Developing Fast and Accurate Water Models for Atomistic Molecular Dynamics Simulations

Yeyue Xiong

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Engineering Mechanics

Alexey V. Onufriev, Chair
Sanket A. Deshmukh
Anne E. Staples
Raffaella De Vita
Eduard F. Valeyev

August 12, 2021
Blacksburg, Virginia

Keywords: Water model, molecular dynamics, global optimization, polarizable, Drude
Copyright 2021, Yeyue Xiong
Developing Fast and Accurate Water Models for Atomistic Molecular Dynamics Simulations

Yeyue Xiong

(ABSTRACT)

Water models are of great importance for different fields of studies such as fluid mechanics, nano materials, and biomolecule simulations. In this dissertation, we focus on the water models applied in atomistic simulations, including those of biomolecules such as proteins and DNA. Despite water’s simple structure and countless studies carried out over the decades, the best water models are still far from perfect. Water models are normally divided into two types—explicit model and implicit model. Here my research is mainly focused on explicit models. In explicit water models, fixed charge n-point models are most widely used in atomistic simulations, but have known accuracy drawbacks. Increasing the number of point charges, as well as adding electronic polarizability, are two common strategies for accuracy improvements. Both strategies come at a considerable computational cost, which weighs heavily against modest possible accuracy improvements in practical simulations. With a careful comparison between the two strategies, results show that adding polarizability is a more favorable path to take. Optimal point charge approximation (OPCA) method is then applied along with a novel global optimization process, leading to a new polarizable water model OPC3-pol that can reproduce bulk liquid properties of water accurately and run at a speed comparable to 3- and 4-point non-polarizable water models. For practical use, OPC3-pol works with existing non-polarizable AMBER force fields for simulations of globular protein or DNA. In addition, for intrinsically disordered protein simulations, OPC3-pol fixes the over-compactness problem of the previous generation non-polarizable water models.
Developing Fast and Accurate Water Models for Atomistic Molecular Dynamics Simulations

Yeyue Xiong

(GENERAL AUDIENCE ABSTRACT)

With the rapid advancements of computer technologies, computer simulation has become increasingly popular in biochemistry research fields. Simulations of microscopic substances that are vital for living creatures such as proteins and DNAs have brought us more and more insights into their structures and functions. Because of the fact that almost all the microscopic substances are immersed in water no matter they are in a human body, a plant, or in bacteria, accurately simulating water is crucial for the success of such simulations. My research is focused on developing accurate and fast water models that can be used by researchers in their biochemical simulations. One particular challenge is that water in nature is very flexible and the properties of water can change drastically when its surroundings change. Many classical water models cannot correctly mimic this flexibility, and some more advanced water models that are able to mimic it can cost several times more computing resources. Our latest water model OPC3-pol, benefited from a new design, accurately mimics the flexibility and runs as fast as a traditional rigid water model.
Contents

List of Figures vii

List of Tables xv

1 Introduction 1

2 Exploring optimization strategies for improving explicit water models: rigid n-point model and polarizable model based on Drude oscillator 7

2.1 Introduction ................................................. 9

2.2 Methods ....................................................... 13

2.2.1 Overall approach ......................................... 14

2.2.2 Reference multipole moment sets ......................... 20

2.2.3 n-point rigid models ..................................... 20

2.2.4 Polarizable water model with Drude oscillator .......... 22

2.3 Results and Discussion ...................................... 25

2.3.1 Rigid n-point water models ............................. 27

2.3.2 Limited optimal 3-point polarizable model—only polarizability optimized .......................... 29
2.3.3 Globally optimal 3-point polarizable model—both the polarizability and the “rigid base” are optimized. 32

2.4 Conclusions 34

2.5 Supporting information 36

2.6 Acknowledgments 36

3 Melting points of OPC and OPC3 water models 37

3.1 Introduction 38

3.2 Results and discussion 40

3.3 Conclusions 44

3.4 Methods 47

3.4.1 Direct Coexistence 47

3.4.2 New anisotropic pressure scaling option in AMBER 52

4 A fast polarizable water model for atomistic simulations 54

4.1 Introduction 55

4.2 Overall approach 57

4.3 Results and discussions 61

4.4 Conclusions 64

4.5 Supplementary information 66

4.5.1 Method details 66
4.5.2 Results ................................................................. 71
4.5.3 Tests in atomistic MD simulations .............................. 73
4.5.4 Further speed-up with 4 fs time step and HMR ................ 77
4.5.5 Calculation of liquid water properties .......................... 79
4.5.6 Temperature dependent properties ............................. 84

5 Conclusions .................................................................. 86

Bibliography ..................................................................... 89
List of Figures

1.1 **Rigid n-point water model geometries for n=3, n=4 and n=5.** Hydrogen, Oxygen and the extra interaction point with a point charge but no mass are represented as white, red and pink spheres respectively. 4, and 5-point water models have zero charge on their oxygen, the negative charge is placed on the extra point(s). (a) 3-point water model. (b) 4-point water model. (c) and (d) Different possible non-planar configurations of 5-point water models. ................................................................. 2

2.1 **Rigid n-point water model geometries for n=3, n=4 and n=5.** Hydrogen, Oxygen and the extra interaction point with a point charge but no mass are represented as white, red and pink spheres respectively. 4, and 5-point water models have zero charge on their oxygen, the negative charge is placed on the extra point(s). (a) 3-point water model. (b) 4-point water model. (c) and (d) Different possible non-planar configurations of 5-point water models. ................................................................. 10

2.2 **Schematic of a water dimer.** \( r(OO) \) is the distance between the two oxygens; \( \theta \) and \( \varphi \) are the angles between water molecule’s \( C_2 \) axis and the O-O line for the two molecules respectively. ......................................................... 14
2.3 **Schematics of 3, 4 and 5-point rigid water models.** Each circle represents an interaction point (H, O atom, or extra point–EP). Thick lines are chemical bonds. The origin point is on the Oxygen; $C_2$ symmetry axis of water is set as the z-axis; y-axis is in the H-O-H plane and passes through O; x-direction is only accounted for in a 5-point model, and the 2 extra interaction points (EP) are placed on the $xOz$ plane. Electric charge on H is $q$, charge $-2q$ is placed on O for 3-point model, on EP for 4-point model, and divided evenly on the two EPs for 5-point model. (a) 3-point water model. (b) 4-point water model. (c) 5-point water model.

2.4 **Flowchart: the process of building a rigid $n$-point water model.**

2.5 **Schematic of a 3-point polarizable water model with Drude oscillator.** This model is composed of three fixed point charges and a Drude oscillator particle attached to the Oxygen atom. Charge $q$ is assigned to each white point as Hydrogen, charge $(−2q − Q_D)$ is assigned to the red point as Oxygen and the black point as the Drude particle has charge $Q_D$ and zero mass. The black point is connected to the red point through a virtual bond with equilibrium length 0, force constant $k$ and no directional restrictions. An external electric field $E$ moves the Drude particle from its equilibrium position.

2.6 **Flowchart: the process of building a polarizable Drude water model.**
2.7 Dimer total dipole moment and total angle error of the limited optimal 3-point polarizable model. With the water monomer dipole fixed at 1.86D, we test different 3-point polarizable models by varying the polarizability from 0.10Å$^3$ to 2.10Å$^3$. The blue diamond symbols represent the total dipole errors and the orange triangle symbols show the total angle errors, where both are compared with the ab initio reference [58, 119].

2.8 Globally optimal polarizable model error distribution. (a) The absolute error of the total angle ($\theta + \varphi$) as a function of the corresponding model’s monomer dipole $\mu$ and polarizability $\alpha$ value. (b) The water dimer total dipole moment error with respect to the model’s $\mu$, $\alpha$ combination. (c) The approximate range of the optimal $\mu$, $\alpha$ combinations, where the “combined relative error” is defined as $\text{Error}_{\text{combined}} = \frac{\text{Error}_{\text{angle}}}{180^\circ} + \frac{\text{Error}_{\text{dipole}}}{p_{dimer}}$ with $p_{dimer} = 2.68$D [119]. The “combined relative error” was not used in the actual optimization process.

2.9 Errors in reproducing the ab initio water dimer geometry by the rigid water models. Our models’ $\theta$ and $\varphi$ angle errors (reference values—$\theta_0 = 57.9^\circ$ and $\varphi_0 = 55.6^\circ$—are from the benchmark ab initio calculation by Klopper et al. [58]) are represented with blue and orange connected circles. In (a), (b) and (c) are the results of models built with gas exp., gas QM and liquid MP2/4MM multipole moments (Table.2.1) respectively. Water dimers of TIP3P, OPC3 and OPC model are energy minimized with O-O distance fixed at 2.91Å, and their angle errors in the dimers are shown as square, diamond and triangle respectively.
2.10 **Absolute errors of the two key angles (θ and φ, Fig.2.2)** in a water dimer made of rigid or Drude polarizable model molecules. These errors are calculated with reference to the *ab initio* calculation values (θ₀ = 57.9°, φ₀ = 55.6°) [58, 119]. Results of the 3, 4 and 5-point rigid water model dimers are built based on the “gas exp.” reference (Table.2.1), same with the polarizable models.

3.1 The overall geometry of classical n-point fixed charge water models. (a) a 3-point model, mimicking H-O-H structure; (b) a 4-point model, offsetting oxygen charge on the 4th point; (c) a 5-point model, with two extra charges emulating lone pairs.

3.2 Simulation box for the direct coexistence method (93Å × 56Å × 45Å).

3.3 Total energy (per mole) of the ice-liquid coexistence system for (a) TIP4P-Ew, (b) TIP4P/2005, (c) OPC, and (d) OPC3 models. The blue, green, and red curves represent total energy recorded for simulations at each model’s low, medium, and high temperature respectively. The peak at the end of a red curve shows the point where the whole simulation box becomes liquid. The flat “tail” region at the end of a blue curve corresponds to the state where the whole box is frozen. In (c), additional curves (yellow and black) at different temperatures are shown. With the temperature farther away from the melting point, the phase transition is faster. In (d), the pink and light blue curves are OPC3 simulations with the ice secondary prism (1210) plane as the interface. Also note in (d) the blue and light blue curves of OPC3 model does not have the flat “tail”, but we still observed the whole box becoming ice at the last few time steps of the simulation.
3.4 The hexagonal unit cell formed by 12 water molecules in ice $I_h$: (a) side view; (b) top view.

3.5 The two fluid-solid interface choices with respect to ice hexagonal structure: pink shows the primary prism plane (10\bar{1}0); blue shows the secondary prism plane (12\bar{1}0).

3.6 The process to prepare the fluid-ice coexistence box.

4.1 Developing OPC3-pol: 3-point Drude polarizable water model, shown schematically on the left. Here $q$ is the electric charge on hydrogen, $q_D$ is the charge on the Drude particle (black dot), $-2q - q_D$ is the oxygen charge, $k$ is the spring constant of the Drude-oxygen bond, $m_O$ is the oxygen mass, and $E$ denotes the external electric field (e.g. generated by the surrounding water molecules). The OPCA approximation maps the conventional “native” charge–distances–angle-polarizability 4D parameter space of a water model onto a reduced, 3D space of two lowest multipole moments and the polarizability. An exhaustive search for the global optimum in the 3D space is performed, with the model quality score (shown by color scale, from 1 to 10) being the deviation of model properties from experimental bulk water properties. Linear regression is applied to refine model parameters near the optimum. 1-1 mass partitioning for the oxygen-Drude pair is applied as a design choice for computational efficiency.

4.2 Global optimal OPC3-pol model. The values of the OH bond length $l_{OH}$, H-O-H angle $\angle \text{HOH}$, partial charge values $q$, $q_D$, and the Drude particle spring force constant $k$ are shown. The LJ interaction center is on the oxygen atom, its parameters are $\sigma_{LJ} = 3.176 \ \text{Å}$ and $\epsilon_{LJ} = 0.2085 \ \text{kJ/mol}$. 
4.3 Relative errors of five liquid water properties simulated with TIP3P [54], TIP4P-Ew, SWM4-NDP [68], iAMOEBA [113], and OPC3-pol water models at 300K.

4.4 Relative simulation speed comparison between TIP3P, TIP4P-Ew, iAMOEBA, SWM4-NDP, and OPC3-pol water models. The speeds (ns/day) for TIP3P, TIP4P-Ew, and OPC3-pol are measured from liquid water simulations of 806 water molecules with 2 fs and 4 fs time step. The speeds of iAMOEBA and SWM4-NDP model are those reported by Wang et al. and Huang et al. respectively [45, 113]. The non-polarizable models TIP3P and TIP4P-Ew are shown in green, while iAMOEBA, SWM4-NDP, and OPC3-pol—polarizable models are shown in blue. Dark color bars show the speeds achieved using each model’s “standard” time step: 2 fs for TIP3P, TIP4P-Ew, and OPC3-pol; 1 fs for iAMOEBA and SWM4-NDP. Light color bars show the maximum speeds if a water model supports 4 fs time step with hydrogen mass repartitioning (HMR).

4.5 Schematic of a rigid 3-point water model.
4.6 Developing a 3-point Drude polarizable water model: the use OPCA approximation replaces the conventional “native” 4D parameter space of a water model with a reduced, 3D space of two lowest multipoles and polarizability. An exhaustive search for the global optimum is performed in the reduced 3D space, and then conduct global optimization. 1-1 mass partitioning for oxygen-Drude pair is applied as a design choice for computational efficiency. $q$ is the electric charge on hydrogen, $q_D$ is the charge on the Drude particle, $-2q - q_D$ is the oxygen charge, $k$ is the bond force constant for the Drude-oxygen bond, $m_O$ is the oxygen mass, and $E$ denotes the external electric field (generated by surrounding molecules in MD simulations).

4.7 Water dimer structure at lowest energy.

4.8 Performance of OPC3pol model in all-atom MD simulation with 2 fs time step of 76-residue globular protein ubiquitin shown in the backbone root-mean-square deviation (RMSD) relative to the X-ray reference, over the course of 8 independent simulations, 2 $\mu$s each. Blue curves: AMBER-ff14SB force field, orange curves: AMBER-ff19SB force field. Top panel: RMSD curves of all of the trajectories (four for ff14SB and four for ff19SB) concatenated. Bottom panel: Probability density distributions of the RMSD values. The trajectories are combined for each of the two force fields respectively.

4.9 Performance of OPC3pol model in all-atom MD simulations of a B-DNA dodecamer with AMBER OL15 force field and 2 fs time step, over the course of 4 independent simulations, 1 $\mu$s each. Left: The all non-hydrogen atom (excluding terminal nucleotides) RMSD to crystal structure reference Right: The widths of major (red) and minor (green) grooves in comparison with the experimental NMR values (black line).
4.10 Radius of gyration $R_g$ of the IDP A$\beta$42 simulated with OPC3-pol water model. The two 8 $\mu$s simulations are under two standard AMBER protein force fields: ff14SB (blue) and ff19SB (orange).

4.11 Self diffusion coefficient $D$ of OPC3-pol water simulated under 4 fs time step and with different repartitioned hydrogen mass values.

4.12 Backbone root-mean-square deviation (RMSD) of ubiquitin, simulated 3 $\mu$s twice with OPC3pol model at 4 fs time step. The blue curve is the trajectory simulated with ff14SB and the orange curve with ff19SB. Both have the hydrogen mass repartitioning (HMR) applied.

4.13 B-DNA dodecamer (1BNA) simulated with OPC3-pol water model and under AMBER OL15 force field. Two independent trajectories are included, each is 2 $\mu$s long, with 4 fs time step and hydrogen mass repartitioning (HMR) applied. **Left:** The backbone RMSD of the simulations are ranging around 2 Å. **Right:** The widths of major (red) and minor (green) grooves match the experimental NMR values (black line).

4.14 Temperature dependency curves (blue curve) of OPC3-pol model's density, static dielectric constant, heat of vaporization, and self diffusion coefficient, in comparison with experiment reference (black curve). Temperature ranges from 220 K to 360 K and the pressure is 1 bar.
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Reference multipole moments of single water molecule.</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>Parameters of 3, 4 and 5-point optimal water models, and their geometry (Fig. 2.1). For comparison, parameters of several existing models are shown in the bottom rows.</td>
<td>22</td>
</tr>
<tr>
<td>2.3</td>
<td>Parameters of the 3-point polarizable water models.</td>
<td>24</td>
</tr>
<tr>
<td>3.1</td>
<td>Calculated melting points (ice $I_h$) of TIP4P/2005, TIP4P-Ew, OPC and OPC3 water models.</td>
<td>41</td>
</tr>
<tr>
<td>4.1</td>
<td>Accuracy of OPC3Pol in simulating ubiquitin (PDB: 1UBQ) and B-DNA dodecamer (PDB: 1BNA) structures at ambient conditions on microsecond time-scale. A comparison with “industry standard” non-polarizable TIP4P-Ew model is under AMBER-ff14SB for ubiquitin and AMBER-OL15 for B-DNA. SWM4-NDP model is accompanied with Drude FF (*B-DNA crystal structure is used in the training of Drude FF with SWM4-NDP) [43, 95]. OPC3-pol model is simulated with both AMBER-ff14SB and ff19SB for ubiquitin and AMBER-OL15 for B-DNA.</td>
<td>63</td>
</tr>
<tr>
<td>4.2</td>
<td>Bulk liquid properties of OPC3pol in comparison with other water models (bold model names referring to polarizable models) and experimental reference: Average Dipole $\mu$, Density $\rho$, Static Dielectric Constant $\epsilon_0$, Self Diffusion Coefficient $D$, Heat of Vaporization $\Delta H_{vap}$, and First Peak Position in the RDF $\rho_{oo1}$.</td>
<td>72</td>
</tr>
</tbody>
</table>
4.3 Water dimer simulated with OPC3pol and other polarizable water models (minimal energy state). The simulated water dimers are compared through their total dipole moments $\mu_{\text{dimer}}$, oxygen-oxygen distances $r(\text{OO})$, and the angle $\varphi$ between the bisector of the acceptor and the O-O vector.

