Development of Polarizable Methods for Molecular Mechanics Simulations

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Jaebeom Han

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jiali Gao, Adviser

January 2014

Acknowledgements

It is a great pleasure for me to thank

- Prof. Jiali Gao for his support and direction as adviser during my graduate course as well as for his patience,
- Former and current Gao group members, including Dr. Lingchun Song, Dr. Isegawa Miho, Dr. Kin-Yiu Wong, Dr. Wangshen Xie, Dr. Jorge Estevez Mews, Dr. Pinsker Yen-lin Lin, Dr. Yan Zhou, Dr. Alessandro Cembran, Dr. Dejun Si, Peng Zhang, Yingjie Wang, Michael Mazack, Makenzie Provorse, Carlos Ramirez-Mondragon, and Pavel Rehak, for their helpful discussions,
- Prof. Donald G. Truhlar and his group member, including Dr. Hannah R. Leverentz and Luke Fiedler, for their helpful discussions on the X-Pol potential projects,
- Korean colleagues in the Department of Chemistry and in the Department of Chemical Engineering and Materials Science, especially the late Dr. Seong Ho Choi,
- National Institute of Health, National Science Foundation, and the Department of Chemistry for funding,
- Minnesota Supercomputing Institute for computing resources for my projects,
- and God for his guidance and protection.

To My Dear, Minsu

Abstract

This dissertation presents the development of two different types of polarization methods for molecular simulation methods, including Monte Carlo and molecular dynamics (MD) simulations. The first model, which is a polarizable intermolecular potential function (PIPF) method, is based on the point dipole method, where polarization energy is obtained from induced dipole moments and is added as correction to a force field. Hydrogen sulfide (H_2S) molecule is studied and parameterized for the PIPF method, and this study displays that the PIPF method reproduces experimental gas-phase dipole moment, molecular polarizability, liquid density, and heat of vaporization very well with a relative error of less than 1.0%. Due to the over-polarization of the model, however, some liquid properties and liquid structure failed to reproduce experimental values, which indicates further improvement is necessary for the PIPF method. The second one is an explicit polarization (X-Pol) method, which is a self-consistent fragment-based electronic structure theory in which molecular orbitals are block-localized within fragments of a cluster, macromolecule, or condensed-phase system. The Lennard-Jone potential function is incorporated into the X-Pol potential in order to express short-range exchange repulsion and long-range dispersion interactions. The X-Pol potential is first developed at the B3LYP hybrid density functional with the 6-31G(d) basis set, and the Lennard-Jones parameters have been optimized on a dataset consisting of 105 hydrogen-bonded bimolecular complexes. It is shown that the X-Pol potential can be optimized to provide a good description of hydrogen bonding interactions; the root mean square deviation (RMSD) of the computed binding energies from CCSD(T)/aug-cc-pVDZ results is 0.8 kcal/mol, and that of the calculated hydrogen bond distances is about 0.1 Å from B3LYP/aug-cc-pVDZ optimizations. In addition, the explicit polarization with three-point-charge potential (XP3P) model is introduced using the polarized molecular orbital model for

water (PMOw). The XP3P model is shown to be suitable for modeling both gas-phase clusters and liquid water, which is demonstrated from simulations of gas-phase water and protonated water clusters, and pure liquid consisting of 267 water molecules in a periodic system. This model is anticipated to be useful for simulating biological system in the condensed phase.

Table of Contents

List of Tables

List of Figures

List of Schemes

Chapter 1

Introduction

Molecular dynamics simulations based on empirical force field methods have been widely used for a variety of chemical and biological systems.[1,2] The Lifson-type, pairwise empirical potential functions have produced reasonable potential energy surfaces for these systems with equitable computational time and memory requirements.[1-3] Shown in eq. (1.1) is the CHARMM force field,[4] developed to investigate biological systems using harmonic-approximations to describe internal motions of biomacromolecules.

$$\begin{aligned} U(\mathbf{R}) = & \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{Urey-Bradley}} K_{\text{UB}} (S - S_0)^2 \\ & + \sum_{\text{dihedrals}} K_\varphi (1 + \cos(n\varphi - \delta)) + \sum_{\text{impropers}} K_\omega (\omega - \omega_0)^2 \\ & + \sum_{\substack{\text{non-bonded} \\ \text{pairs}}} \left\{ \varepsilon_{ij}^{\min} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \right\} \end{aligned} \tag{1.1}$$

As shown above, only nuclear positions are used, while electrons and their effects are modeled by an effective Coulumb potential for electrostatic interaction. This approximation leads to employ fixed atomic partial charges, which overestimate gas-phase dipole moments, in order to mimic the polarization effects in the condensed phase.

The fixed and enhanced atomic partial charges in these "non-polarizable" force fields treat the polarization effects in an effective way. However, they cannot correctly reproduce the dependence of charge distributions on the system's state. In addition, the non-polarizable models are unable to directly respond to the fluctuations in the electric field due to molecular motion.[1,5-8] In order to remedy this problem, many studies have been carried out to implement the polarization effects. Three methods have been developed and used mostly for classical methods,

which are also called "polarizable" force fields; the point dipole,[9,10] the Drude oscillator,[11] and the fluctuating charge methods.[12] The point dipole method is the simplest way among these methods, where the polarization energy is calculated using the induced dipole moments and isotropic atomic polarizabilities, and is added to the typical force field like eq (1.1) above. In the Drude oscillator method, the Drude particles are introduced, which are attached to atoms by harmonic springs in order to mimic the polarization effects. Then, the displacement of the Drude particles leads to the generation of the induced dipole moments, and the electrostatic interaction also includes their effects. The fluctuating charge method, however, is based on the electronegativity equalization,[13] and the electrostatic interaction of a molecule is obtained from the screened Coulomb integrals. Their successful developments and applications are reported in a special issue at Journal of Chemical Theory and Computation in 2007.[14]

However, although the force field methods have successfully applied to various chemical and biological systems, the reorganization of the electronic structures by the bond formation and breaking cannot be handled by these simple potential functions, and quantum mechanics is a necessary tool. Due to computational costs, however, it is impractical to employ quantum mechanical methods in order to determine the potential energy surface of macromolecular systems. This obstruction led to combining quantum mechanical methods with molecular mechanical methods by dividing a system into two regions, which are a quantum mechanical region and a molecular mechanical region. This "multi-scaling" method was first applied to a conjugated system by Warshel and Karplus,[15] where the σ and π electrons are treated by the empirical potential and the semiempirical quantum mechanical methods, respectively. This method was extended to an enzyme system by Warshel and Levitt,[16] where the active site of lysozyme and the substrate are represented by the semiempirical quantum mechanical method and the rest of the system is treated by the empirical potential function, and is generally called the combined quantum mechanical and molecular mechanical (QM/MM) method now. These studies

allow the researches stated above to win the Nobel Prize in Chemistry in 2013.[17] Although combined QM/MM methods have been shown to be a powerful tool in studying chemical and biological systems,[18,19] as the computing power improves, the quantum mechanical regions have been expanded in order to obtain more accurate potential energy surfaces of biological systems, resulting in the development of fragmentation methods. (See Ref. 20 for review on the development of fragmentation methods and their applications.)

This dissertation presents two different polarization methods, which are used for molecular dynamics simulations: the first one is a polarizable intermolecular potential function (PIPF) method, based on the Thole's interaction dipole (TID) method,[10] and the second one is an explicit polarization (X-Pol) potential,[21-23] which is based on quantum mechanics with a fragmentation method.

Chapter 2 presents the parameterization of the PIPF method for hydrogen sulfide (H_2S) molecule in the liquid phase. This method produces the density, the heat of vaporization, and the diffusion coefficient of H_2S in the liquid phase in excellent agreement with experiments. In addition, with the atomic partial charges and atomic polarizability optimized in the gas phase, the liquid phase dipole moment looks quite reasonable, compared to other H_2S models by both non-polarizable and polarizable methods. The liquid structure of H_2S is also investigated by comparing the computed radial distribution functions (RDFs) with the experiment. The PIPF method predicts the peak positions in a good agreement with the experiment, and the RDFs by PIPF also show characteristics of a simple Lennard-Jones fluid, which indicates that H_2S lacks the hydrogen-bonding network in the liquid phase as shown in the experiment. However, the overall shapes of the RDFs by PIPF are more structured than those by the experiment, which requires further study.

In Chapter 3, the X-Pol potential is parameterized at a hybrid density functional theory (DFT) using the B3LYP method with 6-31G(d) basis set. This chapter shows that the

parameterization of the neglected exchange-repulsion and dispersion terms in the X-Pol potential using the Lennard-Jones potential function improves energies and geometries of hydrogen-bonded complexes. Employing several hydrogen-bonded complexes with a variety of functional groups and ions, we discover that the X-Pol calculations at B3LYP/6-31G(d), XP@B3LYP/6-31G(d) in short hereafter, yield energies and geometries in excellent agreement with energies from CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ and geometries from B3LYP/aug-cc-pVDZ calculations. Compared to similar studies using the combined QM/MM method using various electronic structure methods and molecular mechanics force fields, XP@B3LYP/6-31G(d) calculations yield better performance because of the inclusion of mutual polarization effects between two monomers in each complex. This result shows that the proper parameterization of the X-Pol potential is useful for biomolecular simulations in the condensed phase at a moderate quantum mechanical method.

Chapter 4 is the first step for the condensed-phase simulations using the X-Pol potential. The water model for the X-Pol potential, which is called the explicit polarization with three-point-charge potential (XP3P), is developed using a new neglect of diatomic differential overlap (NDDO) method, a polarized molecular orbital (PMO) model. The PMO method adds diffuse *p* functions on hydrogen atom in order to improve molecular polarizabilities, which is underestimated by typical NDDO methods with a minimal basis set. The further parameterized PMO method for water molecule, or called PMOw, was employed for our water model together with the dipole preserving and polarizability consistent (DPPC) charge model. The DPPC charge model was developed for the NDDO methods in order to obtain reasonable atomic partial charges and dipole moments. The PMOw model reproduces several gas-phase properties of hydrogen-, and/or oxygen-containing molecules, binding energies and geometries of water clusters, and proton transfer potential energy surfaces and interaction energies of protonated water clusters from experiments and high-level quantum chemistry calculations very well. In the liquid phase,

the XP3P model produces several liquid-phase properties in excellent agreement with experimental data, and the liquid structure predicted by the XP3P model is much more improved than our previous water model at AM1. In addition, this model shows some interesting features from Monte Carlo simulations at a variety of temperature ranges from -40 ℃ to 100 ℃; especially, there exists a maximum-density temperature with a three-site model although it is not correct. Although further studies are necessary using molecular dynamics method, the XP3P model is believed to be useful for studying biological systems in the condensed phases.

Chapter 2

Development of a Polarizable Intermolecular Potential Function (PIPF) for Hydrogen Sulfide Using Molecular Dynamics Simulations

2.1. Introduction

Computer simulations using quantum mechanics and classical mechanics are important tools for studying chemical and biological systems such as proteins and nucleic acids.[1-5] Molecular dynamics (MD) simulations are typically performed by using additive, pairwise potential energy functions, in which partial atomic charges are fixed and are optimized based on the mean-field approximation to achieve computational efficiency. These fixed partial charges overestimate gas-phase dipole moments in order to properly describe the bulk properties of liquids. These pairwise, effective potentials have been widely used and can yield thermodynamic results in agreement with experiments for biomolecular systems.

However, such fixed charges do not properly describe the instantaneous charge distribution of a system.[4,6] This is particularly important in the study of the solution of ions and processes that involve a large change in the environment. For example, Foresman and Brooks[7] investigated the interaction energies between a chloride ion and water molecules by varying the number of water molecules and geometries using ab initio molecular orbital calculations. The results are compared with those obtained from two-body additive potentials. They found that for large complexes, the deviations of their calculations from two-body additive results are 4.72 kcal/mol in the five-water complex and 7.65 kcal/mol in the six-water system, respectively. The difference was attributed to polarization effects, and it has become an important task to incorporate polarization effects in

potential energy functions in MD simulations. The importance of polarizable force fields has been discussed in several review articles.[4-6,8-10]

We have explored a polarizable intermolecular potential function (PIPF) for liquid alcohols and amides,[11,12] which were chosen to represent the functional groups found in biomolecules such as proteins. In this project, we plan on focusing on the parameterization for sulfur-containing compounds, following the parameterization procedure that has been employed for alcohols, amides and alkanes.[13] Previously, a number of studies have been carried out using a variety of non-polarizable and polarizable force fields in order to parameterize the force fields for sulfur-containing molecules.[14-23] These systems are models for sulfur-containing amino acid residues, including cystein and methionine, and the disulfide bridge. Unlike other organic species such as alcohol and amide, which have been extensively studied and parameterized for polarizable force fields, few studies have been performed on sulfur-containing molecules using polarizable force fields. In this project, hydrogen sulfide (H_2S) was considered as an initial step for modeling sulfur-containing molecules using the PIPF method. H_2S is known for its corrosivity and toxicity in crude oil and natural gas, and studies on its effect on human health in petroleum industry[24] and its physiological roles in the human body[25] have been carried out. The goal of this study is to parameterize the PIPF force field for H_2S molecule, and to check the validity of the optimized parameters in the liquid state by comparison with non-polarizable OPLS,[14] Forester,[16] KL,[17] NERD,[20] and Potoff[21] models and polarizable Delhommelle[19] and Drude[23] models.

In the following, a brief summary of the theoretical backgrounds of the PIPF method will be described in Section 2-2, followed in Section 2-3 by computational details. In Section 2-4, computational results and discussions are provided, and Section 2-5 wraps up this chapter with concluding remarks and summary.

