water Interaction Parameters for a Coarse-Grained Model

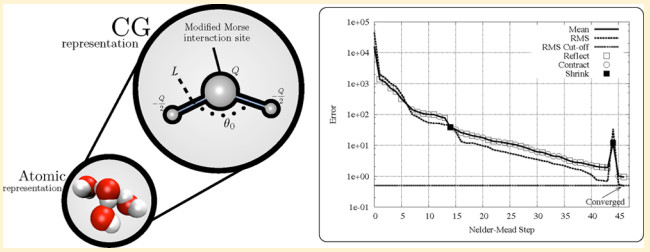
Joseph C. Fogarty,*[,†] See-Wing Chiu,[§] Peter Kirby,[†] Eric Jakobsson,[‡] and Sagar A. Pandit*[,†]

[†]Department of Physics, University of South Florida, Tampa, Florida 33620, United States

[‡]Institute for Advanced Science and Technology, University of Illinois, Urbana, Illinois 61801, United States

[§]Department of Molecular and Integrative Physiology, Beckman Institute for Advanced Science and Technology, Department of Biochemistry, and Center for Biophysics and Computational Biology, University of Illinois, Urbana, Illinois 61801, United States

ABSTRACT: We have developed an automated parameter optimization software framework (ParOpt) that implements the Nelder−Mead simplex algorithm and applied it to a coarse-grained polarizable water model. The model employs a tabulated, modified Morse potential with decoupled short- and long-range interactions incorporating four water molecules per interaction site. Polarizability is introduced by the addition of a harmonic angle term defined among three charged points within each bead. The target function for parameter optimization was based on the experimental density, surface tension, electric field permittivity, and diffusion coefficient. The model was validated by comparison of statistical quantities with experimental observation. We found very good performance of the optimization procedure and good agreement of the model with experiment.



1. INTRODUCTION

Water plays a fundamental role in a wide range of fields of study, including geology, atmospheric sciences, chemistry, biology, and physics.[1] Because of its relevance and ubiquity, water is the most common solvent in experimental and computational studies of biological systems.[2] The hydrophobic effect, the segregation of molecules based on relative interactions with water, governs membrane self-assembly, protein folding, and many other biological processes.[3] Despite its seemingly simple structure, the water molecule forms intricate and dynamic networks of hydrogen bonds that give it unique bulk and interfacial properties.[4,5] The wealth of anomalous characteristics that water exhibits relative to most other liquids underscore its uniqueness. In part because of its enigmatic character, it has been the subject of many theoretical and modeling efforts in the past few decades.[6] Reproducing these anomalous characteristics poses a serious challenge for any effort to attain a detailed understanding of water dynamics and structure.

Quantum mechanical models of water based on first principles (*ab initio*) precipitate the evolution of the electronic structure along with nuclear coordinates over time. The resolution of these methods is at the level of molecular orbitals. These *ab initio* models produce radial distribution functions (RDFs) and rotational/vibrational spectra that strongly agree with experimental data.[7,8] The computational complexity of quantum models restricts systems to a range too limited to determine most statistical bulk quantities. These properties, such as the diffusion coefficient and thermodynamic variables, are more accurately modeled by semiempirical approaches. Water models with atomic-level detail are the most common for simulation of biomolecules and their aggregates. To reproduce a wide range of complex water behavior, several models that consist of four (Bernal−Fowler,[9] TIP4P,[10] and TIPS2[11]) and five (TIP5P[12]) interaction sites have been developed. In the interest of reducing computational cost, most simulations use models (SPC,[13] SPC/E,[14] and TIP3P[15]) that are composed of three partially charged atoms connected with rigid bonds and a Lennard-Jones (6-12) intermolecular potential between oxygen atoms. Each model is reasonably accurate within its domain of applicability, but none reproduce a large portion of the anomalous properties of water. Various models[16] include flexible bonds that allow the water dipole moment to change with the surrounding environment.

Coarse-grained (CG) models are able to achieve longer time scales by averaging over rapid inter- and intramolecular motions, thus permitting a larger time step. CG models are able to effectively represent a larger number of molecules by reducing the number of interaction sites considered. Because of the unusual properties of water, the development of a CG model that reliably describes the bulk and medium-range properties is a challenging project. The development of models of molecular assemblies with lower complexity that reproduce important qualities of water[17,18] is ongoing. For an extensive review, see *Coarse-Graining of Condensed Phase and Biomolecular*

Received: September 24, 2013
Revised: December 31, 2013
Published: January 24, 2014

   dx.doi.org/10.1021/jp409545x | *J. Phys. Chem. B* 2014, 118, 1603−1611

Systems.[19] Johnson et al.[20] have identified limitations of CG models by demonstrating the lack of transferability of a CG pair potential across different states. The work also shows that the pair potential cannot simultaneously resolve all the properties of the reference system for a given state. Despite these limitations, there is extensive development of various CG models for water.[21−32] One common CG protocol is to group a number of water molecules into a single bead with its center as the interaction site.[21−24,26,27,32] This class of CG models carries no charge. Electrostatic interactions are implicitly incorporated into the effective pairwise interaction potential which is either a Lennard-Jones (LJ)[21−24] or Morse-like function.[27] The free parameters in pair potentials are trained against thermodynamic properties such as density, surface tension, and solvation free energy. Molinero and Moore[32] built a coarse-grained water model around the tetrahedral properties shared by water, silicon, and carbon by adapting the Stillinger−Webber potential originally developed for silicon. The model maps one water molecule onto one interacting bead and describes important properties of water across a wide range of temperatures. Using a clustering algorithm, Hadley and McCabe derive the effective potential through fitting to structural data.[17] The models of Shelley et al.[21] and Chiu et al.[27] have large isothermal compressibility. Using a LJ 12-4 potential, Shinoda et al.[24] have designed a CG water that has density and surface tension comparable to those of water. Electrostatic interactions can be introduced into CG water models with multiple charged sites per bead.[25,28−31] These models can describe the dielectric properties of water. Despite the incorporation of electrostatic interactions, both the BMW and polarizable MARTINI models predict density different from experimental measurements.[28,29] The CG representation of Darré et al.[30] associates approximately 11 water molecules with four tetrahedrally interconnected beads (WT4). The model explicitly accounts for long-range electrostatics. The parameter set of the WT4 model is tuned to match the experimental density. On the other hand, the isothermal compressibility and surface tension are modeled less accurately.[30] The recent GROMOS CG[31] is a 5-to-1 mapped water model with two electrostatic interaction sites. It delivers good results for density, surface tension, and dielectric constant as compared to those of real water but yields a coefficient of thermal expansion 1 order of magnitude higher than the experimental value. Adaptive multiscale models allow transfer between simulation domains with differing granularity, from the quantum level up to CG.[33−35] Adaptive methods face unique challenges because of the variability in the degrees of freedom over the span of a simulation. In continuation of our previous CG development based on the Morse potential,[27] we upgrade our previous CG water model (termed CSJ) with polarizability and a Morse-like potential with more flexibility in the landscape of pairwise interaction parameters. We will refer to the new model as modified Morse coarse-grain (MMCG).

