Systematic optimization of water models using liquid/vapor surface tension data

Yudong Qiu,[†] Paul S. Nerenberg,[‡] Teresa Head-Gordon,[¶] and Lee-Ping Wang[*,†]

†*Chemistry Department, The University of California, Davis, Davis, California 95616*

‡*Departments of Physics & Astronomy and Biological Sciences, California State University, Los Angeles, California 90032*

¶*Pitzer Theory Center and Departments of Chemistry, Bioengineering and Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, USA*

E-mail: leeping@ucdavis.edu

Abstract

In this work we investigate whether experimental surface tension measurements, which are less sensitive to quantum and self-polarization corrections, are able to replace the usual reliance on the heat of vaporization as experimental reference data for fitting force field models of molecular liquids. To test this hypothesis we develop the fitting protocol necessary to utilize surface tension measurements in the ForceBalance optimization procedure in order to determine revised parameters for both three-point and four-point water models, TIP3P-ST and TIP4P-ST. We find that the incorporation of surface tension in the fit results in a rigid three-point model that reproduces the correct temperature of maximum density of water for the first time, but also leads to over-structuring of the liquid and less accurate transport properties. The rigid four-point TIP4P-ST model is highly accurate for a broad range of thermodynamic and kinetic properties, with similar performance compared to recently developed four-point water models. The results show surface tension to be a useful fitting property in

general, especially when self-polarization corrections or nuclear quantum corrections are not readily available for correcting the heat of vaporization as is the case for other molecular liquids.

1 Introduction

Empirical force fields are widely used in molecular simulation studies, mostly when chemical reactivity is not operative.[1,2] Due to the availability of plentiful experimental and first-principles quantum mechanical data, water is a popular testing application for developing new force field models and new approaches to developing models.[3–7] Among the most widely used physics-based water models today are the TIP3P and TIP4P models introduced in the 1980s,[8] which employ well-established functional forms dating back to the 1930s[9] consisting of a rigid molecular geometry, fixed atomic partial charges, and Lennard-Jones interactions. In recent years, several new water models have been published that are reparameterizations of the rigid TIP3P and TIP4P models; examples of this include TIP4P-Ew,[10] TIP4P/2005,[11] TIP4P/ϵ,[12] TIP3P-FB,[13] TIP4P-FB,[13] OPC,[14] and OPC3.[15] These new models more accurately reproduce a number of experimentally measured physical properties of water without increasing the computational cost of the simulation. Perhaps more importantly, some of the more recent models were developed using automated parameter optimization tools such as ForceBalance,[13] making possible the systematic optimization of force fields for a wide range of molecular liquids given the availability of experimental data.

One important caveat in force field development is that the fundamental approximations in the functional form could make it impossible or inappropriate to reproduce certain physical properties. For example, it is well-known that classical models cannot reproduce the heat capacity due to the importance of nuclear quantum effects in high-frequency intramolecular and intermolecular degrees of freedom. Another relevant example is that all known simple three-point water models, i.e. those that use fixed partial charges, fail to reproduce the density anomaly at 4 °C even when they are fit to data for the temperature dependence in the density. In modeling the heat of vaporization, it is often necessary to apply post-hoc corrections to account for condensed phase polarization as well as nuclear quantum effects. In the development of the SPC/E water model,[16] the authors argued that because the atomic partial charges include the effects of mean-field polarization in the condensed phase, there

exists an implicit energetic cost of polarization that should increase the potential energy of each water molecule in the liquid simulation. Thus, in order to fit the heat of vaporization, a polarization correction of +5.22 kJ mol^{-1} was added to the simulated potential energy of each molecule in the liquid. The development of the TIP4P-Ew model included a polarization correction and a simple quantum correction derived from making harmonic approximations to the high-frequency vibrational modes of liquid water, as well as some more minor nonideality corrections.[10] In summary, these corrections increase the complexity of the parameterization procedure, require additional experimental data for the compounds being parameterized, and introduce uncertainty because they only approximately model the effects they are supposed to correct. Moreover, classical force fields are not uniform in how or whether the corrections are applied; for example, the OPLS-AA force field for organic liquids was developed by fitting Monte Carlo simulated density and heat of vaporization to experiments without corrections.[17] For these reasons, it is desirable to use physical properties that require fewer post-hoc corrections when fitting parameters to improve agreement with experiment.

The surface tension of the liquid/vapor interface originates from the energetic preference for molecules to be located in the bulk liquid compared to at the surface, thus it is a property that characterizes the cohesive forces in the liquid, similar to the heat of vaporization. Furthermore, because the surface tension calculation does not involve taking any energetic differences with molecules fully in the gas phase, we hypothesize that it can substitute for the heat of vaporization in the force field parameterization without requiring corrections for polarization or nuclear quantum effects. Indeed, the nuclear quantum effects are smaller for the surface tension compared to heat of vaporization, as the heat of vaporization increases by 2.1% from H_2O (40.657 kJ mol^{-1}) to D_2O (41.521 kJ mol^{-1}),[18] while the surface tension only changes by 0.15 % between light and heavy water (from 71.98 mJ m^{-2} to 71.87 mJ m^{-2}).[19] This is further supported by established protocols for calculating surface tension in MD simulations,[20] in which all post-hoc corrections are intended to account only for long-range

dispersion interactions.

Because surface tension data is widely available and can be easily measured,[21] there exists an opportunity to create more accurate models of a wide range of liquids by using surface tension as a training physical property instead of heat of vaporization. Nielsen[22] first demonstrated the use of surface tension as a fitting property to parameterize a coarse-grained mixture of hydrocarbons. Salas and Alejandré[23] developed a procedure that scales the charge and Lennard-Jones parameters to reproduce the dielectric constant, surface tension, and density in stepwise fashion, and applied the approach to build all-atom and coarse grained models for four molecular liquids including methanol and ionic liquids. Martínez-Jiménez and Saint-Martin applied a similar procedure to refine a coarse-grained potential for methanol that included an off-center charge site.[24] As for water, many popular models such as TIP3P,[8] SPC/E,[16] and TIP4P-Ew[10] utilize surface tension as a validation test in the sense that models fitted to some properties should accurately predict other known properties. However, to the best of our knowledge, no water model has been developed by adjusting the parameters to reproduce the surface tension directly;[25] in particular, water was not one of the four liquids studied in Ref. 23. The development of such a water model is needed for testing the hypothesis that surface tension can effectively substitute for the heat of vaporization in the force field parameterization. Moreover, the utility of surface tension as reference data for force field devleopment creates a need for automated tools and procedures that can effectively use this data to generate models for molecular liquids in systematic fashion.

In this article, we describe how the fitting of surface tension is enabled by extending the ForceBalance optimization method to include surface tension as a fitting target. To demonstrate feasibility, we develop and characterize two new water models, namely TIP3P-ST and TIP4P-ST (here ST stands for "surface tension"), where the surface tension property replaces the heat of vaporization in the training data. The resulting TIP4P-ST model confirms our hypothesis by exhibiting high accuracy for thermodynamic properties across a range of temperatures, for both training and validation data that include the density, dielectric constant,

isothermal compressibility, thermal expansion coefficient, and self-diffusion coefficient. The TIP3P-ST model offers moderate agreement with the full range of data, but reproduces the correct temperature of maximum density of water for the first time in models of this form. In both cases, the optimization procedure is able to match the experimental surface tension within 10%, which is highly accurate in the context of existing water models. We conclude that when placed in the context of other fixed charge models with rigid geometries, that the four-point models will yield accurate predictions in studies involving liquid/vapor interfaces and extremes of temperature and pressure, whereas the functional form of three-point rigid models is too limited to simultaneously describe the temperature dependence of density and other structural and kinetic properties with equivalent accuracy across broad temperature ranges. The model parameterization approach of picking alternative properties such as surface tension that require minimal post-hoc corrections is also expected to be broadly useful in developing the next generation of force fields for other molecular liquids and small molecules where such corrections are not easily obtainable.