2.2. Polarizable Intermolecular Potential Function (PIPF)

The PIPF model has been developed based on the induced point dipole method.[26,27] The induced dipole moment at an atomic site *i* is given as

$$\boldsymbol{\mu}_i = \boldsymbol{\alpha}_i \mathbf{E}_i = \boldsymbol{\alpha}_i \left(\mathbf{E}_i^0 - \sum_{j \neq i}^{N} \mathbf{T}_{ij} \boldsymbol{\mu}_j \right) \tag{2.1}$$

where $\boldsymbol{\alpha}_i$ is the atomic polarizability tensor of atom *i*, $\mathbf{E}_i$ is the total electric field at atom site *i*, $\mathbf{E}_i^0$ is the electric field at atomic site *i* due to the permanent charges on all other atoms, *N* is the number of atomic interaction sites, and $\mathbf{T}_{ij}$ is the dipole field tensor defined as

$$\mathbf{T}_{ij} = \frac{1}{r_{ij}^3}\mathbf{I} - \frac{3}{r_{ij}^5}\begin{bmatrix} x^2 & xy & xz \\ yx & y_2 & yz \\ zx & zy & z^2 \end{bmatrix} \tag{2.2}$$

where **I** is the unit matrix, r_{ij} is the distance between atoms *i* and *j*, and *x*, *y*, and *z* are the Cartesian coordinate components of the vector between atom *i* and *j*. Then, the polarization energy E_{pol} is calculated as

$$E_{\text{pol}} = -\frac{1}{2}\sum_{i=1}^{N} \boldsymbol{\mu}_i \cdot \mathbf{E}_i^0 \tag{2.3}$$

which is added to the standard CHARMM force field in eq (1.1).

The induced point dipole method has a problem, called the polarization catastrophe, leading to an infinite polarization when the distance between atoms *i* and *j* is $\left(4\alpha_i\alpha_j\right)^{1/6}$. For a diatomic molecule, the parallel and perpendicular components of the molecular polarizability to the bond axis are derived as follows

$$\alpha_{\parallel} = \frac{\alpha_i + \alpha_j + 4\alpha_i\alpha_j / r_{ij}^3}{1 - 4\alpha_i\alpha_j / r_{ij}^6} \tag{2.5}$$

$$\alpha_{\perp} = \frac{\alpha_i + \alpha_j - 2\alpha_i\alpha_j / r_{ij}^3}{1 - \alpha_i\alpha_j / r_{ij}^6} \tag{2.6}$$

When r_{ij} approaches $\left(4\alpha_i\alpha_j\right)^{1/6}$, the denominator of $\alpha_{\|}$ becomes zero, leading to infinite value for $\alpha_{\|}$. In order to avoid this polarization catastrophe, Thole's interaction dipole (TID) model[27] is employed with the use of a damping scheme, which is independent of atoms, but dependent on a model charge distribution ρ.

Thole tested several forms of charge densities. Of those charge density distributions, we employed the form of the smeared charge distribution in PIPF as in the AMOEBA force field by Ren and Ponder[28,29] as follows

$$\rho(u) = \frac{3a}{4\pi}\exp\left(-au^3\right) \tag{2.7}$$

where a is a dimensionless damping factor that controls the strength of damping, which is set to 0.572 in PIPF, and $u = r_{ij}\Big/\left(\alpha_i\alpha_j\right)^{1/6}$ is an effective distance between atoms. Using the charge distribution in eq (2.7), the damped first-order dipole field matrix can be expressed by

$$\left(\mathbf{T}_{ij}\right)_p^D = -\left\{1-\exp\left(-au^3\right)\right\}\frac{\left(\mathbf{r}_{ij}\right)_p}{r_{ij}^3} \tag{2.8}$$

where p represents a Cartesian component of the vector $\mathbf{r}_{ij}$, and D denotes a damped interaction tensor. Then, the modified higher-order **T** matrix elements can be obtained by the successive derivatives of the preceding lower rank elements

$$\left(\mathbf{T}_{ij}\right)_{pq}^D = \nabla_p\left(\mathbf{T}_{ij}\right)_q^D = \lambda_5\frac{3\left(\mathbf{r}_{ij}\right)_p\left(\mathbf{r}_{ij}\right)_q}{r_{ij}^5} - \lambda_3\frac{\delta_{pq}}{r_{ij}^3} \tag{2.9}$$

$$\left(\mathbf{T}_{ij}\right)_{pqr}^D = \nabla_p\left(\mathbf{T}_{ij}\right)_{qr}^D = -\lambda_7\frac{15\left(\mathbf{r}_{ij}\right)_p\left(\mathbf{r}_{ij}\right)_q\left(\mathbf{r}_{ij}\right)_r}{r_{ij}^7} + \lambda_5\frac{3\left[\left(\mathbf{r}_{ij}\right)_p\delta_{qr} + \left(\mathbf{r}_{ij}\right)_q\delta_{pr} + \left(\mathbf{r}_{ij}\right)_r\delta_{pq}\right]}{r_{ij}^5} \tag{2.10}$$

where λ_i are given as follows

$$\lambda_3 = 1-\exp\left(-au^3\right) \tag{2.11}$$

$$\lambda_5 = 1-\left(1+au^3\right)\exp\left(-au^3\right) \tag{2.12}$$

$$\lambda_7 = 1 - \left(1 + au^3 + \frac{3}{5}a^2u^6\right)\exp\left(-au^3\right) \tag{2.13}$$

The interactions are damped only between interacting induced dipoles, whereas the interactions due to the permanent charges are not affected.

The induced dipole moment in eq (2.1) is typically solved self-consistently because the induced dipole moment of each atomic site is dependent on those of all other sites. An initial guess of induced dipole moments is made in order to obtain a set of new induced dipole moments, and these induced dipoles are used to generate the next set of induced dipoles until a pre-defined convergence criterion is met. Another way to obtain induced dipole moments is by the matrix inversion. Eq (2.1) can be rearranged as follows

$$\boldsymbol{\mu} = \mathbf{A}^{-1}\mathbf{E}^0 \tag{2.14}$$

where $\boldsymbol{\mu}$ is the column induced dipole vector, $\mathbf{E}^0$ is the electric field matrix due to permanent partial charges, and the interaction matrix $\mathbf{A}$ is defined as

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{\alpha}_1^{-1} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1N} \\ \mathbf{T}_{21} & \boldsymbol{\alpha}_2^{-1} & \cdots & \mathbf{T}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}_{N1} & \mathbf{T}_{N2} & \cdots & \boldsymbol{\alpha}_N^{-1} \end{bmatrix} \tag{2.15}$$

The matrix inversion method provides the exact results, but the disadvantage of this method is the high computational cost, scaling as $O(N^3)$ to invert the $\mathbf{A}$ matrix, which is $3N \times 3N$ dimension.[13] Thus, the iterative self-consistent method is useful in the MD simulations, but this method is also known for its numerical instability when intramolecular 1-2 and 1-3 polarization terms are included.[30] In order to include the intramolecular polarizations, a coupled polarization-matrix inversion and iteration (CPII) method was developed, where the matrix inversion is employed for intramolecular interactions and the iterative method is for intermolecular interactions.[30] In this study, however, only the iterative method was employed to obtain the induced dipole moments.

2.3. Computational Details

2.3.1. Parameterization

The parameterization in this study for hydrogen sulfide uses the same process described in Ref. 13, where the PIPF-CHARMM force fields for liquid amides and alkanes were parameterized. In a previous study, we already optimized the atomic polarizability and Lennard-Jones (LJ) parameters for hydrogen. Thus, the atomic polarizability and LJ parameters for sulfur, atomic partial charges of sulfur and hydrogen, and internal parameters related to sulfur were considered in this study.

The nonbonded parameters include the atomic polarizability of sulfur atom, atomic partial charges of sulfur and hydrogen, and Lennard-Jones parameters of sulfur atom. The atomic polarizability of sulfur atom was optimized to fit to the experimental molecular polarizabilities of hydrogen sulfide, ethanethiol, diethyl sulfide, ethyl methyl sulfide, and dimethyl disulfide, and computed molecular polarizabilities of methanethiol and dimethyl sulfide at B3LYP/aug-cc-pVQZ using GAUSSIAN 03 program.[31] The atomic partial charges of hydrogen sulfide were chosen to reproduce the gas-phase dipole moment. The Lennard-Jones parameters of sulfur atom were adjusted to fit to the experimental liquid density and heat of vaporization through liquid-phase simulations using the molecular dynamics method.

The internal parameters, including the equilibrium geometries of hydrogen sulfide and force constants in bond, angle, and Urey-Bradley energy functions, were initially taken from the experimentally determined geometrical parameters in the gas phase, and the OPLS force field.[14] The force constants were refined through the vibrational analysis using MOLVIB module in CHARMM by comparing vibrational frequencies of hydrogen sulfide from the experiment and the quantum chemistry calculation at B3LYP/6-311++G(d,p).

2.3.2. Liquid-Phase Simulations

Liquid-phase simulations with 256 hydrogen sulfide molecules were carried out in order to optimize the Lennard-Jones parameters of sulfur atom using c33a3 version of CHARMM program[32] in the isothermal-isobaric (NPT) ensemble at the boiling point of hydrogen sulfide, 212.81 K, and 1 atm with periodic boundary conditions, where the box size is about 24.8 × 24.8 × 24.8 $Å^3$. Nosé-Hoover thermostat[33,34] and the Langevin piston method[35] were used to control the temperature and the pressure, respectively. The velocity Verlet algorithm was used for the integration with a time step of 1 fs. A spherical cutoff was used to generate the nonbonded list for all pairs within 12.5 Å. The van der Waals interactions beyond the cutoff were corrected using the method explained in Ref. 36, and the particle-mesh Ewald (PME)[37] was used for long-range electrostatic interactions. During simulations, we constrained the geometry of hydrogen sulfide using the SHAKE algorithm.[38] Therefore, the current H_2S model in PIPF employs a rigid geometry. The system was equilibrated for 1 ns, followed by additional 1-ns runs for averaging. Statistical uncertainties ($\pm 1\sigma$) for the computed liquid-phase properties were determined through averages of blocks of 200-ps simulations.

Monte Carlo simulations were carried out to obtain the average energy of a single molecule in the gas phase at the same temperature as in the liquid phase with the Metropolis sampling.[39] 1×10^6 configurations were sampled for equilibration, and additional 5×10^6 configurations were sampled for averaging. This average energy obtained was used in calculating the heat of vaporization of hydrogen sulfide in the liquid phase.

2.4. Results and Discussion

2.4.1. Parameterization

The optimized non-bonded parameters and internal parameters for H_2S are listed in Tables 2-1 and 2-2, respectively. For H atom, the non-bonded parameters were taken from Ref. 13 except

its atomic partial charge. The initial atomic partial charges of H_2S were taken from Ref. 14 by Jorgensen. The resultant dipole moment is calculated to 0.984 D (Table 2-3), which is slightly higher than the experimental dipole moment by 0.6%. In contrast to PIPF, the non-polarizable models produce gas-phase dipole moments larger than the experimental value by 15.6 ~ 114.7% in the relative error,[14-17,20] and this is because the non-polarizable models overestimate the gas-phase dipole moment in order to mimic the polarization effects in the liquid phase in an average way. Other polarizable models such as the Delhommelle model[18] and the Drude model,[23] however, yield gas-phase dipole moments comparable to PIPF. The atomic polarizability of sulfur atom was first taken from Ref. 40, and the final atomic polarizability produces 3.600 $Å^3$ for the molecular polarizability of H_2S, which is similar to the experimental value of 3.631 $Å^3$,[41] smaller by 0.9%.

Table 2-1. Optimized Non-bonded Parameters for H_2S

Atom type	$R_{min}/2$ (Å)	ε (kcal/mol)	q (e)	α ($Å^3$)
S	2.116	-0.521	-0.22	2.985
H(H-S)	0.7577[a]	-0.015[a]	0.11	0.496[a]

[a]Values are taken from Ref. 13.

Table 2-2. Optimized Internal Parameters for H_2S

Internal term type	Force constant	Equilibrium value
S-H, bond (kcal mol^{-1} $Å^{-2}$)	324.440	1.34
H-S-H, angle (kcal mol^{-1} rad^{-2})	48.783	92.0
H-S-H, Urey-Bradley (kcal mol^{-1} $Å^{-2}$)	4.590	1.9278

With atomic partial charges and atomic polarizabilities obtained, the average molecular dipole moment of H_2S in the liquid phase is listed in Table 2-3. In the liquid phase, the induced dipole moment is calculated to be 0.292 D, and the total dipole moment is 1.220 D, which is enhanced from the gas-phase dipole moment by 24.0%. It is found that the obtained total dipole moment in the liquid phase is almost comparable to dipole moments by non-polarizable methods except the OPLS model that produces 2.1 D. (Table 2-3) Although this increase of the dipole moment in the PIPF method cannot be verified due to the lack of experimental data, the overestimated dipole moments from non-polarizable models are consistent with this trend by the PIPF method in the liquid phase. The Lennard-Jones parameters were initially taken from the CHARMM22 force field, and the optimized parameters are also shown in Table 2-1. Compared with the initial corresponding parameters of the CHARMM22 force field, the sulfur radius increases from 2.000 Å, and the depth decreased from 0.000 kcal/mol. That is, although the previous study[13] obtained Lennard-Jones parameters quite similar with their initial CHARMM parameters, S atom in the PIPF method has quite different parameters from the corresponding CHARMM parameters. The equilibrium geometry of H_2S in Table 2-2 was taken from the OPLS model,[14] where the experimental gas-phase geometry was employed. The vibrational frequencies of H_2S were calculated using the MOLVIB module of CHARMM with the optimized force constants. The obtained vibrational frequencies are listed in Table 2-4 along with the vibrational frequencies from the quantum mechanical calculation at B3LYP/6-311+G(d,p) and the experiment in the gas phase.