2. MODEL

A CG mapping of four water molecules to one CG water bead was used for the current model. For interactions between water beads, we have further refined our previous model,[27] which used the standard Morse potential, to meet two goals: (1) to improve compressibility without sacrificing other features of the potential and (2) to build into the interaction potential a smoothing function that permits fairly accurate simulations using a shorter cutoff. The choice of pairwise potential gives flexibility in the functional form without introducing any extra simulation costs. The coarse-grained potential must model the average of several different interatomic and intermolecular forces. Therefore, the functional form is solely based on a phenomenological understanding of the system rather than a first-principle derivation, such as the r^{-6} behavior of the dispersion force included in the Lennard-Jones potential. A modified Morse potential (eq 1) was used to describe interactions between water bead centers.

$$V_{\mathrm{mm}}(r) = \begin{cases} D_{\mathrm{e}}[\mathrm{e}^{\alpha(r)(1-\frac{r}{R})} - 2\mathrm{e}^{\alpha(r)/2(1-\frac{r}{R})}] \text{ if } r \le R & \text{(1a)} \\ D_{\mathrm{e}}[\mathrm{e}^{\beta(r)(1-\frac{r}{R})} - 2\mathrm{e}^{\beta(r)/2(1-\frac{r}{R})}] \text{ if } r > R & \text{(1b)} \end{cases}$$

where

$$\alpha(r) = \alpha_{\mathrm{L}} + (\alpha_0 - \alpha_{\mathrm{L}})\left(\frac{R-r}{R}\right)$$

$$\beta(r) = \beta_{\mathrm{R}} + (\beta_{\mathrm{c}} - \beta_{\mathrm{R}})\left(\frac{r-R}{r_{\mathrm{c}}-R}\right)$$

where R and D_{e} are the potential well location and depth, respectively. Compared to the Morse potential, the MMCG potential replaces α, the scaling parameter in the exponent, with two separate distance-dependent variables, $\alpha(r)$ and $\beta(r)$. This form decouples the repulsive (eq 1a) and attractive (eq 1b) parts of the potential. The shape of the repulsive term as r tends to 0 is determined by α_0. The exponential factor tends toward α_{L} as r approaches R. Parameters β_{R} and β_{c} provide similar behavior for the attractive part of the potential. Beyond the cutoff distance r_{c}, the potential is taken to be 0. At this writing, we have not conducted an exhaustive search for the value for r_{c} that optimally combines accuracy and efficiency, but the conventional value of 1.6 nm gives good results with reasonable efficiency. Comparing systems representing equal numbers of water molecules, the MMCG model achieves a >1 order of magnitude improvement in performance over an atomistic system. The polarizability of CG water beads is modeled like that of the polarizable MARTINI water model,[28] through the addition of a harmonic angle potential term, with spring constant K_0 and equilibrium angle θ_0, among three charged points. This model of polarizability allows for not only dipole but also higher moments of the electrostatic energy. Figure 1 provides an illustration of the MMCG model. The central point, which acts as the interaction site for the modified Morse potential, carries negative charge Q, while the two outer points each carry charge $-Q/2$. A mass equal to that of four water molecules is distributed evenly among the three points. The distance between the outer points and the central point is a tunable parameter I. The change in the angle (θ) between the points represents the polarizable nature of the bead. The set of potential parameters $\{D_{\mathrm{e}},R,\alpha_0,\alpha_{\mathrm{L}},\beta_{\mathrm{R}},\beta_{\mathrm{c}},\theta_0,K_0,Q,I\}$ determines the space on which optimization will be performed.

3. METHODS AND RESULTS

3.1. Parameter Optimization. Parameter optimization is exploration of a parameter space to determine extreme values of a target function. In the case of molecular dynamics force field optimization, the set of potential parameters defines the space and comparison of model predictions with either *ab initio* data or experimental measurements defines the target function to be minimized. We have developed a new software package, Parameter Optimizer (ParOpt), for general optimization problems (ParOpt is available for download under the GNU

dx.doi.org/10.1021/jp409545x | J. Phys. Chem. B 2014, 118, 1603−1611

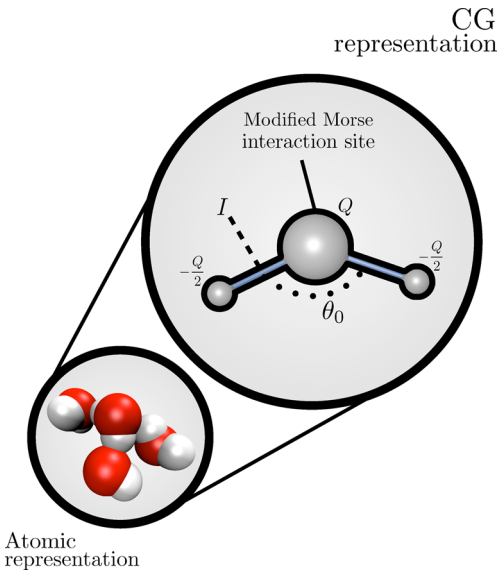


Figure 1. Illustration of the MMCG model. Four water molecules map onto one CG bead. Each CG bead has three charge sites. The central site with charge Q provides the interaction site for the modified Morse potential. Outer charge sites (with charge $-Q/2$) are connected to the central site by a fixed bond of length I. The central angle is a determined by a harmonic potential with equilibrium angle θ_0.

public license at https://csmlabfs1.cas.usf.edu/Sites). The software provides various optimization algorithms, including an orthogonal steepest descent and the Nelder–Mead simplex-based method. See Figure 2 for a schematic representation of ParOpt. Target functions defined for molecular dynamics force field optimization present a challenging problem. Sources of difficulty include high dimensionality, discontinuities, possible stochasticity due to random numerical error and finite sample size effects, and difficulty in defining a derivative. Thus, in the context of molecular dynamics parameter optimization methods such as simplex-based, approaches with the least demanding conditions on the target function are the most suitable. Nelder–Mead, a commonly used simplex-based optimization algorithm, iteratively evolves $M + 1$ points on the M-dimensional space using four basic moves.[36]