4.4 Relative simulation speed comparison between OPC3pol, TIP3P, TIP4P-Ew, SWM4-NDP, and AMOEBA water models. The speeds of TIP3P, OPC and OPC3pol models are compared through 2 fs time step NPT simulations of 806 water molecules with AMBER 2019 on one RTX2080Ti GPU, measured in ns/day. The speeds of iAMOEBA and SWM4-NDP water models relative to TIP3P are from previous published results by Wang et al. and Huang et al. respectively [45, 113].
Chapter 1

Introduction

Water molecule Has a deceptively simple structure H$_2$O, yet many anomalies of liquid water are still hard to explain despite the countless studies [13, 14, 30, 40, 89]—not surprisingly, water in its liquid state is notoriously difficult to model. However, it is important to simulate water accurately for different research fields including both engineering and science, such as fluid mechanics [27, 60], nano materials [10, 59], and biomolecule simulations [85]. Many complex biomolecules (DNA, RNA, proteins, etc.), vital for a living cell’s function, cannot be studied alone without considering their environment—water as the solvent [121]. To study those large biomolecules, atomistic simulations have been widely used, and numerous different water models [85] have been developed to reproduce water properties, including the class of water models most widely used today—simple, rigid, fixed-charge explicit models such as TIP3P [54], TIP4P [1, 52], TIP5P [75] and SPC/E [9]. According to the convention, these classical water models are distinguished by the number of “points”—interaction sites—in them: 3-point models (Fig.1.1a, with 3 point charges) such as TIP3P are the most common due to what is perceived by many as an acceptable balance between accuracy and computational cost (compared with 4 or 5-point models). For water models with more points, 4-point models (Fig.1.1b, with 3 point charges and neutral oxygen) such as TIP4P, and especially 5-point models (Fig.1.1c,1.1d, with 4 point charges and neutral oxygen) models such as TIP5P (Fig.1.1d), the cost considerations become significant, even though these models tend to reproduce water properties better than their 3-point counterparts. However,
despite decades of effort by many groups, none of the existing simple water models is perfect \cite{31, 39, 85, 86, 87, 108}.

Figure 1.1: **Rigid n-point water model geometries for n=3, n=4 and n=5.** Hydrogen, Oxygen and the extra interaction point with a point charge but no mass are represented as white, red and pink spheres respectively. 4, and 5-point water models have zero charge on their oxygen, the negative charge is placed on the extra point(s). (a) 3-point water model. (b) 4-point water model. (c) and (d) Different possible non-planar configurations of 5-point water models.

It is known that liquid water properties are determined by a complex network of hydrogen bonds. In these classical water models, hydrogen bonds are mimicked mostly by the electrostatics, complemented by a Lennard-Jones (LJ) potential. The latter—LJ potential—is generally represented by a single site centered on the oxygen, and its corresponding interaction is isotropic and featureless, in contrast to hydrogen bonding, which is directional. Thus, accurately representing electrostatic interactions is paramount for a classical water model to mimic hydrogen bond interactions and reproduce liquid water properties \cite{47}. Therefore, our first question is how accurately can we describe the reference electrostatics within the unavoidable limitations of $n$-point models, and how far it can get us in terms of water model accuracy, by optimizing the electrostatics alone. This is a non-trivial question since, by construction, these simple $n$-point rigid models miss important physics, such as molecular flexibility, electronic polarizability, and charge transfer.

We note that a good water model intended for biomolecular simulations must reproduce experimental properties of water with high accuracy. There are at least two reasons why existing water models are far from perfect in this respect. One is the limitations of the
optimization strategies used in constructing water models. The other reason is on a more fundamental level—the limited physics built into these simple models. As an example of the first kind of limitation, current widely used rigid 3-point models (TIP3P, SPC, SPC/E, etc.) are based on an assumption that the experimental water molecule geometry is somehow optimal, or near optimal, for a classical water models—consequently, these models place the point charges on, or near, the centers of hydrogen and oxygen atoms. While sophistication of the optimization techniques employed to find the optimum has grown tremendously [2], from essentially “guess-and-test” to the complex, state-of-the-art force balance optimizations [114], one crucial aspect of the over-all procedure has not changed until recently: the search for best fit model is performed in the vicinity of the “canonical” water geometry, thus returning a local optimum in the parameter space. It was recently demonstrated [47] that abandoning the restrictions (except fundamental $C_{2v}$ symmetry) on water model geometry, and performing an exhaustive search for a truly global optimum in the parameter space can result in significant accuracy improvements. The resulting 4-point rigid model—OPC [47] was built without the geometry restriction on point charge placements. It was optimized globally to reproduce bulk water properties as best as possible, without any increase of the computational cost of employing the model in simulations, relative to common 4-point models such as TIP4P-Ew.

Apart from using global optimization, a natural strategy to improve the accuracy of a rigid $n$-point water model is to consider larger “$n$”. The key question with this approach is whether the inevitable and substantial increase in computational cost of employing “larger $n$” water models is justified by significantly better accuracy? If the accuracy gain is large enough, the cost increase may still be well worth it. Answering the question is important both for model developers and practitioners alike. Unfortunately, it is nearly impossible to compare existing $n$-point water models on the same footing to find out exactly how much gain an increase in $n$ brings. For once, these models are optimized using different criteria, and against non-
identical set of water properties. For example, TIP5P-Ew model [91] surpasses many of its 3- and 4-point predecessors in accuracy of describing water structure, but its accuracy of reproducing the self-diffusion is lower than that of TIP4P-Ew. In addition, optimization protocols vary widely, and, except for OPC-family models, there is no guarantee that the model corresponds to the global optimum in the optimization landscape. Thus, it is entirely possible that a global optimum for a smaller $n$ may yield a more accurate model than a local one for a larger $n$.

Even if a hypothetically perfect fixed-charge rigid model reproduced a large subset properties of bulk liquid water exactly, the model would still be inherently incapable of responding to the change of polarity of its micro-environment, relevant to biomolecular simulations. Specifically, water is highly polarizable [103]: the experimentally observed change in the dipole moment of real water molecule upon transfer from gas to liquid phase is as large as 1 Debye, while for any rigid fixed-charge model that change is zero by construction. The polarity of micro-environment near a macromolecule can be quite different from that of bulk water; non-polarizable models cannot properly respond to different micro-environments during the course of a simulation [90], e.g. in cross-membrane transport. However, the vast majority of current water models lack polarizability, which must adversely affect the accuracy of biomolecular simulations. This example illustrates the second type of limitation on water model accuracy — the missing physics. To address the lack of electronic polarization, a number of polarizable water models have been developed, for example POL1 [16, 26], FF12POL [111], SWM model family [72] and AMOEBA model family [70, 73, 113], see recent reviews for a comprehensive account of the field [43, 51, 85]. There is little doubt that, in principle, availability of highly accurate and efficient polarizable models for routine practical simulations should improve the accuracy of biomolecular simulations significantly, by accounting for effects completely missed by rigid fixed-charge models. However, because
of the unavoidable higher complexity of polarizable models relative to the fixed-charge ones, the balance towards wider adoption can only be tipped by distinctly higher accuracy, which does not seem to be the case yet [109, 111]. Since current polarizable water models typically use existing fixed-charge models as their “base”, they likely inherit at least some of the existing flaws of the fixed-charge models, unrelated to their lack of polarizability. That is they are likely represented by local optima in the complex optimization landscape. Indeed, despite undeniably better physical foundation, accuracy of a sophisticated polarizable water model can be lower than that of a globally optimal non-polarizable rigid model [47]. This observation begs the question: how much accuracy gain can the inclusion of polarizability bring, if the above optimization-related limitations were removed, and a truly globally optimal polarizable model was constructed? In other words, is global optimization a potentially useful strategy for polarizable models? For reasons discussed above in the context of comparing non-polarizable models, comparing existing \( n \)-point non-polarizable with polarizable models (even with respect to liquid water properties alone) on the same footing is difficult. Moreover, to the best of our knowledge, a truly globally optimal polarizable model has not be constructed yet.

To address the questions outlined above, in Chapter 2, we set up a tightly controlled computational experiment around water dimer—simplest possible mimic of water in condensed state. In this experiment, we build and evaluate several test water models of two types—rigid \( n \)-point models \((n = 3, 4, 5)\) and polarizable Drude models. Models of the types are globally optimized to reproduce the electrostatics of a water dimer, and are evaluated on their accuracy of reproducing a water dimer’s reference geometry [116]. In Chapter 3, we look into two recent water models OPC [47] and OPC3 [46] that can accurately reproduce liquid water properties under ambient conditions. Applying the direct coexistence method, we evaluate the two water models at an important point on water’s phase diagram—the
melting point, which is away from the temperature range the models are designed for and has been shown to be a challenge to simulate accurately with traditional $n$-point models [117]. In Chapter 4, following the conclusions from Chapter 2, we build a fast and accurate global optimal polarizable Drude water model—OPC3-pol. With a novel design choice different from current existing polarizable Drude water models, OPC3-pol model benefits from the added polarizability without sacrificing its computational efficiency. The global optimal OPC3-pol model is tested to work with latest existing non-polarizable AMBER force fields and it achieves high accuracy in protein and DNA benchmark simulations.
Chapter 2

Exploring optimization strategies for improving explicit water models: rigid n-point model and polarizable model based on Drude oscillator

Authors: Yeyue Xiong and Alexey V. Onufriev


Abstract

Rigid n-point water models are widely used in atomistic simulations, but have known accuracy drawbacks. Increasing the number of point charges, as well as adding electronic polarizability, are two common strategies for accuracy improvements. Both strategies come at considerable computational cost, which weighs heavily against modest possible accuracy improvements in practical simulations. In an effort to provide guidance for model devel-
CHAPTER 2. EXPLORING OPTIMIZATION STRATEGIES FOR IMPROVING EXPLICIT WATER MODELS: RIGID N-POINT MODEL AND POLARIZABLE MODEL BASED ON DRUDE OSCILLATOR

opment, here we have explored the limiting accuracy of “electrostatically globally optimal” $n$-point water models in terms of their ability to reproduce properties of water dimer—a mimic of the condensed state of water. For a given $n$, each model is built upon a set of reference multipole moments (e.g. ab initio) and then optimized to reproduce water dimer total dipole moment. The models are then evaluated with respect to the accuracy of reproducing the geometry of the water dimer. We find that global optimization of the charge distribution alone can deliver high accuracy of the water model: for $n = 4$ or $n = 5$, the geometry of the resulting water dimer can be almost within $5^\circ$ of the ab initio reference, which is half that of the experimental error margin. Thus, global optimization of the charge distribution of classical $n$-point water models can lead to high accuracy models. We also find that while the accuracy improvement in going from $n = 3$ to $n = 4$ is substantial, the additional accuracy increase in going from $n = 4$ to $n = 5$ is marginal. Next, we have explored accuracy limitations of the standard practice of adding electronic polarizability (via a Drude particle) to a “rigid base”—pre-optimization rigid $n$-point water model. The resulting model ($n = 3$) shows a relatively small improvement in accuracy, suggesting that the strategy of merely adding the polarizability to an inferior accuracy water model used as the base cannot fix the defects of the latter. An alternative strategy in which the parameters of the rigid base model are globally optimized along with the polarizability parameter is much more promising: the resulting 3-point polarizable model out-performs even the 5-point optimal rigid model by a large margin. We suggest that future development efforts consider 3- and 4-point polarizable models where global optimization of the “rigid base” is coupled to optimization of the polarizability to deliver globally optimal solutions.
2.1 Introduction

Water molecule has a deceptively simple structure $\text{H}_2\text{O}$, yet many anomalies of liquid water are still hard to explain despite the countless studies [13, 14, 30, 40, 89]—not surprisingly, water in its liquid state is notoriously difficult to model. Many complex biomolecules (DNA, RNA, proteins, etc.), vital for a living cell’s function, cannot be studied alone without considering their environment—water as the solvent [121]. To study those large biomolecules, atomistic simulations have been widely used, and numerous different water models [85] have been developed to reproduce water properties, including the class of water models most widely used today—simple, rigid, fixed-charge explicit models such as TIP3P [54], TIP4P [1, 52], TIP5P [75] and SPC/E [9]. According to the convention, these classical water models are distinguished by the number of “points”—interaction sites—in them: 3-point models (Fig. 2.1a, with 3 point charges) such as TIP3P are the most common due to what is perceived by many as an acceptable balance between accuracy and computational cost (compared with 4 or 5-point models). For water models with more points, 4-point models (Fig. 2.1b, with 3 point charges and neutral oxygen) such as TIP4P, and especially 5-point models (Fig. 2.1c, 2.1d, with 4 point charges and neutral oxygen) models such as TIP5P (Fig. 2.1d), the cost considerations become significant, even though these models tend to reproduce water properties better than their 3-point counterparts. However, despite decades of effort by many groups, none of the existing simple water models is perfect [31, 39, 85, 86, 87, 108].

It is known that liquid water properties are determined by a complex network of hydrogen bonds. In these classical water models, hydrogen bonds are mimicked mostly by the electrostatics, complemented by a Lennard-Jones (LJ) potential. The latter—LJ potential—is generally represented by a single site centered on the oxygen, and its corresponding interaction is isotropic and featureless, in contrast to hydrogen bonding, which is directional. Thus, accurately representing electrostatic interactions is paramount for a classical water model
Figure 2.1: **Rigid n-point water model geometries for n=3, n=4 and n=5.** Hydrogen, Oxygen and the extra interaction point with a point charge but no mass are represented as white, red and pink spheres respectively. 4, and 5-point water models have zero charge on their oxygen, the negative charge is placed on the extra point(s). (a) 3-point water model. (b) 4-point water model. (c) and (d) Different possible non-planar configurations of 5-point water models.

to mimic hydrogen bond interactions and reproduce liquid water properties [47]. Therefore, our first question is how accurately can we describe the reference electrostatics within the unavoidable limitations of n-point models, and how far it can get us in terms of water model accuracy, by optimizing the electrostatics alone. This is a non-trivial question since, by construction, these simple n-point rigid models miss important physics, such as molecular flexibility, electronic polarizability, and charge transfer.

We note that a good water model intended for biomolecular simulations must reproduce experimental properties of water with high accuracy. There are at least two reasons why existing water models are far from perfect in this respect. One is the limitations of the optimization strategies used in constructing water models. The other reason is on a more fundamental level—the limited physics built into these simple models. As an example of the first kind of limitation, current widely used rigid 3-point models (TIP3P, SPC, SPC/E, etc.) are based on an assumption that the experimental water molecule geometry is somehow optimal, or near optimal, for a classical water models—consequently, these models place the point charges on, or near, the centers of hydrogen and oxygen atoms. While sophistication of the optimization techniques employed to find the optimum has grown tremendously [2], from essentially “guess-and-test” to the complex, state-of-the-art force balance optimizations [114],
one crucial aspect of the overall procedure has not changed until recently: the search for best fit model is performed in the vicinity of the “canonical” water geometry, thus returning a \textit{local optimum} in the parameter space. It was recently demonstrated \cite{47} that abandoning the restrictions (except fundamental $C_{2v}$ symmetry) on water model geometry, and performing an exhaustive search for a truly \textit{global} optimum in the parameter space can result in significant accuracy improvements. The resulting 4-point rigid model—OPC \cite{47} was built without the geometry restriction on point charge placements. It was optimized globally to reproduce bulk water properties as best as possible, without any increase of the computational cost of employing the model in simulations, relative to common 4-point models such as TIP4P-Ew.

Apart from using global optimization, a natural strategy to improve the accuracy of a rigid $n$-point water model is to consider larger “$n$”. The key question with this approach is whether the inevitable and substantial increase in computational cost of employing “larger n” water models is justified by significantly better accuracy? If the accuracy gain is large enough, the cost increase may still be well worth it. Answering the question is important both for model developers and practitioners alike. Unfortunately, it is nearly impossible to compare existing $n$-point water models on the same footing to find out exactly how much gain an increase in $n$ brings. For once, these models are optimized using different criteria, and against non-identical set of water properties. For example, TIP5P-Ew model \cite{91} surpasses many of its 3- and 4- point predecessors in accuracy of describing water structure, but its accuracy of reproducing the self-diffusion is lower than that of TIP4P-Ew. In addition, optimization protocols vary widely, and, except for OPC-family models, there is no guarantee that the model corresponds to the global optimum in the optimization landscape. Thus, it is entirely possible that a global optimum for a smaller $n$ may yield a more accurate model than a local one for a larger $n$.

Even if a hypothetically perfect fixed-charge rigid model reproduced a large subset proper-
ties of bulk liquid water exactly, the model would still be inherently incapable of responding to the change of polarity of its micro-environment, relevant to biomolecular simulations. Specifically, water is highly polarizable \([103]\): the experimentally observed change in the dipole moment of real water molecule upon transfer from gas to liquid phase is as large as 1 Debye, while for any rigid fixed-charge model that change is zero by construction. The polarity of micro-environment near a macromolecule can be quite different from that of bulk water; non-polarizable models cannot properly respond to different micro-environments during the course of a simulation \([90]\), e.g. in cross-membrane transport. However, the vast majority of current water models lack polarizability, which must adversely affect the accuracy of biomolecular simulations. This example illustrates the second type of limitation on water model accuracy — the missing physics. To address the lack of electronic polarization, a number of polarizable water models have been developed, for example POL1 \([16, 26]\), FF12POL \([111]\), SWM model family \([72]\) and AMOEBA model family \([70, 73, 113]\), see recent reviews for a comprehensive account of the field \([43, 51, 85]\). There is little doubt that, in principle, availability of highly accurate and efficient polarizable models for routine practical simulations should improve the accuracy of biomolecular simulations significantly, by accounting for effects completely missed by rigid fixed-charge models. However, because of the unavoidable higher complexity of polarizable models relative to the fixed-charge ones, the balance towards wider adoption can only be tipped by distinctly higher accuracy, which does not seem to be the case yet \([109, 111]\). Since current polarizable water models typically use existing fixed-charge models as their “base”, they likely inherit at least some of the existing flaws of the fixed-charge models, unrelated to their lack of polarizability. That is they are likely represented by local optima in the complex optimization landscape. Indeed, despite undeniably better physical foundation, accuracy of a sophisticated polarizable water model can be lower than that of a globally optimal non-polarizable rigid model \([47]\). This observation begs the question: how much accuracy gain can the inclusion of polarizability
bring, if the above optimization-related limitations were removed, and a truly globally optimal polarizable model was constructed? In other words, is global optimization a potentially useful strategy for polarizable models? For reasons discussed above in the context of comparing non-polarizable models, comparing existing \( n \)-point non-polarizable with polarizable models (even with respect to liquid water properties alone) on the same footing is difficult. Moreover, to the best of our knowledge, a truly globally optimal polarizable model has not be constructed yet.

The main motivation of this work is to address the questions outlined above in a tightly controlled computational experiment. To this end, here we build and evaluate several test water models of two types—rigid \( n \)-point globally optimal models \((n = 3, 4, 5)\) and also polarizable models based on these rigid models—that illustrate the two limitations discussed above. We also construct a truly globally optimal polarizable test model. All models are optimized using the same type of protocols, and to the same accuracy level in reproducing the training set. The models are evaluated by examining their ability to reproduce water dimer, which can be considered as the simplest possible mimic of water in condensed state.