2 Methods

2.1 Parameterization

The TIP3P-ST three-point model was optimized using the same functional form as TIP3P. Five individual parameters were optimized: two weight parameters w_{O} and w_{H} that control the molecular geometry, the charge on hydrogen q_{H}, and the Lennard-Jones parameters for oxygen σ_{LJ} and ϵ_{LJ}. In order to optimize the geometry of the rigid water model, all interactions are defined in terms of off-center interaction sites (virtual sites) whose positions $\mathbf{r}_{\mathrm{O}}, \mathbf{r}_{\mathrm{H1}}, \mathbf{r}_{\mathrm{H2}}$ are defined in terms of the rigid TIP3P molecular geometry $\mathbf{r}'_{\mathrm{O}}, \mathbf{r}'_{\mathrm{H1}}, \mathbf{r}'_{\mathrm{H2}}$ and the

weight parameters w_{O} and w_{H} as:

$$\begin{aligned}
\mathbf{r}_{\mathrm{O}} &= w_{\mathrm{O}} \cdot \mathbf{r}'_{\mathrm{O}} + \left(\frac{1 - w_{\mathrm{O}}}{2}\right) \cdot (\mathbf{r}'_{\mathrm{H1}} + \mathbf{r}'_{\mathrm{H2}}) \\
\mathbf{r}_{\mathrm{H1}} &= (1 - w_{\mathrm{H}}) \cdot \mathbf{r}'_{\mathrm{O}} + w_{\mathrm{H}} \cdot (\mathbf{r}'_{\mathrm{H1}}) \\
\mathbf{r}_{\mathrm{H2}} &= (1 - w_{\mathrm{H}}) \cdot \mathbf{r}'_{\mathrm{O}} + w_{\mathrm{H}} \cdot (\mathbf{r}'_{\mathrm{H2}})
\end{aligned} \tag{1}$$

As a consequence of parameterizing the geometry of the interaction sites, the interaction sites are distinct from the dynamical degrees of freedom during the parameter optimization. Afterward, the final three-point model is defined by setting r_{OH} and Θ_{HOH} equal to the distance and angle formed by the interaction sites, which restores the model to having only three sites per molecule with identical thermodynamic properties to the optimized model. The same procedure was previously used to optimize the TIP3P-FB three-point model.[13]

The TIP4P-ST four-point model used the TIP4P-Ew functional form and four parameters were optimized: the virtual-site position that carries the negative charge, the hydrogen charge q_{H}, and the Lennard-Jones parameters for oxygen σ_{LJ} and ϵ_{LJ}. Starting values of the parameters are given in Table 1.

Reference data for the parameterization obtained from experimental thermodynamic properties are shown in Table 1. The objective function computed in the parameterization has the formula:

$$L_{\mathrm{tot}}(\mathbf{k}) = \sum_{T \in \mathrm{targets}} w_T L_T(\mathbf{k}) + w_{\mathrm{reg}} |\mathbf{k}|^2 \tag{2}$$

where the total objective function L_{tot} depends on the optimization variables or "mathematical parameters" $\mathbf{k}$, and is equal to the sum of contributions from the parameterization *targets* $L_T(\mathbf{k})$ weighted by w_T, plus a regularization term. A parameterization target consists of a collection of weighted least-squares residuals between the force field predictions and a training data set. In this study, all of the liquid thermodynamic properties including surface tension are included in a single target with a weight of 1.0.

In general L_{tot} may contain many least-squares residuals, thus the objective function

is organized in hierarchical fashion with each target containing ≥ 1 properties, and each property containing ≥ 1 data points. The objective function for a target is a weighted sum of contributions for one or more individual properties:

$$L_T(\mathbf{k}) = \sum_{j \in \text{properties}} w_j^{(T)} L_j^{(T)}(\mathbf{k}) \tag{3}$$

where $w_j^{(T)}$ and $L_j^{(T)}(\mathbf{k})$ represent the weight for property j and the contribution from each property within the target T respectively. In this study, $w_j^{(T)}$ was set to 1.0 for all properties being fitted. $L_j^{(T)}(\mathbf{k})$ is given by a weighted and normalized sum over individual data points:

$$L_j^{(T)}(\mathbf{k}) = \frac{1}{\left(d_j^{(T)}\right)^2} \frac{\sum_{p \in \text{points}} w_{jp}^{(T)} \left| y_{jp}^{(T)}(\mathbf{k}) - y_{jp,\text{ref}}^{(T)} \right|^2}{\sum_{p \in \text{points}} w_{jp}^{(T)}} \tag{4}$$

where $y_{jp}^{(T)}$ and $y_{jp,\text{ref}}^{(T)}$ are respectively the simulated and reference data point for property j and point p within target T. $d_j^{(T)}$ is the scaling factor used to normalize and remove physical units for property j, with values given in Table 1.

In order to evaluate $y_{jp}(\mathbf{k})$, the mathematical parameters are first mapped to a set of "physical parameters" $\mathbf{K}$ by a linear transformation:

$$K_\lambda = K_\lambda^{(0)} + p_\lambda k_\lambda \tag{5}$$

where λ is the index for the force field parameter being optimized, $K_\lambda^{(0)}$ represents the original parameter value, and p_λ is the *prior width* that represents the expected magnitude of variation of the parameter over the course of the optimization. Table 1 shows the values of p_λ for different parameter types. In the case of TIP3P-ST, all values of p_λ were set equal to $K_\lambda^{(0)}$, which effectively makes k_λ into a scaling of the original parameter. In cases where parameters need to satisfy functional relationships such as a constraint on the total charge of a residue or molecule, the parameters used directly in the energy expression may be specified

as functions of $\mathbf{K}$; the charge on oxygen was defined in this way as $q_{\mathrm{O}} = -2q_{\mathrm{H}}$.

The regularization term for preventing overfitting may be expressed in terms of the physical parameters as:

$$w_{\mathrm{reg}} \left|\mathbf{k}\right|^2 = w_{\mathrm{reg}} \sum_{\lambda \in \mathrm{params}} k_\lambda^2 = w_{\mathrm{reg}} \sum_{\lambda \in \mathrm{params}} \left| \frac{K_\lambda - K_\lambda^{(0)}}{p_\lambda} \right|^2 \tag{6}$$

Thus, increasing the prior width p_λ allows the physical parameter K_λ to have greater variations for the same contribution to the penalty function. Although the optimization result depends on the choice of p_λ, in practice these values may be varied within a factor of 2 without incurring significant changes in the performance of the optimized model.[13]

Table 1: Top: Data references for parameterization and validation of water model. The first five properties comprise the training data set whereas the last four properties were used as validation. All experimental data values used in the parameterization are listed in Supporting Information Table S1. Middle: The starting values and prior widths for the parameterization of TIP3P-ST. Bottom: The starting values and prior widths for the parameterization of TIP4P-ST.

Reference Property	Scaling Factor	No. Data Points
Density ρ	2 kg m^{-3}	39
Thermal Expansion Coefficient α	10^{-4} K^{-1}	39
Isothermal Compressibility κ_{T}	5×10^{-5} bar^{-1}	39
Dielectric Constant $\epsilon(0)$	2	39
Surface Tension γ	10^{-3} J m^{-2}	26
Enthalpy of Vaporization ΔH_{vap}		31
Isobaric Heat Capacity c_{P}	.	39
Self-diffusion Coefficient D_0		16
Shear Viscosity η		16
TIP3P-ST Parameter	**Initial Value**	**Prior Width**
w_{O}	0.999	0.999
w_{H}	0.999	0.999
q_{H} (e)	0.4238	0.4238
σ_{LJ} (Å)	3.16557	3.16557
ϵ_{LJ} (kJ mol^{-1})	0.650194	0.650194
TIP4P-ST Parameter	**Initial Value**	**Prior Width**
w_{O}	0.78664	0.999
q_{H} (e)	0.52422	0.4238
σ_{LJ} (Å)	3.16435	3.16557
ϵ_{LJ} (kJ mol^{-1})	0.680946	0.650194

Five physical properties were included in the parameterization. The evaluation of density ρ, thermal expansion coefficient α, isothermal compressibility κ_{T}, and dielectric constant $\epsilon(0)$ followed previous simulation procedures for the parameterization of the TIP3P-FB and TIP4P-FB models. The simulation of these bulk properties consisted of 216 water molecules in a periodic cubic box in the isothermal-isobaric NPT ensemble. A Langevin integrator

with a time step of 1.0 fs and collision frequency of 1.0 ps^{-1} was used for integrating the equations of motion with added temperature control, and a Monte Carlo barostat was added with an attempt interval of 25 MD steps. Simulated temperature values ranged from 249 K to 373 K and pressures ranged from 1.0 atm to 2000 bar. The particle mesh Ewald (PME) method[26] is used to treat the electrostatic interactions with a real-space cutoff of 9 Å, and the same cutoff was used for Lennard-Jones (LJ) interactions. The system was first equilibrated for 1 ns, followed by an 8 ns production run. Thermodynamic averages were obtained by averaging over trajectory frames spaced 0.1 ps apart for a total of 80,000 samples.

The surface tension γ was evaluated separately using a simulation setup consisting of a water film in the NVT ensemble with two liquid-vapor interfaces. We used a tetragonal simulation cell with dimensions 3 nm $\times$ 3 nm $\times$ 10 nm containing a 3 nm thick water layer normal to the z-dimension with 1024 water molecules in total. Figure S1 shows that this setup preserves the stable geometry of the water film, which is an important consideration in these types of simulations.[27] A real-space cutoff distance of 15 Å was chosen for nonbonded interactions because the surface tension calculations required accounting for Lennard-Jones interactions at large distances. The other simulation parameters matched the NPT simulations. To evaluate the surface tension for a trajectory frame, we adopted the test-area method[28] with the formula

$$\gamma = \lim_{\Delta S \to 0} \frac{-1}{2\beta \Delta S} \left[\ln \left\langle \exp\left(-\beta \Delta E^{+}\right) \right\rangle - \ln \left\langle \exp\left(-\beta \Delta E^{-}\right) \right\rangle \right] \tag{7}$$

where E is the potential energy, $\beta \equiv \frac{1}{k_B T}$ the inverse temperature, k_B Boltzmann's constant, and T the temperature. ΔE^{+} and ΔE^{-} are calculated by making two perturbations to the surface area $S \equiv L_x L_y$, by $\Delta S = \pm 0.0005 S$ as suggested in Ref. 28. In each perturbation, the x and y dimensions of the simulation box are scaled proportionally, while the z dimension is scaled in the opposite direction to keep the total volume constant. The scaling operation is also applied to the molecular centroids, and the molecules are rigidly translated without

modifying the molecular geometry. The ensemble averages in the formula are evaluated as the arithmetic average over trajectory frames.