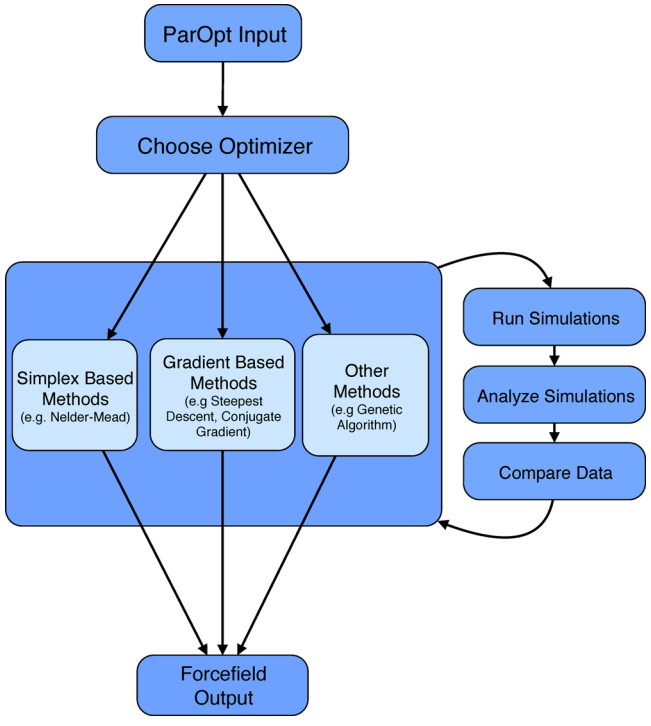


Figure 2. Schematic representation of the ParOpt optimization framework. ParOpt provides an environment in which different optimization methods can be accessed by defining the generation of the target function.

Reflect. Replace the highest point with a point reflected about the centroid of the remaining points.

$$P_r = (1 + \alpha)\bar{P} - \alpha P_h \tag{2}$$

Expand. If the reflected point is lower in value than all points in the centroid, consider a point further from the centroid.

$$P_e = \gamma P_r + (1 - \gamma)\bar{P} \tag{3}$$

Contract. If the reflected point would remain the highest, consider a point nearer to the simplex.

$$P_c = \beta P_h + (1 - \beta)\bar{P} \tag{4}$$

Shrink. Move all points nearer to the lowest point.

$$P_i = \delta P_i + P_l(1 - \delta) \text{ for all } P_i \text{ with } i \neq l \tag{5}$$

The $M + 1$ individual points in the M-dimensional space are denoted by P_i, while subscripts h and l denote the lowest and highest points, respectively. The centroid is defined by

$$\bar{P} = \frac{1}{M}\sum_{i\neq h}^{M+1} P_i \tag{6}$$

Values used for optimization of water parameters are those suggested by Nelder–Mead:[23] $\alpha = 1.0$, $\gamma = 2.0$, $\beta = 0.5$, and $\delta = 0.5$. Figure 3 graphically illustrates the Nelder–Mead simplex transformations.

The Nelder–Mead algorithm replaces the highest point in the simplex with one with a lower target function value by applying these moves. A Nelder–Mead step begins with

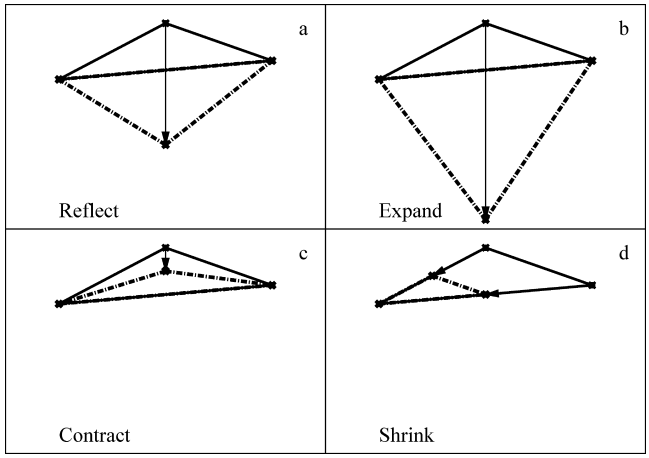


Figure 3. Schematic illustration of Nelder–Mead steps. Arrows denote points that will be replaced and the new point to be considered: (a) the reflection of the highest point about the centroid of the remaining points, (b) the further expansion of the reflection move, (c) the contraction along the line joining the centroid and the highest point, and (d) the shrinking of the simplex by shifting all points nearer to the lowest point.

dx.doi.org/10.1021/jp409545x | J. Phys. Chem. B 2014, 118, 1603–1611

calculation of a reflect move. If the target function value for this new reflected point is lower than all points in the simplex, an expansion move is attempted. If this expanded point is lower than all points in the simplex, it is accepted. If the reflected point is not higher than the highest point in the new simplex, it is accepted and replaces the highest point in the simplex. If the newly accepted reflected point is still the highest point in the simplex, a contraction step is attempted. If the contracted point is higher than the highest point in the simplex, then a shrink move is applied, replacing all points except the lowest. If the contracted point is lower than at least one point, it is accepted into the simplex. This process is repeated until the stopping criterion is met. The stopping criterion for the Nelder–Mead method is a cutoff on the root-mean-square target function value over the entire simplex. The number of occurrences of each move in an optimization depends on the characteristics of both the simplex and the target function and is therefore problem specific. The parameter space for the MMCG water model is a 10-dimensional bounded space. Minimal and maximal values (see Table 1) for each parameter were chosen

Table 1. Optimal Parameters for a Set of Final Parameters from Nelder–Mead Optimization Compared with Results from the Previous Model (CSJ)

parameter	CSJ value[27]	optimized value (this work)	minimal value	maximal value
D_c (kJ mol^{-1})	3.4	3.742	2.0	5.0
R (nm)	0.629	0.560	0.55	0.61
α_0	7	15.545	7.0	20.0
α_L	7	13.323	7.0	20.0
β_R	7	10.976	7.0	20.0
β_c	7	3.198	1.0	7.0
θ_0	–	131.406	105.0	145.0
K_0 (kJ mol^{-1})	–	93.918	20.0	180.0
Q (e)	–	−1.126	−1.6	−1.06
I (nm)	–	0.141	0.1	0.25

heuristically to restrict target function evaluations to reasonable combinations of parameters. The original description of the Nelder–Mead algorithm has no set method for boundary conditions on the parameter space, though suggestions are made for modification of functional forms or assignment of large "penalty" numbers for points outside of the boundary. We chose to replace points outside the boundary with the nearest point within the allowable space.

For optimization of the MMCG water model, the target function was defined by the weighted percent error in comparison of CG simulation results with four experimental quantities: density, dielectric constant, diffusion coefficient, and surface tension. To ensure that all comparison data contribute to the target function and thus the evolution of the simplex, percent errors are assigned scalar weights to yield similar orders of magnitude among all quantities. With this goal in mind, density was weighted 100 times higher than surface tension, while permittivity and diffusion coefficient were weighted 10 times higher. Experimental values used in training are listed in Table 2.