Table 2-3. Comparison of Dipole Moments and Molecular Polarizability of H_2S from PIPF and Other Models and from Experiements

Model	μ_g (D)[a]	μ_{tot} (D)[b]	μ_{ind} (D)[b]	α (Å^3)
PIPF	0.984	1.220	0.292	3.600
OPLS[c]	2.1			
Forester[d]	1.357			
KL[e]	1.4			
Delhommelle[f]	1.0			3.78
NERD[g]	1.134			
Drude[h]	0.98			2.5
Expt.	0.9783[i]			3.631[j]

[a]Gas-phase dipole moment. [b]Calculated in the liquid phase. [c]Ref. 14. [d]Ref. 16. [e]Ref. 17. [f]Ref. 18. [g]Ref. 20. [h]Ref. 23. [i]Ref. 42. [j]Ref. 41.

Table 2-4. Vibrational frequencies of H_2S (unit in cm^{-1})

Vibrational Mode	CHARMM	B3LYP/6-311+G(d,p)	Expt.[a]
Bending	1190.6	1207.6	1183
Sym. Stretching	2799.6	2679.6	2615
Anti-sym. Stretching	2816.2	2695.1	2626

[a]Ref. 43.

2.4.2. Liquid Properties

The energetic results are also shown in Table 2-5, including the total intermolecular interaction energy of the liquid, ΔE_i, which is the sum of the electrostatic energy, ΔE_{elec}, the van der Waals energy, ΔE_{vdW}, and the polarization energy, ΔE_{pol}, and the intramolecular energy,

ΔE_{intra}. The sum of ΔE_i and ΔE_{intra} is the total energy of the liquid system, ΔE_{tot}. Because the current model employs a rigid geometry, leading to 0.0 kcal/mol for ΔE_{intra}, ΔE_i and ΔE_{tot} are the same as shown in Table 2.5. The heat of vaporization, ΔH_{vap}, is related to the total intermolecular energy in the liquid, $\Delta E_i(l)$, the intramolecular energies in the liquid, $\Delta E_{\text{intra}}(l)$, and in the gas phase, $\Delta E_{\text{intra}}(g)$ as below[13]

$$\begin{aligned} \Delta H_{\text{vap}} &= -\left(\Delta E_i(l) + \Delta E_{\text{intra}}(l)\right) + \Delta E_{\text{intra}}(g) + RT \\ &= -\Delta E_{\text{tot}}(l) + \Delta E_{\text{intra}}(g) + RT \end{aligned} \tag{2.15}$$

where R is the gas constant, and T is the temperature, which is 212.81 K in this study. Because this study employs a rigid geometry for H_2S as stated above, $\Delta E_{\text{intra}}(l)$ and $\Delta E_{\text{intra}}(g)$ are reduced to zero, resulting in $-\Delta E_i(l) + RT$ for ΔH_{vap} in eq. (2.15). The PIPF method produces the best heat of vaporization with less relative error than 0.1%, compared to other models. (Table 2-6)

Table 2-5. Computed Energetic Results for H_2S (unit in kcal/mol)

	$-\Delta E_i$	$-\Delta E_{\text{elec}}$	$-\Delta E_{\text{vdW}}$	$-\Delta E_{\text{pol}}$	$-\Delta E_{\text{intra}}$	$-\Delta E_{\text{tot}}$
H_2S	4.04	0.31	3.59	0.14	0.00	4.04

Together with the heat of vaporization, all liquid properties of H_2S calculated from the PIPF method are listed in Table 2-6 in comparison with other H_2S models. The liquid density of H_2S is obtained using the equation below

$$\rho = \frac{M}{N_A \bar{V}} \tag{2.14}$$

where M is the molar mass, N_A is Avogadro's number, and $\bar{V}$ is the molecular volume. The liquid density by PIPF is in excellent agreement with experiment, and other models also produce comparable results to experiment within 0.2% error. Because the Lennard-Jones parameters were

adjusted to fit to these two liquid properties, density and heat of vaporization, it is not surprising for the PIPF method to produce these properties in excellent agreement with experimental data.

Table 2-6. Computed and Experimental Liquid Properties of H_2S Compared with Other Models

	ΔH_{vap} (kcal/mol)	d (g/cm^3)	10^5 D (cm^2/s)	ε
PIPF	4.465 ± 0.003	0.9495 ± 0.0004	3.77 ± 0.01	12.0 ± 0.2
OPLS[a]	4.36 ± 0.01	0.948 ± 0.005	9.25 ± 0.21[b]	48.60 ± 1.14[b]
Forester[b]	4.873 ± 0.002	1.0165 ± 0.0004	2.88 ± 0.11	9.73 ± 0.15
KL[c]	3.91 ± 0.06	0.949 ± 0.002	3.73 ± 0.09[b]	11.59 ± 0.27[b]
Potoff[b]	4.718 ± 0.004	0.962 ± 0.002	4.43 ± 0.15	25.09 ± 0.75
Drude[b]	4.26 ± 0.00	0.9514 ± 0.0004	3.27 ± 0.09	8.18 ± 0.13
Expt.	4.462[d]	0.949[e]	3.7 (at 206.5 K)[f]	8.01[d]

[a]Ref. 14. [b]Ref. 23. [c]Ref. 17. [d]Ref. 44. [e]Ref. 45. [f]Ref. 46.

The self-diffusion constant, D, can be determined from the particle flux related to a concentration gradient; therefore, it can be evaluated from the velocity autocorrelation function. In MD simulations, however, the velocity autocorrelation function can be replaced by the mean-squared displacement using the Einstein relation. Therefore, the self-diffusion constant can be obtained from the slope of the mean-squared displacement versus simulation time as follows[47]

$$D = \lim_{t\to\infty} \frac{1}{6t} \left\langle \frac{1}{N} \sum_{i=1}^{N} \left| \mathbf{r}_i(t) - \mathbf{r}_i(0) \right|^2 \right\rangle \tag{2.16}$$

where N is the number of molecules, and $\langle\ \rangle$ specifies an ensemble average. The PIPF method produces 3.77×10^{-5} cm^2/s for the diffusion coefficient, which is in excellent agreement with the experimental value of 3.7×10^{-5} cm^2/s at 206.5 K. Of other H_2S models, the KL model yields the

best diffusion coefficient value, and the Drude model also produces reasonable diffusion coefficient. (Table 2-6)

The last liquid property calculated is the dielectric constant, ε, and it is calculated using the method developed by Neumann and Steinhauser in the periodic boundary conditions as follows[48]

$$\varepsilon = \varepsilon_\infty + \frac{4\pi}{3\langle V\rangle k_B T}\left(\langle \boldsymbol{\mu}^2 \rangle - \langle \boldsymbol{\mu} \rangle^2\right) \tag{2.17}$$

where $\langle V \rangle$ is the average volume of the simulation box, k_B is the Boltzmann constant, T is the specified temperature, and $\boldsymbol{\mu}$ is the total dipole moment of the simulation box. The high frequency dielectric constant, ε_∞, can be estimated from Clausius-Mossotti equation below

$$\frac{\varepsilon_\infty - 1}{\varepsilon_\infty + 2} = \frac{4\pi\alpha}{3\bar{V}} \tag{2.18}$$

where α is the gas-phase molecular polarizability, and $\bar{V}$ is the molecular volume from the simulation. The dielectric constant calculated from the PIPF method is higher than the experimental one with the relative error of 49.3%. It is known that $\langle \boldsymbol{\mu} \rangle$ and $\langle \boldsymbol{\mu}^2 \rangle$ are slow to converge, and the 1-ns trajectories for this system would not be enough to converge this property.

2.4.3. Liquid Structure

The structure of liquid can be characterized by the radial distribution function (RDF), $g_{xy}(r)$. RDF provides the probability of finding an atom y at a distance of r from a given atom x, and it is normalized to the bulk density. X-ray[49] and neutron[50,51] diffraction experiments were performed for liquid H_2S at several thermodynamic states. Of the experiments, we compare the RDFs from the neutron experiment[51] obtained at 298 K and 30.6 atm with those from the PIPF, which are obtained both at 212.81 K and 1 atm, and at 298 K and 30.6 atm. (Figure 2-1)

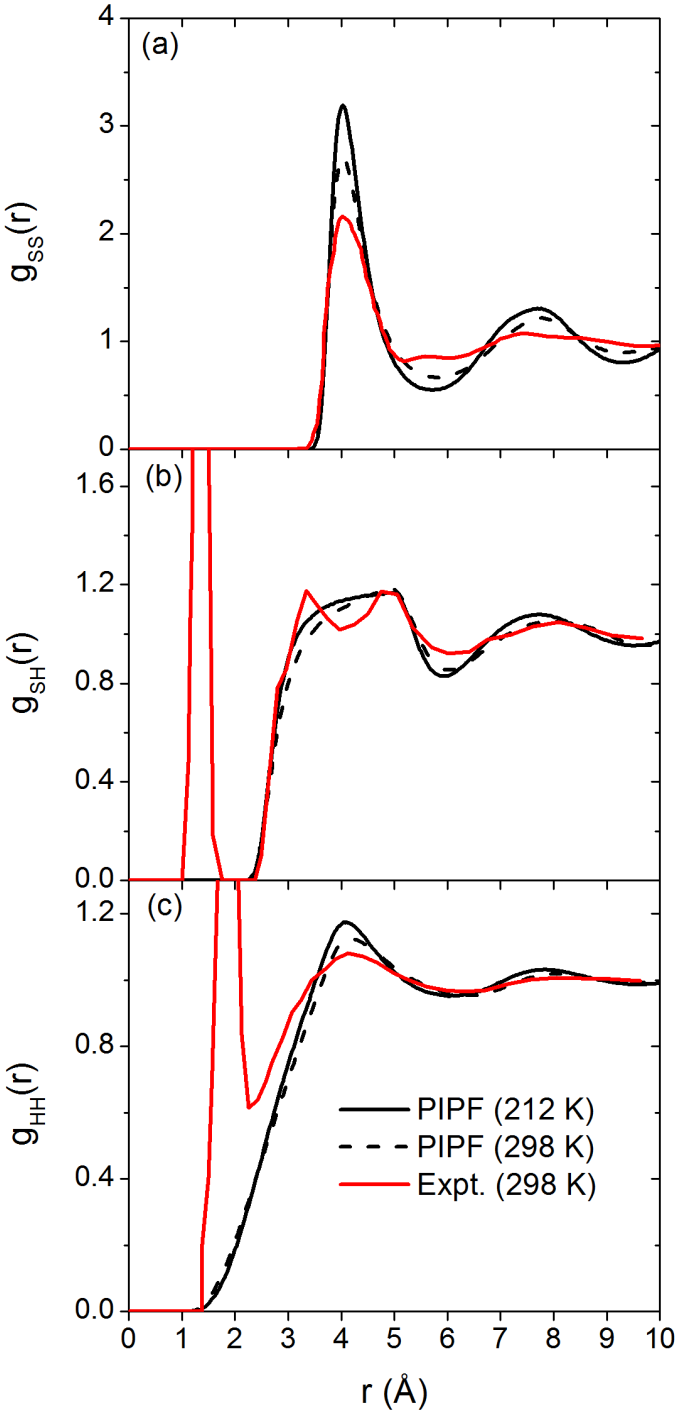


Figure 2-1. Computed (black and dashed black) and experimental (red) sulfur-sulfur (a), sulfur-hydrogen (b), and hydrogen-hydrogen (c) radial distribution functions of H_2S.

Figure 2-1(a) displays the sulfur-sulfur RDF, $g_{SS}(r)$. The first peaks from PIPF are located at 4.03 Å (212.81 K and 1 atom) and at 4.05 Å (298 K and 30.6 atom), respectively, and the experiment produces at 4.02 Å. However, the PIPF method produces more intense first peak than the experiment at both of two different thermodynamic states than the experiment. In addition, while the experiment g_{SS}(r) has a broad second peak at 7.53 Å, the PIPF yields more clear and structured second peaks at 7.68 Å (212.81 K and 1 atm) and 7.75 Å (298 K and 30.6 atm). Overall, g_{SS}(r) from the PIPF is more structured than that from the experiment.

Riahi and Rowley[23] also reported information of $g_{SS}(r)$ calculated from their Drude model together with other H_2S models, including the OPLS,[14] Forester,[16] KL,[17] and Potoff[21] models, and ab initio molecular dynamics (AIMD)simulation at 212 K and 1 atm. (Table 2-7) Almost all of the H_2S models except the OPLS model (3.75 Å) display the first peak of g_{SS}(r) between at 3.95 Å and at 4.05 Å, which is comparable with the position by PIPF at 4.03 Å. For the second peaks, most models have at 7.65 Å, which is similar with PIPF at 7.68 Å, while the OPLS and Forster models display the second peak at shorter locations, 7.35 Å and 7.55 Å, respectively. The ratio of the position of the second peak to the position of the first peak is also shown in Table 2-7. All of the H_2S models produce the similar ratio, which is near 2. The experimental g_{SS}(r) also yields the ratio of 2, and this is a characteristic of a simple Lennard-Jones fluid,[51] where no network-like correlation is present. The coordination number, n_c, of the first solvation shell was calculated from the calculated g_{SS}(r) by PIPF using the equation below

$$n_c = 4\pi \int_0^{r_{\min}} \rho g(r) r^2 \, dr \tag{2.19}$$

where $r_{\min}$ is the first minimum position of $g_{SS}(r)$, and ρ is the density of H_2S. The PIPF method yields 11.7 for the coordination number, similar with other models (11.1 – 12.1). The original coordination number by Jorgensen is 11.2 from the OPLS method using the Monte Carlo simulation,[14] whereas Ref. 23 reports 9.0 for the OPLS model from their MD simulation. (Table 2-7) This high coordination number is also a representative of a simple Lennard-Jones fluid.[14]

Table 2-7. Positions of the first peak and the second peak of the sulfur-sulfur RDF, their ratios, and the coordination number (n_c) of the first solvation shell at 212 K and 1 atm[a]

	First peak	Second peak	Ratio	n_c
PIPF	4.03	7.68	1.91	11.7
OPLS	3.75	7.35	1.96	9.0
Forster	3.95	7.55	1.91	11.9
KL	4.05	7.65	1.89	12.0
Potoff	3.95	7.65	1.94	11.1
Drude	4.05	7.65	1.89	12.1
AIMD	3.95	7.75	1.96	12.0

[a]Values are taken from Ref. 23.