The procedure for evaluating surface tension was implemented into the ForceBalance automated parameter optimization software, which uses the OpenMM library[29,30] to carry out the NVT and NPT MD simulations, thus allowing the entire optimization procedure to be carried out in a single reproducible calculation. Although thermodynamic fluctuation formulas were used to estimate the parametric derivatives of thermodynamic properties simulated in the NPT ensemble in previous parameterization of TIP3P-FB and TIP4P-FB models, we found that in the case of surface tension, the parametric derivatives estimated in this way contained such high levels of statistical noise that it was more efficient to calculate parametric derivatives numerically via a 3-point finite difference formula, which involved running two separate simulations for each parameter being optimized. Details of the error analysis are described in Section 3.3.

2.2 Validation

Among the properties for validation, the enthalpy of vaporization ΔH_{vap} and isobaric heat capacity c_{P} were obtained from analysis of the NPT simulation trajectories described above. ΔH_{vap} is calculated as:

$$\begin{aligned}\Delta H_{\mathrm{vap}} = \langle H\rangle_g - \langle H\rangle_l =& (\langle E_{\mathrm{pot}} + E_{\mathrm{kin}}\rangle_g + k_{\mathrm{B}}T) - (\langle E_{\mathrm{pot}} + E_{\mathrm{kin}}\rangle_l + P\langle V\rangle_l + E_{\mathrm{sp}}) \\ &+ C_{vib} + C_{ni}\end{aligned} \tag{8}$$

where $\langle\cdot\rangle_{\{g,l\}}$ indicate ensemble averages in the gas and liquid phases respectively. The gas phase potential energy $\langle E_{\mathrm{pot}}\rangle_g$ is exactly zero for a rigid water model, and the $E_{\mathrm{kin}\{g,l\}}$ terms are analytically equal in classical mechanics and cancel each other out. E_{sp} is the self-polarization correction that represents the potential energy increase of molecules in the

liquid due to polarization, and is computed as:

$$E_{\mathrm{sp}} = \frac{(\mu - \mu_0)^2}{2\alpha} \tag{9}$$

where μ is the molecular dipole moment of the water model, $\mu_0 = 1.855D$ is the gas-phase dipole moment of water, and $\alpha = 1.470$ Å^3 is the isotropic molecular polarizability of water. The quantum vibrational and nonideality corrections C_{vib} and C_{ni} are computed following Ref. 10; their values are given in Supporting Table S1. The remaining terms $\langle E_{\mathrm{pot}} \rangle_l$ and $\langle V \rangle_l$ are computed from simulations. The isobaric heat capacity was calculated using a fluctuation formula as $c_p = \langle H^2 \rangle_l - \langle H \rangle_l^2 + C'$ where C' is a quantum vibrational correction also listed in Supporting Table S1. These validation properties were evaluated automatically from the NPT simulations in the course of parameter optimization but excluded from the objective function by setting their weights equal to 0 in ForceBalance.

To evaluate the self-diffusion coefficient D_0, we first carried out a 1 ns equilibration and 1 ns production simulation in the NPT ensemble, and saved 100 trajectory frames containing position and velocity information with 10 ps time resolution as initial conditions for energy-conserving simulations. From each simulation snapshot, an energy-conserving simulation was propagated for 10 ps using the Verlet integrator and 1.0 fs time step to generate a trajectory of 100 frames with a 0.1 ps time interval.

The self-diffusion coefficient D_0 is then estimated as:

$$D_0 = \frac{1}{6N} \lim_{t \to \infty} \frac{\left\langle |\mathbf{r}_{t_0+t} - \mathbf{r}_{t_0}|^2 \right\rangle}{t} \tag{10}$$

The numerator on the RHS is the mean square displacement of the coordinates $\mathbf{r}$ from the initial conditions after time t and ensemble-averaged over 100 initial conditions. N is the total number of atoms.

The diffusion coefficient contains a known dependence on the size of the periodic box.[31] To estimate the intrinsic diffusion coefficient at infinite box sizes, the diffusion coefficient

calculation is repeated for six box sizes, 25, 30, 40, 50, 60, and 90 Å. The final self-diffusion coefficient for each temperature point is then computed as an extrapolation of the inverse box size towards infinity. The shear viscosity η is also obtained from the slope of the linear fit of self-diffusion coefficient against the inverse box size L:[31]

$$D_{\mathrm{PBC}} = D_0 - \frac{k_B T \xi}{6\pi\eta L} \tag{11}$$

In order to compute the hydration (self-solvation) free energy G_{hyd}, we ran a series of 21 simulations of alchemical intermediates where the interactions between solute (i.e. 1 chosen water molecule) and solvent were gradually decoupled. The electrostatic interactions were decoupled by scaling the Coulomb interactions in 11 steps corresponding to (Coulomb, LJ) coupling parameters of (1.0, 1.0), (0.9, 1.0) ... (0.0, 1.0). This was followed by decoupling the LJ parameters in 10 additional steps as (0.0, 1.0) ... (0.0, 0.0) where a soft-core potential was used to improve thermodynamic overlap.[32] Each of these simulations consisted of a cubic water box of 3 nm in each dimension containing 887 molecules in the NPT ensemble using a Langevin integrator with a 1.0 fs time step, 298.15 K temperature and 1.0 ps^{-}1 friction coefficient, and a Monte Carlo barostat with 1.0 atm pressure and an attempt interval of 25 steps. The simulations were equilibrated for 1 ns followed by a 10 ns production run, saving one frame per 1 ps for a total of 10,000 frames. After the simulations were completed, multistate Bennett acceptance ratio (MBAR) analysis was carried out to estimate the free energy difference between the fully interacting and fully decoupled states.[33] MBAR analysis requires computing the ratio of Boltzmann factors between each pair of alchemical intermediate Hamiltonians for each sampled frame. We constructed a dimensionless energy tensor $\mathbf{U}$ of shape [21 x 21 x 10000], where U_{ijk} corresponds to trajectory frame k of alchemical intermediate j evaluated using the Hamiltonian of intermediate i divided by $k_B T$. This quantity was used as input to the *pymbar* software package, which implements MBAR and provides the estimates of the free energy differences as output. Our method reached good

agreement with the literature[34] for available models, and we found no dependence on the choice of non-bonded cutoff distance and simulation box size. (SI Section 3)

3 Results and Discussion

3.1 Optimized force field Parameters

Table 2: Optimized force field parameters for TIP3P-ST and TIP4P-ST compared to existing water models

	r_{OH} (Å)	Θ_{HOH} (deg)	ω_{v} (Å)	$q_{\mathrm{H}}(e)$	σ_{LJ}	ϵ_{LJ}
TIP3P[8]	0.9572	104.52	-	0.41700	3.15075	0.63597
TIP4P-Ew[10]	0.9572	104.52	0.1250	0.52422	3.16435	0.68095
TIP3P-FB[13]	1.0118	108.15	-	0.42422	3.17796	0.65214
TIP4P-FB[13]	0.9572	104.52	0.1052	0.52587	3.16555	0.74928
TIP3P-ST	1.0230	108.11	-	0.42556	3.19257	0.60190
TIP4P-ST	0.9572	104.52	0.0989	0.52172	3.16610	0.74030

ω_{v}: Oxygen Virtual-site displacement.

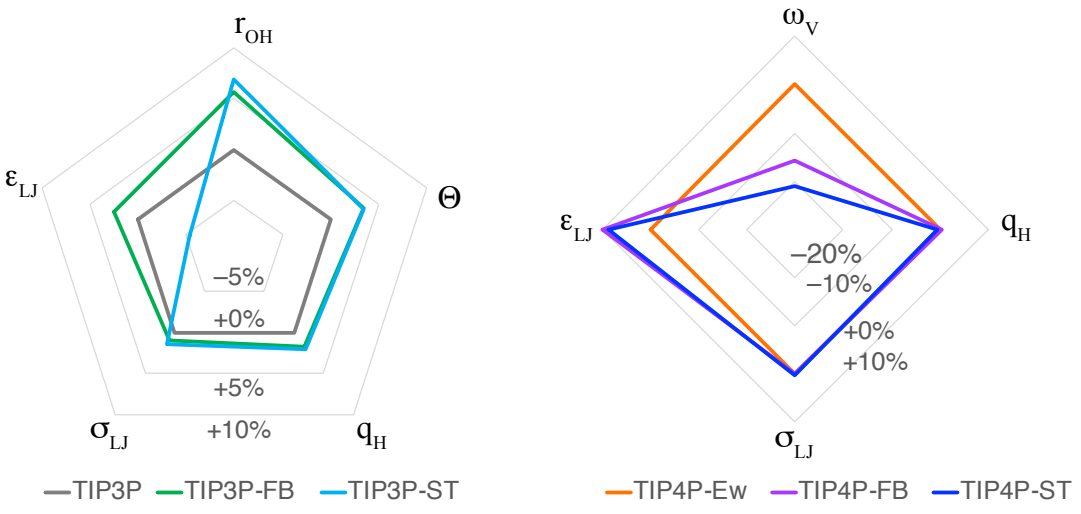


Figure 1: Comparison of 3-point (left) and 4-point (right) model parameters as percentage differences with respect to TIP3P and TIP4P-Ew respectively.