### 2.2 Methods

As mentioned above in the Introduction, liquid water properties are determined by a complex network of hydrogen bonds, which is the most challenging part a water model needs to mimic in order to simulate real water. Water dimer, involving only two water molecules and their interactions, is a good minimal starting point to reveal the secrets of hydrogen bonding, and has been studied over half a century [79, 80]. Due to its significance to hydrogen bonding studies, and being the simplest possible mimic of the liquid phase water, the water dimer structure is utilized here to evaluate the quality of the water models.
2.2.1 Overall approach

The overall approach is as follows. Individual water model’s geometry and values of point charges are optimized to match as closely as possible a set of reference multipole moments (Table 2.1, below). The matching is done in the following precise sense. Through optimal point charge approximation (OPCA, a method to approximate electrostatic charge distributions with a small number of point charges to optimally represent the original charge distribution) \[4\], each water model has its quadrupole and octupole moments fitted to the reference, and the monomer dipole moment is optimized so that when the two water models form a dimer, the dimer total dipole has a smaller than 0.1% error relative to \textit{ab initio} calculation—2.68D \[119\]. Then the water model’s accuracy is evaluated by comparing its dimer’s geometry (Fig.2.2) with the one obtained from an \textit{ab initio} calculation \[58\].

![Figure 2.2: Schematic of a water dimer.](image)

Figure 2.2: \textbf{Schematic of a water dimer.} $r$(OO) is the distance between the two oxygens; \(\theta\) and \(\phi\) are the angles between water molecule’s $C_2$ axis and the O-O line for the two molecules respectively.

As discussed in the introduction, a water model’s accuracy of simulating liquid water is largely determined by its electrostatic properties, small variations in these properties can lead to large differences in simulation outcomes \[5\]. Also, the “electrostatic” parameters are where different types of water models differ the most from each other, while the Van der Waals potential is typically simulated in the same manner: by a Lennard-Jones (LJ) site on oxygen. All of model types studied here follow that approach. In reality, optimization of the LJ parameters can be largely decoupled from optimization of the “electrostatic” parameters.
2.2. Methods

such as partial charges and their distribution [47]. An accepted approach to construct realistic water models involves optimizing the LJ parameters against reference O-O RDF (radial distribution function), to obtain a close match [47]; we follow essentially the same strategy here, except that we use *ab-initio* reference oxygen-oxygen distance \( r(OO) = 2.91 \) [58, 119] as the target reference for optimization of LJ parameters. With the \( r(OO) \) distance used in LJ optimization, the remaining parameters of the dimer geometry—the two angles \( \theta \) and \( \varphi \) in Fig.2.2)—are left as convenient metrics to evaluate the model accuracy, which is mainly determined by the electrostatics.

From the perspective of electrostatics, a model with larger \( n \) (number of interaction points) is expected to better reproduce given multipole moments, and thus improve the accuracy of the model [28, 36, 93, 100].

For a set of \( N \) point charges, in this case—the water model, the Coulomb potential can be written as:

\[
\varphi(R) = \frac{1}{4\pi \epsilon_0} \sum_{n=1}^{N} \frac{q_n}{||R - r_n||} \tag{2.1}
\]

In a Cartesian system, this equation becomes:

\[
\varphi(R) = \frac{1}{4\pi \epsilon_0} \left( \frac{1}{R} q + \frac{1}{R^2} \sum_{i=x,y,z} \hat{R}_i \hat{\mu}_i + \frac{1}{R^3} \sum_{i,j=x,y,z} \hat{R}_i \hat{R}_j Q_{ij} + \frac{1}{R^4} \sum_{i,j,k=x,y,z} \hat{R}_i \hat{R}_j \hat{R}_k O_{ijk} + \ldots \right) \tag{2.2}
\]

\[
q = \sum_{n=1}^{N} q_n \tag{2.3}
\]
CHAPTER 2. EXPLORING OPTIMIZATION STRATEGIES FOR IMPROVING EXPLICIT WATER MODELS: RIGID N-POINT MODEL AND POLARIZABLE MODEL BASED ON DRUDE OSCILLATOR

\[ \mu_i = \sum_{n=1}^{N} q_n r_{n,i} \]  
(2.4)

\[ Q_{ij} = \frac{1}{2} \sum_{n=1}^{N} q_n (3r_{n,i}r_{n,j} - (r_n)^2 \delta_{ij}) \]  
(2.5)

\[ O_{ijk} = \frac{1}{6} \sum_{n=1}^{N} q_n \left[ 15r_{n,i}r_{n,j}r_{n,k} - 3(r_n)^2 (r_{n,i} \delta_{jk} + r_{n,j} \delta_{ik} + r_{n,k} \delta_{ij}) \right] \]  
(2.6)

Because of the \( C_2 \) symmetry of a water molecule, the quadrupole moments and the z-plane elements of octupole moments can be written as:

\[
Q = \begin{bmatrix}
-Q_t - Q_0/2 & 0 & 0 \\
0 & Q_t - Q_0/2 & 0 \\
0 & 0 & Q_0 \\
-\Omega_t - \Omega_0/2 & 0 & 0 \\
0 & \Omega_t - \Omega_0/2 & 0 \\
0 & 0 & \Omega_0 
\end{bmatrix}
\]  
(2.7)

\[
O_{ijz} = \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & \Omega_0 
\end{bmatrix}
\]  
where \( i, j \) range over \( x, y \) and \( z \)  
(2.8)

With the equations set up, we can then rearrange them to the form where water model parameters (coordinates and charge) are explicitly expressed via multipole moments [47].

For a 3 or 4-point water model, the rearranged equations become:
2.2. Methods

\[ z_{1,2} = \frac{2Q_t + 3Q_0}{6\mu} \pm \frac{\mu}{4q} \] \hfill (2.9)

\[ y = \sqrt{\frac{2Q_t}{3q}} \] \hfill (2.10)

\[ q = -3\sqrt{\frac{\mu^4(256Q_t^2 + \xi) + 16Q_t\mu^2}{2\xi}} \] \hfill (2.11)

where \[ \xi = 52Q_t^2 + 60Q_tQ_0 - 9[3Q_0^2 + 8(\Omega_t - \Omega_0/2)]\mu \] \hfill (2.12)

A 4-point water model is shown in Fig. 2.3b. With the \( C_2 \) symmetry, only 4 parameters are needed to construct the 4-point water model—\( y_1, z_1, z_2 \) and electric charge \( q \). Both 3 and 4-point models have 3 point charges, so the calculation process are the same (Eq. 2.9-2.12) except for 3-point models the condition \( z_1 = 0 \) needs to be imposed (Fig. 2.3a).

Figure 2.3: Schematics of 3, 4 and 5-point rigid water models. Each circle represents an interaction point (H, O atom, or extra point–EP). Thick lines are chemical bonds. The origin point is on the Oxygen; \( C_2 \) symmetry axis of water is set as the z-axis; y-axis is in the H-O-H plane and passes through O; x-direction is only accounted for in a 5-point model, and the 2 extra interaction points (EP) are placed on the \( xOz \) plane. Electric charge on H is \( q \), charge \(-2q\) is placed on O for 3-point model, on EP for 4-point model, and divided evenly on the two EPs for 5-point model. (a) 3-point water model. (b) 4-point water model. (c) 5-point water model.

For a 5-point model, there are 5 parameters to be determined (\( q, x_1, y_1, z_1, z_2 \), see Fig. 2.3c).
From Eq. (2.3-2.6), we have:

\[ \mu = 2q(z_2 - z_1) \]  
\[ Q_t = \frac{3}{2}q(x_1^2 + y_1^2) \]  
\[ Q_0 = q(x_1^2 - y_1^2 - 2z_1^2 + 2z_2^2) \]  
\[ \Omega_t = \frac{5}{2}q(x_1^2z_1 + y_1^2z_2) \]  
\[ \Omega_0 = q(3x_1^2z_1 - 3y_1^2z_2 - 2z_1^3 + 2z_2^3) \]

With Eq. (2.13-2.16), a series of expression \((x_1, y_1, z_1, z_2) = f(q, \mu, Q_t, Q_0, \Omega_t)\) can be derived:

\[ x_1 = \sqrt{-36\Omega_t\mu - 30Q_0Q_tq + 20Q_t^2q + 15Q_t\mu^2 \over q(60Q_tq + 45\mu^2)} \]  
\[ y_1 = \sqrt{36\Omega_t\mu - 30Q_0Q_tq + 20Q_t^2q + 15Q_t\mu^2 \over q(60Q_tq + 45\mu^2)} \]  
\[ z_1 = {48\Omega_tq^2 + 30Q_0\mu q - 20Q_t\mu q - 15\mu^3 \over q(80Q_tq + 60\mu^2)} \]  
\[ z_2 = {48\Omega_tq^2 + 30Q_0\mu q + 20Q_t\mu q + 15\mu^3 \over q(80Q_tq + 60\mu^2)} \]

After choosing a set of multipole data \((\mu, Q_t, Q_0, \Omega_t, \Omega_0)\), insert \(\mu, Q_t, Q_0, \Omega_t\) values into Eq. (2.18-2.21) so that we can then express the coordinates \((x, y, z_1, z_2)\) with one argument—\(q\) the charge value. And now, what is left is a very simple optimization problem: finding the optimal \(q\) that makes the \(\Omega_0\) error of the model as small as possible with exhaustive search. Then with the optimal \(q\) and its corresponding coordinates \((x, y, z_1, z_2)\), a 5-point water model is then constructed and is ready for evaluation.
2.2. Methods

Evaluation of the built water models starts with the energy minimization calculation of two identical water molecules that form a water dimer, using AMBER 2019 [20]. Steepest descent, conjugate gradient method (for 3, 4-point rigid models) and Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) quasi-Newton algorithm (for 5-point rigid models and 3-point Drude models) are used in this minimization calculation. With the energy minimization process, the LJ parameters on the oxygen of each model are determined so that the minimized water dimer has an oxygen-oxygen distance matching the reference (2.91Å). With the oxygen-oxygen distance controlled, the dimer system has a very limited number of variables (rotation angles), therefore finding the global minimum is straightforward. To assert it, we ran multiple minimizations with random starting coordinates, and found that all the minimized states were the same. Thus, after minimization, the dimer system of the constructed water model reaches its global lowest energy state. The total dipole moment of the minimized water dimer is then calculated and compared with the ab initio result (2.68D [119]). If the relative error of the total dipole moment is greater than 0.1%, the monomer dipole moment will be changed, we then redo the model building and the dimer minimization. This optimization process is continued until a monomer dipole value is found when the total dipole moment of the minimized dimer has a relative error within 0.1%.

As a metric to evaluate the geometry of water dimer, two parameters are used—θ and φ. θ and φ are the angles between water molecule’s C₂ axis and the O-O line for the two molecules respectively. θ is of the water molecule coplanar with O-O, φ is of the other one non-coplanar with O-O (Fig.2.2). We use an ab initio calculation of a water dimer as the reference geometry, where \( \theta_0 = 57.9^\circ, \varphi_0 = 55.6^\circ \) and \( r(OO) = 2.91 \). [58, 119] This ab initio calculation is done by Klopper et al. in 2000 [58], optimized at the level of CCSD(T) theory with IO249 basis. It is seen as a benchmark result of theoretical studies on water dimer. [80]
2.2.2 Reference multipole moment sets

To ascertain robustness of our methodological conclusions, we use three different sets of multipole moments as references to which we fit our water models (Table 2.1). The monomer dipole, quadrupole and octupole moments of gas phase QM calculation (CCSD with aug-cc-pCVTZ basis) are from Anandakrishnan et al [4]. The gas phase experimental data are from Clough et al [22]. When used to construct water models, the octupole moments from the gas QM set are used in the gas phase experimental set (whose octupole moments are not available). The third set of reference multipole moments is of liquid phase water, a QM calculation done by Niu et al [83].

Table 2.1: Reference multipole moments of single water molecule.

<table>
<thead>
<tr>
<th>Data sets</th>
<th>( \mu / \text{D} )</th>
<th>( Q_t / \text{DÅ} )</th>
<th>( Q_0 / \text{DÅ} )</th>
<th>( \Omega_t / \text{DÅ}^2 )</th>
<th>( \Omega_0 / \text{DÅ}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas phase experimental [4]</td>
<td>1.86</td>
<td>2.57</td>
<td>0.11</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gas phase QM calculation [4]</td>
<td>1.81</td>
<td>2.49</td>
<td>0.08</td>
<td>1.93</td>
<td>-1.35</td>
</tr>
<tr>
<td>Liquid phase MP2/4MM calculation [83]</td>
<td>2.49</td>
<td>2.93</td>
<td>0.13</td>
<td>2.09</td>
<td>-1.73</td>
</tr>
</tbody>
</table>

Each set of reference multipole moments is used to construct a set of optimal water models comprising of 3-point, 4-point and 5-point rigid models. These three sets of water models are called “gas exp.”, “gas QM” and “liquid MP2/4MM” respectively, based on the dataset used for model construction. The “gas exp” reference multipole moments are also used to construct polarizable models in this study.

2.2.3 n-point rigid models

Optimized against each of the three multipole moment data sets (“gas exp.”, “gas QM” and “liquid MP2/4MM”), three rigid models were constructed with 3, 4 and 5 interaction points respectively, following the process shown in Fig. 2.4. Parameters of these rigid models are listed in Table. 2.2.
For each $n$-point rigid water model built, its monomer dipole moment $\mu$ value is varied for optimization. When the monomer dipole moment changes, the resulting water model parameters (coordinates and charges) change, as a result, the water dimer of this model will be different. An optimal monomer dipole moment is chosen when the total dipole of the corresponding dimer is within 0.1% relative error from the \textit{ab initio} calculation [119].
Table 2.2: Parameters of 3, 4 and 5-point optimal water models, and their geometry (Fig. 2.1). For comparison, parameters of several existing models are shown in the bottom rows.

<table>
<thead>
<tr>
<th>Name or reference</th>
<th>n</th>
<th>Type (Fig. 2.1)</th>
<th>( q [e] )</th>
<th>( z_1 [\text{Å}] )</th>
<th>( z_2 / [\text{Å}] )</th>
<th>( y_1 [\text{Å}] )</th>
<th>( x_1 [\text{Å}] )</th>
<th>( \mu [\text{D}] )</th>
<th>( \sigma_{LJ} [\text{Å}] )</th>
<th>( \epsilon_{LJ} [\text{kJ/mol}] )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gas exp.</strong></td>
<td>3</td>
<td>a</td>
<td>0.2202</td>
<td>—</td>
<td>0.9284</td>
<td>1.2727</td>
<td>—</td>
<td>1.964</td>
<td>3.3320</td>
<td>0.7113</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>b</td>
<td>1.9094</td>
<td>0.3351</td>
<td>0.4601</td>
<td>0.4322</td>
<td>—</td>
<td>2.293</td>
<td>3.2999</td>
<td>0.7196</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>c</td>
<td>0.8633</td>
<td>0.1862</td>
<td>0.4723</td>
<td>0.6179</td>
<td>0.1773</td>
<td>2.373</td>
<td>3.2945</td>
<td>0.7740</td>
</tr>
<tr>
<td><strong>Gas QM</strong></td>
<td>3</td>
<td>a</td>
<td>0.2217</td>
<td>—</td>
<td>0.9039</td>
<td>1.2486</td>
<td>—</td>
<td>1.925</td>
<td>3.2999</td>
<td>0.7573</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>b</td>
<td>1.4346</td>
<td>0.3073</td>
<td>0.4698</td>
<td>0.4908</td>
<td>—</td>
<td>2.239</td>
<td>3.2910</td>
<td>0.7071</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>c</td>
<td>0.9194</td>
<td>0.2184</td>
<td>0.4769</td>
<td>0.5989</td>
<td>0.1314</td>
<td>2.283</td>
<td>3.2749</td>
<td>0.7782</td>
</tr>
<tr>
<td><strong>Liquid MP2/4MM</strong></td>
<td>3</td>
<td>a</td>
<td>0.2261</td>
<td>—</td>
<td>0.9795</td>
<td>1.3413</td>
<td>—</td>
<td>2.127</td>
<td>3.3854</td>
<td>0.7448</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>b</td>
<td>2.7937</td>
<td>0.3679</td>
<td>0.4615</td>
<td>0.3815</td>
<td>—</td>
<td>2.512</td>
<td>3.3516</td>
<td>0.7364</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>c</td>
<td>0.5833</td>
<td>0.0234</td>
<td>0.5076</td>
<td>0.7632</td>
<td>0.3386</td>
<td>2.713</td>
<td>3.3801</td>
<td>0.7699</td>
</tr>
<tr>
<td><strong>TIP3P</strong></td>
<td>3</td>
<td>a</td>
<td>0.4170</td>
<td>—</td>
<td>0.5859</td>
<td>0.7570</td>
<td>—</td>
<td>2.35</td>
<td>3.1506</td>
<td>0.6364</td>
</tr>
<tr>
<td><strong>OPC3</strong></td>
<td>3</td>
<td>a</td>
<td>0.4476</td>
<td>—</td>
<td>0.5652</td>
<td>0.7992</td>
<td>—</td>
<td>2.43</td>
<td>3.1743</td>
<td>0.6837</td>
</tr>
<tr>
<td><strong>OPC</strong></td>
<td>4</td>
<td>b</td>
<td>0.6791</td>
<td>0.1594</td>
<td>0.5395</td>
<td>0.6856</td>
<td>—</td>
<td>2.48</td>
<td>3.1666</td>
<td>0.8904</td>
</tr>
</tbody>
</table>

2.2.4 Polarizable water model with Drude oscillator

There are several different approaches available to add electronic polarizability to a “rigid base” water model, adding the Drude oscillator [29, 99] being one of the simplest methods. The Drude oscillator is easy to implement in most MD simulation packages. Critically, it only adds one more degree of freedom for optimization, which will not impair the search for the global optimum. Compared with other approaches, the Drude approach also has the benefit of high computational efficiency [43, 45, 50, 66, 67, 71]. Here the polarizable water model is constructed by adding a Drude particle to the oxygen in a 3-point rigid water model (Fig. 2.5, the Drude particle has no interactions with the 2 hydrogens in the same molecule). It is referred to as a “3-point polarizable model”. The polarizability of this model is

\[
\alpha = \frac{Q_D^2}{k} \quad (2.22)
\]
where \( k \sim 1000 \text{ kcal/mol}^2 \) is the force constant of the virtual bond connecting the Drude particle and the Oxygen [69], \( Q_D \) is the charge of the Drude particle. In this study, we apply \( k = 1000 \text{ kcal/mol}/\text{Å}^2 \) and keep it constant. The charge on the Drude particle is set to

\[
Q_D = \sqrt{\alpha k} \times 18.2223 \text{ e}
\] (2.23)

where the unit of \( \alpha \) is \(^3\). The coordinates and charges of the fixed three points are calculated with the same OPCA method used in constructing 3-point rigid models.