The sulfur-hydrogen, $g_{SH}(r)$, and hydrogen-hydrogen, $g_{HH}(r)$, RDFs are also displayed in Figure 2-1(b) and (c), respectively. The PIPF produces broad first peaks at both thermodynamic states, and their positions are located at 4.95 Å (212.81 K and 1 atom) and 4.90 Å (298 K and 30.6 atm), which are comparable with the experimental peak position at 4.75 Å. In addition, the peak heights are in good agreement with experiment. Although the experimental $g_{SH}(r)$ has two distinct peaks at 3.35 Å and 4.75 Å, the authors insist that it be due to the instrumental instabilities. For the second peak of $g_{SH}(r)$, however, the PIPF predicts more distinct peaks at shorter positions, 7.83 Å (212.81 K and 1 atm) and 8.00 Å (298 K and 30.6 atm), respectively, than the experiment at 8.10 Å. The computed $g_{HH}(r)$ also produces first peaks (4.05 Å at 212.81 K and 1 atm, and 4.15 Å at 298 K and 30.6 atm) at the similar positions with the experiment (4.13 Å). The peak heights by PIPF are more intense than the experiment at both thermodynamic states.

Overall, experimental RDFs show that the liquid H_2S lacks hydrogen-bonding networks, the authors from the experiments insist that this view the RDFs of H_2S resemble those of a simple

Lennard-Jones system such Ar[49,51] Although the PIPF method produces similar peak positions with the experiment, the shapes by PIPF is a little more structured than experiment as shown in Figure 2-1. However, the RDFs by PIPF also produce some characteristics of a simple fluid in $g_{SS}(r)$ as stated above.

2.5. Conclusions

The Thole's interaction dipole (TID) based polarizable intermolecular potential function (PIPF) has been developed for H_2S molecule. This study is the first step for parameterization of the PIPF method for sulfur-containing molecules, and finally sulfur-containing amino acid residues for biomolecular simulations. The PIPF method provides promising results for dipole moments both in the gas-phase and liquid phase with the parameters optimized in the gas phase. In addition, the fundamental liquid phase properties such as the density, the heat of vaporization, and the diffusion coefficient computed by the PIPF method are also in excellent agreement with experimental values compared to other existing non-polarizable and polarizable H_2S models. The liquid structure probed by radial distribution functions (RDFs) shows a similar feature with other H_2S models and experiment as simple liquid that lacks hydrogen-bonding network, but more structured shapes represent that the liquid H_2S by the PIPF method is a more polar system than other models and the real system. This observation shows that the PIPF method needs to be improved more for third-row atoms such as sulfur and phosphorous.

Chapter 3

Optimization of the Explicit Polarization (X-Pol) Potential Using a Hybrid Density Functional[a]

3.1. Introduction

The basic approach used in the most popular current parametrizations of molecular mechanics (MM) was established in the 1960s by Lifson, and this approach continues to play an essential role in providing force fields for dynamical simulations of macromolecular systems such as proteins and nucleic acids[1] as well as other nanomaterials. Despite its success, which was promoted by careful and laborious parameterization by many research groups over the past half century, there are also a number of well-known shortcomings, including redundancy of energy terms and parameters, the widespread use of harmonic approximations for bond stretching and angle bending, and the difficulty of treating electronic polarization and charge transfer (for a recent special issue on polarizable force fields, see Jorgensen[2]). Furthermore, molecular mechanics is not designed to treat chemical reactions and photochemical processes.[3] With continuing advances in computer architecture, it is natural to ask what type of force fields will be used for biomolecular and materials simulations in the future. To this end, we have introduced the explicit polarization (X-Pol) potential,[4-8] which is an electronic structure method based on block localization of molecular orbitals.[4,5,9] The X-Pol method differs from the effective fragment potential (EFP)[10,11] and SIBFA (sum of interactions between fragments computed ab initio)[12] potentials in that the latter models are derived by fitting results to ab initio results in terms of a

[a] Springer and *Theoretical Chemistry Accounts*, 131, 2012, 1161, Optimization of the Explicit Polarization (X-Pol) Potential Using a Hybrid Density Functional, J. Han, D. G. Truhlar, and J. Gao, original copyright is given to the publication in which the material was originally published; reproduced with kind permission from Springer Science and Business Media.

multipole expansion of the electrostatics along with other energy terms. In the X-Pol method, a macromolecular system is partitioned into constituent blocks, also called fragments, each of which can be, for example, an individual solvent or solute molecule, an enzyme cofactor, a ligand or molecular fragment, or a peptide unit of a protein. The internal energies of the fragments are determined by an explicitly quantum mechanical method, and interfragment interactions are approximated in a way akin to a combined quantum mechanical and molecular mechanical (QM/MM)[3,13-15] method. However, the electrostatic field in which each individual fragment is embedded is obtained from the corresponding instantaneous wave functions of all other fragments in the system, and the mutual electronic polarization among fragments is included self-consistently.[4-8]

X-Pol can also be used as an electronic structure method such that any quantum chemical model, e.g., Hartree–Fock (HF) theory (or semiempirical models of HF), second-order Møller-Plesset perturbation theory (MP2), coupled cluster theory, or density functional theory (DFT), may be adopted to represent the individual fragment blocks. In this regard, one can treat all fragments by using the same method, or by mixing different electronic structure methods for different fragments (for example, MP2 for one fragment and DFT for all other fragments). Because a large system is partitioned into fragments, the X-Pol method can be made to scale well for fast calculations, and therefore, it can be used to establish a framework for the development of a next-generation force field[4] that goes beyond the conventional molecular mechanics by explicitly including a quantum mechanical treatment of electronic polarization and possibly charge transfer effects (which can be included, for example, by a recently proposed method[16] involving ensemble DFT). When X-Pol is used as a force field, we introduce a set of empirical terms to account for the missing exchange repulsion[17] and dispersion-like attractive, noncovalent interactions. Because these terms are empirical, they can increase the accuracy and, at the same time, reduce computational costs by using parameterization to compensate for errors introduced

by using a low or modest level of electronic structure theory.[8] In the present study, we illustrate this by showing how we can use a modestly accurate density functional with a small basis set to treat the individual fragments in the X-Pol method. In particular, we employ the hybrid B3LYP model and a fairly small 6-31G(d) basis set, and we show that X-Pol with this choice can be parameterized to model hydrogen bonding interactions in good agreement with the results from full CCSD(T) calculations. Here, we emphasize that our goal is not to reproduce the geometries and energies at the B3LYP/6-31G(d) level that is used to represent the X-Pol fragment, but rather to obtain agreement with the higher-level CCSD(T) results by optimization of the parameters introduced in the X-Pol quantum force field.[5,8]

There are many other fragment-based molecular orbital methods.[18] For example, Zhang et al.[19,20] developed a molecular fractionation with conjugated caps (MFCC) approach to treat proteins and protein–ligand interactions. In this method, the individual fragments are capped with a structure representative of the local structure of the original system, and the total energy is obtained by subtracting the energies of the common fragments used in the "caps". The method provides a good means to evaluate interfragment interactions and a straightforward procedure to incorporate the local electronic structure into a fragment-based molecular orbital approach.[21,22] Another way of separating the total energy into fragmental contributions is the general interaction energy expansion approach described by Stoll and Preuss.[23] The key to achieve fast convergence in this method, in contrast to early schemes,[24] is to optimize the monomer, dimer and many-body fragmental molecular orbitals in the presence of all other fragments, rather than using isolated gas-phase fragment terms. There are a number of applications of this strategy, including the fragment molecular orbital (FMO) method[25,26] and the electrostatically embedded many-body (EE-MB) expansion method.[27-30] The SCF procedure used in the FMO model is identical to that developed in the X-Pol method,[25,26] whereas two-body and three-body exchange and charge transfer effects are included in the FMO2 and FMO3 implementations.[23]

In Section 3.2, we briefly review the theoretical background of the X-Pol potential, and in Section 3.3, we present the computational details. In Section 3.4, we describe the optimization of parameters and compare the computed hydrogen bonding energies and geometries obtained from the X-Pol method with higher-level results. Finally, Section 3.5 summarizes the main findings from this work and presents concluding remarks.

3.2. Theoretical Background

The X-Pol method has been described in detail elsewhere.[4,7,8] For completeness, we briefly describe the key aspects and approximations made in the X-Pol potential and the empirical parameters introduced to correct for these approximations. We note that the X-Pol method was developed based on block localization of the molecular wave function of the system, which includes a hierarchy of approximations.[8] There are numerous other methods based on localized molecular orbitals or molecular fragments. A recent review, which appeared online after the submission of this manuscript, contains an account of these methodologies.[18]

First, we partition a macromolecular system into structural blocks, also called fragments. The molecular wave function, $\mathbf{\Phi}$, is approximated as a Hartree product of antisymmetric wave functions of the individual fragments, $\{\mathbf{\Psi}^A\}$:

$$\mathbf{\Phi} = \prod_{A=1}^{N} \mathbf{\Psi}^A \tag{3.1}$$

where N is the number of fragments in the system, and $\mathbf{\Psi}^A$ is a Slater determinant of occupied molecular orbitals (MOs) that are constructed using an atomic orbital basis located on the atoms of fragment A. Thus, these MOs are block-localized by construction. In the present work, density functional theory is used to represent the molecular fragments, and the block-localized molecular orbitals (BLMO) are block-localized Kohn–Sham (BLKS) orbitals, in terms of which the electron density $\rho^A(\mathbf{r})$ of fragment A is given by

$$\rho^A(r) = 2\sum_i \left|\phi_i^A\right|^2 \tag{3.2}$$

where ϕ_i^A is the *i*th doubly occupied Kohn–Sham orbital of fragment *A*. In the present work, the molecular fragments are closed-shell molecules.

The X-Pol total energy of the system can be written as follows: [4,8]

$$E_{\text{tot}} = \sum_{A=1}^{N}\left(E^A + \frac{1}{2}E_{\text{int}}^A\right) + E_{\text{XD}} \tag{3.3}$$

where E^A is the energy of fragment *A* with the wave function Ψ^A, which can be calculated at any given theoretical level, including Hartree–Fock (HF), density functional theory (DFT), or post-HF theories such as Møller-Plesset perturbation theory or coupled cluster theory, E_{int}^A is the Coulomb interaction energy between fragment *A* and other fragments, and E_{XD} accounts for the exchange-repulsion (X) and dispersion-correlation (D) interactions between the fragments. It should be pointed out that the wave function Ψ^A in Eq. 3.3 corresponds to that of fragment *A* polarized by the remaining fragments in the system, and it differs from the wave function of an isolated fragment in the gas phase (Ψ_0^A). The energy difference between the two states, Ψ^A and Ψ_0^A, is the energy penalty paid for distorting the fragmental wave function due to many-body polarization.[13]

The use of the Hartree-product wave function in Eq. 1 implies that the short-range exchange repulsion and long-range and medium-range dispersion and dispersion-like interactions (for brevity, will just call these dispersion in the rest of the article) as well as charge transfer among fragments are neglected.[4] The exchange repulsion and dispersion energies can be determined in various ways, for example by antisymmetrizing the block-localized (i.e., fragmental) orbitals in Eq. 1[9,17,31,32] or by perturbation methods such as symmetry-adapted perturbation theory (SAPT).[33-35] However, these methods are not suitable for the construction of a fast quantum mechanical

force field for large systems due to their high computational cost as compared to the method adopted here, which is discussed next.

Because the exchange repulsion is short-ranged and approximately pairwise additive[17] and the dispersion interactions can also be adequately modeled by pairwise potentials,[4] such as those used in dispersion-corrected density functional theory (DFT-D),[36] we have used the Lennard-Jones potential to parametrically model the exchange-repulsion and dispersion interactions between each pair of fragments, *A* and *B*:

$$E_{\mathrm{ErD}}^{AB} = 4\sum_{a}^{A}\sum_{b}^{B}\varepsilon_{ab}^{AB}\left[\left(\frac{\sigma_{ab}^{AB}}{R_{ab}}\right)^{12} - \left(\frac{\sigma_{ab}^{AB}}{R_{ab}}\right)^{6}\right] \tag{3.4}$$

where the *A* over the sum means that the sum is restricted to orbitals *a* on center *A*, and where ε_{ab}^{AB} and σ_{ab}^{AB} are parameters. These parameters are determined from atomic parameters by using standard combining rules: $\varepsilon_{ab}^{AB} = \sqrt{\varepsilon_a^A \varepsilon_b^B}$ and $\sigma_{ab}^{AB} = \sqrt{\sigma_a^A \sigma_b^B}$. The values of ε_a^A and σ_a^A depend on the atomic number of the atom and sometimes also on its hybridization. These parameters can be optimized for a particular electronic structure method used in the X-Pol potential, and the main objective of the present study is to illustrate the optimization of these parameters and the performance of the X-Pol potential with the B3LYP hybrid density functional and the modest 6-31G(d) basis set for calculating binding energies of bimolecular complexes. We will judge the accuracy by comparing to the results of higher-level CCSD(T) calculations.