Density. Density was computed by taking the ratio of the mass of the system and the ensemble average volume.

Diffusion Coefficient. The standard practice in comparing diffusion coefficients between CG models and experiment is to

Table 2. Training Data Showing Results for the Optimized Point Compared with Experimental Target Data

property	experiment	training result
ρ (g cm^{-3})	0.996[39]	0.993
D ($\times 10^{-9}$ m s^{-2})	2.597[40]	2.61
ε	76.8[41]	76.24
γ (mN m^{-1})	71.2[42]	78.43

assume that diffusion for N-sized clusters scales as $1/N$; i.e., $D_N = D_1/N$, where D_N is the diffusion coefficient for an N-sized cluster. This assumption neglects any interatomic correlation. The GROMOS CG work[31] compares the CG bead diffusion coefficient with the diffusion coefficient for the center of mass of clusters in the SPC water model. In that work, static clusters in the atomistic representation were produced by adding distance restraints between oxygen atoms, which produces diffusion coefficient scaling close to the inverse behavior usually assumed. In the work presented here, clusters are constructed similarly, though unrestrained. To calculate diffusion coefficient scaling, clusters were constructed from each molecule and its nearest $N - 1$ neighbors. The MSD for such clusters was calculated using a window averaging method with eq 7. Though the method presented here may not be generalizable to all types of clusters, it provides an improvement over current assumptions for homogeneous systems. Figure 4 shows the

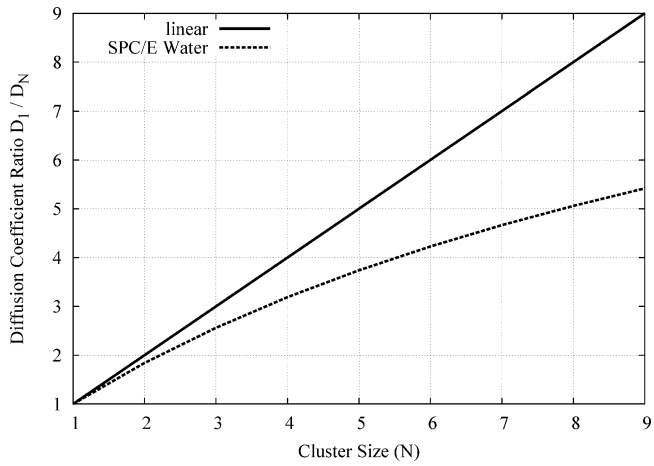


Figure 4. MSD scaling for SPC/E water. The line shows the usual practice of taking the diffusion of an N cluster beginning N times slower than the single molecule. SPC/E was determined by simulation of SPC/E at 303 K. Clusters were chosen by the $N - 1$ nearest neighbors for each water molecule, and the center of mass diffusion was calculated.

diffusion coefficient scaling for SPC/E water. The plot for atomistic water shows significant deviation from the scaling behavior of uncorrelated clusters, especially at larger values of N. Therefore, approximations leading to $1/N$ scaling are more appropriate at smaller cluster sizes. At a bead size of four water molecules, the difference is significant. On the basis of this analysis, we use a scale factor (s) of 3.16 to compare the CG diffusion coefficient with experiment, instead of the usual factor of 4.

$$D_1 = sD_N = s \lim_{t\to\infty} \frac{\langle [x(t) - x(0)]^2 \rangle}{6t} \quad (7)$$

dx.doi.org/10.1021/jp409545x | J. Phys. Chem. B 2014, 118, 1603–1611

where $x(t)$ is the position of the central bead at time t and the broken brackets denote an ensemble average.

Relative Permittivity. The relative permittivity (ε) was calculated using a Clausius−Mossotti-like equation with a reaction field with an infinite dielectric constant (conducting boundary conditions).[37]

$$\varepsilon = 1 + \frac{\langle M^2 \rangle - \langle M \rangle^2}{3\varepsilon_0 \langle V \rangle T} \tag{8}$$

where M is the total system dipole, ε_0 is the permittivity of free space, $\langle V \rangle$ is the ensemble average system volume, T is the target temperature for the thermostat, and ε_0 is the permittivity of the vacuum.

Surface Tension. The surface tension was taken from the GROMACS internal calculation:

$$\gamma(t) = \frac{L_z}{N_s}\left[P_{zz}(t) - \frac{P_{xx}(t) + P_{yy}(t)}{2}\right] \tag{9}$$

where the z-axis is normal to the interface, P_{nn} terms are the pressure tensor diagonal elements, N_s is the number of interfaces, and L_z is the size of the system along the z-axis.

The optimization procedure involves many target function evaluations. These evaluations, at a point in parameter space, involve the execution of two MD simulations. This places practical limits on the size and length of simulations used for optimization. Therefore, systems used for determination of the diffusion coefficient, density, and electric field permittivity contained 512 CG beads, corresponding to 2048 water molecules. This system is denoted as S_1. Because of large fluctuations, the calculation of surface tension requires a larger system: 4096 CG beads, implying 16384 water molecules (system S_2). System S_2 was constructed, at every evaluation of the target function, with a slab near the average density computed from S_1 in contact with a vacuum.[38] System S_2 was simulated under the constant number, volume, and temperature (NVT) ensemble. More details on simulation methods are found in Appendix A.

The first step in optimization is the construction of an initial simplex. In this work, optimization started with a simplex consisting of random points within the domain of parameter space that satisfied constraints. This simplex was optimized until convergence. To further improve the accuracy of the final point, the optimization was restarted with an initial simplex consisting of the converged point and random points around it. This procedure was then iterated two more times, with the final optimization resulting in the point presented in Table 1.

Figure 5 shows the mean and root-mean-square (rms) target function value versus the Nelder−Mead step for the final optimization iteration. Because of the roughness of the hypersurface generated by the target function, the simplex may occasionally include a vertex with an abnormally high value, as observed in the mean and rms of the second shrink step of Figure 5. Figure 6 shows the evolution of parameter values over the final run, while Figure 7 shows the change in computed physical properties. The figures demonstrate the convergence of both the data and parameter values. Further, Figure 6 demonstrates that although parameters did encounter the boundaries over the course of the optimization, the final simplex converged away from the boundary. Thus, the parameter value boundaries chosen were not unreasonable.