The optimized force field parameters for TIP3P-ST and TIP4P-ST are listed in Table 2. To assist with model comparison, Figure 1 displays the parameters of each model in terms of percentage differences from TIP3P and TIP4P-Ew for 3-point and 4-point models respectively. In both 3-point and 4-point models, the σ_{LJ} parameter has the least variation among the three models, which may be expected given its important role in determining the excluded volume, and consequently the liquid density. TIP3P-FB and TIP3P-ST feature larger values of r_{OH}, Θ_{HOH} and q_{H} compared to TIP3P, consistent with increasing the hydrogen bonding strength by decreasing the intermolecular O—H distance, increasing the Coulomb interaction strength, and bringing the bond angle closer to the ideal tetrahedral angle. The parameter with the largest variation is ϵ_{LJ} where TIP3P-ST has a smaller value than the other three models. Among the other four parameters, the TIP3P-ST and TIP3P-FB parameter are closer, though we note the former has a slightly higher value of r_{OH}. This indicates TIP3P-ST has stronger directional character in its intermolecular interactions, and could be further understood by examining the thermodynamic properties. On the other hand, the TIP4P-FB and TIP4P-ST are highly similar in the q_{H}, σ_{LJ} and ϵ_{LJ} parameters, and both models place the virtual site closer to the O atom than TIP4P-Ew. The value of ω_{v} in TIP4P-ST is smaller than TIP4P-FB, but the accuracy of these two models are highly similar.

3.2 Thermodynamic Properties

The comparison of thermodynamic properties at room temperature and standard pressure for six models vs. experiment are listed in Table 3. The temperature dependence of fitted thermodynamic properties are plotted in Figure 2, while the validation properties are plotted in Figure 3. The three-point TIP3P-ST model accurately reproduces experimental thermodynamic properties with a level of accuracy that well exceeds the widely adopted TIP3P model. Notably, TIP3P-ST correctly reproduces the temperature of maximum density, which could not be accomplished by the other rigid three-point water models in our

comparisons. The closer agreement with the experimental density curve in TIP3P-ST is a significant difference from TIP3P-FB, and possibly caused by stronger directional interactions resulting from reduced ϵ_{LJ} and increased r_{OH}. TIP3P-ST reproduces the experimental thermal expansion coefficient and surface tension more closely than TIP3P-FB, but also has a lower self-diffusion coefficient and higher viscosity compared to experiment. The four-point TIP4P-ST model agrees within 5% of the experimental value for most properties, and the fitted surface tension is surprisingly close to the TIP4P-FB model which did not include surface tension in the fitting targets. Generally speaking, the performance of TIP4P-ST is nearly identical to TIP4P-FB, except that TIP4P-ST achieves an even closer fit to the density amounting to $< 0.1\%$ deviations across the whole temperature range.

Table 3: Comparison of water model performance at 298.15 K, 1.0 atm. Numbers in parentheses represent one standard error in the least significant digit. The standard error of the density, ΔH_{vap} and E_{pot} are smaller than the least significant digit provided. Error estimates were not computed for γ_∞ and η.

Property	Experiment[a]	TIP3P	TIP4P-Ew	TIP3P-FB	TIP4P-FB	TIP3P-ST	TIP4P-ST
ρ / g cm^{-3}	0.997	0.985	0.994	0.994	0.996	0.996	0.997
α / 10^{-4} K^{-1}	2.572	9.0(2)	3.2(2)	4.0(2)	2.3(2)	2.4(2)	2.4(2)
κ_{T} / 10^{-6} bar^{-1}	45.247	57.4(4)	47.6(3)	44.3(3)	44.8(3)	39.7(3)	45.2(3)
$\epsilon(0)$	78.409	97.2(8)	64.1(7)	81(1)	76(1)	81(1)	82(1)
γ / mJ m^{-2}	71.990	49.6(4)	61.8(5)	64.0(6)	67.5(6)	67.9(7)	67.7(6)
γ_∞ / mJ m^{-2}	71.990	53.9	66.9	67.8	73.1	71.3	73.1
ΔH_{vap} / kcal mol^{-1}	10.513	8.92	10.57	10.74	10.84	11.33	10.84
c_{P} / cal mol^{-1} K^{-1}	18.002	16.9(2)	19.3(2)	18.9(2)	19.2(2)	19.9(2)	18.9(2)
D_0 / 10^{-5} cm^2 s^{-1}	2.29	6.10(9)	2.78(6)	2.42(6)	2.36(9)	1.48(4)	2.33(4)
η / mPa s	0.896	0.43	0.90	0.96	0.95	1.44	0.81
TMD / K	277	(182)	273	261	281	277	277
ΔG_{hyd} / kcal mol^{-1}	−6.33	−4.82(1)	−5.82(1)	−5.88(1)	−5.96(1)	−6.17(1)	−5.93(1)
$\langle E_{\mathrm{pot}} \rangle_l$ / kcal mol^{-1}		−9.57	−11.10	−11.77	−11.92	−12.57	−12.00

a. Experimental data source: surface tension;[35] hydration free energy;[34] all others.[36] The TMD for TIP3P model was from reference.[13] ρ: Density; α: Thermal expansion coefficient; κ_{T}: Isothermal compressibility; $\epsilon(0)$: Dielectric constant; γ: Liquid/vapor surface tension; ΔH_{vap}: Enthalpy of vaporization; c_{P}: Isobaric heat capacity; D_0: Self-diffusion Coefficient; η: shear viscosity; TMD: Temperature of maximum density; ΔG_{hyd}: Hydration (self-solvation) free energy; $\langle E_{\mathrm{pot}} \rangle$: Average total potential energy per water molecule in simulation.

The validation properties provide insights into the predictive power of models fitted to

surface tension. The TIP3P-ST model yields a higher ΔH_{vap} than experiment, and also has a relatively large self-polarization correction of 7.38 kJ mol^{-1} indicating a large dipole moment. The TIP4P-ST model also predicts ΔH_{vap} slightly higher than the experimental value with a self-polarization correction of 7.03 kJ mol^{-1}. The correction for nuclear quantum effects is relatively small at -0.27 kJ mol^{-1} at 298 K. Notably, the corrected ΔH_{vap} is almost identical to TIP4P-FB, again indicating they are close in terms of performance. The self-diffusion coefficient is another property where TIP3P-ST is different from the other models included in our comparison. The lower self-diffusion coefficient indicates a slightly more structured liquid, with stronger hydrogen bonds needed to reproduce the surface tension. This behavior is also reflected in the radial distribution plot, where the TIP3P-ST curve shows a higher first peak and lower first trough.

There is a notable trend in the rigid three-point water models where TIP3P-ST has the highest surface tension, temperature of maximum density and heat of vaporization, the most highly structured O—O RDF, and the lowest self-diffusion coefficient. All of these properties correspond to stronger cohesion and a highly structured hydrogen-bonding network. TIP3P on the other hand has the lowest surface tension, temperature of maximum density and heat of vaporization, the least structured O—O RDF, and the highest self-diffusion coefficient, whereas TIP3P-FB is intermediate between TIP3P-ST and TIP3P for all of these properties. The physically motivated correspondence between all of these properties, coupled with the observation that none of the rigid three-point models can reproduce all of the experimental properties equally accurately across the whole temperature range, reveals a potential limitation of the functional form of rigid three-point rigid water models. Despite these limitations, the high accuracy of TIP3P-FB for all tested thermodynamic, structural and kinetic properties except for the temperature dependence of the density (Table 3) indicates that it is suitable for simulating biomolecular systems near ambient conditions, especially in applications that benefit from the lower computational cost of three-point models.

We additionally found that TIP3P-FB and TIP3P-ST are both able to fit the dielectric

constants accurately independent of the trends discussed above. More generally, the dielectric constant appears relatively "orthogonal" to the other thermodynamic properties and can be accurately fitted if the geometric parameters of the three-point model are optimized. The four-point models have one fewer parameter because the molecular geometry is not being optimized, but more accurate results are obtained for the validation properties; in particular, the diffusion coefficient of TIP4P-ST agrees closely with experiment, and the O-O radial distribution function of TIP4P-ST agrees with experiment at a similar level as the TIP4P-Ew, TIP3P-FB and TIP4P-FB models. The improved ability of four-point models to reproduce experimental properties has previously been attributed to the model's ability to predict the correct quadrupole moment of the water molecule.