Based on “gas exp.” data set (Table.2.1), we built two polarizable models: the “limited optimal” one with only polarizability optimized for fair comparison with rigid models; and the “globally optimal” one with both polarizability and gas phase monomer dipole moment optimized for practical evaluation.
Optimizing the 3-point polarizable water model

For the limited optimal 3-point polarizable model, we fixed the monomer dipole moment to experimental value–1.86D (thus the rigid base is fixed), and optimized it by varying polarizability until its dimer dipole reproduces experimental result with in 0.1% error, the procedure is shown in Fig. 2.6. The optimization and the resulting model are shown in Fig. 2.7 and Table. 2.3, respectively. The resulting polarizability of the limited optimal polarizable model is 0.94Å³, 35% lower than the gas phase experiment value–1.44Å³ [69].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>3</td>
<td>0.1975</td>
<td>1.6825</td>
<td>0.9803</td>
<td>1.3439</td>
<td>0.94</td>
<td>1.86</td>
<td>3.3836</td>
<td>0.7113</td>
</tr>
<tr>
<td>α &amp; μ</td>
<td>3</td>
<td>0.1700</td>
<td>2.6261</td>
<td>1.0565</td>
<td>1.4484</td>
<td>2.2900</td>
<td>1.7258</td>
<td>3.3786</td>
<td>1.0435</td>
</tr>
</tbody>
</table>

For the globally optimal 3-point polarizable model, we modify the model by changing two input parameters—monomer dipole moment μ and polarizability α (both in the vicinity of their gas phase experimental value, \(μ_{exp} = 1.86D\) and \(α_{exp} = 1.44Å³\)), to make the model’s dimer total dipole moment fall within 0.1% error from reference [58]. Error distributions of the intermediate models during this optimization process are shown in Fig. 2.8, with respect to the two varying parameters.

The resulting globally optimal 3-point polarizable model is constructed with polarizability \(α = 2.2900Å³\) and monomer dipole moment \(μ = 1.7258D\) (Table. 2.3). This model has a total angle error of 3.19° when simulating a water dimer. The monomer dipole moment (1.7258D) of this polarizable model is relatively close to the gas phase experimental data (1.86D), while its polarizability (2.275Å³) is larger than the gas phase experimental value (1.44Å³).
2.3. Results and Discussion

For biomolecular simulations, the most important state of water is the liquid state at ambient temperature and pressure. However, the number of experimental parameters used to characterize liquid water is vast, and existing practical procedures used to build water models to match a subset of these parameters are highly diverse, non-trivial and expensive [85]. The
Figure 2.7: Dimer total dipole moment and total angle error of the limited optimal 3-point polarizable model. With the water monomer dipole fixed at 1.86D, we test different 3-point polarizable models by varying the polarizability from 0.10Å³ to 2.10Å³. The blue diamond symbols represent the total dipole errors and the orange triangle symbols show the total angle errors, where both are compared with the \textit{ab initio} reference [58, 119].

Figure 2.8: Globally optimal polarizable model error distribution. (a) The absolute error of the total angle ($\theta + \varphi$) as a function of the corresponding model’s monomer dipole $\mu$ and polarizability $\alpha$ value. (b) The water dimer total dipole moment error with respect to the model’s $\mu$, $\alpha$ combination. (c) The approximate range of the optimal $\mu$, $\alpha$ combinations, where the “combined relative error” is defined as $Error_{combined} = Error_{angle}/180^\circ + Error_{dipole}/p_{dimer}$ with $p_{dimer} = 2.68D$ [119]. The “combined relative error” was not used in the actual optimization process.

process is still somewhat of an art. For a strict, limiting-case apples-to-apples comparison we want to present here, using liquid state properties is thus not ideal, if not completely impossible. We therefore resort to using a mimic of condensed state of water—water dimer. Water dimer is the smallest form of a water cluster. It has a hydrogen bond, which greatly contributes to numerous anomalies in water properties. As a starting point to study water
properties, water dimer has been used extensively [74, 79, 80]. It has also been employed as a standard reference to test water models [9, 47, 52, 54, 70, 72, 75, 111, 113].

Our over-all strategy is as follows (see Methods for details). For n-point rigid models, we build each model to match, as closely as possible, the reference quadrupole, octupole moments of water monomer (Table 2.1), and a monomer dipole moment optimized for water dimer. For polarizable models, its rigid “base” is built first to match the reference multipole moments including dipole, and then the polarizable model is built by adding a Drude oscillator with an optimal polarizability value—optimized for water dimer—to the rigid “base”. Parameters of the “base” are not re-optimized. The optimization of the last 3-point polarizable model is slightly different: it involves varying both the monomer dipole and the polarizability, such that both its “base” and Drude oscillator are optimized. Regardless of what parameters are optimized, the standard of the optimization is the same for all these models—the total dipole moment of the water model’s dimer matches the ab initio reference value for water dimer dipole [58] within a given tolerance.

We then test each model’s ability to reproduce the geometry of a water dimer formed by the two water model molecules. The reference dimer geometry is from the benchmark ab initio calculation by Klopper et al. [58, 119] For reasons detailed in “Methods”, we keep the O-O distance in the dimer fixed to its reference value, thereby excluding oxygen LJ parameters from the consideration.

2.3.1 Rigid n-point water models

First we explore accuracy of rigid non-polarizable water models with 3, 4 and 5 point charges. The model parameters are given in Table 2.2. Three sets of n-point (n = 3, 4, 5) rigid water models are included. These three sets of water models are constructed based on 3 different
Figure 2.9: **Errors in reproducing the ab initio water dimer geometry by the rigid water models.** Our models’ $\theta$ and $\varphi$ angle errors (reference values—$\theta_0 = 57.9^\circ$ and $\varphi_0 = 55.6^\circ$—are from the benchmark ab initio calculation by Klopper et al. [58]) are represented with blue and orange connected circles. In (a), (b) and (c) are the results of models built with gas exp., gas QM and liquid MP2/4MM multipole moments (Table. 2.1) respectively. Water dimers of TIP3P, OPC3 and OPC model are energy minimized with O-O distance fixed at 2.91Å, and their angle errors in the dimers are shown as square, diamond and triangle respectively.

sets of multipole moments data, see “Methods”.

Accuracies of these models with respect to reproducing our main reference are compared in Fig. 2.9. For reference, TIP3P-, OPC3- and OPC-based dimers are also tested and included in Fig. 2.9. The water dimer geometry error values of the 4- and 5-point models are considerably smaller than those of 3-point models. The remaining errors are at $\sim 5^\circ$ level, well below the $\pm 10^\circ$ experimental error margin of the reference values [119]. Note that this is the result of optimizing the electrostatics only. Additionally, our models have smaller errors than the existing models (TIP3P, OPC3 and OPC, Fig. 2.9), which is not unexpected because the latter ones were optimized for liquid water instead of the water dimer. This relatively high accuracy of the “electrostatics only” optimization emphasizes the critical importance of getting right the distribution of charge in a water model. The result also suggests that one can improve a rigid water model significantly by optimizing its electrostatics, even though some physics is still missing from the model.

The dimers of 4-point water model show notably better agreement with the reference data.
than the dimers based on 3-point model—error of the two key angles reduced to $\sim 5^\circ$ from $\sim 20^\circ$. This improvement of accuracy from 3-point to 4-point model is robust, seen for all reference sets used to build these models. However, further increasing $n$ from 4 to 5 only has minimal effect on improving the agreement with the reference dimer geometry. The 5-point models with “gas exp.” and “gas QM” data improve from their 4-point counterparts with marginal changes ($<1^\circ$). The “liquid MP2/4MM” 5-point model even has a slightly inferior accuracy compared with its 4-point model (error is $1^\circ$ larger). Thus, adding more points beyond $n = 4$, which increases the accuracy of reproducing multipole moments, $n$-point rigid models still cannot improve much from what a 4-point model can already do. As seen from Fig.2.9, all of the above conclusions are independent of the specific reference multipole set used.

### 2.3.2 Limited optimal 3-point polarizable model—only polarizability optimized

One obvious physical feature missing from rigid $n$-point water models is electronic polarizability [7, 62, 63, 70, 78, 82, 99, 102]. This missing physics may be the reason why there is little, if any, improvement in the accuracy of the rigid model in reproducing water dimer as $n$ is increased beyond $n = 4$, Fig.2.9. How much of an accuracy improvement can be achieved by adding the missing polarizability to the model? To answer this question, we compare two different approaches of improving rigid water models.

So far, in the rigid models we built, the geometry and point charge values are optimized (with OPCA method [4], by varying the monomer dipole moment of the water model) specifically for water dimer to compensate the missing gas-to-condensed-phase polarization. Now, we add polarizability to a 3-point model, whose monomer multipole moments (including the
dipole) fit the reference for the monomer, then optimize the added polarizability against wa-
ter dimer to the same standard—error of dimer total dipole moment smaller than 0.1%. This way, an apple-to-apple comparison is established between the two approaches to improve a rigid water model: 1) adding the missing physics of electronic polarizability; 2) compensating for its absence by optimizing the electrostatics. Either approach has only one parameter for optimization so they are competing on a level field.

Parameters of the resulting polarizable model are shown in Table 2.3. The accuracy of the resulting 3-point limited optimal polarizable model is compared to that of our rigid models in Fig. 2.10.

As illustrated in Fig. 2.10, this limited globally optimal 3-point polarizable model has $\sim 5^\circ$ smaller total angle error than the 3-point rigid model. This means that under the same optimization standard, adding and optimizing polarizability (while keeping the monomer dipole moment fixed to its gas phase value) is more beneficial than compensating the lack of polarization by only optimizing the dipole moment of the rigid model.

Moreover, because the monomer dipole moment is fixed to the gas phase value in this limited optimization process, the resulting polarizable model reproduces correct gas phase multipole moments when in gas phase, unlike rigid models whose dipole moment is changed for condensed phase optimization. Also, the optimized polarizability is 0.94Å$^3$, 35% lower, but still qualitatively similar to the gas phase experimental value—1.44Å$^3$.

However, this limited global optimal 3-point polarizable model is still inferior to the 4- and 5-point rigid models with quite a margin ($\sim 10^\circ$ higher total angle error). Therefore, adding polarizability alone cannot fix the defects of suboptimal electrostatics of a 3-point model.

With the computational cost in mind (in real simulations, 3-point polarizable model is even slower than the corresponding 4-point rigid models), this result is not very satisfactory. It
2.3. Results and Discussion

Figure 2.10: Absolute errors of the two key angles ($\theta$ and $\varphi$, Fig.2.2) in a water dimer made of rigid or Drude polarizable model molecules. These errors are calculated with reference to the ab initio calculation values ($\theta_0 = 57.9^\circ$, $\varphi_0 = 55.6^\circ$) [58, 119]. Results of the 3, 4 and 5-point rigid water model dimers are built based on the “gas exp.” reference (Table.2.1), same with the polarizable models.

suggests that the optimization protocol in which the fixed-charge base model is ignored and only polarizability is optimized has significant limitations.
2.3.3 Globally optimal 3-point polarizable model—both the polarizability and the “rigid base” are optimized.

So far, we have focused on purely conceptual questions about water models’ limitations in reproducing liquid water, mimicked here by water dimer. In practical MD simulations, water models are mostly used to simulate the solvent in liquid phase. Therefore from practical perspective of water model design, we do not have to keep the rigid base of a polarizable model unoptimized. Also, the results of the limited optimal 3-point polarizable model make it clear that only optimizing polarizability is not sufficient to improve a water model’s accuracy beyond the rigid \( n \)-point model’s limitations. Therefore, here we build a globally optimal 3-point polarizable model, in which two parts are both optimized for the water dimer scenario: the geometry and point charge values of the rigid base (through OPCA \([4]\), varying the monomer dipole); and the polarizability introduced by the Drude oscillator.

As discussed in “Methods” section, under the same optimization standard (reproducing dimer total dipole) a polarizable model has one more degree of freedom—the added polarizability—than the rigid models. By optimizing both its rigid base and its polarizability, we are utilizing this additional degree of freedom afforded by the polarizable model. In other words, this 3-point polarizable model is optimized with two parameters: polarizability(\( \alpha \)) and gas phase monomer dipole moment(\( \mu \)). The result is shown in Fig.2.10. In comparison with the 3, 4, 5-point rigid models and the limited optimal polarizable model, this globally optimal polarizable model’s result is shown along side with them as the rightmost column.

With both polarizability and gas phase monomer dipole moment optimized, the dimer geometry error of this fully optimal polarizable model shows a significant improvement from the limited optimal polarizable model, and is more accurate than the corresponding 4- and 5-point rigid models—the \( \varphi \) error is very close to zero and the \( \theta \) error is at the same level.
or smaller than that of all other models. Not only the accuracy of the globally optimal polarizable model is considerably higher than that of all rigid models and the other limited optimal polarizable model, it is worth noting that this globally optimal polarizable model has a gas phase monomer dipole moment of 1.7258D, not far from the gas phase experimental value of 1.86D. The resulting polarizability is 2.29Å$^3$, which is considerably larger than the experimental 1.44Å$^3$. This deviation of the optimized polarizability value points to remaining deficiencies of the specific polarizable water model, likely that the model is too simplistic or/and that other physical effects not explicitly considered here (e.g. charge transfer) play a noticeable role. For the 3-point polarizable water model optimized for water dimer, decreasing the monomer dipole moment by 7% (from 1.86D to 1.7258D) resulted in a 144% polarizability increase (from 0.94Å$^3$ to 2.29Å$^3$). This implies that the water dimer properties are less sensitive to the model’s polarizability than to its monomer dipole moment, which agrees with the study of Soetens and Millot [98]. Although the water dimer properties are less sensitive to the water model’s polarizability, the polarizability is still very important in improving the model’s accuracy, since a 3-point polarizable model outperforms a 5-point non-polarizable model by quite a margin.

The combined results show the full potential of a polarizable water model when properly optimized. Under the same optimization standard (reproducing dimer total dipole) and with the additional degree of freedom utilized, the 3-point polarizable model outperforms 3, 4 and even 5-point rigid models by achieving the smallest error in reproducing water dimer geometry, while still having electrostatic properties in gas phase close to experimental values. At the same time, even this globally optimal polarizable model is not perfect.
2.4 Conclusions

In order to study possible novel avenues for optimization strategies aimed at improving accuracy of explicit water models for biomolecular simulations, different “toy” water models were constructed and examined with respect to their ability to reproduce properties of water dimer—a mimic of the condensed state of water. Specifically, we constructed rigid models with 3, 4 and 5 points, and two different 3-point polarizable models with a single Drude particle to mimic electronic polarizability. The models were built to match reference multipole moments, including ab-initio, and were then optimized to reproduce, as closely as possible, the total dipole moment of water dimer. The ability of the models to reproduce the water dimer geometry was used as the metric of the model accuracy.

First, we conclude that optimizing the “electrostatics” (charge distribution) alone can deliver high accuracy of the water model: the geometry of the resulting water dimer is essentially within 5° of the ab initio reference, this remaining error is almost half of the reported experimental error margin (±10°) on the dimer geometry. Not unexpectedly, the resulting water models show smaller errors than the existing models (TIP3P, OPC3, OPC), which were optimized for liquid water instead of the water dimer. This result reinforces the notion that optimizing electrostatics is key to water model quality, and an “electrostatically globally optimal” water model can be quite accurate.

Second, we conclude that for rigid n-point water models, increasing the number of interaction points from n=3 to n=4 can easily bring better accuracy, while by further increasing n from 4 to 5 only a marginal improvement can be achieved. Considering the steep increase with n of the computational cost associated with performing simulations based on n-point water model, a very convincing justification is needed for renewed efforts to build rigid water models with 5 or even more point charges. Justification for such models may need to be
based on expected or demonstrated improvements in specific areas, e.g. where the possibility that non-planarity of a water models may be critical, such as phase transition between ice and liquid water that TIP5P model has a melting point much closer to experiment than most 4-point models [106]. The above conclusions are robust to the specific type—out of three different ones—of the reference multipole set used to build the models.

Next, we have investigated the effect of adding electronic polarizability to the rigid base model. Our first exercise was aimed at quantifying how much extra accuracy one can gain by having a dedicated mechanism (Drude model) to account for the polarizability. To make an apples-to-apples comparison, we compared two optimization options that employed just one adjustable parameter. The first option is the method we already applied to build our rigid models where point charge placements are adjusted to compensate for the polarizability in an average sense; the second option is to add and optimize polarizability of the Drude oscillator without changing the rigid “base” model. The second option bears similarity to optimization strategies used in practice to construct Drude polarizable models. The same reference multipole moments and dimer geometry accuracy metric were used in both options. The results show that for the same 3-point “base”, having the dedicated polarizability component does achieve a better accuracy than the purely rigid model, which is an argument in favor of polarizable models, despite their extra complexity. At the same time, the accuracy of this limited optimal 3-point polarizable model falls significantly short of the optimal 4-point rigid non-polarizable model. Apparently, the correct physics added to the rigid model via electronic polarizability is mostly “wasted” on correcting the deficiencies of the 3-point “base” model.

To reveal full potential of polarizable models, we have explored a 3-point polarizable model in which both the base charge distribution and the polarizability are globally optimized simultaneously—to the best of our knowledge, this strategy has not yet been used to build
existing polarizable models used in practical simulations. The resulting truly globally optimal 3-point polarizable model easily outperforms 4- and 5-point globally optimal rigid models. Not only that the globally optimal polarizable model is more accurate in reproducing the water dimer, its unpolarized monomer dipole moment (1.7258D) is quite close to the correct gas phase value (1.86D), indicating that the model is a decent mimic for both the condensed phase and gas phase of water.

We stress that all of the water models developed in this work are not intended for use in actual atomistic simulations. Instead, these “toys” models serve only one purpose—to compare various model optimization strategies on the same footing and suggest avenues to consider for future practical optimization efforts to improve realistic water models. In the future, we plan to build and explore a truly electrostatically globally optimal polarizable model, following the over-all optimization strategy outlined in this work.

2.5 Supporting information

S1 Protocol AMBER files for each water model we built in this study. Water dimer minimization results and input files that can be used for reproducing the results are included.