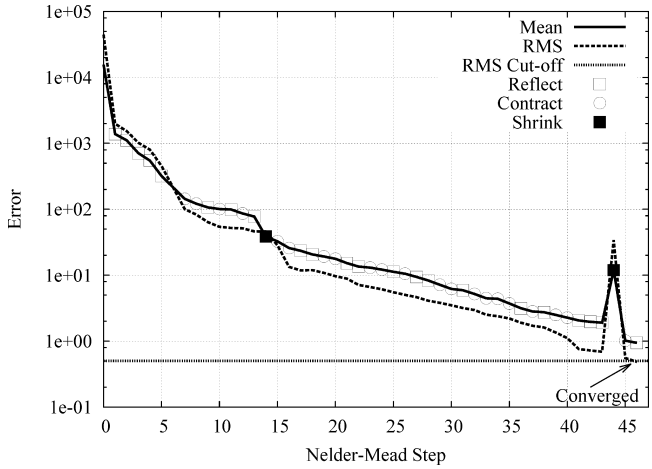


Figure 5. Optimization results. The mean error, rms error, and rms cutoff as a function of Nelder−Mead step are shown. Reflection (□), contraction (○), and shrink (■) steps are labeled at the corresponding mean error point.

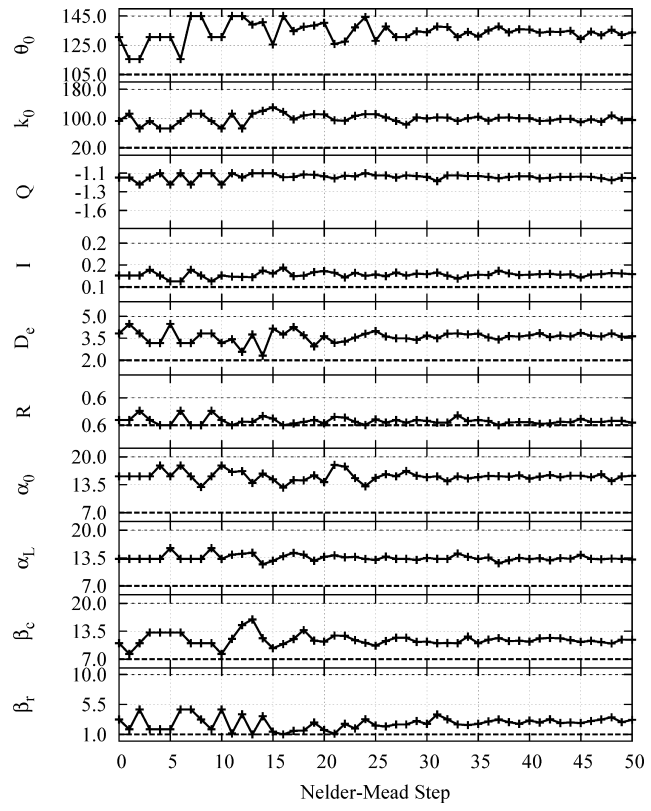


Figure 6. Optimization results. Parameter trajectory. Variation in parameters as a function of Nelder−Mead step. Dashed lines indicate minimal and maximal values.

The final set of parameters was chosen as the point in the final converged simplex with the lowest target function value (see Table 1). The decoupling of repulsive and attractive parameters, the added distance dependence of the exponential factors, and the added electrostatic interaction have allowed for a decrease in the equilibrium distance, while maintaining the experimental density. The large differences between exponential parameters in the new and original models hint that the extensions have added important flexibility. The simulation

dx.doi.org/10.1021/jp409545x | J. Phys. Chem. B 2014, 118, 1603−1611

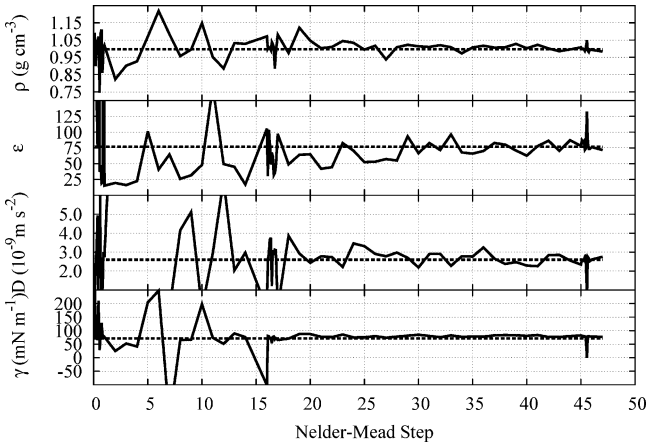


Figure 7. Optimization results. Training data [density (ρ), electric field permittivity (ε), diffusion coefficient (D), and surface tension (γ)] (—) as a function of Nelder–Mead step compared with target values (---).

results corresponding to the final set of parameters are listed in Table 2.

3.2. Validation. When validating the results of an optimization procedure, one cannot consider the training data alone. The suitability of the force field will depend on matching data that the model was not explicitly trained to reproduce. Therefore, validation simulations were performed to verify the reliability of the optimized parameters. Because validation is not performed iteratively, it is not subject to the same size constraints as optimization. All systems for validation consisted of 110592 CG beads, which corresponds to 442368 water molecules. See Appendix A for more details on simulation methods.

The density, diffusion coefficient, relative permittivity, and surface tension were determined from the larger validation systems using the same methods that were used during optimization [see section 3.1; in addition, bulk thermodynamic quantities were also calculated (see Table 3)].

Coefficient of Thermal Expansion. The coefficient of thermal expansion, α, was calculated using the finite-difference method.[43]

$$\alpha = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P \approx -\left[\frac{\ln\left(\frac{\rho_2}{\rho_1}\right)}{T_2 - T_1}\right]_P \tag{10}$$

Two constant number, pressure, and temperature (NPT) ensemble simulations at temperatures of 303.15 K (T_1) and 308.0 K (T_2) were performed. The ensemble averages of the resulting densities were used for ρ_1 and ρ_2, respectively.

Isothermal Compressibility. Isothermal compressibility κ_T was calculated using two methods:[44] (i) a finite difference method for NVT simulations at different densities (ρ_i)

$$\kappa_T = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial P}\right) \approx \frac{\ln\left(\frac{\rho_2}{\rho_1}\right)}{\langle P_2\rangle - \langle P_1\rangle} \tag{11}$$

and (ii) volume fluctuations in an NPT ensemble simulation

$$\kappa_T = \frac{\Delta V^2}{\langle V\rangle k_B T} \tag{12}$$

where P_i is the pressure on system i and ΔV^2 is the variance of the volume. For eq 11, the following densities were used: ρ_1 = 0.993 g cm^{-3}, and ρ_2 = 0.963 g cm^{-3}.