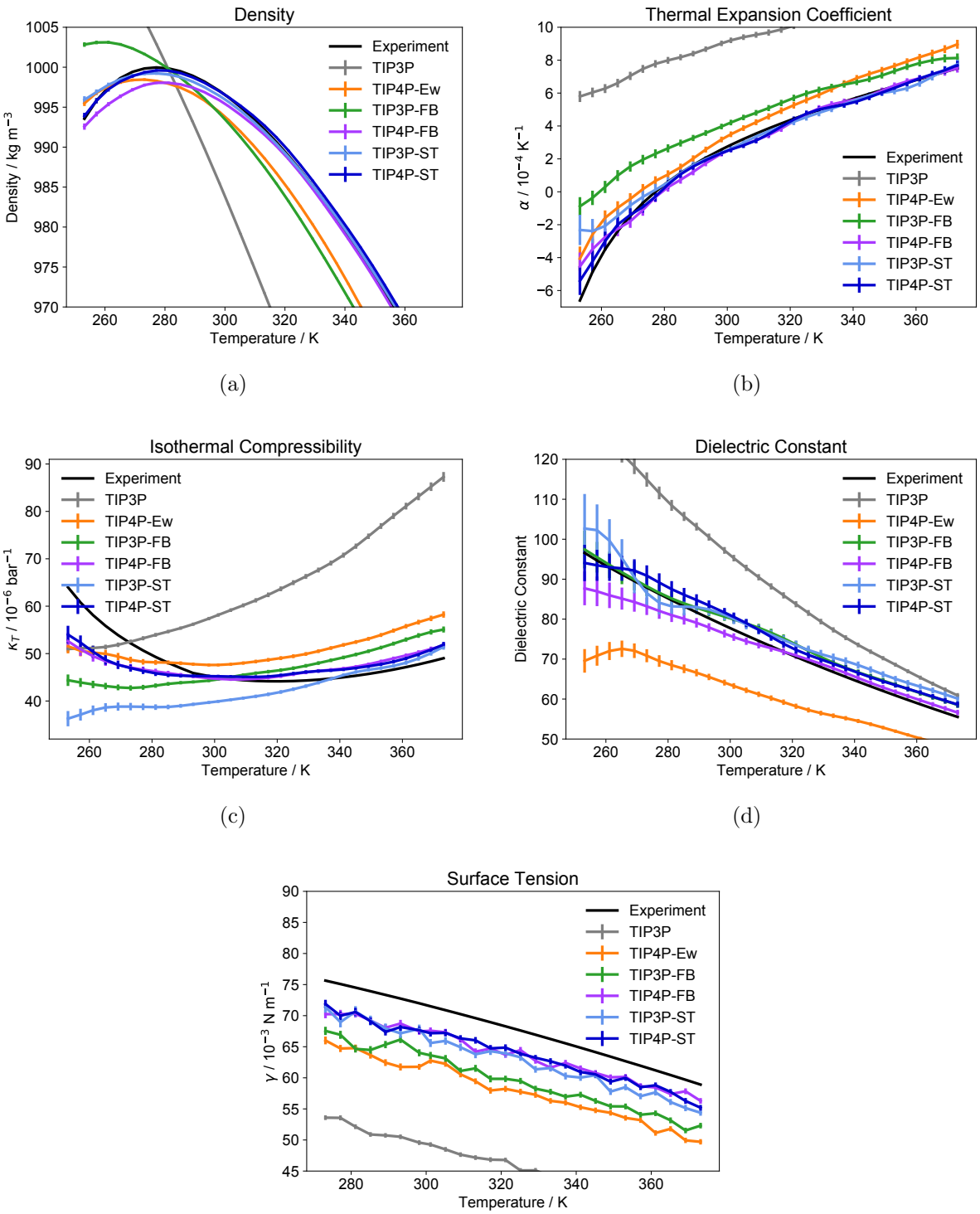


Figure 2: Performance of TIP3P-ST and TIP4P-ST compared to existing water models on fitted properties.

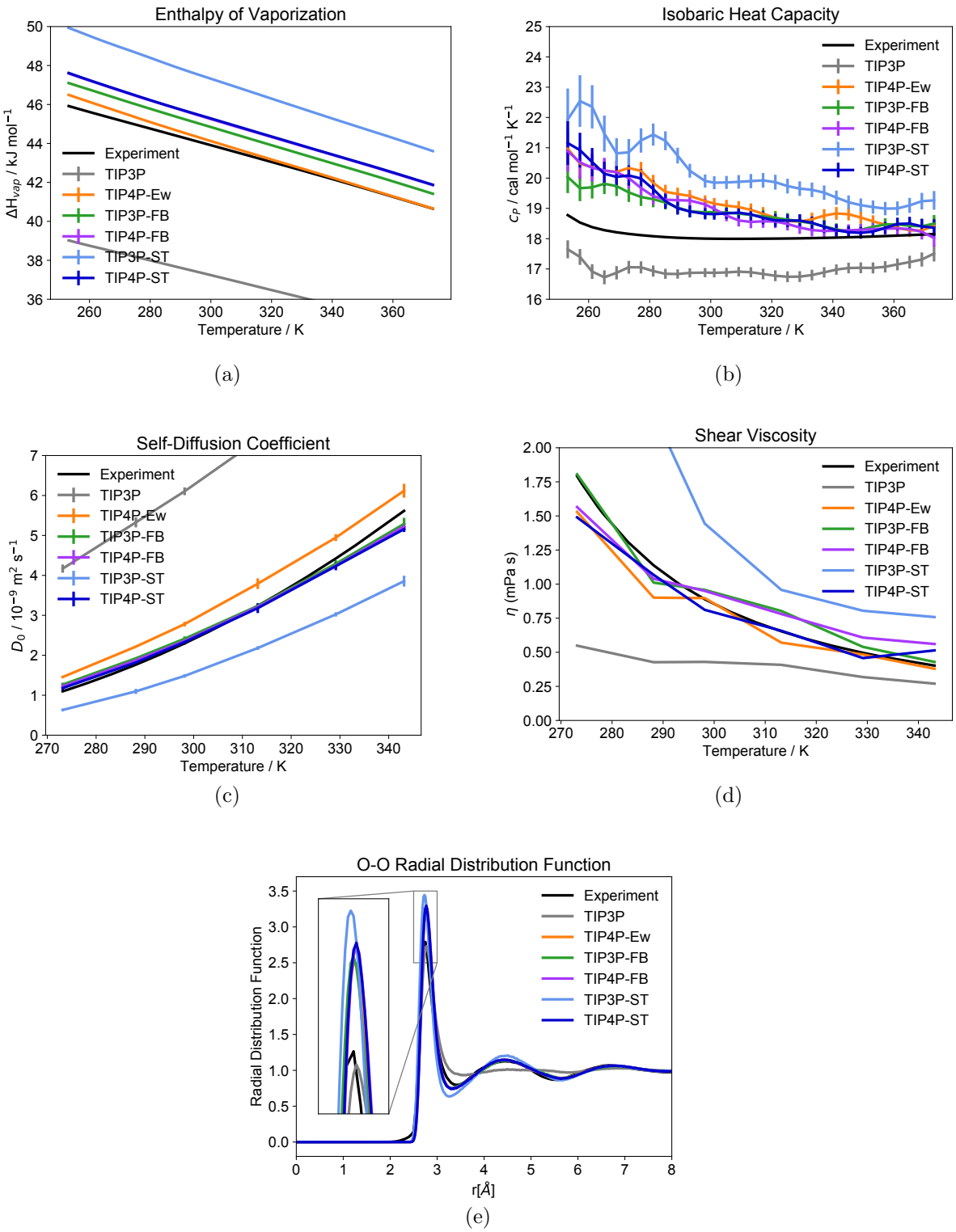


Figure 3: Performance of TIP3P-ST and TIP4P-ST compared to existing water models on properties not used in fitting.

3.3 Fitting with Reduced Reference Dielectric Constants

Recent studies on electrostatic models have raised questions regarding whether the simulated dielectric constant requires post-hoc corrections.[37,38] These studies posit that the *effective* electrostatic moments used to compute the MM interactions of ions and polar species should be reduced with respect to the *physical* charges used to compute electrostatic properties, due to the dielectric screening caused by the electronic polarization of the medium. This implies that the dielectric constant computed from the partial charges in the force field should be increased by a correction prior to comparing with experiment, or conversely, the experimental value should be reduced prior to making the comparison with the force field.

In Reference 37, the authors concluded that the missing polarizability in non-polarizable models scales the dielectric constant by a factor of 1.78. Under the assumption that the same correction factor would apply to our models, the reference dielectric constant should be reduced by a factor of $1/1.78 = 0.5618$. Here we test the effective charge hypothesis by reducing the reference dielectric constants by a factor of 0.56 in the fitting of three-point water models. If the effective charge hypothesis is correct, then we expect the model fitted to a reduced dielectric constant should produce improved agreement with experiment for validation properties.