2.6 Acknowledgments

The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this paper.
Chapter 3

Melting points of OPC and OPC3 water models

Authors: Yeyue Xiong, Parviz Seifpanahi Shabane, and Alexey V. Onufriev

Author contributions:

Study conception and design: Y. Xiong, P. S. Shabane, A. V. Onufriev; Data collection: Y. Xiong; Analysis and interpretation of results: Y. Xiong; Draft manuscript preparation: Y. Xiong, A. V. Onufriev.


Link: https://pubs.acs.org/doi/full/10.1021/acsomega.0c02638

Abstract

A recently introduced family of globally optimal water models, OPC, has shown promise in a variety of bio-molecular simulations, but properties of these water models outside of the liquid phase remain mostly unexplored. Here we contribute to filling the gap by reporting melting temperatures of ice $I_h$ of OPC and OPC3 water models. Through the direct coexistence method, which we make available in AMBER package, the melting points of OPC
and OPC3 are estimated as 242 K and 210 K, similar to TIP4P-Ew and SPC/E models respectively, and appreciably below the experimental value of 273.15 K under 1 bar pressure. Water models of the OPC family were optimized to best reproduce water properties in the liquid phase, where these models offer noteworthy accuracy advantages over many models of previous generations. It is not surprising that the accuracy of OPC models in describing the phase transition to the solid state does not appear to offer similar improvements. The new anisotropic barostat option implemented in AMBER may benefit system preparation and simulation outside of the direct coexistence simulations, such as membranes or very long DNA strands.

### 3.1 Introduction

Water is one of the most important substances on earth. The water molecule has a deceptively simple structure H$_2$O, yet many anomalies remain unexplained despite decades of efforts [13, 14, 30, 40, 89]. Computational water models are helpful, but water is known to be difficult to simulate accurately with simplified atomistic water models in a mechanical force field, which tend to be most computationally efficient. Many different water models have been designed dating back at least 50 years ago, yet none is perfect. [53] Phase transitions of water are one of the properties of interest, and water has a rather complex phase diagram, considering the existence of its various solid phases such as ice-Ih, ice-II, and ice-III.

Even a single point on the phase diagram—the melting point under 1 bar pressure—is not reproduced accurately by most of the existing simple atomistic water models. For example, among the widely used n-point fixed charge models (Fig.3.1), the classical TIP3P water model [54] (Fig.3.1a) has a melting point (of ice I$_h$) as low as 146 K [106], while TIP4P model [52] (Fig.3.1b) is not much better in that respect—232 K [106]—compared with the
3.1. Introduction

Experimental 273.15 K ice $I_h$ melting point. On the one hand, these discrepancies may not be too critical for the main purpose for which these water models were originally designed, namely simulations at around 300 K and 1 bar pressure. However, if one wants to model the aqueous environment outside of the “biological range” of ambient conditions, the melting point of a water model can be a very important aspect of its performance. More generally, the fact that a water model cannot reproduce some of the most basic properties of real water should be a concern, as it points to potential problems lurking in the dark, even when the model is used under standard conditions.

Figure 3.1: The overall geometry of classical n-point fixed charge water models. (a) a 3-point model, mimicking H-O-H structure; (b) a 4-point model, offsetting oxygen charge on the 4th point; (c) a 5-point model, with two extra charges emulating lone pairs.

Efforts to estimate the melting temperature of a water model have a long history [6, 15, 23, 24, 33, 34, 35, 55, 56, 61, 81, 105, 106, 107, 110]. Among the relatively modern approaches are Hamiltonian Gibbs-Duhem integration and direct coexistence method. In Hamiltonian Gibbs-Duhem integration, the energy of liquid-solid coexistence state can be calculated, and thus the melting point is found [107]. Direct coexistence method [18, 64, 65], which is applied in this work, is a method where molecular dynamics simulations (NPT ensemble) are carried out for a box of ice and a box of water in direct contact with each other. The upper and lower bound of the water models melting point can be determined by observing at what temperature the box turns completely solid or completely liquid. This method is straightforward and has become popular in recent years with the increasing computing power.
available [23, 24, 25, 34]. In this work, we apply the direct coexistence method to calculate the melting points of two novel water models developed in our group—OPC (a 4-point model, Fig.3.1b) and OPC3 (a 3-point model, Fig.3.1a). OPC and OPC3 water models were developed differently from most other common water models. Utilizing the electric multipole moments parameter space, OPC water model was constructed as a global optimum in reproducing 6 important bulk water properties; similar global optimization was used to develop OPC3 model, which has 3 interaction sites instead of 4 for OPC model [46, 47]. With better liquid water properties, better accuracy has been observed in simulations with OPC or OPC3 used as the solvent in several types of bio-molecular simulations at ambient temperature [11, 32, 48, 96, 101, 104]. Still, very little is known about the performance of these models outside of the liquid phase. There is a recent study where OPC and OPC3 were benchmarked in surface tension calculations and the results show good agreement with the experiment result [49]. To add a small, but important piece to the overall picture, here we examine the performance of OPC family models in simulating the liquid-solid phase change.

3.2 Results and discussion

Through the direct coexistence method (Fig.3.2), the melting point of a water model is calculated within an uncertainty range. In what follows we consider ice to be $I_h$ phase, unless otherwise specified. The temperature at which the whole simulation box becomes ice is the lower bound of the melting point while the temperature where the whole box melts is the upper bound. The temperature difference between the two systems is the uncertainty for the calculated melting point.

The melting or freezing process is monitored through the total energy of the simulation system. (Fig.3.3). During the direct coexistence simulation, if the temperature is higher
3.2. RESULTS AND DISCUSSION

Figure 3.2: Simulation box for the direct coexistence method (93Å × 56Å × 45Å).

Table 3.1: Calculated melting points (ice $I_h$) of TIP4P/2005, TIP4P-Ew, OPC and OPC3 water models.

<table>
<thead>
<tr>
<th>Water model</th>
<th>melting point (K) (this work)</th>
<th>melting point (K) (previous work [34])</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIP4P-Ew</td>
<td>241.0±1.0</td>
<td>242±1</td>
</tr>
<tr>
<td>TIP4P/2005</td>
<td>248.25±0.75</td>
<td>249±1</td>
</tr>
<tr>
<td>OPC</td>
<td>242.15±0.85</td>
<td>-</td>
</tr>
<tr>
<td>OPC3</td>
<td>210 ±10</td>
<td>-</td>
</tr>
</tbody>
</table>

than the actual melting point of the water model, the total energy of the system will increase until the whole box melts. Similarly, for the simulations with a temperature lower than the water model’s melting points, the total energy decreases until the whole box becomes ice.

We begin by reproducing previously published melting points for TIP4P-Ew and TIP4P/2005 water models (Table 3.1). From an earlier direct coexistence method study by Vega et al., the estimated melting point of model TIP4P-Ew and TIP4P/2005 are 242±2 K and 249±2 K, respectively. [34] These numbers match closely the melting points of the same two water models as estimated in this work (241.0±1.0 K and 248.25±0.75 K, Table 3.1, Fig.3.3a,3.3b), even though fine details of the corresponding simulation protocols were not the same, see “Methods”.

Here we demonstrate that (despite being quite accurate in reproducing bulk water properties) OPC and OPC3 models still have melting points far away from the experimental value—273.15 K. The melting point of OPC model is estimated as 242.15 K (Table 3.1, Fig.3.3c),
Figure 3.3: Total energy (per mole) of the ice-liquid coexistence system for (a) TIP4P-Ew, (b) TIP4P/2005, (c) OPC, and (d) OPC3 models. The blue, green, and red curves represent total energy recorded for simulations at each model’s low, medium, and high temperature respectively. The peak at the end of a red curve shows the point where the whole simulation box becomes liquid. The flat “tail” region at the end of a blue curve corresponds to the state where the whole box is frozen. In (c), additional curves (yellow and black) at different temperatures are shown. With the temperature farther away from the melting point, the phase transition is faster. In (d), the pink and light blue curves are OPC3 simulations with the ice secondary prism (1210) plane as the interface. Also note in (d) the blue and light blue curves of OPC3 model does not have the flat “tail”, but we still observed the whole box becoming ice at the last few time steps of the simulation.

31 K lower than the experiment, and close to that of TIP4P-Ew model, which is 241 K. For OPC3 model the estimated melting point is 210 K (Table.3.1, Fig.3.3d), 63.15 K lower than the experiment, and is in the vicinity of the melting points of SPC (190 K) and SPC/E
3.2. Results and discussion

(215 K) models [106], but still closer to the experiment than TIP3P model whose melting point is even lower—146 K [106]. The two comparisons above are within the same types of water models: OPC and TIP4P-Ew are both 4-point fixed charge models while OPC3, SPC, and SPC/E are all 3-point fixed charge models. Therefore, we can argue that there may exist intrinsic defects in the 3- and 4-point fixed charge water models such that even globally optimal models still show relatively large errors in their melting points.

One possible explanation for the lack of accuracy of 3- and 4-point fixed charge rigid models in reproducing the melting point temperature may have to do with a particular feature of these models’ geometry. In ice $I_h$, water molecules organize in a well-structured form (Fig. 3.4), and the melting point reflects the amount of energy needed to break the formation. Compared to a real water molecule’s three-dimensional electron cloud, both 3- and 4-point fixed charge water models have their point charges confined to the H-O-H plane, facilitating movements of the entire structure along directions out of the H-O-H plane. The relative ease of these movements means that less energy is needed to break the fairly rigid ice structure, thus lowering the melting point. We stress that the above discussion applies to “general purpose” models not specifically fitted to reproduce water-ice transition. On the other hand, TIP5P, a 5-point fixed charge “general purpose” water model (developed without training for solid phase), has two out-of-plane extra point charges forming a tetrahedron with the two hydrogen atoms [76]. This 5-point structure makes up for the lack of out-of-plane charge distribution in 3- and 4-point models, and it was estimated that TIP5P model’s melting point is 274 K [106]—matching the experiment value 273.15 K very well.

It is also worth noting that the time it takes for the whole direct coexistence box to completely freeze or melt is different for each water model studied here, especially when comparing the 3-point model (OPC3) and the 4-point models (OPC, TIP4P-Ew and TIP4P/2005). For the three 4-point water models studies here, the simulation box completely freezes or melts within
1000 ns, with a no larger than 2 K temperature range around the melting point. The 3-point model OPC, however, takes longer for the complete phase transition to occur, especially the freezing process, which takes 3400 ns. To observe the complete range of behaviors, from the freezing to the melting, a much larger temperature range is needed—20 K instead of 2 K for the 4-point models, making the error bar for OPC3 model’s melting point considerably larger than the 4-point models estimated in this work. Even with the secondary prism ($\{1\bar{2}10\}$ plane (fastest ice growth plane [24, 84]) as the solid-liquid interface, the complete melting or freezing process of OPC3 box is only marginally faster (Fig. 3.3d). This phenomenon—phase changes are much slower with 3-point models than with 4-point ones—was also observed in a previous study by García Fernández et al. with SPC/E model [34]. A plausible explanation is that OPC3 model and SPC/E model both have very low melting points, at which simulation systems are of low kinetic energy, leading to a much slower speed of the phase transition. It may be possible that much longer simulation times—tens of microseconds—might reduce the uncertainty of the calculated melting point.

### 3.3 Conclusions

In this work, we applied the direct coexistence method to calculate melting points of ice $I_h$ of two novel fixed charge water models—OPC and OPC3. The calculations were performed within AMBER molecular dynamics simulation software suite, where we implemented a new option for anisotropic pressure scaling along a specified axis. This added feature may benefit preparation and simulation of certain molecular systems, such as the direct coexistence simulations, membrane systems, and infinitely long DNA strands. Because implementations of some relevant algorithms (e.g. thermostat options) in AMBER are not exactly the same as those in other packages used in previous melting point studies, we first tested TIP4P-Ew
model and TIP4P/2005 model and the results match previous findings. This agreement has validated the direct coexistence method in AMBER. With OPC3 model, we also compared the phase transition speed between two solid-liquid interface choices—the primary prism plane \((10\bar{1}0)\) and the secondary prism plane \((1\bar{2}10)\), where the latter leads to only marginally faster phase transition than the former.

The calculated melting points of OPC and OPC3 models at 1 bar are \(242.15 \pm 0.85\) K and \(210 \pm 10\) K respectively. These clearly deviate from the experimental value of \(273.15\) K. Despite having better liquid properties and relatively more accurate simulation outcomes in various contexts, the accuracies in the melting points of OPC and OPC3 are on par with their respective 4- and 3- point counterparts TIP4P-Ew and SPC/E, and inferior to that of a 5-point model TIP5P whose estimated melting point (274 K) is very close to the experimental value. Since OPC and OPC3 models are globally optimal, the remaining errors imply the existence of an intrinsic limitation in the ability of 3- and 4-point fixed charge water models to reproduce water melting point (unless optimized specifically to reproduce the ice phase). Note that many properties of super-cooled liquid OPC3 and OPC are quite accurate, including density, heat of vaporization, self-diffusion, and dielectric constant. Thus, it is likely that it is the solid state where these models are in error (which is not surprising, given that they were optimized to best reproduce the liquid state). From the results, we can conclude that OPC and OPC3 may not perform as well in simulating liquid-solid transition as they do in describing the pure liquid state. Also, considering that the 5-point model TIP5P has a rather accurate melting point, one likely intrinsic defect of the 3- and 4-point fixed charge models with respect to reproducing the liquid to solid transition is that these models lack any out-of-HOH-plane charge distribution, which is present in a real water molecule, as well as in the 5-point model. Without the out-of-plane charges, it is easier for the ice structure of 3- and 4-point water models to break due to weaker out-of-plane interaction.
between adjacent water molecules, hence the lower melting temperature.

At the same time, TIP5P model, while being able to simulate the experimental melting point accurately, does not perform as well when evaluated on its bulk liquid properties [85]. Thus, TIP5P may not be optimal to simulate the liquid-solid phase change either. However, with the out-of-plane charge distribution accounted for, we conjecture that a globally optimal 5-point fixed charge water model may be found that performs well in both liquid-solid phase change simulations and liquid bulk simulations.

Polarizable water models [44, 51, 120] is another promising direction, although the relative computational cost of the corresponding simulations remains a concern, especially in light of the very long equilibration time that may be needed to study phase transitions. The computational expense is arguably lowest for algorithmically efficient Drude-based polarizable water models [72]. Not only that the Drude water models are relatively efficient, our recent study has demonstrated that a globally optimal 3-point Drude polarizable water model can, in principle, be as accurate as a 5-point non-polarizable model (over liquid bulk properties), when compared in a carefully controlled optimization procedure [116]. However, it is not yet proven whether the added polarizability via Drude oscillator can make a 3- or 4-point water model accurately reproduce the liquid-solid phase change. Although the Drude charge can move out of the H-O-H plane, which should be beneficial for reproducing the melting point, the distance it moves out-of-plane is very small and likely not enough to alleviate the intrinsic defect of 3- and 4-point models. For example, two well established Drude polarizable water models, COS/G2 [119] and SWM-DP [67], have rather poor melting points—215 K and 186 K respectively [57]. More sophisticated polarizable models such as those of AMOEBA family [73, 90] deliver better performance in that respect (e.g. iAMOEBA model: $T_m = 261 \pm 1$ K) [113], but at a considerable computational cost.

It is still yet to be seen whether a globally optimal Drude polarizable water model or a
3.4 Methods

3.4.1 Direct Coexistence

We apply the direct coexistence method to calculate the melting point of water models. In this method, a fluid-solid interface is created by combining a liquid water box and an ice box together. With the fluid-solid interface, the molecular dynamics simulation won’t suffer from super-cooling or super-heating like simulating pure liquid water or ice.

First, we prepare a box of ice, which satisfies the ice rules [12]. The ice rules describe how water molecules are arranged in the “ordinary” ice phase—ice \( I_h \): each oxygen accepts at most two hydrogen atoms from other water molecules through hydrogen bonding, forming a tetrahedron where a water molecule is surrounded by four others. Further, every 12 water molecules form a hexagonal unit cell that can be stacked together periodically, which is the building block to construct an ice box (Fig.3.4). By construction, the ice box built in this manner has a highly ordered proton arrangement, thus it is different from ice \( I_h \) that exists in nature, which is not “perfect”. Fortunately, it was shown earlier that the initial ice configuration with different proton arrangements does not significantly affect the estimate of the melting point. [24] We note that as the ice is being heated in the simulation, a certain amount of disorder is introduced naturally.

A recent study on the direct coexistence method by Conde et al. [24], has shown that the accuracy (error bar size) of the calculated melting point is affected by the simulation box size. Here we chose the box size of 3456 (liquid) + 3456 (solid) water molecules for all the...
four water models calculated in this work, which according to Conde’s study, results in a reasonable accuracy of $\sim 1$ K error bar for a 4-point water model. We consider this box size an acceptable compromise between accuracy and speed for the purpose of this work.

To create the liquid part of the coexistence box, we heat the ice $I_h$ box to a complete melt, then cool it down to a pre-selected temperature that’s near the estimated water model’s melting point (temperatures where the ice part doesn’t melt during its equilibration process, see below). For the ice part, we equilibrate a same ice $I_h$ box at the same temperature near melting point. The fluid-solid interface is then created by attaching the liquid water and ice box together, where the interface is on the primary prism plane (1010) of the ice part (Fig.3.5,3.6). Ice grows fast on the primary prism plane (1010), but not as fast as the secondary prism plane (1210). [84] We compared the full phase change time of OPC3 model between the two prism plane interface choices, the results show only marginal speed difference where using 1210 interface is about 10% faster (Fig.3.3d).

The simulations were carried out using a modified version of AMBER 2018 [19, 21], with a new barostat feature for anisotropic box re-scaling, see below. Monte-Carlo barostat was
Figure 3.5: The two fluid-solid interface choices with respect to ice hexagonal structure: pink shows the primary prism plane (10\bar{1}0); blue shows the secondary prism plane (1\bar{2}10).