For closed-shell fragments, the Kohn–Sham DFT energy of fragment *A* in the presence of the rest of the system is

$$E^A\left[\rho^A(\mathbf{r})\right] = \sum_i 2H_i^A + \sum_{i,j} 2J_{ij}^A + E_{xc}^A\left[\rho^A(\mathbf{r})\right] + E_{\mathrm{nuc}}^A \tag{3.5}$$

where the superscript *A* labels the energies and densities for monomer fragment *A*, the indices *i* and *j* run through the doubly occupied, BLKS molecular orbitals of fragment *A*, H_i^A

and J_{ij}^{A} are respectively the one-electron Hamiltonian integrals and the Coulomb integrals, $E_{xc}^{A}\left[\rho^{A}(\mathbf{r})\right]$ is the exchange–correlation functional, and E_{nuc}^{A} is the nuclear repulsion energy.

The Coulomb interaction energy, E_{int}^{A}, in Eq. 3 is given by:

$$E_{\text{int}}^{A} = -2\sum_{i}\left\langle \psi_{i}^{A}\middle|V^{A}(\mathbf{r})\middle|\psi_{i}^{A}\right\rangle + \sum_{a}^{A} Z_{a}^{A} V^{A}\left(\mathbf{R}_{a}^{A}\right) \tag{3.6}$$

where Z_{a}^{A} is the nuclear charge of atom *a* of fragment *A*, and $V^{A}(\mathbf{r})$ is the total external electrostatic potential (ESP) due to all other fragments in the system. The external ESP is defined as follows:

$$V^{A}(r) = -\sum_{B\neq A}\left(\int\frac{\rho^{B}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dr' + \sum_{b}^{B}\frac{Z_{b}^{B}}{\left|\mathbf{r}-\mathbf{R}_{b}^{B}\right|}\right) \tag{3.7}$$

where Z_{b}^{B} is the nuclear charge of atom *b* of fragment *B* located at $\mathbf{R}_{b}^{B}$. The potential $V^{A}(\mathbf{r})$ could be determined analytically and used to compute the two-electron integrals in Eq. 6 (the terms in the first summation), but this is time-consuming and not a useful choice for fast calculations on macromolecular systems. Alternatively, the ESP in Eq. 7 can be treated by a distributed multipole expansion, and the simplest approximation is to retain only the distributed monopole terms, possibly with scaling to make up for this approximation. This yields the following expression.

$$V^{A}(\mathbf{r}) = \lambda\sum_{B\neq A}\sum_{b}^{B}\frac{q_{b}^{B}}{\left|\mathbf{r}-\mathbf{R}_{b}^{B}\right|} \tag{3.8}$$

where λ is a scaling parameter, and q_{b}^{B} is the partial atomic charge on atom *b* of fragment *B*. The partial atomic charges can be determined in various ways, for example by fitting electrostatic potentials or by using a charge population analysis method.[5] We make the latter choice[8] for the

present calculations, using Mulliken population charges from the BLKS orbitals, and the single parameter λ is set to unity.

In calculating the total energy of the system by Eq. 3.3, we use the double self-consistent-field (DSCF) method.[6,7] With an initial guess of the one-electron density matrix for each fragment, the electronic structure calculations for each fragment are performed in the presence of the Mulliken charges of all the other fragments until the change in the total electronic energy or density matrix reaches a predefined tolerance. Although the X-Pol theory has been formulated variationally to allow efficient calculations of energy gradients,[7] here we use the older, nonvariational sequential optimization energy formulation.

3.3. Computational Details

For the bimolecular complexes, in both the "high-level" reference calculations and the X-Pol calculations, partial geometry optimizations were performed in which the monomer geometries are held fixed at the corresponding level of theory. Thus, in each bimolecular complex, the hydrogen bond distance and angle between the donor and acceptor molecules, as illustrated in Figures 3-1, 3-2, 3-3, 3-4 and 3-5, are optimized. (Only one angle is involved because we adopt a high symmetry for each hydrogen bond, as illustrated). In all cases, the monomer geometries that were optimized at the corresponding level of theory were held fixed.

The reference geometries were calculated by full (i.e., nonfragmental) B3LYP[37-39] calculations with the aug-cc-pVDZ basis set. The reference energies were obtained by full CCSD(T) single-point calculations with the aug-cc-pVDZ basis set at the geometries of the complexes optimized using B3LYP/aug-cc-pVDZ.

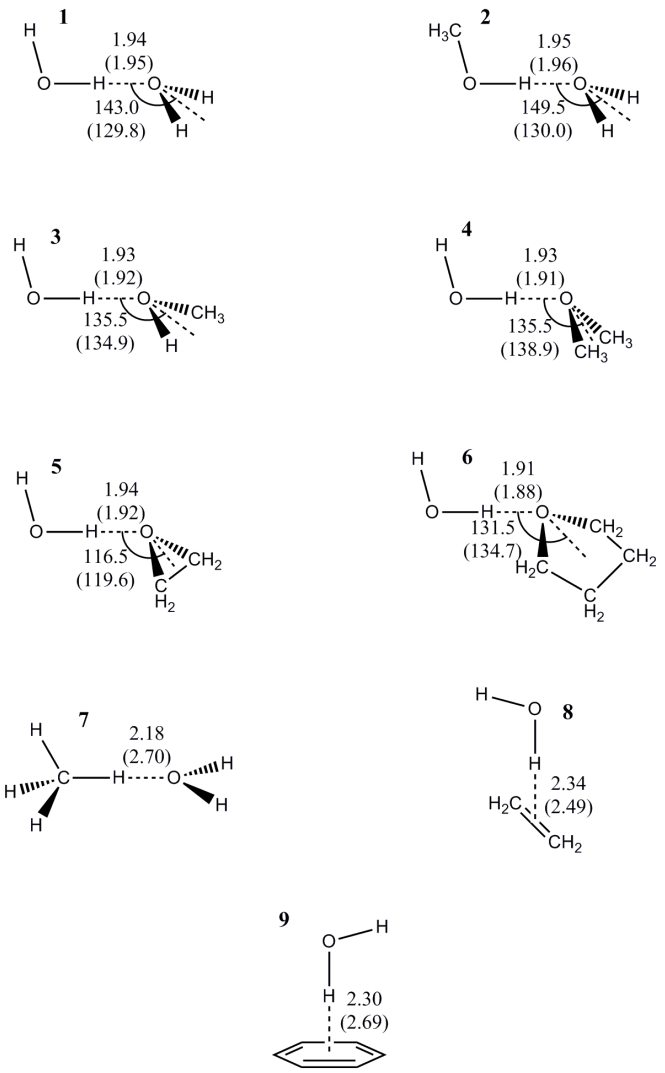


Figure 3-1. Schematic illustration of the bimolecular complexes between water and small molecules. Optimized hydrogen bond distances and angles from the X-Pol potential, XP@B3LYP/6-31G(d), are first given, followed by values in parentheses by B3LYP/aug-cc-pVDZ. Distances are given in angstroms and angles in degrees.

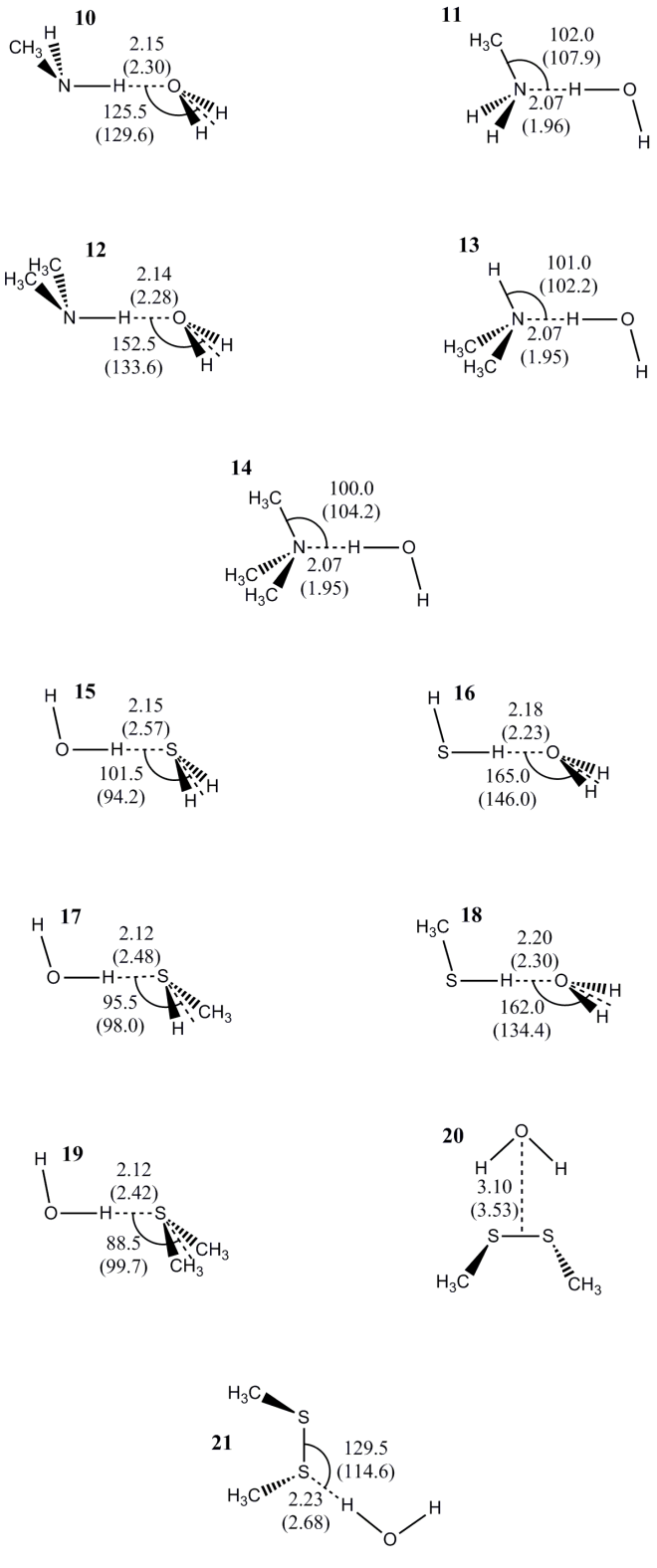


Figure 3-1. *Continued.*

The X-Pol calculations were carried out using the B3LYP/6-31G(d) method as the quantum mechanical level with the geometries optimized by the same level of X-Pol calculation. The hybrid B3LYP functional was chosen in the present study because it is a popular model that has been used widely; one can certainly select a more accurate and recent functional, but the goal here is not to compare the quality and performance of different functionals. We sometimes use the notation XP@B3LYP/6-31G(d) to specify such an X-Pol calculation.

The binding energy for a bimolecular complex, including the empirical Lennard-Jones terms to account for the exchange repulsion and dispersion contributions, is calculated by:

$$\Delta E_b(A\cdots B) = E(A) + E(B) - E_{\text{tot}}^{\text{X-Pol}}(A\cdots B) \tag{3.9}$$

where $E_{\text{tot}}^{\text{X-Pol}}(A\cdots B)$ is the X-Pol energy (Eq. 3.3) of the bimolecular complex in which each monomer, *A* or *B*, is treated as an individual fragment, and $E(A)$ and $E(B)$ are the B3LYP/6-31G(d) energies of the optimized monomer structures. All binding energies in the present article are zero-point-exclusive.

The Lennard-Jones parameters in Eq. 3.4 have been adjusted so that the XP@B3LYP/6-31G(d)-binding energies best reproduce the results calculated using CCSD(T)/aug-cc-pVDZ// B3LYP/aug-cc-pVDZ. The full quantum mechanical calculations for all systems were performed using *Gaussian03*,[40] whereas all X-Pol calculations were carried out using a local program that is coupled to a modified version of the GAMESS package.[41]

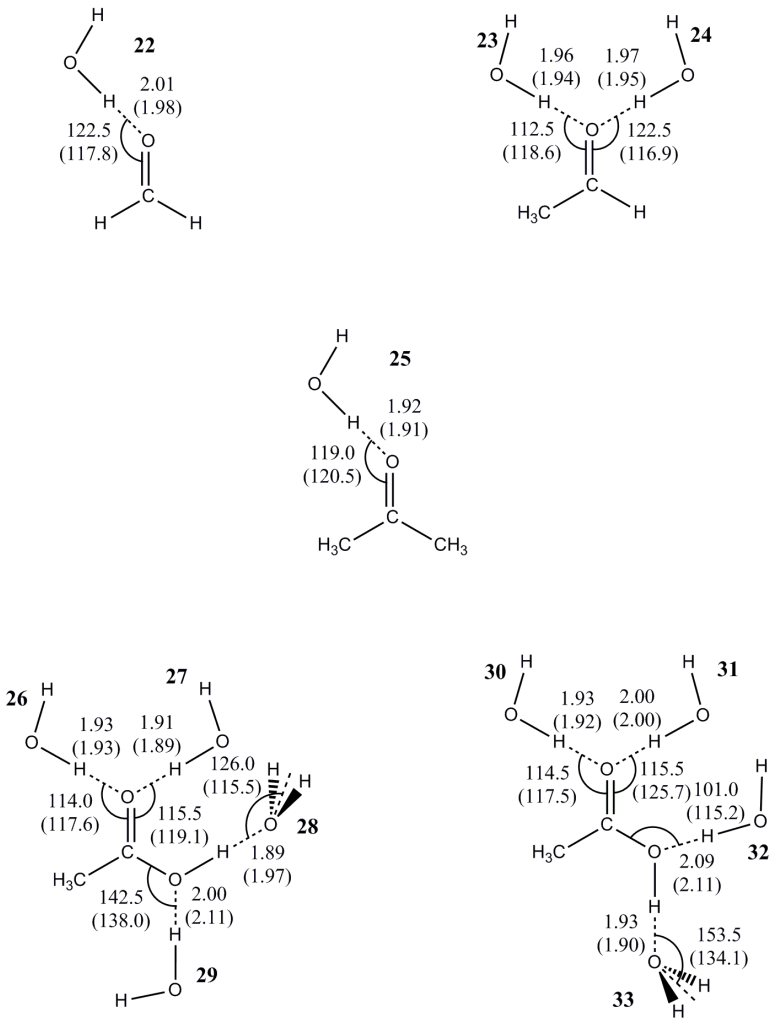


Figure 3-2. Bimolecular complexes depicting the interactions between water and a series of carbonyl-containing compounds. Optimized hydrogen bond distances and angles from the X-Pol potential, XP@B3LYP/6-31G(d), are given first, followed by values in parentheses by B3LYP/aug-cc-pVDZ. Distances are given in angstroms and angles in degrees.