Table 4: Optimized force field parameters for TIP3P-ST and TIP3P-ST-0.56$\epsilon(0)$, i.e. fitted to dielectric constant reduced by factor of 0.56

	r_{OH} (Å)	Θ_{HOH} (deg)	$q(e)$	σ_{LJ}	ϵ_{LJ}
TIP3P[8]	0.9572	104.52	0.41700	3.15075	0.63597
TIP3P-ST	1.0230	108.11	0.42556	3.19257	0.60190
TIP3P-ST-0.56$\epsilon(0)$	1.0534	114.89	0.41037	3.17463	0.64649

Table 5: Comparison of water models fitted to original and reduced dielectric constants at 298.15 K, 1.0 atm

Property	Experiment	TIP3P-ST	TIP3P-ST-0.56$\epsilon(0)$
ρ / g cm^{-3}	0.997	0.996(0)	0.997(0)
α / 10^{-4} K^{-1}	2.572	2.4(2)	2.3(2)
κ_{T} / 10^{-6} bar^{-1}	45.247	39.7(3)	39.8(3)
$\epsilon(0)$	78.409/44.050	81(1)	47(1)
γ / mJ m^{-2}	71.990	67.9(7)	66.6(7)
ΔH_{vap} / kcal mol^{-1}	10.513	11.333(4)	11.823(4)
c_{P} / cal mol^{-1} K^{-1}	18.002	19.9(2)	20.8(2)
D_0 / 10^{-5} cm^2 s^{-1}	2.29	1.48(4)	1.45(3)
η / mPa s	0.896	1.44	1.30
TMD / K	277	277	277
ΔG_{hyd} / kcal mol^{-1}	−6.33	−6.17(1)	−6.63(1)
$\langle E_{\mathrm{pot}} \rangle$ / kJ mol^{-1}		−12.57	−12.00

ρ: Density; α: Thermal expansion coefficient; κ_{T}: Isothermal compressibility; $\epsilon(0)$: Dielectric constant; γ: Liquid/vapor surface tension; ΔH_{vap}: Enthalpy of vaporization; c_{P}: Isobaric heat capacity; D_0: Self-diffusion Coefficient; η: shear viscosity; TMD: Temperature of maximum density; ΔG_{hyd}: Hydration (self-solvation) free energy; $\langle E_{\mathrm{pot}} \rangle$: Average total potential energy per water molecule in simulation.

The comparison of optimized parameters between TIP3P-ST and the model fitted to reduced dielectric constant, denoted, as TIP3P-ST-0.56$\epsilon(0)$, is shown in Table 4. A main difference is that the atomic charges q_{H} increase and the H-O-H angle widens to accommodate the reduced dielectric constants. The molecular dipole moment is 2.24 D and the self-polarization correction is smaller at 2.951kJ mol^{-1}, compared to TIP3P-ST which has a dipole moment of 2.46 D and self-polarization correction of 7.498kJ mol^{-1}. Table 5 shows the effect on preperty predictions by reducing the reference dielectric constant. The TIP3P-ST-0.56$\epsilon(0)$ model is able to reach similar levels of agreement with experiment as the original TIP3P-ST. The heat of vaporization increases further with respect to both experiment and TIP3P-ST. These observations support our earlier assertion that the quality of fitting for dielectric constants mainly depends on the molecular structure parameters and does not have major impact on the ability to fit other thermodynamic properties. However, due to the mixed results in relative accuracy of the models fitted to the original and reduced dielectric constants, we cannot conclude from this study whether correction of the dielectric

constant is necessary in general.

3.4 Statistical Uncertainty of Surface Tension Analytic Gradients

The accurate computation of gradients of simulated thermodynamic properties with respect to force field parameters is highly important for efficient model optimization. The thermodynamic property being differentiated contains statistical noise due to the finite length of the simulation, so we expect the parametric gradients to contain statistical noise as well. Moreover, different methods for computing the parametric gradient may exhibit varying levels of statistical noise for the same computational resources used in the calculation. Thus, we decided to compare the statistical noise in the surface tension gradients for two calculation methods: "semi-analytic" (i.e. the property gradient is computed from a thermodynamic fluctuation formula using finite-difference potential energy gradients), and "pure numerical" (i.e. by running separate simulations for each parameter).

The gradient of the simulated surface tension with respect to force field parameter may be obtained by analytic differentiation of the test-area formula resulting in a thermodynamic fluctuation formula, similar to the procedure for other thermodynamic properties:

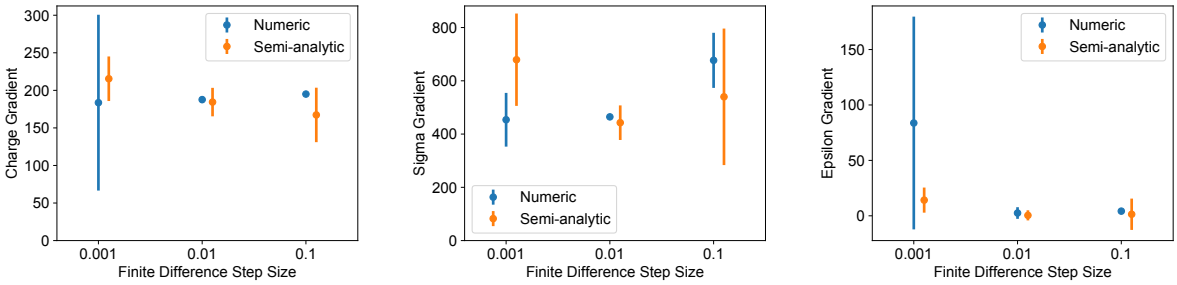
$$
\begin{aligned}
\frac{\partial \gamma}{\partial k} &= \lim_{\Delta S \to 0} \frac{-1}{2\beta \Delta S} \left[\frac{\partial \ln \langle \exp\left(-\beta \Delta E^{+}\right) \rangle}{\partial k} - \frac{\partial \ln \langle \exp\left(-\beta \Delta E^{-}\right) \rangle}{\partial k} \right] \\
&= \lim_{\Delta S \to 0} \frac{-1}{2\beta \Delta S} \left[\frac{1}{\langle \exp\left(-\beta \Delta E^{+}\right) \rangle} \frac{\partial \frac{\int e^{-\beta E^{+}} dr}{\int e^{-\beta E} dr}}{\partial k} - \frac{1}{\langle \exp\left(-\beta \Delta E^{-}\right) \rangle} \frac{\partial \frac{\int e^{-\beta E^{-}} dr}{\int e^{-\beta E} dr}}{\partial k} \right] \\
&= \lim_{\Delta S \to 0} \frac{-1}{2\Delta S} \left[\frac{\left\langle -\frac{\partial E^{+}}{\partial k} \exp\left(-\beta \Delta E^{+}\right) \right\rangle}{\langle \exp\left(-\beta \Delta E^{+}\right) \rangle} - \left\langle \frac{\partial E}{\partial k} \right\rangle - \frac{\left\langle -\frac{\partial E^{-}}{\partial k} \exp\left(-\beta \Delta E^{-}\right) \right\rangle}{\langle \exp\left(-\beta \Delta E^{-}\right) \rangle} + \left\langle \frac{\partial E}{\partial k} \right\rangle \right] \\
&= \lim_{\Delta S \to 0} \frac{-1}{2\Delta S} \left[\frac{\left\langle -\frac{\partial E^{+}}{\partial k} \exp\left(-\beta \Delta E^{+}\right) \right\rangle}{\langle \exp\left(-\beta \Delta E^{+}\right) \rangle} - \frac{\left\langle -\frac{\partial E^{-}}{\partial k} \exp\left(-\beta \Delta E^{-}\right) \right\rangle}{\langle \exp\left(-\beta \Delta E^{-}\right) \rangle} \right]
\end{aligned}
\tag{12}
$$

Here, the new terms in the formula $\frac{\partial E^{+}}{\partial k}$ and $\frac{\partial E^{-}}{\partial k}$ may be recognized as the potential energy derivatives of the surface-perturbed trajectory frames. These potential derivatives

are evaluated numerically using sufficiently small steps in k to avoid incurring machine-precision errors. All quantities in angle brackets representing ensemble averages are then evaluated as arithmetic averages over the trajectory frames. The computational cost of evaluating the full set of potential energy derivatives scales linearly with the number of parameters, and the added cost per parameter is significantly less than the simulation itself. In practice, the calculation of a single gradient element is roughly equal to 20% of the original MD simulation. On the other hand, pure numerical gradients of the surface tension involve running separate simulations where the parameter is perturbed by a small step, and repeating this procedure for each parameter being optimized. We used a central difference approximation, which implies the computational cost of the gradient is $2N_{\mathrm{param}}$ times the cost of simulating the property itself. Compared to the semi-analytic gradients, the numerical gradients involve running separate simulations with nearly fully independent samples (save for the same initial condition). The noise in the gradients also increases with decreasing parameter step size because the statistical error in the property is roughly independent of parameter size, resulting in large numerical errors for steps that are too small. It is also important to avoid step sizes that are too large and no longer within the linear regime.