3.4. METHODS

Figure 3.5: The two fluid-solid interface choices with respect to ice hexagonal structure: pink shows the primary prism plane (10\bar{1}0); blue shows the secondary prism plane (1\bar{2}10).

used with this new anisotropic pressure scaling setting. The temperature was controlled by Langevin thermostat with the coupling constant $\gamma_{ln} = 1.0$. For the liquid part of the box, the heating process is a 2 ns NVT simulation at 340 K, the cooling process is a 2 ns NPT simulation at a lower temperature near estimated melting point (varies for different runs). During the cooling process, we apply the anisotropic pressure scaling to the simulation box, where the box size is allowed to adjust to maintain the pressure only along the $x$ direction (normal to the primary prismatic plane (10\bar{1}0) of the ice before heating). By applying this anisotropic pressure scaling, the $y$ and $z$ dimensions of the liquid box remain unchanged, and thus the liquid water and ice boxes can be easily put in contact along the $x$-axis, Fig.3.6, so that their $y$-$z$ faces match exactly. After combining the liquid box and the ice box, we perform a short equilibration of the direct coexistence box at the selected temperature (for example, the simulation runs shown in Fig.3.3c are at temperature 241.3 K, 242.0 K and 243.0 K, respectively) for a short period (10 ns). Then, we start a longer simulation under the same conditions, in increments of 100ns, until the whole box melts or freezes. We found that the farther away the temperature is from the melting point, the faster the phase transition completes, (Fig.3.3c). To find the upper and lower bounds of the melting temperature, we start with testing temperatures with 10 K intervals (e.g. 240 K, 250 K, 260 K, ...), then stop
when, at the given temperature, the whole box freezes or melts within 100 ns; we refine that temperature by exploring the transition with 1 K temperature intervals. We stop when the complete phase change takes longer than 1000 ns. If the temperature found in the previous step still has fast phase change, we further refine the last step by testing 0.1 K intervals. For OPC3 model’s lower bound, temperatures lower than 200 K do not result in faster freezing, and we settled on 200 K where the whole box becomes frozen after 3400 ns of simulation. We ran the simulation three times for each of the boundary temperatures and only accepted the one at which a complete phase change was observed in all of the three runs.

![Figure 3.6: The process to prepare the fluid-ice coexistence box.](image-url)
For the direct coexistence box, the simulations are NPT with the default anisotropic pressure scaling option, where the three dimensions of the box can freely adjust independently to each other. The independence between the box size changes along three dimensions keeps the system from imposing external force on the liquid-ice interface. The integration time-step was 2 fs. Particle Mesh Ewald (PME, GPU module: pmemd.cuda) [94] was used to handle electrostatic interactions, with $cut = 8.5\text{Å}$ (AMBER input parameter, nonbonded cutoff distance) and all other settings set to their default values. It is worth noting that in AMBER default settings, the option to account for the van der Waals interactions beyond the cutoff ($8.5\text{Å}$ in this study) via a continuum model is enabled ($vdw\text{meth} = 1$ [19, 20]), which is critical for OPC and OPC3 [3, 46, 47]. The AMBER input files are included in the supporting information.

The exact barostat and thermostat algorithms employed in this work are different from some of the previous studies using the same direct coexistence method, such as Vega et al. [34], where Parrinello-Rahman barostat and Nosé-Hoover thermostat were used. Despite these methodological differences, our implementation yields essentially the same results as in Ref. [34] for TIP4P-Ew and TIP4P/2005 water models. Thus, we argue that the modifications we used in AMBER are acceptable for the direct coexistence method. Although the results calculated with AMBER are consistent with previous studies, the simulations take longer for the system to change phase completely. This may be related to the different thermostat and barostat used in this study, or to the different integrator implementations between different programs, and will require further exploration to explain fully.
3.4.2 New anisotropic pressure scaling option in AMBER

This anisotropic pressure scaling feature where a user can choose one dimension to allow the simulation box to extend or shrink was not available in AMBER 2018 or any earlier version. The first author of this work implemented the pressure scaling option based on AMBER 2018 \[19\], and it is now available in AMBER 2019 \[20\] and newer releases (AMBER input parameter “baroscalingdir”). With this new option, the user can choose x-, y-, or z-direction such that the simulation box only adjusts its size along the chosen direction to maintain pressure. For certain molecular systems and simulation types, such as the direct coexistence, infinitely long DNA strands, and membranes, this option should provide the convenience in system preparations, and is very important for correctly reflecting the natural anisotropy of these simulations.

Acknowledgement

The authors thank Dr. Igor Tolokh for sharing his insight during the results discussions and thank Virginia Tech Advanced Research Computing for providing the computational resources. Funding from NIH R21GM134404 is acknowledged.

Support information

AMBER input files with the simulation control parameters, AMBER topology and coordinate files for the co-existence box for each water model at the reported temperatures.
Author Information

Corresponding author: Alexey V. Onufriev

Email: alexey@vt.cs.edu
Chapter 4

A fast polarizable water model for atomistic simulations

Authors: Yeyue Xiong and Alexey V. Onufriev

In preparation for submission.

Abstract

Simulating water accurately has been a major challenge in atomistic simulations for decades. Inclusion of electronic polarizability effects holds considerable promise, yet existing approaches suffer from significant computational overhead compared to the widely used non-polarizable water models. We have developed a globally optimal polarizable water model, OPC3-pol, that explicitly accounts for electronic polarizability with minimal impact on computational efficiency. OPC3-pol reproduces five key bulk water properties at room temperature with an average relative error of 0.6%. In atomistic simulations, OPC3-pol’s computational efficiency is in-between that of 3- and 4-point non-polarizable models; the model supports increased (4 fs) integration time step. OPC3-pol is tested in simulations of a globular protein and a B-DNA dodecamer with recent AMBER force-fields ff14SB, ff19SB, and OL15, demonstrating structure stability close to reference at multi-microsecond time-scale. The model can be trivially adopted by any package that supports standard non-polarizable
models; its intended use is in long classical atomistic simulations where water polarization effects are expected to be important.

4.1 Introduction

Water molecule has a deceptively simple structure \( \text{H}_2\text{O} \), yet many anomalies of liquid water remain hard to explain despite countless studies [13, 14, 30, 40, 89]—not surprisingly, water in its liquid state is notoriously difficult to model. At the same time, simulating water accurately is critical in different fields, including engineering and science; examples include fluid mechanics [27, 60], nano materials [10, 59], and biomolecular simulations [85]. Many complex biomolecules (DNA, RNA, proteins, etc.), vital for a living cell’s function, cannot be studied alone without considering their aqueous environment [121]. To study those large biomolecules, atomistic simulations have been widely utilized; numerous different water models [85] have been developed to reproduce water properties, including the class of water models most prevalent today—simple, rigid, fixed-charge explicit models such as TIP3P [54], SPC/E [9] or TIP4P [1, 52]. According to the convention, these classical water models are distinguished by the number of “points”—interaction sites—they consist of: 3-point models such as TIP3P or SPC/E are still the most common due to what is perceived by many as an acceptable balance between accuracy and computational cost. For water models with just one more interaction site, such as TIP4P (a 4-point model with 3 point charges and neutral oxygen), computational cost considerations often become significant, especially in long simulations, even though these models tend to reproduce water properties better than their 3-point counterparts [85]. However, despite impressive recent progress [47, 112], models of the fixed-charge class are still far from perfect [31, 39, 85, 86, 87, 108]. One possible reason is that these simple models do not account, explicitly, for many physical effects present
in real water. These physical effects missing in widely used rigid n-point models include many quantum effects, water flexibility, and very importantly, many-body effects such as electronic polarizability. Without polarizability, a rigid n-point model is incapable of responding to the change of polarity of its micro-environment, which is very relevant to many types of simulations. For example, on crossing the water-vapor interface, the dipole moment of a real water molecule changes by about 40%, while the change is zero by definition for a fixed-charge water model [37, 38]. Unlike non-polarizable n-point water models, polarizable water models such as AMOEBA [90], SWM4-NDP [68], POL3 [17] are designed to mimic water molecule’s electronic polarization; the hope is that these models can bring the accuracy of biomolecular simulations to a qualitatively new level. Sophisticated polarizable water models [45, 51] can indeed achieve high accuracy in reproducing many liquid water, yet the associated computational costs are many times higher than those of classical non-polarizable water models, discouraging wider adoption. Another barrier to adoption by wider molecular simulation community is the perceived need to use a polarizable water model together with the matching polarizable force-field describing the solute itself. The strategy has own issues and associated computational overheads, including limited choice of such force-fields and the need for specialized algorithms not supported by many popular simulation packages.

In this work, we aim to design a polarizable water model that can deliver to the community full benefits of electronic polarization with minimal impact on computational efficiency compared to industry standard fixed-charge models. The goal is not to deliver the most accurate water model ever—which is highly unlikely for a 3-point water model [116]—but to create a polarizable water model that runs as fast as the widely efficient non-polarizable models such as TIP3P or TIP4P, while achieving comparable or better accuracy. Another key consideration is to encourage immediate adoption and testing by multiple simulation communities by delivering a “plug-and-play” model to be used with existing, widely available non-polarizable
4.2 Overall approach

In this study, our goal is to design a water model for the purpose of accurately simulating water in biomolecular simulations. To achieve this goal, the relevant parameter space needs to be fully explored without any parameter limited to experimental observations. However, globally optimizing even the simplest 3-point water model in its commonly used “native” parameter space of the hydrogen charge $q$, H-O bond length $l$ and H-O-H angle is a herculean task, because key bulk properties of liquid water are extraordinarily sensitive [115] to the charge–distances–angles parameters. For example, a change of the $|\text{OH}|$ distance by as little as $10^{-2} \, \text{Å}$ can change the calculated self-diffusion constant by as much as 30 % [115]. As a consequence, only an extremely fine-grain search can succeed in finding the true global optimum, but that can become prohibitively expensive in this parameter space. The difficulty of performing such global optimization for a polarizable model is even more difficult, since an additional polarizability parameter $\alpha$ needs to be included.

With the optimal point charge approximation (OPCA) method developed earlier by Anandakrishnan et al [4], a complex charge distribution can be simplified into a set of limited number of point charges, reproducing electronic multipole moments to a certain order. Using the OPCA method for building a water model, the number of search parameters can be reduced, and the liquid properties become much less sensitive to the new “indirect” parameters—multipole moments, which makes global optimization possible. A subspace of low order multipole moments that is electrostatically most relevant can be used for the global optimization of a water model. This method has been applied in the successful development of the non-polarizable OPC and OPC3 water model, where both models have shown high
accuracy in pure liquid water and practical biomolecule MD simulations [46, 47]. Here we extend the OPCA method to the development of OPC3-pol Drude polarizable water model. Unlike in the optimization process of OPC and OPC3 models, where only two degrees of freedom was included—dipole moment \( \mu \) and square quadrupole \( Q_t \), here we bring in one additional parameter—the polarizability \( \alpha \)—to form a 3-dimensional parameter space (Fig. 4.1). The added polarizability is introduced via a Drude oscillator attached to the oxygen atom; the charge of the Drude particle is determined by \( \alpha \) value using equation: 
\[
q_D = \sqrt{\alpha k} \times 18.223 \text{ e.}
\]
The charge on the oxygen is then set accordingly: 
\[
q_O = -2q - q_D.
\]
For better integrator stability, we chose a relatively low force constant \( k = 500 \text{ kcal/mol/Å}^2 \).

In previous implementations of Drude polarizable water models, in order to mimic the essentially instantaneous polarization response of the water molecule electrons, the motion of the Drude particle is calculated separately from the rest of the system, and the Drude particle is set to be massless or near massless [66, 120]. To this end, the two widely used methods are the self-consistent field (SCF) approach and the extended Lagrangian dynamics [67, 72]. While achieving the goal of handling the very low mass of the Drude particle, both methods have the same drawback that they increase the computational cost of the simulation considerably. Even with the faster method—extended Lagrangian, simulations with Drude polarizable water models are at least 4 times slower than those relying on classical non-polarizable 4-point rigid models [45]. Here, we argue that, for classical atomistic simulations that do not consider explicit electron transfer, the very fast response time afforded by massless or near massless electrons is not needed. We therefore treat the Drude particle as an “ordinary” atom in the molecular dynamics system: the water oxygen mass is split equally between the oxygen atom and the Drude particle (Fig. 4.1). This choice is critical for computational efficiency of the model, permitting standard integration time step, even with the relatively stiff Drude spring constant needed to keep the Drude dipole minimally exten-
4.2. Overall approach

Figure 4.1: Developing OPC3-pol: 3-point Drude polarizable water model, shown schematically on the left. Here \( q \) is the electric charge on hydrogen, \( q_D \) is the charge on the Drude particle (black dot), \(-2q - q_D\) is the oxygen charge, \( k \) is the spring constant of the Drude-oxygen bond, \( m_O \) is the oxygen mass, and \( E \) denotes the external electric field (e.g. generated by the surrounding water molecules). The OPCA approximation maps the conventional “native” charge–distances–angle-polarizability 4D parameter space of a water model onto a reduced, 3D space of two lowest multipole moments and the polarizability. An exhaustive search for the global optimum in the 3D space is performed, with the model quality score (shown by color scale, from 1 to 10) being the deviation of model properties from experimental bulk water properties. Linear regression is applied to refine model parameters near the optimum. 1-1 mass partitioning for the oxygen-Drude pair is applied as a design choice for computational efficiency.

There are also extra steps [45] in previous Drude model implementations for avoiding the so-called “polarization catastrophe” [120], that may result from the Drude particle extending too far away from its parent atom towards an oppositely charged atom of another molecules, causing energy and force divergence, leading to computational instability. We omit these extra steps, and rely instead on van der Waals repulsion between oxygen atoms along with a strong Drude-oxygen bond force constant to avoid the polarization catastrophe,
as originally suggested by Yu et al. [120]. We have verified the approach in OPC3-pol model by multi-microsecond long simulations. With this new method, the Drude polarizable water model can run as fast as classical non-polarizable water models, see below.

Similarly to previously developed globally optimal OPC and OPC3 fixed-charge models, OPC3-pol has a single Lennard-Jones (LJ) site on its oxygen. The two LJ parameters ($\sigma$ and $\epsilon$) of OPC3-pol model’s oxygen is determined separately from the global optimization in the $(\mu, Q_t, \alpha)$ 3D parameter space, by fitting the liquid water density and the oxygen-oxygen radial distribution function (O-O RDF). Each trial model built during the optimization process is evaluated via a predefined quality score based on experiment values of five key liquid water properties [46, 47, 109].

In order to find global optimum, the first step is to divide the 3-D parameter space $(\mu, Q_t,$ and $\alpha$) into a series of 2-dimensional “slices” with fixed polarizability values and calculate quality scores of water models built from the parameters on these “slices” (Fig. 4.1). With the quality score results of several slices collected, the next step is to apply regression analysis on the data to narrow down the “hot regions” where the quality scores are high. A finer exhaustive search is then carried out in the immediate vicinity of the “hot regions”, among the sampled models only the one with the highest average quality score is chosen.

After the globally optimal parameters of OPC3-pol model have been obtained, and its reasonable quality for reproducing bulk water properties at room temperature is established (Fig.4.3), the largest remaining concern is whether OPC3-pol can also support practical MD simulations of biomolecules. This is a nontrivial question, as previously developed polarizable water models such as SWM4-NDP or AMOEBA were used along with their matching, specialized polarizable force fields [43, 90]. Below we test whether OPC3-pol can be used with existing non-polarizable gas-phase force fields. Biomolecules tested include a ubiquitin, a B-DNA dodecamer, and an intrinsically disordered amyloid $\beta$-peptide.
4.3 Results and discussions

Through an exhaustive search and regression analysis, globally optimal OPC3-pol model is found (Fig. 4.2). The search has yielded the following model parameter: $\mu = 2.05 \text{D}$, $Q_t = 2.15 \text{DÅ}$, and $\alpha = 1.16 \text{Å}^3$; the Lennard Jones center is placed on the oxygen atom and the parameters are $\sigma_{LJ} = 3.176 \text{ Å}$ and $\epsilon_{LJ} = 0.2085 \text{ kJ/mol}$. Among five bulk liquid properties for which performance of an existing Drude-based model SWM4-NDP was reported previously by Lamoureux et al. [68], OPC3-pol model reproduces the reference experimental values quite accurately (Fig. 4.3), with the average quality score of 9.3 out of 10. Several other parameters and properties are reported in the Supplementary Information (Supplementary Fig. 4.2, Supplementary Table. 4.2).

Figure 4.2: Global optimal OPC3-pol model. The values of the OH bond length $l_{OH}$, H-O-H angle $\angle \text{HOH}$, partial charge values $q$, $q_D$, and the Drude particle spring force constant $k$ are shown. The LJ interaction center is on the oxygen atom, its parameters are $\sigma_{LJ} = 3.176 \text{ Å}$ and $\epsilon_{LJ} = 0.2085 \text{ kJ/mol}$.

To test the potential of OPC3-pol for MD simulations of globular proteins, we use ubiquitin—a 76-amino-acid $\alpha/\beta$ mixed globular protein commonly used in testing of force fields [47]. First, we have tested OPC3-pol with the standard 2 fs integration time step. As shown in Supplementary Fig. 4.8, with both the AMBER ff14SB [77] and ff19SB [101] force fields,
during the four 2 µs independent MD simulations the protein stays stable, and close to the reference crystal structure, with the RMSD to X-ray structure ranging from 1 to 2 Å. With ff14SB, the ubiquitin is slightly more stable than with ff19SB, with an average RMSD at 1.0 Å compared to 1.2 Å for the latter. To test OPC3-pol on a nucleic acid, we used a B-DNA dodecamer; current AMBER nucleic acid force field (OL15) is used. The simulated B-DNA structure matches the crystal structure within 1.5 Å backbone RMSD (excluding terminal nucleotides, following Ref. [95]), and the widths of its major and minor grooves agree with experimental reference as well (S.I., Supplementary Fig.4.9). These results clearly demonstrate stability of OPC3-pol in multi-microsecond MD simulations of two typical biomolecules. The RMSD deviation from the reference resulting from the use of OPC3-pol water model is comparable to non-polarizable traditional water model TIP4P-Ew, while being notably and curiously less than that of the polarizable Drude model SWM4-NDP used together with the matching polarizable force-field, when applied to the same structures (Table.4.1).

Different from globular proteins such as ubiquitin that keep their compact structure under
4.3. Results and discussions

Figure 4.4: Relative simulation speed comparison between TIP3P, TIP4P-Ew, iAMOEBA, SWM4-NDP, and OPC3-pol water models. The speeds (ns/day) for TIP3P, TIP4P-Ew, and OPC3-pol are measured from liquid water simulations of 806 water molecules with 2 fs and 4 fs time step. The speeds of iAMOEBA and SWM4-NDP model are those reported by Wang et al. and Huang et al. respectively [45, 113]. The non-polarizable models TIP3P and TIP4P-Ew are shown in green, while iAMOEBA, SWM4-NDP, and OPC3-pol—polarizable models are shown in blue. Dark color bars show the speeds achieved using each model’s “standard” time step: 2 fs for TIP3P, TIP4P-Ew, and OPC3-pol; 1 fs for iAMOEBA and SWM4-NDP. Light color bars show the maximum speeds if a water model supports 4 fs time step with hydrogen mass repartitioning (HMR).