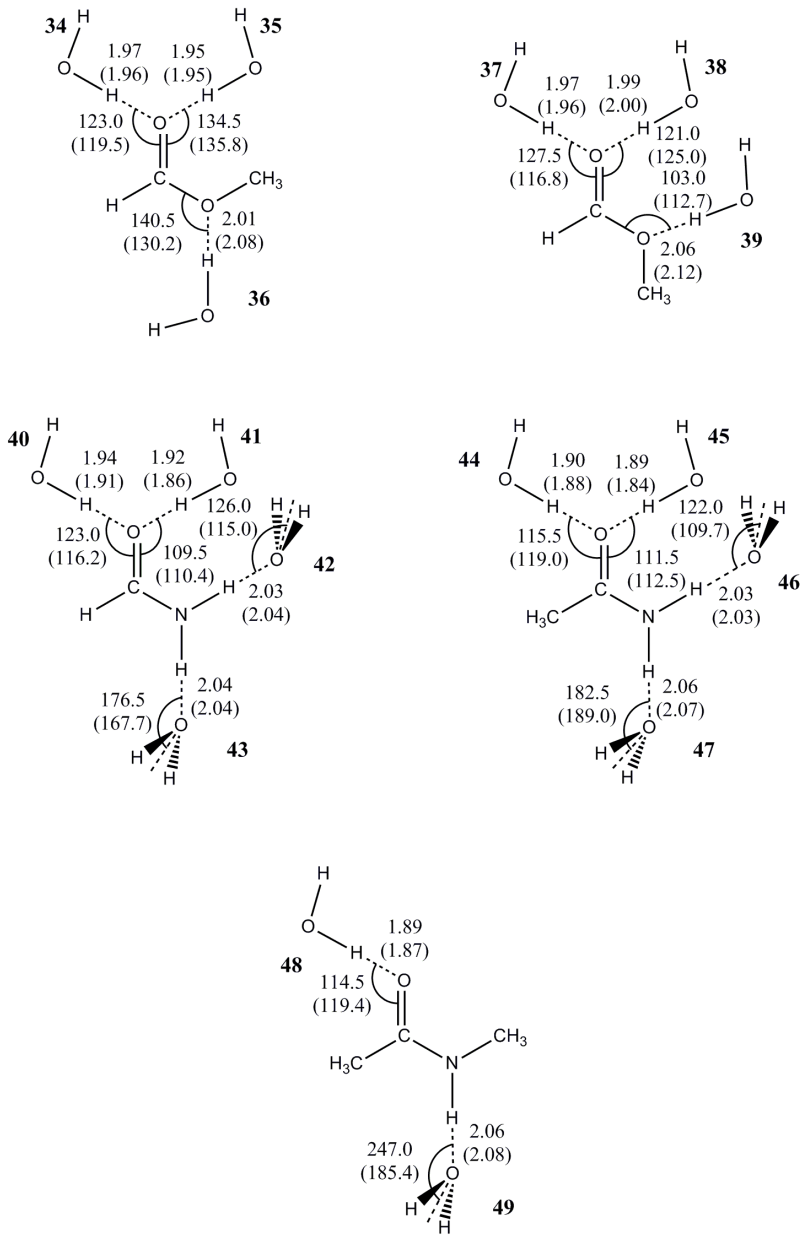


Figure 3-2. *Continued.*

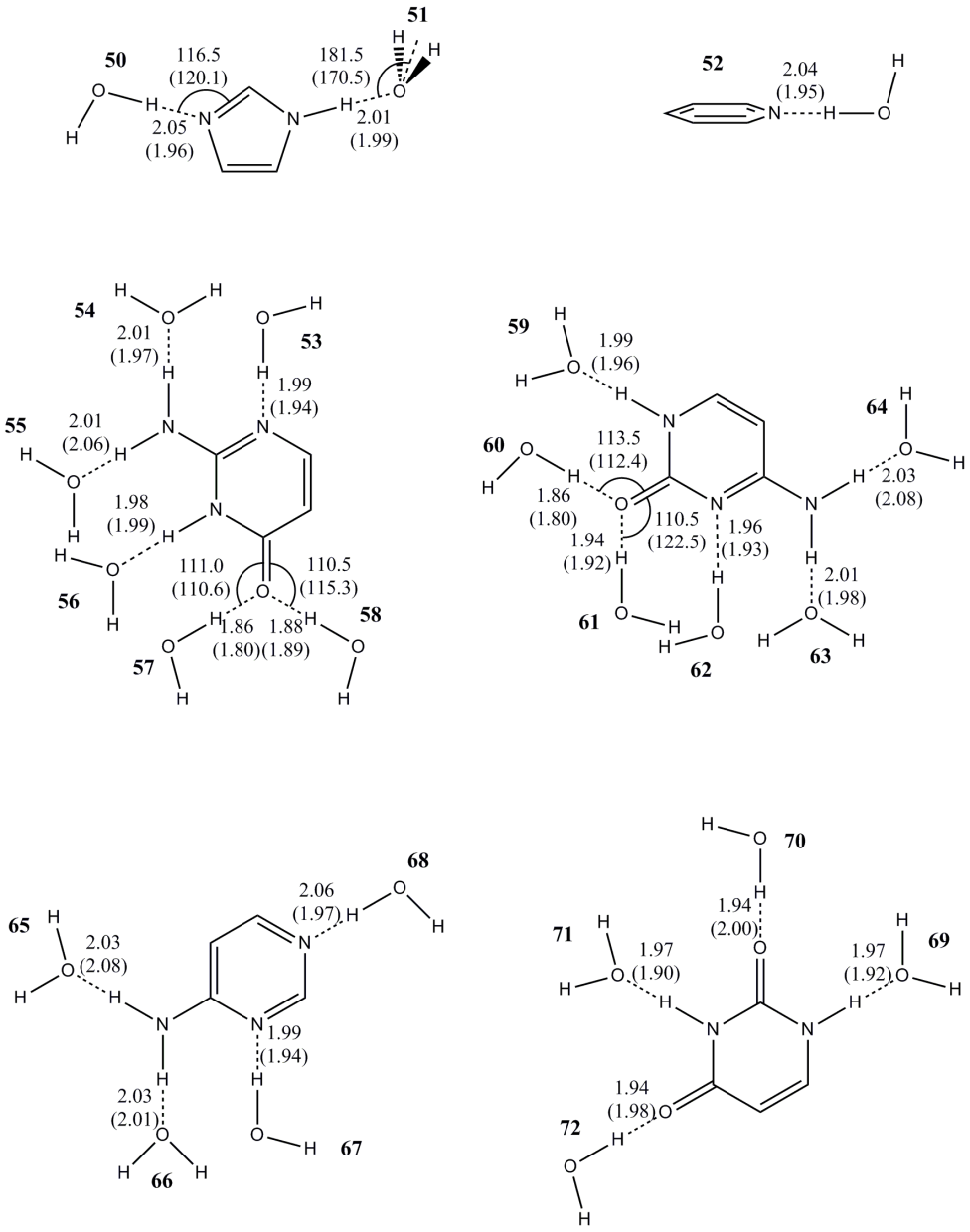


Figure 3-3. Bimolecular complexes depicting the interactions between water and heterocyclic compounds. Optimized hydrogen bond distances and angles from the X-Pol potential, XP@B3LYP/6-31G(d), are given first, followed by values in parentheses by B3LYP/aug-cc-pVDZ. Distances are given in angstroms and angles in degree.

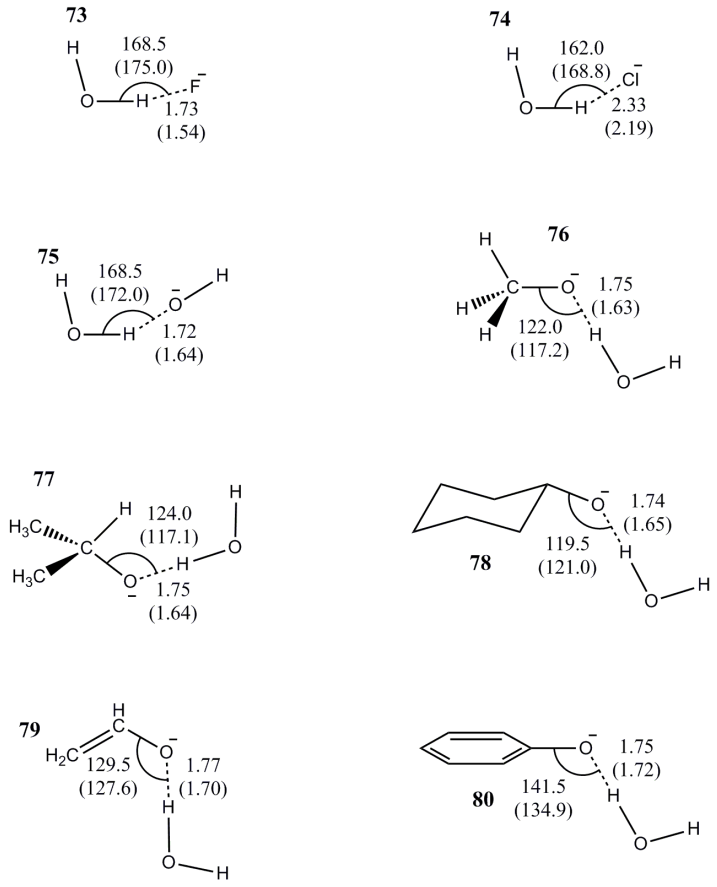


Figure 3-4. Bimolecular complexes depicting the interactions between water and anionic species. Optimized hydrogen bond distances and angles from the X-Pol potential, XP@B3LYP/6-31G(d), are given first, followed by values in parentheses by B3LYP/aug-cc-pVDZ. Distances are given in angstroms and angles in degrees.