Figure 4 compares the accuracy of the semi-analytic and numerical methods with a fixed simulation run length. The mean and standard error for each gradient is computed from five independent runs using the TIP3P parameters with a simulation length of 20ns. Finite difference step sizes of $\delta k_\lambda = 0.001$, 0.01 and 0.1 in the mathematical parameters were tested for both methods. When numeric gradients were used, the statistical errors were largest for step sizes of 0.001 and smallest for 0.01. For σ_{LJ}, increasing the step size to 0.1 resulted in a different mean and larger standard error, indicating this step size was outside the linear regime; we did not observe this for q_{H} and ϵ_{LJ}. The semi-analytic gradients are computationally less costly but also have higher uncertainty than the numerical gradients, thus we concluded that numerical gradients with a step size of 0.01 provide the most statistically precise surface tension gradients for a fixed simulation length. These conclusions are based

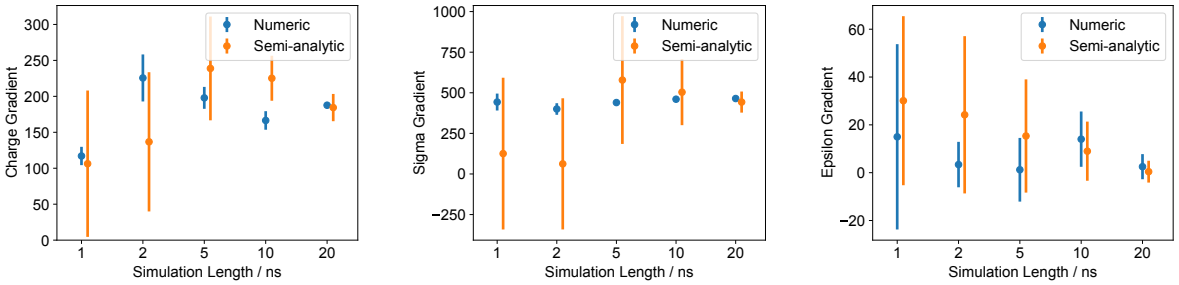
on our choice of the prior widths for the parameters; for a different choice of prior width, the recommended step size may be obtained by ensuring the step size in physical parameters, i.e. $\delta K_\lambda = \delta k_\lambda p_\lambda$ matches the current results.



(a) (b) (c)

Figure 4: Comparison between ensemble-averaged semi-analytic surface tension gradients, and pure numeric surface tension gradients. (a) charge q_{H} parameter; (b) σ_{LJ} parameter; (c) ϵ_{LJ} parameter; The error bar shows one standard error. Each point is computed from five independent runs, with the simulation length of 20 ns.

The benchmark of gradients computed with various simulation lengths are plotted in Figure 5. With the same finite difference step size of 0.01, we found that longer simulation lengths reduced the error bars on both numeric and semi-analytic surface tension gradients as expected. The semi-analytic gradients evaluated with the longest 20 ns simulation has error bars comparable to the numeric gradients evaluated with 5 ns simulation, indicating that numeric gradients can provide statistically more reliable results with comparable computational cost.

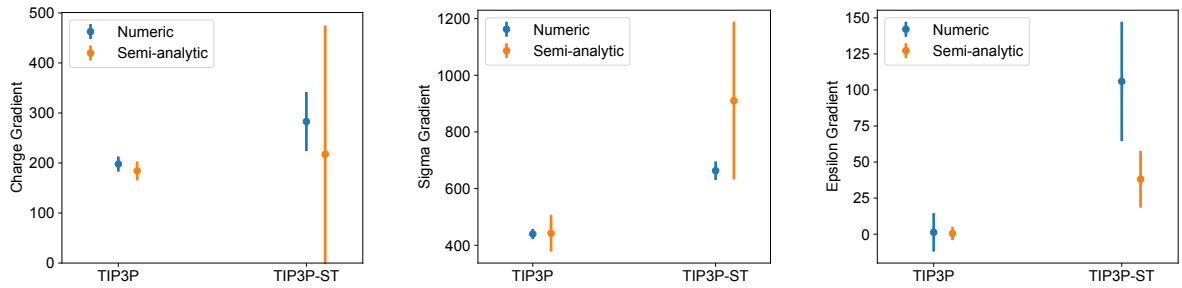


(a) (b) (c)

Figure 5: Comparison between ensemble-averaged semi-analytic gradients, and pure numeric gradients. (a) charge q_{H} parameter; (b) σ_{LJ} parameter; (c) ϵ_{LJ} parameter; The error bar shows one standard error. Each point is computed from five independent runs, with the relative finite difference step size of 0.01.

Taking the total computational cost into account, we compared the numeric gradients from 5 ns simulation with the semi-analytic gradients from 20 ns simulation, both with the optimal step size 0.01, as shown in Figure 6. Two sets of parameters were used, namely the TIP3P parameters and the TIP3P-ST parameters. The results show that for the TIP3P parameters, the numeric and semi-analytic gradients agree relatively well with comparable standard errors. However, when evaluated at the final TIP3P-ST parameters, the semi-analytic gradients have larger errors than the numeric gradients for the q_{H} and σ_{LJ} parameters, while ϵ_{LJ} exhibits the opposite behavior. The small errors for the semi-analytic gradients of ϵ_{LJ} may be due to the intrinsically small value of the gradient (i.e. in the limit of infinite simulation time). An intrinsically small gradient would reduce the error bars of the analytic gradient but not the numerical gradient, as the latter contains statistical noise from independent estimations of the surface tension and contributes a constant term to the error. The scale-independent behavior of the numerical gradient error is confirmed by comparing the standard error across parameters; for TIP3P these errors are $(15.2, 18.0, 13.3)$ for $q_{\mathrm{H}}, \sigma_{\mathrm{LJ}}, \epsilon_{\mathrm{LJ}}$ respectively, and for TIP3P-ST the errors are $(59.0, 32.9, 41.4)$. The standard error for surface tension gradients are larger overall for TIP3P-ST compared to TIP3P, which may be due to the slower dynamics of the model causing slower convergence of the property. Based

on our observation that the statistical errors were mostly smaller using numeric gradients, we decided to use numeric gradients for optimizing the water models in this study.



(a) (b) (c)

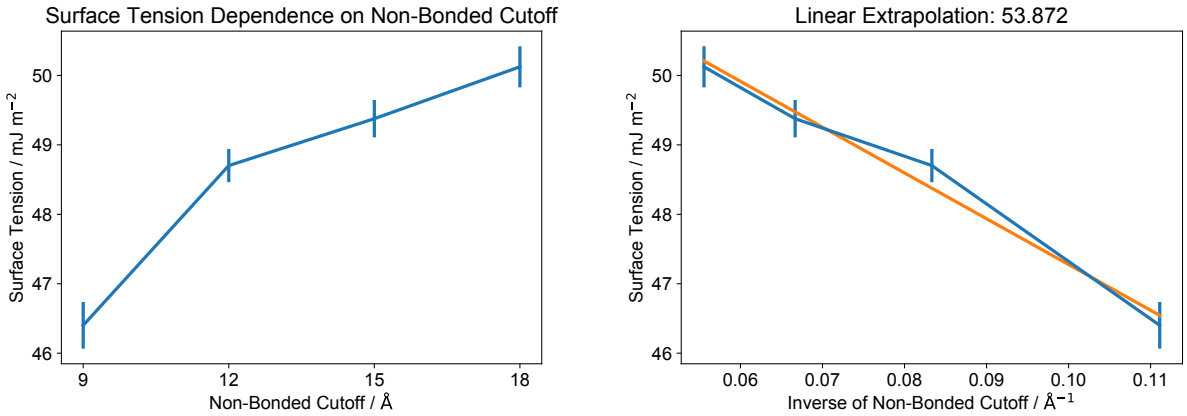
Figure 6: Comparison between ensemble-averaged analytic gradients, and pure numeric gradients. (a) charge q_{H} parameter; (b) σ_{LJ} parameter; (c) ϵ_{LJ} parameter; The error bar shows one standard error. The ensemble-averaged analytic gradients are calculated from 20 ns simulations, and the pure numeric gradients are calculated from 5 ns simulations. Each point is computed from five independent runs, with the relative finite difference step size of 0.01.

3.5 Surface Tension Dependence on Non-Bonded Cutoff

We evaluated the surface tension using the test-area method, for TIP3P at 298 K, 1.0 atm, using various van der Waals cutoff distances. Figure 7a shows that as the cutoff distance increases, the simulated surface tension continues to increase even at the distance of 18 Å. To estimate the surface tension at infinite cutoff distance, we performed an empirical linear extrapolation of the surface tension vs. the inverse of the cutoff value, as shown in Figure 7b. The final value of 53.87 mJ m^{-2} obtained from the intercept of the linear extrapolation is about 4 mJ m^{-2} higher than the value computed with cutoff at 15 Å. This indicates that our surface tensions shown in Figure 2e, which were evaluated with the cutoff distance of 15 Å, may have been underestimated by about 5% (4 mJ m^{-2}).

In Table 3, the surface tension extrapolated to infinite cutoff are reported as γ_∞; using these extrapolated numbers, TIP3P-ST achieved the best agreement of within 1 mJ m^{-2} to the experiment. Although we were aware of this source of error during our parameterization,

we could not use larger nonbonded cutoffs or extrapolations to infinity due to the increased computational expense. Instead, we decreased the weight of the surface tension property in the objective function, which led to an underestimation of the experimental surface tension by about 5% during fitting. It should be noted here that the implementation of PME for the van der Waals force could improve the behavior.[39] In addition, utilizing special long-range corrections for the Lennard-Jones potential in anisotropic systems could also significantly reduce the effect of the truncation.[20]



(a) (b)

Figure 7: Dependence of simulated surface tension on the non-bonded cutoff parameter.