Enthalpy of Vaporization. The enthalpy of vaporization was calculated from the interaction energy between CG beads (V_{Inter}) and the intermolecular interactions that occur within the bead V_S of 128.5 kJ mol^{-1} determined from quantum mechanical calculations of the water tetramer binding energy.[27,45]

$$\Delta H_{vap} = -\frac{V_{Inter} + V_S}{4} + RT \tag{13}$$

where the factor of 4 is due to the level of CG.

The average dipole moment of an individual bead was calculated for the MMCG water model. Because a CG bead represents four water molecules, the total dipole moment for an atomistically detailed SPC/E water molecule along with its three nearest neighbors was computed for comparison. The MMCG model yields an average dipole moment of 4.1 ± 0.25 D, whereas the four-water cluster dipole moment from the SPC/E water model was 5.3 ± 1.0 D; density function theory (DFT) calculations for an isolated cluster produce no dipole moment.[46] The effective polarizability (α_0) for a CG bead was calculated using eq 14.[47] The MMCG model yields a value of 63 au compared with a value of 39.33 au given by DFT for an isolated four-water cluster.[48] The disparity in these values is most likely a result of differences between clusters in the bulk and in isolation.

Table 3. Validation Data Comparing Experimental Data, an Existing CG Model (polarizable MARTINI), an Atomistic Model (SPC/E), and the Previous Model (CSJ) to This Work (MMCG)[a]

property	experiment	polarizable MARTINI[28]	SPC/E	CSJ[27]	MMCG (this work)
ρ (g cm^{-3})	0.996[39]	1.043	0.998[14]	0.998	0.993 ± 0.0005
D (×10^{-9} m s^{-2})	2.597[40]	2.5	2.5[14]	4.3	3.07 ± 0.03
ε	76.8[41]	75.6	70.7 (at 298 K)[49]		74.17 ± 0.06
γ (mN m^{-1})	71.2[42]	30.5		71.0	78.73 ± 0.2
α (×10^4 K^{-1})	3.21[50]				6.89 ± 1.3
κ_T (×10^{-6} bar^{-1}) (NPT)	44.75[51]		34.07[44]	170.0	68.95 ± 10
κ_T (×10^{-6} bar^{-1}) (NVT)	44.75[51]		41.41[44]		57.16 ± 0.5
H_{vap} (kJ mol^{-1})	44.0[52]			38.4	48.52

[a]Quantities computed are density ρ, diffusion coefficient D, dielectric permittivity ε, surface tension γ, coefficient of thermal expansion α, isothermal compressibility κ_T, and enthalpy of vaporization H_{vap}. Values computed both during optimization and for validation differ because of the size differences between simulated systems.

dx.doi.org/10.1021/jp409545x | J. Phys. Chem. B 2014, 118, 1603–1611

$$\alpha_0 = \frac{(Q/2)^2 I^2}{K_0} \tag{14}$$

4. CONCLUSION

To automate force field optimization, we have developed a general optimization framework ParOpt. The software utilizes various methods to locate minima of the defined target function on parameter space. Parameters are systematically varied to determine the most accurate set of parameters. We have applied this general framework to optimize the parameters of a polarizable CG water model with nonbonded interactions given by a modified Morse potential via the Nelder–Mead algorithm. The model has more flexibility in the functional form than most CG models, allowing it to match a wide range of experimental properties (see Table 3).

Any model in a classical molecular dynamics framework is a simplification of a physical system and therefore involves a reduction in the degrees of freedom in the description of the interactions. Any such modeling attempt would be nonunique. Though the set of parameters that govern the model are mathematically independent, their effects on observable data and therefore the target function are not necessarily uncorrelated. Therefore, because of this interdependence of the parameters, the optimization is not necessarily underdetermined despite an apparent abundance of free parameters to match the target data. The question of uniqueness is nontrivial to answer and is in fact a matter of ongoing research (e.g., in the context of economic models[53]). For complex systems in general, there is no straightforward, computationally tractable, and general solution to this problem; rather, heuristic approaches must be developed for the particular system being studied. In light of these factors, it may be the case that our parameter set presented in this paper is one of many equally good sets. The point we have chosen was a suitable point that matched the training data well. Additionally, it can be seen from the good agreement with the validation data, which is independent of the training data, that the resulting force field parameters are well-tuned.

There are fundamental limitations to the explanatory power of CG water models. RDFs are notoriously difficult for to match. An artifact common to most CG models is the long-range correlation between CG water beads not seen in experiment or atomistic models. Figure 8 shows that the current model suffers from that same limitation. Likely as a result of the high degree of interbead correlation at long distances, the model has the propensity to spontaneously solidify at room temperature, a weakness found in other CG water models.[23] This phenomenon would not be expected in a heterogeneous system, because long-range correlations would be broken by the presence of other interactions. Additionally, spontaneous freezing can be avoided by periodically assigning random velocities chosen from a Maxwell distribution, a method similar to one suggested by Harvey et al. to prevent the overpopulation of low-frequency modes of a molecular dynamics system.[54] Because dynamic and structural properties of water clusters vary with temperature, the internal states of CG beads would, as well. Thus, CG models cannot necessarily be assumed to transfer to different temperatures. We believe that the interaction parameters for CG models would be more accurately treated as being temperature-dependent. The set of potential parameters has been tuned for one temperature and is not expected to accurately simulate behavior far from the target temperature or in thermodynamic ensembles without constant temperature control.

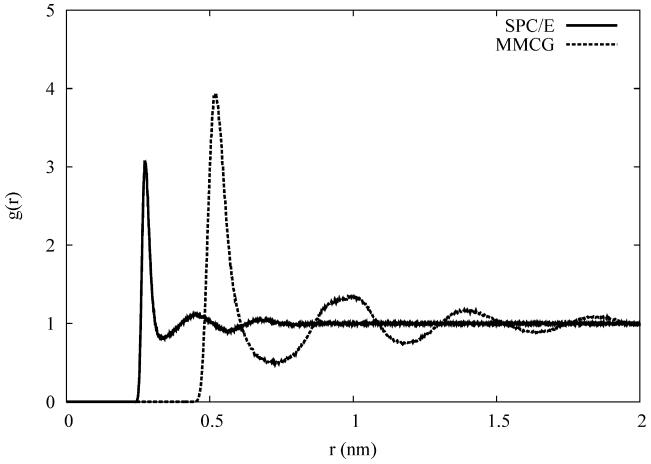


Figure 8. RDF. Comparison between $g(r)$ for the atomistic SPC/E water model (—) and the MMCG water model (---). The CG model has greater long-range coordination than the atomistic model.

The predictions of the model agree well with experiment for structural, dynamic, and bulk properties. The diffusion constant shows that the model has accurate dynamics. The dielectric constant indicates the validity of the parameters that determine electrostatics and polarizability. Density ensures the correct spatial scale of the CG beads. Accurate surface tension indicates an accurate representation of the strength of the interaction between beads. The thermodynamic quantities predicted by the model are in good agreement with those of experiment, validating the fitness of the parameters.