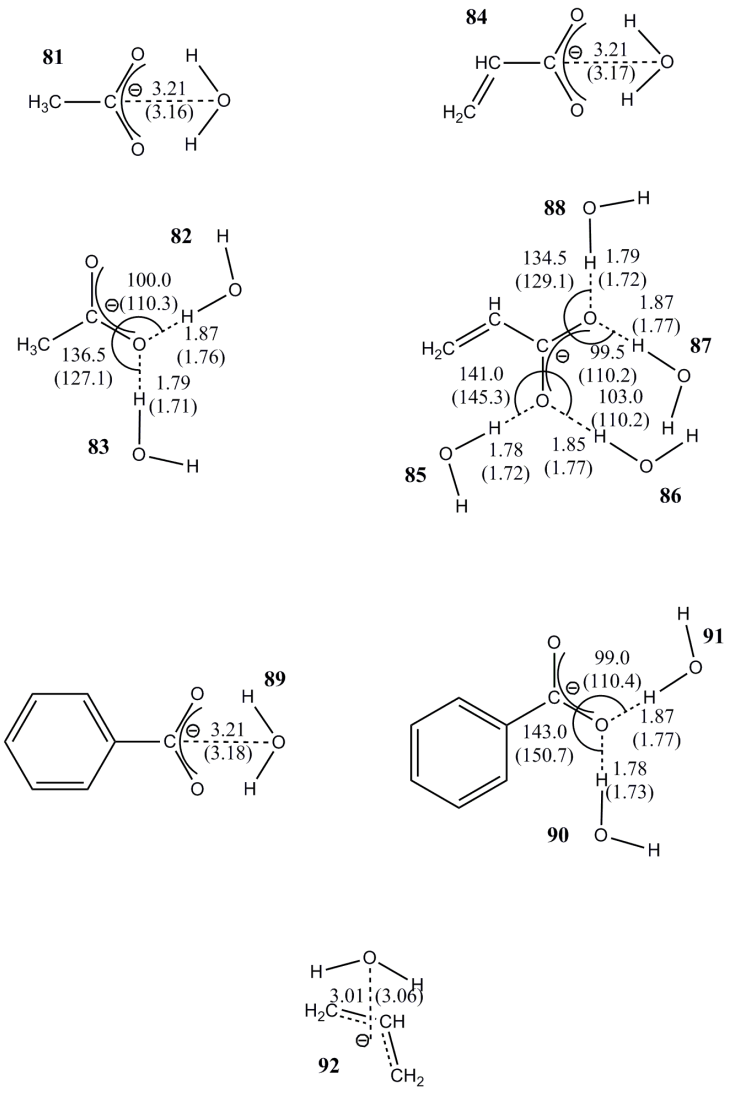


Figure 3-4. *Continued.*

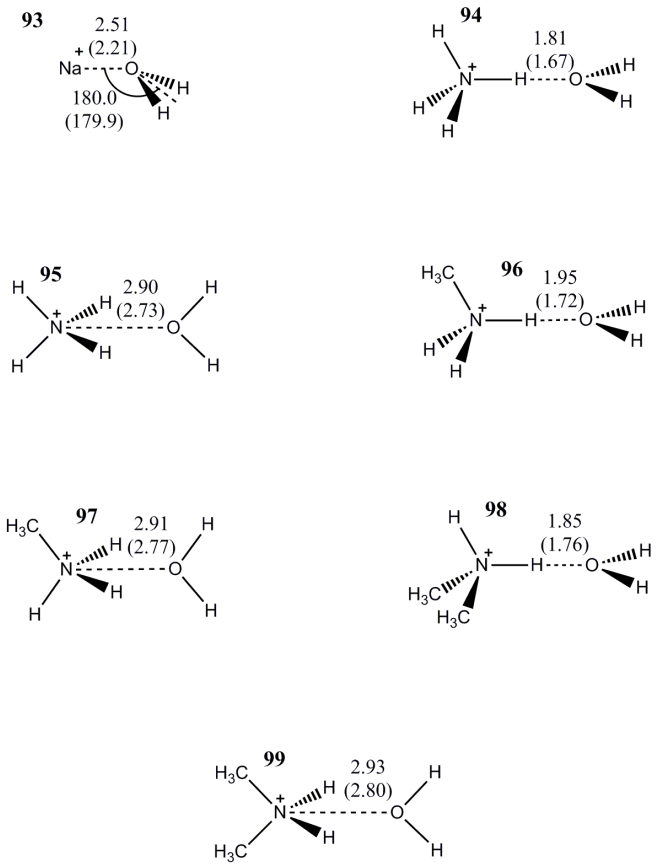


Figure 3-5. Bimolecular complexes depicting the interactions between water and cationic species. Optimized hydrogen bond distances and angles from the X-Pol potential, XP@B3LYP/6-31G(d), are given first, followed by values in parentheses by B3LYP/aug-cc-pVDZ. Distances are given in angstroms and angles in degrees.

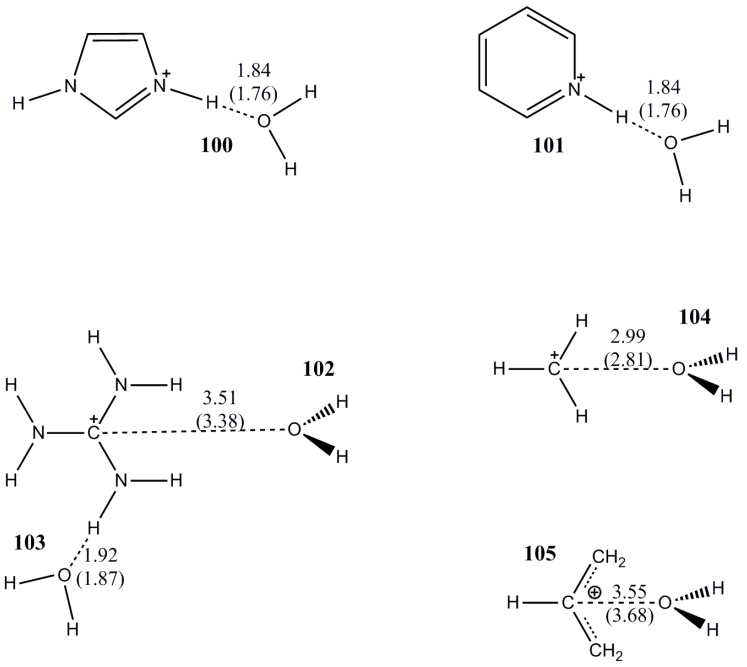


Figure 3-5. *Continued*.

3.4. Results and Discussion

3.4.1 Optimization of the Repulsion and Dispersion Interactions between Fragments

We considered a total of 105 bimolecular complexes, each of which involves one water molecule and an organic or inorganic compound or ion; the organic compounds include ionic and neutral functional groups found in amino acids and nucleobases. Although experimental results for a number of hydrogen bonding complexes are available, we wish to examine a much larger dataset, and therefore, we used theoretical results as reference data, as explained in Section 3. The reference data for these complexes were used to optimize the Lennard-Jones parameters by an iterative procedure for the case where X-Pol fragments are treated by B3LYP/6-31G(d). In this process, we placed greater emphasis on the performance for binding energies than on hydrogen bond distances and angles. The resulting parameters for H, C, O, N, and S atoms and for F^-, Cl^-, and Na^+ ions are listed in Table 3-1.

Table 3-1. Optimized Lennard-Jones Parameters in the X-Pol Potential with B3LYP/6-31G(d)

Atom	σ (Å)	ε (kcal/mol)
H	1.31	0.04
H (–SH)	1.81	0.04
C	3.67	0.16
N (neutral)	3.60	0.20
N (cation)	3.47	0.20
O (both sp^2 and sp^3)	3.25	0.15
O^- (RO^-)	3.21	0.15
O^- (RCO_2^-)	3.24	0.15
S	3.11	0.56
Na^+	2.51	0.30
Cl^-	4.37	0.21
F^-	2.97	0.45

Previously, we examined a small set of 14 bimolecular complexes and reported a set of parameters for several atoms,[8] and the values listed in Table 1 for these atoms are very similar to those obtained in that work. In the present work, we introduced a new atom type for hydrogen attached to a sulfur atom (thiols and H_2S), and this atom type has a greater σ value than that used in other situations. Three atom types are assigned to oxygen, corresponding to an oxygen type in neutral functional groups and two types for anionic species. Previously, different Lennard-Jones parameters were used for sp^2 and sp^3 oxygen atoms,[8] but a single oxygen type for both hybridizations is adequate here. The Lennard-Jones parameters for the carboxylate oxygen and neutral oxygen atoms are very similar; although it would be possible to use the same oxygen parameters in both cases, we kept the two atom types to increase flexibility. For nitrogen atoms,

we distinguish atom types for neutral and protonated cases. For other elements, including carbon, a single set of parameters for each is sufficient for the present data set.

3.4.2 Energies and Geometries of Hydrogen Bonded Complexes

Figures 3-1, 3-2, 3-3, 3-4 and 3-5 depict the structural arrangements used in the present calculations, along with the optimized geometrical parameters from both the reference calculations and the X-Pol calculations. In many cases, more than one structure is considered for a given chemical species, corresponding to placing water molecules at different positions or in different orientations. Each structure is assigned a number for discussion purposes, and the computed binding energies are given in Tables 3-2, 3-3, 3-4, 3-5 and 3-6. The figures and tables are organized roughly according to functional groups.

Both the B3LYP/aug-cc-pVDZ and CCSD(T)/aug-cc-pVDZ-binding energies are given in the tables for comparison; however, the B3LYP-binding energies are inaccurate due to a poor treatment of dispersion contributions, and only the CCSD(T) values should be considered as reference values.

3.4.2.1 Small Molecules and Simple Functional Groups

Figure 3-1 and Table 3-2 give the results for water complexes with small molecules. Six sp^3 oxygen structures are included in our study, including water, two structures of methanol, and three structures of ethers (structures **1**–**6**).

The interaction energy for a water dimer (**1**) is calculated to be 5.7 kcal/mol by the X-Pol method, which yields a hydrogen bond length and bond angle of 1.93 Å and 135.5^o (Fig. 3-1). These may be compared with the corresponding reference values of 5.2 kcal/mol, 1.91 Å, and 138.9°. The best estimate of the water dimer interaction energy is 5.0 kcal/mol using CCSD(T) with extrapolation to a complete basis set.[42,43]

For the methanol–water complexes (**2** and **3**), both XP@B3LYP/6-31G(d) and CCSD(T) calculations predict that methanol is a better hydrogen bond acceptor (structure **3**) by 0.4 –

0.5 kcal/mol. As the number of alkyl groups on the oxygen increases in going from water to alcohol to ether, the calculated hydrogen bond strength is also enhanced, due to the electron donating effect of an alkyl group, to a final value of about 7 kcal/mol for the complex with tetrahydrofuran (**6**). The average unsigned errors in hydrogen bond lengths and angles are, respectively, 0.02 Å and 7° for the sp^3 oxygen-containing compounds.

Table 3-2. Binding Energies of Bimolecular Complexes between Water and Simple Functional Groups Containing Oxygen, Nitrogen, and Sulfur Atoms Computed Using the XP@B3LYP/6-31G(d), B3LYP/aug-cc-pVDZ, and CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ Methods[a]

Complex	XP@B3LYP/6-31G(d)	B3LYP/aug-cc-pVDZ	CCSD(T)/aug-cc-pVDZ
1	5.7	4.6	5.2
2	5.5	4.4	5.4
3	5.9	5.0	5.9
4	6.1	5.0	6.3
5	6.0	4.9	6.0
6	7.0	5.6	7.2
7	1.3	0.4	0.9
8	2.4	1.9	3.0
9	4.2	1.4	3.9
10	2.6	1.9	2.8
11	5.9	6.6	7.4
12	3.0	2.0	2.8
13	6.0	6.5	7.8
14	6.2	6.5	8.4

Continued on next page

Table 3-2. Binding Energies of Bimolecular Complexes between Water and Simple Functional Groups Containing Oxygen, Nitrogen, and Sulfur Atoms Computed Using the XP@B3LYP/6-31G(d), B3LYP/aug-cc-pVDZ, and CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ Methods[a] – *Continued*

Complex	XP@B3LYP/6-31G(d)	B3LYP/aug-cc-pVDZ	CCSD(T)/aug-cc-pVDZ
15	3.7	2.6	3.3
16	3.1	2.3	3.1
17	4.4	3.3	4.2
18	2.5	1.8	2.8
19	5.0	4.0	5.2
20	3.3	1.8	3.3
21	2.6	1.6	2.7

[a]Energies are given in kilocalories per mole.

The binding energies of methane (**7**) and benzene (**9**) with a single water molecule from the X-Pol optimizations are 0.3–0.4 kcal/mol greater than the CCSD(T) results, but the binding energy between ethane and water is 0.6 kcal/mol smaller (**8**).

We examined five complexes involving simple methyl amines (**10**–**14**). The primary and secondary amines are much better hydrogen bond acceptors than donors,[44] both from the XP@B3LYP/6-31G(d) and CCSD(T)/aug-cc-pVDZ calculations. Although the X-Pol-binding energies for the donor complexes (**10** and **12**) are in good agreement with the reference data, the binding energies for the acceptor complexes are underestimated by 1.5–2.2 kcal/mol, and the deviation increases as the basicity of the amines increases with more methyl substitutions.

Figure 3-1 includes four sulfur compounds: hydrogen disulfide, methanethiol, dimethyl sulfide, and dimethyl disulfide (**15**–**21**). Similar to alkyl amine complexes, the binding energies

for sulfur compounds are stronger when the sulfur atom acts as a hydrogen bond acceptor than a donor for H_2S and thiols; however, the difference is smaller than in the corresponding nitrogen compounds. These trends are correctly reproduced in the XP@B3LYP/6-31G(d) model in comparison with the reference data (Table 3-2). In fact, the X-Pol method performs very well, having an average unsigned error of less than 0.2 kcal/mol in binding energy. A somewhat less satisfactory finding is that an additional hydrogen atom type for H_2S and thiols needs to be introduced, whereas the hydrogen bond distance for the acceptor complexes is significantly shorter in the X-Pol calculations than the values optimized using B3LYP/aug-cc-pVDZ.

3.4.2.2 Carbonyl-containing Compounds

Figure 3-2 and Table 3-3 list results for carbonyl compounds, including aldehydes, ketones, carboxylic acids, esters, and amides. The X-Pol results for the aldehydes and acetone (**22**–**25**) are in reasonable agreement with the reference data, resulting in an average unsigned deviation in binding energy of 0.4 kcal/mol. It is especially encouraging that the X-Pol model correctly distinguishes the relative interaction energies between the two complexes of acetaldehyde with water, favoring the configuration with water oriented toward the methyl group. The optimized hydrogen bond geometries using XP@B3LYP/6-31G(d) are also in good agreement with the reference data.

For carboxylic acid systems, both the syn and anti conformations are considered (**26**–**33**). We note that the carboxylic acids are particularly good hydrogen bond donors, with computed binding energies of 7.6 and 8.4 kcal/mol for the syn (**28**) and anti (**33**) conformations of acetic acid using XP@B3LYP/6-31G(d), which may be compared with the reference values of 7.9 and 8.2 kcal/mol. The interaction energies on the carbonyl sites are of similar magnitude as those found in aldehyde and ketone complexes. The hydrogen bond accepting ability of the hydroxyl oxygen is relatively weak (3.3 kcal/mol) in the syn (**29**) configuration, while the structure in the anti conformer (**32**) enjoys a secondary hydrogen bonding interaction[45] to the carbonyl oxygen,

increasing the XP@B3LYP-binding energy to 5.6 kcal/mol. The latter is 0.9 kcal/mol greater than the reference energy. The optimized geometrical parameters are also in excellent agreement between the two computational approaches. Analogously, both the syn and anti conformations for methyl formate are considered (**34**–**39**), and similar trends as the corresponding acids are found for these complexes. Overall, the average unsigned errors for all acid and ester complexes are just under 0.3 kcal/mol in binding energy and 0.1 Å in hydrogen bond distance.