Table 4.1: Accuracy of OPC3Pol in simulating ubiquitin (PDB: 1UBQ) and B-DNA decamer (PDB: 1BNA) structures at ambient conditions on microsecond time-scale. A comparison with “industry standard” non-polarizable TIP4P-Ew model is under AMBER-ff14SB for ubiquitin and AMBER-OL15 for B-DNA. SWM4-NDP model is accompanied with Drude FF (*B-DNA crystal structure is used in the training of Drude FF with SWM4-NDP) [43, 95]. OPC3-pol model is simulated with both AMBER-ff14SB and ff19SB for ubiquitin and AMBER-OL15 for B-DNA.

<table>
<thead>
<tr>
<th>Water model</th>
<th>RMSD to X-ray reference (force field applied)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ubiquitin</td>
</tr>
<tr>
<td>TIP4P-Ew</td>
<td>0.9 Å (ff14SB)</td>
</tr>
<tr>
<td>SWM4-NDP</td>
<td>∼ 2.2 Å (Drude FF) [43]</td>
</tr>
<tr>
<td>OPC3-pol</td>
<td>1.0 Å (ff14SB)</td>
</tr>
<tr>
<td></td>
<td>1.2 Å (ff19SB)</td>
</tr>
</tbody>
</table>
ambient conditions, are intrinsically disordered proteins (IDP), which interconvert between multiple conformations of various degree of compactness [96]. IDPs remain a challenge to simulate for classical non-polarizable water models, where traditional fixed-charge water model tend to favor overly compact structures compared to experiment [88]. Here, we test OPC3-pol model with one of the common IDP benchmarks: amyloid $\beta$-peptide (A$\beta$42). On 8 microsecond time scale, A$\beta$42 is constantly converting between different conformations of various degree of compactness, which is qualitatively consistent with the expectation. Quantitatively, OPC3-pol model with two standard AMBER protein force fields results in A$\beta$42 conformations that are less compact than expected from experiment: the predicted radius of gyration $R_g$ fluctuates between 11 Å and 25 Å, and averages at 19 Å, higher than the experimental estimate of 11-15 Å [8] (Supplementary Fig. 4.10).

With the model speed being the critical factor in practice, we have explored the possibility of increasing the integration time step to further speed up the simulations. Results are promising: with the hydrogen mass repartitioning (HMR) [41] applied, both ubiquitin and B-DNA trajectories generated with a 4 fs time step are as stable as those produced with 2 fs time step (Supplementary Fig. 4.12, 4.13).

### 4.4 Conclusions

In this work, we have developed a new global optimal polarizable explicit water model—OPC3-pol, which runs as fast as a non-polarizable 4-site water model. OPC3-pol model has a simple composition which contains three fixed sites mimicking the H-O-H structure and an additional charged particle that is bonded to the Oxygen atom, forming a Drude oscillator that emulates polarization. We have also made an important design choice for Drude oscillator implementation—splitting half of the oxygen mass onto the Drude particle
and treat it as a regular “atom”. This way, the location of the Drude particle can be determined during each dynamic step without the need of extra calculations required in the SCF or extended Lagrangian implementation.

The OPC3-pol model has shown good agreements to experimental data of liquid water properties under ambient condition. In addition to its accurate reproduction of liquid water properties, OPC3-pol is also very fast, with a speed in-between 3- and 4-point non-polarizable water models (measured with AMBER GPU code).

As solvent, we have tested OPC3-pol model for a ubiquitin protein and a B-DNA dodecamer. Both the simulated structures match their reference crystal structure with average root-mean-square deviation (RMSD) around 1-1.5 Å over microsecond long trajectories. We have also tested simulating an intrinsically disordered protein amyloid β-peptide (Aβ42) and the results qualitatively match experiment observations without being stuck in overly compact conformations. These results are achieved rather stably using a 2 fs time step, which is common for simulations with a non-polarizable explicit solvent but not particularly advisable for those with a polarizable solvent due to computational stability concerns. More interestingly, OPC3-pol model can handle a time step as large as 4 fs, with the help of hydrogen mass repartitioning (HMR).

Although the global optimal OPC3-pol model is tested to be accurate in both pure liquid water and several biomolecule simulations, limitations do exist. OPC3-pol is not perfect, for our goal is to make a fast and accurate polarizable water model, easily accessible for use in biomolecular simulations, thus compromises are made. Firstly, OPC3-pol is optimized under ambient condition, and some temperature dependent properties of water are not reproduced when it is used under temperatures other than 300 K. Secondly, liquid water’s density maximum with respect to temperature is not reproduced by OPC3-pol, similar to some other water models both non-polarizable like TIP3P and polarizable such as SWM4-
CHAPTER 4. A fast polarizable water model for atomistic simulations

NDP and POL3. Lastly, OPC3-pol works with at least two latest non-polarizable AMBER force fields, but without a specialized force field, it is expected that more subtle behavior may still deviate from experimental observations.

Overall, we believe we have achieved the goal to develop a polarizable water model with good accuracy and is as fast as non-polarizable model. With OPC3-pol model’s fast speed, we anticipate it to benefit some but not all simulations. Globular proteins and DNA where polarizability is expected to have major effect can be such examples.

4.5 Supplementary information

4.5.1 Method details

Overall approach

Following the optimal point charge approximation (OPCA) method, the coordinates and charge values of the point charges in a water model are calculated from given multiple moments. For example, during the development of the OPC3 model \cite{46}, the “native” model parameters—coordinates \((y_1, z_2)\) and charge values \(q\) of the three point charges (Fig. 4.5) can be written with respect to the dipole \(\mu\), the square quadrupole \(Q_t\), and the linear quadrupole \(Q_0\):
For OPC3-pol polarizable model, the same representation (Eq. 4.1-4.5) of the “native” parameters of the model’s rigid part—also consists of three point charges—is applied. Unlike the previously reported optimization of fixed charge OPC and OPC3 models [46, 47], where only two degrees of freedom were included—the dipole moment $\mu$ and square quadrupole $Q_t$—here we bring in one additional parameter—the polarizability $\alpha$—to form a 3-dimensional
parameter space. The added Drude particle is attached to the oxygen and its charge is determined by $\alpha$ value using equation: $q_D = \sqrt{\alpha k} \times 18.2223$ e. The charge on the oxygen is then changed accordingly: $q_O = -2q - q_D$. Also, to mitigate potential issues with the integrator stability using the standard 2 fs time step, we chose a force constant $k = 500$ kcal/mol/Å$^2$, which is lower than $k = 1000$ kcal/mol/Å$^2$ suggested previously for a Drude model [67].

For the previously developed OPC and OPC3 water models, the global optimization was carried out via an exhaustive search within 2-dimensional parameter space ($\mu$, $Q_t$), and the quality scores of the models built with different parameter values were shown on a quality score 2D map. Here, in order to globally optimize the OPC3pol model, we apply a similar exhaustive search in the 3-dimensional parameter space ($\mu$, $Q_t$, and $\alpha$).

In the parameter space, each water model built is evaluated by comparing its liquid water properties to a set of experimental reference. The evaluations are performed through the calculation of a predefined quality score [46, 47, 109]. For a calculated property $x$ of the evaluated model and its corresponding experimental reference $x_{\text{exp}}$, the quality score is calculated as:

$$M = \max\{[10 - |(x - x_{\text{exp}}) \times 100/(x_{\text{exp}} \text{tol})]|, 0\} \quad (4.6)$$

where the $\text{tol}$ is the assigned tolerance. The final quality score is averaged over five key properties at ambient conditions (298.16K, 1bar): density $\rho$, static dielectric constant $\epsilon_0$, self diffusion coefficient $D$, heat of vaporization $\Delta H_{\text{vap}}$, and the first peak position of the RDF $\text{Roo}_1$. The tolerance $\text{tol}$ is set to 0.5% for density and RDF first peak position, and 2.5% for the remaining properties.

The first step is to divide this 3-dimensional parameter space ($\mu$, $Q_t$, and $\alpha$) into a series
Figure 4.6: Developing a 3-point Drude polarizable water model: the use OPCA approximation replaces the conventional “native” 4D parameter space of a water model with a reduced, 3D space of two lowest multipoles and polarizability. An exhaustive search for the global optimum is performed in the reduced 3D space, and then conduct global optimization. 1-1 mass partitioning for oxygen-Drude pair is applied as a design choice for computational efficiency. \( q \) is the electric charge on hydrogen, \( q_D \) is the charge on the Drude particle, \(-2q-q_D\) is the oxygen charge, \( k \) is the bond force constant for the Drude-oxygen bond, \( m_O \) is the oxygen mass, and \( E \) denotes the external electric field (generated by surrounding molecules in MD simulations).

of 2-dimensional slices with fixed polarizability values, and calculate quality scores of water models on these slices (Fig.4.6). With the quality score results of several slices, next step is to apply regression analysis on the data. We have found that the quality scores of static dielectric constant \( \epsilon_0 \), self diffusion coefficient \( D \), and heat of vaporization \( \Delta H_{vap} \) are linear to the three parameters \( (\mu, Q_t, \text{and } \alpha) \):

\[
\epsilon_0 = 115.59620\mu/D - 17.45936Q_t/\text{DA} + 41.94769\alpha/\text{Å}^3 - 169.31086 \quad (4.7)
\]

\[
D/(10^9\text{m}^2/\text{s}) = -3.616164\mu/D - 7.301098Q_t/\text{DA} - 1.819891\alpha/\text{Å}^3 + 27.525128 \quad (4.8)
\]

\[
\Delta H_{vap}/(\text{kcal/mol}) = 5.917774\mu/D + 6.175890Q_t/\text{DA} + 1.225421\alpha/\text{Å}^3 - 16.305646 \quad (4.9)
\]
Based on the regression result, three planes can be found where the water models on each plane have $\epsilon_0$, $D$, and $\Delta H_{vap}$ values matching the reference respectively. The intersect of these three planes is where we can find the global optimal OPC3pol model. A finer exhaustive search is then carried out in the immediate vicinity of the intersect location; among the sampled models only the one with the highest average quality score is chosen.

**Design choices for computation efficiency**

Traditional Drude polarizable models (such as SWM4-NDP [68]) despite being more efficient than other types of polarizable water models, are still very slow for practical atomistic simulations (4 times slower than classical non-polarizable explicit water models [45]). The main reason is that in the traditional Drude polarizable models, the motions of Drude particles are treated separately from the rest of the simulation system. Based on the fact that an electron is around 2000 times lighter than the hydrogen atom, let alone the oxygen atom, theoretically the Drude particle (as a mimic of the moving electrons) should have near 0 mass and move instantly in reaction to the water molecule’s environment, which cannot be handled as is in common molecular dynamics programs. The special treatment of Drude particles require additional calculations on top of the molecular dynamics. There are two major methods of treating the Drude motions: self-consistent field (SCF) approach and extended Lagrangian dynamics method. In the SCF approach, the displacements of the Drude particles (massless) are calculated through an energy minimization calculation at each time step of the dynamics of the whole system, making it the much slower option of the two. The extended Lagrangian dynamics method, however, gives the Drude particle a small mass (< 1 amu), and “partially” incorporates it into the dynamics of the whole system. In this method, the water model’s rigid part and its Drude particle are seen as a two-body system, where the motion of its center of mass is simulated in the regular molecular dynamics calculations and the relative
motion of the Drude particle is then determined in a separate extended Lagrangian dynamics step. The extended Lagrangian dynamics method still involves additional calculations, and runs around 4 times slower than classical non-polarizable water models.

The new approach we applied here is to add more mass to the Drude particle (8 amu) and treat it as a “regular atom” in the dynamics of the whole system. As a two-body system, the oscillator’s natural period is calculated to be 260 fs, which is large enough for usual time step length (1-4 fs), and small enough for Drude particle to move faster than the re-orientation of water molecules (ps time-scale). Also importantly, without the additional calculations for separate Drude treatments, our new polarizable model runs as fast as classical non-polarizable models.

4.5.2 Results

Through the exhaustive search and regression analysis described above, the resulting global optimal OPC3-pol model’s parameters are: dipole moment $\mu = 2.05 \text{D}$, square quadrupole $Q_t = 2.15 \text{DÅ}$, polarizability $\alpha = 1.16 \text{Å}^3$, Lennard Jones parameters $\sigma_{LJ} = 3.176 \text{Å}$ and $\epsilon_{LJ} = 0.2085 \text{kJ/mol}$. Also, as shown in Fig.4.2 in the main text, OPC3-pol model’s “native” parameters are: OH bond length $l_{OH} = 1.2110 \text{Å}$, H-O-H angle $\angle \text{HOH} = 109.47^\circ$, hydrogen charge $q = 0.305e$, Drude particle charge $q_D = 1.321e$.

**Bulk liquid water properties**

Detailed bulk liquid water properties of OPC3-pol model in comparison with other existing water models and experiment reference are shown in Table.4.2.
Table 4.2: Bulk liquid properties of OPC3pol in comparison with other water models (bold model names referring to polarizable models) and experimental reference: Average Dipole $\mu$, Density $\rho$, Static Dielectric Constant $\epsilon_0$, Self Diffusion Coefficient $D$, Heat of Vaporization $\Delta H_{vap}$, and First Peak Position in the RDF $r_{oo1}$.

<table>
<thead>
<tr>
<th>Property</th>
<th>TIP3P</th>
<th>TIP4P-Ew</th>
<th>SWM4-NDP</th>
<th>iAMOEBA</th>
<th>OPC3-pol</th>
<th>EXP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$ (D)</td>
<td>2.348</td>
<td>2.32</td>
<td>2.461</td>
<td>2.78</td>
<td>2.42</td>
<td>2.5–3</td>
</tr>
<tr>
<td>$\rho$ (g/cm$^3$)</td>
<td>0.980</td>
<td>0.995</td>
<td>0.998</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>$\epsilon_0$</td>
<td>94</td>
<td>63.90</td>
<td>79 ± 3</td>
<td>80.7 ± 1.1</td>
<td>78 ± 1</td>
<td>78.4</td>
</tr>
<tr>
<td>$D$ ($10^9$ m$^2$/s)</td>
<td>5.5</td>
<td>2.44</td>
<td>2.33 ± 0.02</td>
<td>2.54 ± 0.02</td>
<td>2.30 ± 0.01</td>
<td>2.3</td>
</tr>
<tr>
<td>$\Delta H_{vap}$ (kcal/mol)</td>
<td>10.26</td>
<td>10.58</td>
<td>10.52</td>
<td>10.94</td>
<td>10.64</td>
<td>10.52</td>
</tr>
<tr>
<td>$r_{oo1}$ (Å)</td>
<td>2.77</td>
<td>2.755</td>
<td>2.83</td>
<td>2.8</td>
<td>2.78</td>
<td>2.80</td>
</tr>
<tr>
<td>$C_p$ (cal/(K·mol))</td>
<td>18.74</td>
<td>19.2</td>
<td>n/a</td>
<td>18.5</td>
<td>20.0</td>
<td>18</td>
</tr>
<tr>
<td>$\alpha_p$ ($10^{-4}$ K$^{-1}$)</td>
<td>9.2</td>
<td>3.2</td>
<td>n/a</td>
<td>2.5</td>
<td>6.8</td>
<td>2.56</td>
</tr>
<tr>
<td>$\kappa_T$ ($10^{-6}$ bar$^{-1}$)</td>
<td>57.4</td>
<td>48.1</td>
<td>n/a</td>
<td>41.1</td>
<td>40.5</td>
<td>45.3</td>
</tr>
</tbody>
</table>

Figure 4.7: Water dimer structure at lowest energy.

Water dimer

Water dimer, a structure formed with two water molecules through hydrogen bonding (Supplementary Fig. 4.7), is also tested with OPC3pol model (Table 4.3).

Computational efficiency

Not only OPC3-pol model reproduces the bulk water properties accurately, OPC3-pol’s computational efficiency in atomistic simulations is in-between 3- and 4-point non-polarizable
4.5. Supplementary information

Table 4.3: Water dimer simulated with OPC3pol and other polarizable water models (minimal energy state). The simulated water dimers are compared through their total dipole moments $\mu_{dimer}$, oxygen-oxygen distances $r(\text{OO})$, and the angle $\varphi$ between the bisector of the acceptor and the O-O vector.

<table>
<thead>
<tr>
<th>Water model</th>
<th>$\mu_{dimer}$ (D)</th>
<th>$r(\text{OO})$ (Å)</th>
<th>$\varphi$ (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPC3pol</td>
<td>3.55</td>
<td>2.790</td>
<td>24.82</td>
</tr>
<tr>
<td>SWM4-NDP</td>
<td>2.062</td>
<td>2.83</td>
<td>71</td>
</tr>
<tr>
<td>iAMOEBA</td>
<td>n/a</td>
<td>2.836</td>
<td>54.9</td>
</tr>
<tr>
<td>QM</td>
<td>2.68</td>
<td>2.91</td>
<td>55.6</td>
</tr>
</tbody>
</table>

models, and supports increased (4 fs) time-step (Table 4.4).

Table 4.4: Relative simulation speed comparison between OPC3pol, TIP3P, TIP4P-Ew, SWM4-NDP, and AMOEBA water models. The speeds of TIP3P, OPC and OPC3pol models are compared through 2 fs time step NPT simulations of 806 water molecules with AMBER 2019 on one RTX2080Ti GPU, measured in ns/day. The speeds of iAMOEBA and SWM4-NDP water models relative to TIP3P are from previous published results by Wang et al. and Huang et al. respectively [45, 113].

<table>
<thead>
<tr>
<th>Water model</th>
<th>Relative speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIP3P (3-point non-polarizable)</td>
<td>1</td>
</tr>
<tr>
<td>TIP4P-Ew (4-point non-polarizable)</td>
<td>0.60</td>
</tr>
<tr>
<td>iAMOEBA (3-point w. polarizable multipole)</td>
<td>~0.09</td>
</tr>
<tr>
<td>SWM4-NDP (4-point w. Drude)</td>
<td>~0.25</td>
</tr>
<tr>
<td>OPC3pol (3-point w. Drude)</td>
<td>0.75</td>
</tr>
</tbody>
</table>

4.5.3 Tests in atomistic MD simulations

AMBER 20 is used for the MD simulations. Langevin dynamics with a collision frequency of 2 ps$^{-1}$ is used as the thermostat. Monte Carlo barostat is used for constant pressure dynamics. Nonbonded cutoff distance is set to 8 Angstrom. Beyond the cutoff distance the van der Waals interactions are accounted for via a continuum model (vdwmeth = 1). Default values were used for the other common parameters.
A globular protein test case

We tested OPC3-pol with the regular 2 fs time step setting. As shown in Fig. 4.8, with both the AMBER ff14SB and ff19SB force fields, during the four 2 µs simulations the ubiquitin stays stable and close to the reference crystal structure, with the RMSD ranges from 1 to 2 Å. Under ff14SB, the ubiquitin is slightly more stable than under ff19SB, with an average RMSD at 1.0 Å in comparison with 1.2 Å from the latter.