APPENDIX A

Simulation Details

GROMACS 4.0 modeling software[55] was used for all MD simulations performed in this work. User-custom look-up tables for the Morse potentials (1) were prepared as described in a previous work.[27] We find that replacing the standard Lennard-Jones interaction form with a table look-up does not add significantly to computational time. The neighbor search cutoff (rlist) was set to 1.6 nm. When an energy look-up table is used, GROMACS 4.0 uses rlist as the real cutoff for the interaction potential. A time step (Δt) of 20 fs was used in all simulations. The pair list was updated, and the center of mass motion was removed every five steps. All simulations were performed at 303.15 K and at 1 atm of pressure unless otherwise explicitly specified. Electrostatic interactions were computed using the particle mesh Ewald (PME) method.[56,57] The Nosé–Hoover temperature coupling method[58] and Parrinello–Rahman pressure coupling algorithm[59] were used for setting temperature and pressure boundary conditions, respectively. Temperature and pressure coupling constants of 2.1 and 2.5 ps, respectively, were used for all simulations.

APPENDIX B

Continuity and Differentiability of $V(r)$

The modified Morse potential (eq 1) is continuous and differentiable over the range $r \in [0, r_c)$.

$$\lim_{r \to R^-} \alpha(r) = \alpha_L \tag{15}$$

dx.doi.org/10.1021/jp409545x | J. Phys. Chem. B 2014, 118, 1603–1611

$$\lim_{r\to R^-} V_{mm}(r) = \lim_{r\to R^-} D_e[e^{\alpha_L(1-\frac{r}{R})} - 2e^{\alpha_L/2(1-\frac{r}{R})}] = D_e$$

$$\lim_{r\to R^+} \beta(r) = \beta_R \tag{16}$$

$$\lim_{r\to R^+} V_{mm}(r) = \lim_{r\to R^+} D_e[e^{\beta_R(1-\frac{r}{R})} - 2e^{\beta_R/2(1-\frac{r}{R})}] = D_e$$

Therefore, the limit exists at R, and $V_{mm}(r)$ is a continuous function. The function is also differentiable at $r = R$:

$$\frac{\partial V_{mm}(r)}{\partial r} = -D_e\left[h(r) - \left(1 - \frac{r}{R}\right)\frac{\partial h(r)}{\partial r}\right] [e^{h(r)(1-\frac{r}{R})} - e^{h(r)/2(1-\frac{r}{R})}] \tag{17}$$

where

$$h(r) = \begin{cases} \alpha(r) \text{ if } r \le R \\ \beta(r) \text{ if } r \le R \end{cases}$$

In the limit of r approaching R, the final term in eq 17 is vanishing. Because the first and second terms approach constant values, $\partial V_{mm}(r)/\partial r$ also vanishes, and the derivative of $V_{mm}(r)$ is continuous.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: jcfogart@mail.usf.edu.

*E-mail: pandit@usf.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

J.C.F., P.K., and S.A.P. thank the National Institutes of Health for partial support at various stages of this work via Grant 1R01GM086707-01A1. We thank Prof. H. L. Scott and Prof. S. Varma for careful reading of the manuscript and valuable suggestions.

REFERENCES

(1) Ball, P. Water: An Enduring Mystery. *Nature* **2008**, *452*, 291−292.

(2) Guillot, B. A Reappraisal of What we Have Learnt During Three Decades of Computer Simulations on Water. *J. Mol. Liq.* **2002**, *101*, 219−260.

(3) Darré, L.; Machado, M. R.; Pantano, S. Coarse-grained Models of Water. *WIREs Comput. Mol. Sci.* **2012**, *2*, 921−930.

(4) Sedlmeier, F.; Janecek, J.; Sendner, C.; Bocquet, L.; Netz, R.; Horinek, D. Water at Polar and Nonpolar Solid Walls (Review). *Biointerphases* **2008**, *3*, FC23−FC39.

(5) Berkowitz, M. L.; Bostick, D. L.; Pandit, S. Aqueous Solutions next to Phospholipid Membrane Surfaces: Insights from Simulations. *Chem. Rev.* **2006**, *106*, 1527−1539.

(6) Dill, K. A.; Truskett, T. M.; Vlachy, V.; Hribar-Lee, B. Modeling Water, the Hydrophobic Effect, and Ion Solation. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 173−199.

(7) Szalewicz, K.; Leforestier, C.; van der Avoird, A. Towards the Complete Understanding of Water by a First-principles Computational Approach. *Chem. Phys. Lett.* **2009**, *482*, 1−14.

(8) Kuo, I.-F. W.; Mundy, C. J.; McGrath, M. J.; Siepmann, J. I.; VandeVondele, J.; Sprik, M.; Hutter, J.; Chen, B.; Klein, M. L.; Mohamed, F.; et al. Liquid Water from First Principles: Investigation of Different Sampling Approaches. *J. Phys. Chem. B* **2004**, *108*, 12990−12998.

(9) Bernal, J. D.; Fowler, R. H. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *J. Chem. Phys.* **1933**, *1*, 515−548.

(10) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(11) Jorgensen, W. L. Revised TIPS for Simulations of Liquid Water and Aqueous Solutions. *J. Chem. Phys.* **1982**, *77*, 4156−4163.

(12) Mahoney, M. W.; Jorgensen, W. L. A Five-site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. *J. Chem. Phys.* **2000**, *112*, 8910−8922.

(13) Berendsen, H.; Postma, J.; Gunsteren, W.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; The Jerusalem Symposia on Quantum Chemistry and Biochemistry; Springer: Dordrecht, The Netherlands, 1981; Vol. *14*, pp 331−342.

(14) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.

(15) Jorgensen, W. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103*, 335−340.

(16) Schropp, B.; Tavan, P. The Polarizability of Point-Polarizable Water Models: Density Functional Theory/Molecular Mechanics Results. *J. Phys. Chem. B* **2008**, *112*, 6233−6240.

(17) Klein, M. L.; Shinoda, W. Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems. *Science* **2008**, *321*, 798−800.

(18) Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. Multiscale Modeling of Emergent Materials: Biological and Soft Matter. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869−1892.

(19) Voth, G., Ed. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Taylor & Francis Group: Boca Raton, FL, 2009.

(20) Johnson, M. E.; Head-Gordon, T.; Louis, A. A. Representability Problems for Coarse-grained Water Potentials. *J. Chem. Phys.* **2007**, *126*, 144509.

(21) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. A Coarse Grain Model for Phospholipid Simulations. *J. Phys. Chem. B* **2001**, *105*, 4464−4470.