4 Conclusions

In this work we apply parameteric derivatives of the surface tension calculated using the test-area method to optimize two water models, TIP3P-ST and TIP4P-ST. The gradients are implemented using a semi-analytic approach and a pure numerical approach, both of which are implemented in ForceBalance. We tested the statistical precision of semi-analytic parametric derivatives vs. pure numerical derivatives and found that pure numerical derivatives provide improved statistical precision for the same computational cost, provided an appropriate finite-difference step size is used. While the statistical error in semi-analytic gradients is relative to the intrinsic size of the gradient itself, the error in pure numerical

gradients contains a constant contribution that is essentially independent of which parameter is being differentiated. The effect of truncation of the van der Waals interactions are estimated by a linear extrapolation, which leads to better agreement to the experiment.

The overall results point to the validity of using surface tension as a replacement for heat of vaporization in force field development. Both water models correctly reproduce the temperature of maximum density, which in particular is notable for the three-point model, TIP3P-ST, because models of this functional form have had difficulty in accurately reproducing the density anomaly at ambient pressures. Whereas TIP4P-ST can accurately reproduce a broader range of kinetic and structural properties, consistent with more recent well-optimized 4-point rigid models, the TIP3P-ST generalizes more poorly to the validation set by producing somewhat over-structured radial distribution functions and lower diffusion coefficients. This indicates that rigid 3-point models need to make a compromise between accurate depictions of cohesion vs. structural and kinetic properties due to their limited functional form. We additionally found that the dielectric constant could be independently adjusted without impacting the quality of fit of other training parameters, leading to differences in the molecular geometry and mixed impacts on the validation properties.

Recent work by Milne and Jorge suggests that polarization corrections of the form utilized by Berendsen, this work, and many others is unnecessary – and perhaps undesirable – in order to reproduce experimental observables such as the enthalpy of vaporization and hydration free energy of water and other polar liquids.[34] Interestingly, our results suggest that when these properties are not used in the parameterization of the water model, the resulting enthalpy of vaporization will still be significantly greater than the experimentally measured quantity. Specifically we note that the enthalpies of vaporization of TIP3P-ST and TIP4P-ST are both somewhat greater than experiment (by approximately 0.57 and 0.33 kcal/mol, respectively) even after correction. If the polarization correction were not included, then the simulated ΔH_{vap} would be even more positive and further increase disagreement with experiment, as the polarization correction for moving from the condensed

phase to the gas phase is always favorable. Rivera and coworkers carried out simulations of polarizable water and found that induced dipole interactions contributed significantly to the surface tension;[40,41] this indicates that the physical origin of surface tension may be different in nonpolarizable vs. explicitly polarizable models, a question worthy of further study. Overall, we are optimistic that the procedure described in this study can be applied broadly to develop future generations of force fields for organic liquids and the nonbonded energy terms in biomolecular and general small molecule force fields.

Acknowledgement

Paul S. Nerenberg acknowledges the support of NASA Minority University Research and Education Project (MUREP) Institutional Research Opportunity grant NNX15AQ06A. Teresa Head-Gordon acknowledges the support of grant CHE-1665315 from the U.S. National Science Foundation. Lee-Ping Wang acknowledges the support of award 58158-DNI6 from the American Chemical Society Petroleum Research Fund. We would like to acknowledge Nanhao Chen for helpful discussions regarding the calculation of hydration free energies.

Supporting Information Available

Tables of thermodynamic properties of water used for fitting the models in this work.

References

(1) Nerenberg, P. S.; Head-Gordon, T. New developments in force fields for biomolecular simulations. *Curr. Opin. Struc. Biol.* **2018**, *49*, 129–138.

(2) Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *J. Chem. Inf. Model.* **2018**, *58*, 565–578.

(3) Guillot, B.; Guissani, Y. How to build a better pair potential for water. *J. Chem. Phys.* **2001**, *114*, 6720–6733.

(4) Pinnick, E. R.; Erramilli, S.; Wang, F. Predicting the melting temperature of ice-Ih with only electronic structure information as input. *The Journal of Chemical Physics* **2012**, *137*, 014510.

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

(6) Demerdash, O.; Wang, L.-P.; Head-Gordon, T. Advanced models for water simulations. *WIRES Comput. Mol. Sci.* **2018**, *8*, e1355.

(7) Onufriev, A. V.; Izadi, S. Water models for biomolecular simulations. *WIRES Comput. Mol. Sci.* **2018**, *8*, e1347.

(8) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(9) Bernal, J.; Fowler, R. A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *J. Chem. Phys.* **1933**, *1*, 515–548.

(10) Horn, H.; Swope, W.; Pitera, J.; Madura, J.; Dick, T.; Hura, G.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(11) Abascal, J.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(12) Fuentes-Azcatl, R.; Alejandre, J. Non-Polarizable Force Field of Water Based on the Dielectric Constant: TIP4P/ϵ. *J. Phys. Chem. B* **2014**, *118*, 1263–1272.

(13) Wang, L.-P.; Martínez, T.; Pande, V. Building force fields: an automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885–1891.

(14) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.

(15) Izadi, S.; Onufriev, A. Accuracy limit of rigid 3-point water models. *J. Chem. Phys.* **2016**, *145*, 074501–11.

(16) Berendsen, H.; Grigera, J.; Straatsma, T. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(17) Jorgensen, W.; Maxwell, D.; Tirado-Rives, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(18) Crabtree, A.; Siman-Tov, M. Thermophysical properties of saturated light and heavy water for Advanced Neutron Source applications. *International Nuclear Information System* **1993**, *25*, 4.

(19) Cooper, J. *IAPWS Release on Surface Tension of Heavy Water Substance, R5-85*; International Association for the Properties of Water and Steam, 1994.

(20) Ghoufi, A.; Malfreyt, P.; Tildesley, D. Computer modelling of the surface tension of the gas-liquid and liquid-liquid interface. *Chem. Soc. Rev.* **2016**, *45*, 1387–1409.

(21) Dean, J. *Lange's Handbook of Chemistry*; New York; London: McGraw-Hill, Inc., 1999; Chapter 5.6.

(22) Nielsen, S.; Lopez, C.; Srinivas, G.; Klein, M. A coarse grain model for n-alkanes parameterized from surface tension data. *J. Chem. Phys.* **2003**, *119*, 7043–7049.

(23) Salas, F.; Mndez-Maldonado, G.; Nez-Rojas, E.; Aguilar-Pineda, G.; Domnguez, H.; Alejandre, J. Systematic procedure to parametrize force fields for molecular fluids. *J. Chem. Theory Comput.* **2015**, *11*, 683–693.

(24) Martnez-Jimnez, M.; Saint-Martin, H. A four-site molecular model for simulations of liquid methanol and water-methanol mixtures: MeOH-4P. *J. Chem. Theory Comput.* **2018**, *14*, 2526–2537.

(25) Ghoufi, A.; Malfreyt, P. Calculation of the surface tension of water: 40 years of molecular simulations. *Mol. Simulat.* **2019**, *45*, 295–303.

(26) 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.

(27) Rivera, J. L.; Douglas, J. F. Influence of film thickness on the stability of free-standing Lennard-Jones fluid films. *J. Chem. Phys.* **2019**, *150*, 144705.

(28) Vega, C.; de Miguel, E. Surface tension of the most popular models of water by using the test-area simulation method. *J. Chem. Phys.* **2007**, *126*, 154707.

(29) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 461–469.

(30) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comp. Biol.* **2017**, *13*, e1005659.

(31) Yeh, I.; Hummer, G. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *J. Phys. Chem. B* **2004**, *108*, 15873–15879.

(32) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.

(33) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.

(34) Milne, A. W.; Jorge, M. Polarization Corrections and the Hydration Free Energy of Water. *J. Chem. Theory Comput.* **2019**, *15*, 1065–1078.

(35) 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.

(36) Wagner, W.; Pruß, A. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387–535.

(37) Farahvash, A.; Leontyev, I.; Stuchebrukhov, A. Dynamic and electronic polarization corrections to the dielectric constant of water. *J. Phys. Chem. A* **2018**, *122*, 9243–9250.

(38) Visscher, K.; Swope, W.; Geerke, D. A QM/MM derived polarizable water model for molecular simulation. *Molecules* **2018**, *23*, 3131.

(39) Alejandre, J.; Chapela, G. The surface tension of TIP4P/2005 water model using the Ewald sums for the dispersion interactions. *J. Chem. Phys.* **2010**, *132*, 014701.

(40) Rivera, J.; Predota, M.; Chialvo, A.; Cummings, P. Vaporliquid equilibrium simulations of the SCPDP model of water. *Chem. Phys. Lett.* **2002**, *357*, 189–194.

(41) Rivera, J. L.; Starr, F. W.; Paricaud, P.; Cummings, P. T. Polarizable contributions to the surface tension of liquid water. *J. Chem. Phys.* **2006**, *125*, 094712.