A B-DNA test case

A B-DNA dodecamer (PDB: 1BNA) was simulated with OPC3-pol model under AMBER OL15 force field. Four independent simulations were performed, each 1 µs long. The simulated structure shows a 2 Å backbone RMSD from the crystal structure and its widths of major and minor grooves match the experimental reference well (Fig. 4.9).

An intrinsically disordered protein (IDP) Aβ42

Unlike globular proteins such as the ubiquitin, intrinsically disordered proteins (IDP) usually do not have a stably folded structure: instead, they constantly change structure between multiple conformations [8, 92, 96]. OPC3-pol fixes the over-compactness problem of the previous generation non-polarizable models such as TIP3P [88].

A common intrinsically disordered protein (IDP) benchmark, amyloid β-peptide (Aβ42) is simulated with OPC3-pol model and standard AMBER protein force fields ff14SB and ff19SB. Here we measure the radius of gyration $R_g$ of the simulated IDP Aβ42, and compare it with experimental observations. Under both the ff14SB and ff19SB force fields, the simulated Aβ42’s radius of gyration $R_g$ averages at 19 Å and fluctuates between 11 Å and 25 Å (Fig. 4.10).
4.5. SUPPLEMENTARY INFORMATION

(a) RMSD

Figure 4.8: Performance of OPC3pol model in all-atom MD simulation with 2 fs time step of 76-residue globular protein ubiquitin shown in the backbone root-mean-square deviation (RMSD) relative to the X-ray reference, over the course of 8 independent simulations, 2 µs each. Blue curves: AMBER-ff14SB force field, orange curves: AMBER-ff19SB force field. Top panel: RMSD curves of all of the trajectories (four for ff14SB and four for ff19SB) concatenated. Bottom panel: Probability density distributions of the RMSD values. The trajectories are combined for each of the two force fields respectively.

(b) Probability density of RMSD
Figure 4.9: Performance of OPC3pol model in all-atom MD simulations of a B-DNA dodecamer with AMBER OL15 force field and 2 fs time step, over the course of 4 independent simulations, 1 µs each. **Left:** The all non-hydrogen atom (excluding terminal nucleotides) RMSD to crystal structure reference. **Right:** The widths of major (red) and minor (green) grooves in comparison with the experimental NMR values (black line).

Figure 4.10: Radius of gyration $R_g$ of the IDP Aβ42 simulated with OPC3-pol water model. The two 8 µs simulations are under two standard AMBER protein force fields: ff14SB (blue) and ff19SB (orange).

With its radius of gyration fluctuating between 11 Å and 25 Å, the simulated Aβ42 protein constantly changes conformation, qualitatively matching the experimental observations. However, from the quantitative perspective, the conformational balance is shifted towards
less compact conformations compared to experiment since the experimental average radius of gyration is observed to be between 11 Å and 15 Å. While the results are promising, more detailed investigations are needed for OPC3-pol model’s application in IDP simulations.

### 4.5.4 Further speed-up with 4 fs time step and HMR

In addition to the traditional 2 fs time step size, we explored the possibility of using OPC3-pol model under a 4 fs time step with the help of hydrogen mass repartitioning (HMR) \[^{[41]}\].

**Bulk liquid water properties**

The default setting for hydrogen mass repartitioning (HMR) is to increase the mass of hydrogen atoms, excluding those in water molecules, from 1.008 Da to 3.024 Da and subtract the mass difference from the heavy atoms bonded to the hydrogen atoms. However, the default setting to leave water unchanged was tested for non-polarizable n-point water models such as TIP3P \[^{[41]}\], and it is unknown whether it is best for polarizable OPC3-pol model. Therefore, we simulate pure liquid water of OPC3-pol under 4 fs time step and with the water hydrogen mass repartitioned between 1.008 Da and 3.024 Da, then measure the five liquid water properties same as during the optimization of OPC3-pol. Out of the five liquid properties, only the self diffusion coefficient $D$ deviates considerably from the value measured under 2 fs time step, and the amount of deviation is highly correlated with the repartitioned hydrogen mass (Fig.4.11).

From the self diffusion coefficient versus hydrogen mass result, we propose that it is best to set OPC3-pol’s hydrogen mass to 2.008 Da for simulations with 4 fs time step and HMR.
CHAPTER 4. A FAST POLARIZABLE WATER MODEL FOR ATOMISTIC SIMULATIONS

Figure 4.11: Self diffusion coefficient $D$ of OPC3-pol water simulated under 4 fs time step and with different repartitioned hydrogen mass values.

Ubiquitin and B-DNA dodecamer

The performance of OPC3-pol model under 4 fs time step along with HMR was also tested in the same ubiquitin and B-DNA systems discussed earlier. If it can handle the large time step of 4 fs, without suffering too much on accuracy, OPC3-pol would enable even faster simulations (in terms of ns per day). Based on the 4 fs time step HMR pure water simulation results, we repartition OPC3-pol model’s hydrogen mass to 2.008 Da, and other hydrogen mass to 3.024 Da for both the ubiquitin and B-DNA dodecamer systems.

For ubiquitin protein, results are promising under 4 fs time step. With hydrogen mass repartitioning (HMR) applied, ubiquitin is simulated stably close to its crystal structure, with 1-2 Å RMSD (Fig.4.12) for both ff14SB and ff19SB force fields.

The 4 fs time step simulations of the B-DNA with OPC3-pol result in similar results and are as stable as the ones with 2 fs time step. As shown in Fig.4.13, the simulated B-DNA has a 1.5 Å average RMSD from crystal structure reference and the radii of its major and minor grooves match the experiment values.
4.5. SUPPLEMENTARY INFORMATION

Figure 4.12: Backbone root-mean-square deviation (RMSD) of ubiquitin, simulated 3 µs twice with OPC3pol model at 4 fs time step. The blue curve is the trajectory simulated with ff14SB and the orange curve with ff19SB. Both have the hydrogen mass repartitioning (HMR) applied.

Figure 4.13: B-DNA dodecamer (1BNA) simulated with OPC3-pol water model and under AMBER OL15 force field. Two independent trajectories are included, each is 2 µs long, with 4 fs time step and hydrogen mass repartitioning (HMR) applied. **Left:** The backbone RMSD of the simulations are ranging around 2 Å. **Right:** The widths of major (red) and minor (green) grooves match the experimental NMR values (black line).

4.5.5 Calculation of liquid water properties

During the global optimization process, liquid water properties of OPC3-pol model are calculated based on molecular dynamics (MD) simulations. Below are the details of the approaches to calculate each of the liquid water properties.
Average dipole moment

The Average dipole moment is estimated through the calculation of each water molecule’s dipole moment during the last 10 ns of the 50 ns NPT simulation of pure water in ambient condition. Evenly distributed 200 frames of the 10 ns trajectory are included in the dipole moment calculation. In each frame, all atom coordinates are recorded and the instantaneous dipole moment of each water molecule is calculated through the equation:

$$\mu = \sum_{j=1}^{N} q_j r_j$$  \hspace{1cm} (4.10)

Where $N$ is the number of charges in a water molecule and equals four in terms of OPC3Pol model. The final reported average dipole moment is the average of the dipole moments over all water molecules and 200 time-frames.

Static dielectric constant

The static dielectric constant $\epsilon_0$ is determined through [42, 46, 47]

$$\epsilon_0 = 1 + \frac{4\pi}{3k_B T V} (\langle M^2 \rangle - \langle M \rangle^2)$$  \hspace{1cm} (4.11)

where $M = \sum_i q_i r_i$, $r_i$ is the position of atom i, $k_B$ is the Boltzmann constant, $T$ is the absolute temperature and $V$ is the average volume of the simulation box. Convergence require 50 ns of NPT simulation, and the results are calculated by excluding the first 25 ns of the trajectories.
4.5. Supplementary Information

Self diffusion coefficient

The self-diffusion coefficient $D$ is obtained using the Einstein relation [42, 46, 47]

$$D = \lim_{t \to \infty} \frac{1}{6t} < |r(t) - r(0)|^2 > \quad (4.12)$$

After a well equilibrated NPT simulation, 80 successive short NVE simulations (20 ps) each followed by a short NPT simulation (5 ps) are carried out. The self diffusion coefficient $D_1$ is calculated by averaging $D$ values over all the NVE runs.

Previous studies show that the $D$ value calculated through this method is affected by the size of the simulation box [118]:

$$D = D_0 - \frac{k_B T \xi}{6\pi \eta L} \quad (4.13)$$

where $L$ is the edge length of the cubic simulation box.

To correct the volume effect on the calculated self diffusion coefficient, we apply the same simulation protocol on a larger box (6448 water molecules), and calculate its self diffusion coefficient $D_2$. The corrected $D_0$ can be then calculated through

$$D_0 = \frac{D_2 L_2 - D_1 L_1}{L_2 - L_1} \quad (4.14)$$

where $L_1$ and $L_2$ are the edge lengths of the cubic simulation boxes corresponding to $D_1$ and $D_2$ respectively.
Heat of vaporization

The heat of vaporization $\Delta H_{vap}$ is calculated as

$$\Delta H_{vap} \approx \frac{3}{2}RT - U_{liq}/N + RT - pV/N + C$$

(4.15)

where $U_{liq}$ is the potential energy of the liquid with $N$ molecules at a given external pressure $p$ and a temperature $T$, and $V$ is the average volume of the simulation box. $R$ is the ideal gas constant and the first $3/2RT$ term is the vibrational potential energy from the Drude oscillator in gas phase. $C$ is a correction term to account for polarization effect ($C_{pol}$), change in intra- and inter- molecular vibrational modes ($C_{vib}$) and for non-ideal gas behavior ($C_{ni}$). $C_{vib}$ and $C_{ni}$ are set to the value reported by Horn et al [42]. $C_{pol}$ is calculated as:

$$C_{pol} = - \left( \frac{\left( \mu_{liq}^{model} - \mu_{gas}^{exp} \right)^2}{2\alpha_{exp}} - \frac{\left( \mu_{liq}^{model} - \mu_{gas}^{model} \right)^2}{2\alpha_{model}} \right)$$

(4.16)

where $\mu_{liq}^{model}$ is the model’s average dipole moment in liquid phase, $\mu_{gas}^{exp}$ is the experimental dipole of gas phase water, $\mu_{gas}^{model}$ is the model’s dipole moment in gas phase, $\alpha_{exp}$ and $\alpha_{model}$ are experimental gas phase water polarizability and the model’s polarizability respectively.

This is an extended form of the polarization correction proposed by Berendsen et al. [9] The first term is the total estimated polarization energy and the second term is the portion of the energy stored in the polarizable model in liquid phase.

Isobaric Heat Capacity $C_p$

The isobaric heat capacity is defined as the derivative of the enthalpy $H$ with respect to temperature $T$ under constant pressure $p$:
4.5. Supplementary Information

\[ C_p = \left( \frac{\partial H}{\partial T} \right)_p \]  

(4.17)

To calculate the derivative, we ran a series of NPT simulations of pure water under 1 bar constant pressure and constant temperatures ranging from 250 K to 370 K in 5 K intervals. Each simulation is pre-equilibrated and lasts 50 ns long. Each Enthalpy is calculated as an average over the 50 ns trajectory and the derivative is approximated as the slope of the linear regression of enthalpies over temperatures. With correction terms added, the final \( C_p \) is calculated as:

\[ C_p = \left( \frac{dH}{dT} \right)_p + \Delta C_{QM} - 3R \]

(4.18)

where \( \Delta C_{QM} \approx -2.2408 \) at \( T = 298 \) K and \( R \) is the ideal gas constant. The \( \Delta C_{QM} \) is the quantum correction to account for the quantized character of the neglected intramolecular vibrations. As a classical oscillator, the Drude oscillator stores internal energy of \( 3RT \), but the quantum polarization of the electrons mimicked by it does not. Therefore, we subtract \( 3R \)—the extra heat capacity of the Drude oscillator—from \( C_p \).

Thermal Expansion Coefficient \( \alpha_p \)

The thermal expansion coefficient is defined as the derivative of the logarithm of the volume \( \ln V \) with respect to temperature \( T \) under constant pressure \( p \):

\[ \alpha_p = \left( \frac{\partial \ln V}{\partial T} \right)_p = \frac{1}{V} \left( \frac{\partial V}{\partial T} \right)_p \]

(4.19)

Similar to the calculation of the isobaric heat capacity, here the same series of NPT simula-
tions of pure water under 1 bar constant pressure and constant temperatures ranging from 250 K to 370 K in 5 K intervals. Average volume of each simulation is recorded along with its temperature, and the derivative is approximated numerically.

**Isothermal Compressibility** $\kappa_T$

The isothermal compressibility is defined as the negative derivative of the logarithm of the volume $\ln V$ with respect to pressure $p$ under constant temperature $T$:

$$\kappa_T = -\left(\frac{\partial \ln V}{\partial p}\right)_T = -\frac{1}{V} \left(\frac{\partial V}{\partial p}\right)_T = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B \langle T \rangle \langle V \rangle} \quad (4.20)$$

This calculation is carried out on 50 ns NPT simulation trajectories’ volume fluctuations.

### 4.5.6 Temperature dependent properties

Calculated liquid water properties of OPC3-pol, including density, static dielectric constant, heat of vaporization, and self diffusion coefficient, in a temperature range of 220-360 K, are shown in Fig.4.14. Over the 200-360 K temperature range, OPC3-pol model’s static dielectric constant, heat of vaporization, and self diffusion coefficient all match the experimental values. The density curve of OPC3-pol model deviates from the experiment when the temperature is not at 300 K. This artifact is observed from many water models designed for molecular dynamics simulations, including polarizable models such as SWM4-NDP [97] and POL3 [111].
4.5. Supplementary information

(a) Liquid density $\rho$.

(b) Static dielectric constant $\epsilon_0$.

(c) Heat of vaporization $\Delta H_{\text{vap}}$.

(d) Self diffusion coefficient $D$.

Figure 4.14: Temperature dependency curves (blue curve) of OPC3-pol model’s density, static dielectric constant, heat of vaporization, and self diffusion coefficient, in comparison with experiment reference (black curve). Temperature ranges from 220 K to 360 K and the pressure is 1 bar.
Chapter 5

Conclusions

In conclusion, we first start with studying possible novel avenues for optimization strategies aimed at improving accuracy of explicit water models for biomolecular simulations. Several different types of water models are constructed and examined with respect to their ability to reproduce properties of water dimer—a simple structure formed with two water molecules and a mimic of the condensed state of water. Water models constructed include 3-point, 4-point, 5-point, and 3-point polarizable Drude models. These water models are optimized to reproduce the total dipole moment of water dimer, and then evaluated on their accuracy of the simulated structure of water dimer. For rigid \( n \)-point water models, increasing the number of interaction points from \( n = 3 \) to \( n = 4 \) brings considerably better accuracy, while further increasing \( n \) to 5 the improvement is marginal. Then, the effect of adding electronic polarizability to water model is examined. When fully optimized, the 3-point Drude polarizable model achieved slightly better accuracy than the 5-point non-polarizable model, revealing the potential benefit of a polarizable water model over a non-polarizable one.

Secondly, in addition to the main project of developing a novel water model, we investigate one of the non-liquid properties of OPC and OPC3 water model—the melting point. OPC and OPC3 water models are global optimal classical non-polarizable 4-point and 3-point models respectively, developed in our group through the approach also applied for OPC3-pol model. Both OPC and OPC3 reproduce liquid water properties accurately and perform
well in real-world applications, but their accuracy outside of liquid phase was unknown. Through the direct coexistence method, along with a new pressure scaling option we added to AMBER, the melting points of OPC and OPC3 model were measured. Although both OPC and OPC3 are accurate simulating liquid water, their melting points have considerable deviations from the experiment value and close to their 4-point and 3-point counterparts such as TIP4P-Ew and SPC/E.

Lastly, as a continuation of the first water dimer study, we have chosen the 3-point Drude polarizable model type to develop a global optimal general purpose water model for use in atomistic molecular dynamics simulations. In this global optimal polarizable water model OPC3-pol, we made an important design choice—splitting half of the oxygen mass onto the Drude particle and treat it as a regular “atom”. With this new design, the location of the Drude particle can be determined during each dynamics step without the need of extra calculations required in previous Drude implementations. OPC3-pol model accurately reproduces liquid water properties under ambient conditions, and is tested to simulate ubiquitin protein and B-DNA dodecamer with low errors to reference crystal structures. With the help of hydrogen mass repartitioning (HMR), OPC3-pol also supports a time step as high as 4 fs, this allows OPC3-pol to run roughly 5 times faster than the already fast SWN4-NDP Drude model and in-between 3- and 4-point non-polarizable water models such as TIP3P and TIP4P. Despite being accurate and fast, OPC3-pol is not perfect. OPC3-pol is optimized under ambient condition, and some temperature dependent properties of water are not reproduced when the temperatures is not close to 300 K. Also, the 1:1 mass split of Drude oscillator brings the high computational efficiency of OPC3-pol model, but potentially this design may cause mass polarization effects, which will need further fine-tuning to alleviate. OPC3-pol works with at least two latest non-polarizable AMBER force fields, which makes it easy to access for users, but without a specialized force field some subtle behavior
may deviate from experimental observations. Overall, we achieved the goal of developing a polarizable water model with good accuracy and runs as fast as non-polarizable models.
Bibliography


[45] Jing Huang, Justin A. Lemkul, Peter K. Eastman, and Alexander D. MacKerell Jr. Molecular dynamics simulations using the drude polarizable force field on gpus
with openmm: Implementation, validation, and benchmarks. *Journal of Com-


[51] Zhifeng Jing, Chengwen Liu, Sara Y. Cheng, Rui Qi, Brandon D. Walker, Jean-Philip Piquemal, and Pengyu Ren. Polarizable force fields for biomolecular simulations:


[92] J. Roche, Y. Shen, J. H. Lee, J. Ying, and A. Bax. Monomeric abeta(1-40) and abeta(1-42) peptides in solution adopt very similar ramachandran map distributions


[109] Carlos Vega and Jose L. F. Abascal. Simulating water with rigid non-polarizable