(22) Marrink, S.; de Vries, A.; Mark, A. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108*, 750−760.

(23) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.

(24) Shinoda, W.; DeVane, R.; Klein, M. L. Multi-property Fitting and Parameterization of a Coarse Grained Model for Aqueous Surfactants. *Mol. Simul.* **2007**, *33*, 27−36.

(25) Ha-Duong, T.; Basdevant, N.; Borgis, D. A polarizable Coarse-grained Water Model for Coarse-grained Proteins Simulations. *Chem. Phys. Lett.* **2009**, *468*, 79−82.

(26) Hadley, K. R.; McCabe, C. On the Investigation of Coarse-Grained Models for Water: Balancing Computational Efficiency and the Retention of Structural Properties. *J. Phys. Chem. B* **2010**, *114*, 4590−4599.

(27) Chiu, S.-W.; Scott, H. L.; Jakobsson, E. A Coarse-Grained Model Based on Morse Potential for Water and n-Alkanes. *J. Chem. Theory Comput.* **2010**, *6*, 851−863.

(28) Yesylevskyy, S. O.; Schafer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, *6*, No. e1000810.

(29) Wu, Z.; Cui, Q.; Yethiraj, A. A New Coarse-Grained Model for Water: The Importance of Electrostatic Interactions. *J. Phys. Chem. B* **2010**, *114*, 10524−10529.

(30) Darré, L.; Machado, M. R.; Dans, P. D.; Herrera, F. E.; Pantano, S. Another Coarse Grain Model for Aqueous Solvation: WAT FOUR? *J. Chem. Theory Comput.* **2010**, *6*, 3793−3807.

(31) Riniker, S.; van Gunsteren, W. F. A Simple, Efficient Polarizable Coarse-grained Water Model for Molecular Dynamics Simulations. *J. Chem. Phys.* **2011**, *134*, 084110.

(32) Molinero, V.; Moore, E. B. Water Modeled as an Intermediate Element between Carbon and Silicon. *J. Phys. Chem. B* **2009**, *113*, 4008−4016.
(33) Praprotnik, M.; Site, L. D.; Kremer, K. Adaptive Resolution Molecular-Dynamics Simulation: Changing the Degrees of Freedom on the Fly. *J. Chem. Phys.* **2005**, *123*, 224106.
(34) Ensing, B.; Nielsen, S. O.; Moore, P. B.; Klein, M. L.; Parrinello, M. Energy Conservation in Adaptive Hybrid Atomistic/Coarse-Grain Molecular Dynamics. *J. Chem. Theory Comput.* **2007**, *3*, 1100−1105.
(35) Nielsen, S. O.; Bulo, R. E.; Moore, P. B.; Ensing, B. Recent Progress in Adaptive Multiscale Molecular Dynamics Simulations of Soft Matter. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12401−12414.
(36) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308−313.
(37) Neumann, M. Dipole Moment Fluctuation Formulas in Computer Simulations of Polar Systems. *Mol. Phys.* **1983**, *50*, 841−858.
(38) Alejandre, J.; Tildesley, D. J.; Chapela, G. A. Molecular Dynamics Simulation of the Orthobaric Densities and Surface Tension of Water. *J. Chem. Phys.* **1995**, *102*, 4574.
(39) Weast, R. *Handbook of Chemistry and Physics*, 53rd ed.; Chemical Rubber Publishing, 1972.
(40) Holz, M.; Heil, S.; Sacco, A. Temperature-dependent Self-diffusion Coefficients of Water and Six Selected Molecular Liquids for Calibration in Accurate H-1 NMR PFG Measurements. *Phys. Chem. Chem. Phys.* **2000**, *2*, 4740−4742.
(41) Kaye, G.; Laby, T. *Tables of Physical and Chemical Constants*; Longman: Essex, U.K., 1995.
(42) Vargaftik, N. B.; Volkov, B. N.; Voljak, L. D. International Tables of the Surface Tension of Water. *J. Phys. Chem. Ref. Data* **1983**, *12*, 817−820.
(43) Tironi, I. G.; Van Gunsteren, W. F. A Molecular Dynamics Simulation Study of Chloroform. *Mol. Phys.* **1994**, *83*, 381−403.
(44) Motakabbir, K. A.; Berkowitz, M. Isothermal Compressibility of SPC/E Water. *J. Phys. Chem.* **1990**, *94*, 8359−8362.
(45) Pérez, J. F.; Hadad, C. Z.; Restrepo, A. Structural Studies of the Water Tetramer. *Int. J. Quantum Chem.* **2008**, *108*, 1653−1659.
(46) Yang, M.; Senet, P.; Van Alsenoy, C. DFT Study of Polarizabilities and Dipole Moments of Water Clusters. *Int. J. Quantum Chem.* **2005**, *101*, 535−542.
(47) Rusnak, A. J.; Pinnick, E. R.; Calderon, C. E.; Wang, F. Static Dielectric Constants and Molecular Dipole Distributions of Liquid Water and Ice-Ih Investigated by the PAW-PBE Exchange-correlation Functional. *J. Chem. Phys.* **2012**, *137*, 034510.
(48) Ghanty, T. K.; Ghosh, S. K. Polarizability of Water Clusters: An ab Initio Investigation. *J. Chem. Phys.* **2003**, *118*, 8547−8550.
(49) Reddy, M.; Berkowitz, M. The Dielectric Constant of SPC/E Water. *Chem. Phys. Lett.* **1989**, *155*, 173−176.
(50) Kell, G. S. Density, Thermal Expansivity, and Compressibility of Liquid Water From 0° to 150°. Correlations and Tables for Atmospheric Pressure and Saturation Reviewed and Expressed on 1968 Temperature Scale. *J. Chem. Eng. Data* **1975**, *20*, 97−105.
(51) Kell, G. S. Precise Representation of Volume Properties of Water at One Atmosphere. *J. Chem. Eng. Data* **1967**, *12*, 66−69.
(52) Lide, D., Ed. *Handbook of Chemistry and Physics*; CRC Press: Boca Raton, FL, 1992.
(53) Fanelli, L. Determinacy, Indeterminacy and Dynamic Misspecification in Linear Rational Expectations Models. *Journal of Econometrics* **2012**, *170*, 153−163.
(54) Harvey, S.; Tan, R.; Cheatham, T. The Flying Ice Cube: Velocity Rescaling in Molecular Dynamics Leads to Violation of Energy Equipartition. *J. Comput. Chem.* **1998**, *19*, 726−740.
(55) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(56) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(57) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(58) Evans, D. J.; Holian, B. L. The Nose-Hoover Thermostat. *J. Chem. Phys.* **1985**, *83*, 4069−4074.
(59) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.