Table 3-3. Binding energies of bimolecular complexes between water and carbonyl-containing compounds computed using the XP@B3LYP/6-31G(d), B3LYP/aug-cc-pVDZ, and CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ methods[a]

Complex	XP@B3LYP/6-31G(d)	B3LYP/aug-cc-pVDZ	CCSD(T)/aug-cc-pVDZ
22	4.2	4.2	4.8
23	5.8	5.2	6.4
24	5.0	4.9	5.5
25	6.6	5.7	6.6
26	6.1	5.2	6.2
27	6.0	5.1	5.9
28	7.6	7.9	7.9
29	3.3	2.1	3.2
30	6.4	5.5	6.6
31	5.4	4.6	5.4
32	5.6	3.5	4.7
33	8.4	6.6	8.2
34	4.9	4.6	5.3

Continued on next page

Table 3-3. Binding energies of bimolecular complexes between water and carbonyl-containing compounds computed using the XP@B3LYP/6-31G(d), B3LYP/aug-cc-pVDZ, and CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ methods[a] – *Continued*

Complex	XP@B3LYP/6-31G(d)	B3LYP/aug-cc-pVDZ	CCSD(T)/aug-cc-pVDZ
35	6.2	5.1	6.5
36	3.6	2.2	3.5
37	5.2	4.9	5.6
38	5.2	4.5	5.3
39	5.5	3.5	4.8
40	6.0	5.8	6.4
41	7.4	6.7	7.6
42	5.7	5.0	6.0
43	5.9	4.7	5.6
44	7.5	6.5	7.5
45	8.0	7.1	8.1
46	5.6	4.7	5.9
47	5.6	4.3	5.6
48	7.9	6.8	7.9
49	5.6	4.1	5.7

[a]Energies are given in kilocalories per mole.

The amides complexes are structures **40**–**49**. For the formamide-water complexes (**40**–**43**), the XP@B3LYP/6-31G(d) method predicts that the carbonyl group is a better hydrogen bond acceptor (**40** and**41**) than the amide group as a hydrogen bond donor (**42** and **43**) in agreement with the reference data, and the same trend is found in the *N*-methyl formamide and *N*-methyl

acetamide complexes. However, for the donor complexes, XP@B3LYP/6-31G(d) yields a larger binding energy for **43** than complex **42**, due to the alignment of the carbonyl dipole in the direction of the N–H bond, but the opposite is found in the reference data. In the full QM calculation, there is apparently an overlap interaction between water and the carbonyl group,[46] suggested by the smaller hydrogen bond angles (**42**). This is absent in the present X-Pol method.[47] In all amide complexes, the optimized hydrogen bond lengths using XP@B3LYP/6-31G(d) are in excellent agreement with those at the B3LYP/aug-cc-pVDZ level.

3.4.2.3 Heterocyclic Compounds

We considered a number of heterocyclic compounds, which are displayed in Figure 3-3, and the corresponding interaction energies are given in Table 3-4. For both imidazole and pyridine complexes with water, the XP@B3LYP/6-31G(d) method yields weaker binding energies by about 1 kcal/mol than the corresponding reference data. In these cases, the hydrogen bond lengths from the XP@B3LYP/6-31G(d) optimization are about 0.1 Å longer than the B3LYP/aug-cc-pVDZ results. However, good agreement is obtained between the X-Pol- and CCSD(T)-binding energies for the remaining heterocyclic compounds, even though we restricted the nitrogen atom type to just one for all neutral compounds.

Structures **53**–**72** represent hydrogen bonding interactions between a water molecule and the functionalities of nucleobases; cytosine and uracil are depicted in **59**–**64** and **69**–**72**, respectively, whereas only the six-member-ring portions of guanine (**53**–**58**) and adenine (**65**–**68**) are studied. In these complexes, each organic compound contains both hydrogen bond donor sites and acceptor sites, and the latter can be either an oxygen or a nitrogen atom. Thus, these species cover a large range of hydrogen bonding strengths. Although there are some variations, on average, the binding energies are about 1.3 kcal/mol (XP@B3LYP) and 1.4 kcal/mol (CCSD(T)) larger for structures that accept a hydrogen bond than for those that donate one to water. An exception is found for the two carbonyl groups in uracil, which have binding energies of just over 5 kcal/mol,

without which the above difference would be even greater (2 kcal/mol). The binding energies for several complexes are particularly strong, with values greater than 9 kcal/mol (**57**, **60**, and **62)**; this can be attributed to contributions from secondary hydrogen bonding interactions.[45] The X-Pol method correctly reproduces these features, in good agreement with the reference binding energies; however, structure **60** is predicted to have a stronger hydrogen bond than **62** from CCSD(T) calculations, but the opposite is obtained using XP@B3LYP. For the whole set of 23 heterocyclic complexes, the mean unsigned error in binding energy is 0.5 kcal/mol, and the differences in hydrogen bond distance are all under 0.1 Å.

Table 3-4. Binding energies for bimolecular complexes between water and heterocyclic compounds computed with the XP@B3LYP/6-31G(d), B3LYP/aug-cc-pVDZ, and CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ methods[a]

Complex	XP@B3LYP/6-31G(d)	B3LYP/aug-cc-pVDZ	CCSD(T)/aug-cc-pVDZ
50	5.8	6.6	7.4
51	5.6	5.4	6.6
52	6.4	6.4	7.4
53	7.4	6.1	7.4
54	6.2	4.8	5.9
55	7.5	5.6	6.7
56	7.1	5.3	6.9
57	9.9	8.2	9.4
58	8.1	6.4	7.4
59	6.4	5.0	6.3
60	9.4	8.3	9.5

Continued on next page

Table 3-4. Binding energies for bimolecular complexes between water and heterocyclic compounds computed with the XP@B3LYP/6-31G(d), B3LYP/aug-cc-pVDZ, and CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ methods[a] – *Continued*

Complex	XP@B3LYP/6-31G(d)	B3LYP/aug-cc-pVDZ	CCSD(T)/aug-cc-pVDZ
61	8.1	6.6	7.6
62	10.3	7.9	9.2
63	5.7	4.5	5.7
64	6.3	4.3	5.6
65	6.4	4.2	5.4
66	5.3	3.9	5.1
67	7.5	6.3	7.4
68	5.9	6.1	7.0
69	7.7	6.2	7.3
70	5.5	4.1	5.1
71	6.5	5.3	6.7
72	5.7	4.3	5.1

[a]Energies are given in kilocalories per mole.

3.4.2.4 Ions

Complexes involving anions and cations, ranging from simple monatomic ions to delocalized organic species, are shown in Figures 3-4 and 3-5, and the corresponding binding energies are given in Tables 3-5 and 3-6. The binding energies for the monatomic F^- (**73**) and Cl^- (**74**) ions are fitted in exact agreement with the reference data, but the hydrogen bond distances are 0.19 and 0.14 Å longer in the X-Pol model than the reference values.

Table 3-5. Binding energies for bimolecular complexes between water and anions computed using the XP@B3LYP/6-31G(d), B3LYP/aug-cc-pVDZ, and CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ methods[a]

Complex	XP@B3LYP/6-31G(d)	B3LYP/aug-cc-pVDZ	CCSD(T)/aug-cc-pVDZ
73	24.9	24.8	24.9
74	14.1	13.8	14.1
75	25.6	21.2	21.1
76	19.4	21.1	21.7
77	19.2	19.5	20.6
78	19.1	19.3	20.5
79	16.4	15.8	16.4
80	16.3	14.0	14.7
81	18.0	17.1	18.3
82	17.8	15.3	18.0
83	15.3	17.0	16.2
84	17.7	16.4	17.7
85	15.4	14.2	15.4
86	17.4	16.4	17.4
87	17.4	16.4	17.4
88	15.0	14.7	15.7
89	17.3	15.7	17.2
90	15.0	13.3	14.7
91	17.1	15.7	16.8
92	15.2	13.9	15.9

[a]Energies are given in kilocalories per mole.

For the oxyanions, we consider hydroxide ion, (**75**) alkoxide ions (**76**–**78**), and conjugated species (**79**–**80**). In these complexes, there is strong electronic overlap between the two fragments, particularly for the smaller ions. Thus, the block localization of the fragment orbitals in the X-Pol method tends to introduce greater errors as reflected in the hydroxide–water complex (**75**), for which the X-Pol-binding energy is 4.5 kcal/mol greater than in the CCSD(T) calculations. The agreement for the larger and delocalized oxyanions is much improved, with an average error of 1.3 kcal/mol.

A total of twelve carboxylate–water plus allyl anion–water complexes are shown in Figure 3-4. We found that a different set of Lennard-Jones parameters than those used for the alkoxide anions has to be adopted for the carboxylate anions, perhaps due to the more electron-delocalized nature of the carboxylate group. The binding energy ranges from 14.7 to 18.3 kcal/mol from CCSD(T) calculations. The agreement between these results and the XP@B3LYP ones is generally good with a mean unsigned deviation in binding energy of 0.3 kcal/mol. The XP@B3LYP/6-31G(d) method correctly predicts that the bifurcated forms of the complexes with water (**81**, **84** and **89**) are the most stable in each case,[48] and the differences from the least stable complexes are from 2.3 to 2.7 kcal/mol. This agrees with CCSD(T) calculations except that the least stable complexes are reversed for structures **85** and **88** in the XP@B3LYP/6-31G(d) method.

The computed binding energies for cation–water complexes are given in Table 3-6. For the ammonium ions, the single-site hydrogen bonding complex (**94**, **96** and **98**) yields stronger interactions than the symmetric two-site structure (**95**, **97** and **99**) both from the XP/B3LYP and CCSD(T) models. However, the reference calculations predict that the latter complexes are 0.7 – 0.9 kcal/mol more stable than the predictions of the X-Pol method. In the alkyl ammonium series, binding energies decrease progressively as the number of methyl substituents increases, primarily due to charge delocalization. For imidazolium and pyridinium ions, the agreement between

XP@B3LYP/6-31G(d) and CCSD(T) is also reasonable, although the hydrogen bond distances from the X-Pol optimizations are 0.08 Å longer. Two structures are considered for the guanidium ion–water complex. In this case, the energy difference between the two structures is predicted to be smaller than that from CCSD(T) calculations, but the average of the two binding energies is consistent with the ab initio data. Finally, two carbocations are considered, both of which are found to be adequately modeled by the present X-Pol potential.

Table 3-6. Binding energies for bimolecular complexes between water and cations computed using the XP@B3LYP/6-31G(d), B3LYP/aug-cc-pVDZ, and CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ methods[a]

Complex	XP@B3LYP/6-31G(d)	B3LYP/aug-cc-pVDZ	CCSD(T)/aug-cc-pVDZ
93	22.0	23.9	22.0
94	20.0	20.4	20.1
95	15.4	15.8	16.3
96	18.4	17.9	18.3
97	14.5	14.3	15.2
98	17.0	16.3	17.2
99	13.7	13.2	14.4
100	16.2	14.8	15.6
101	16.6	14.9	16.0
102	16.5	15.8	17.4
103	14.3	12.3	13.3
104	14.7	12.8	14.7
105	16.9	13.9	14.0

[a]Energies are given in kilocalories per mole.

3.4.3 Overall Assessment

The performance of the present XP@B3LYP/6-31G(d) method, based on comparisons to the results of CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ calculations, is shown in Figure 3-6. Overall, the root mean square deviation (RMSD) in binding energy between the XP@B3LYP/6-31G(d) predictions and the CCSD(T)/aug-ccc-pVDZ//B3LYP/aug-cc-pVDZ reference data for all 105 bimolecular complexes, covering a range of binding energies of more than 20 kcal/mol, is 0.8 kcal/mol. For a similar set of bimolecular systems, combined QM/MM calculations using the AM1[13,49] and HF/3-21G[50] methods along with the three-point charge TIP3P model for water yielded RMSDs of 1.2 and 0.5 kcal/mol, respectively. In those studies, the optimization target was obtained from HF/6-31+G(d) calculations. In other studies, Riccardi et al.[51] calculated the binding energies for a series of bimolecular complexes of water and organic compounds representing amino acid side chains using the SCC-DFTB/CHARMM potential and obtained an RMSD of 1.2 kcal/mol with respect to B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) dataset. Freindorf et al.[52] obtained an RMSD of 1.5 kcal/mol for small organic molecule/water complexes using the B3LYP/6-31+G(d)/AMBER potential. The present X-Pol potential, making use of B3LYP/6-31G(d) for each fragment, yields slightly better agreement with the target dataset than several combined QM/MM methods that employed the same strategy of optimizing the van der Waals parameters. Although the origin of the good performance of the X-Pol method needs to be more carefully investigated, a main difference from these QM/MM approaches is that the mutual electronic polarization effects between the two monomers in each complex are included in X-Pol, whereas fixed charge MM force fields are used in the QM/MM calculations.

The RMSD of the optimized hydrogen bond distance between the XP@B3LYP/6-31G(d) optimization and that of the B3LYP/aug-cc-pVDZ method is 0.13 Å (Figure 3-7), which is slightly greater than a value of 0.08 Å for a much smaller set of 14 structures in a previous study.[8] This may also be compared with the AI-3/MM,[50] SCC-DFTB/MM,[51] and B3LYP/6-

31+G(d)/AMBER methods,[52] which have RMSDs of 0.07 Å, 0.08 Å, and 0.11 Å relative to the respective datasets. The RMSD errors for hydrogen bond angles are 12° when comparing XP@B3LYP/6-31G(d) to B3LYP/aug-cc-pVDZ optimizations (Fig. 8). Although the errors in the optimized hydrogen bond angles are relatively large, the potential energy surfaces for these interactions are typically flat and they do not affect the binding energies significantly. Overall, the present XP@B3LYP/6-31G(d) model yields reasonable hydrogen bond geometries for a variety of organic functional groups interacting with a water molecule.

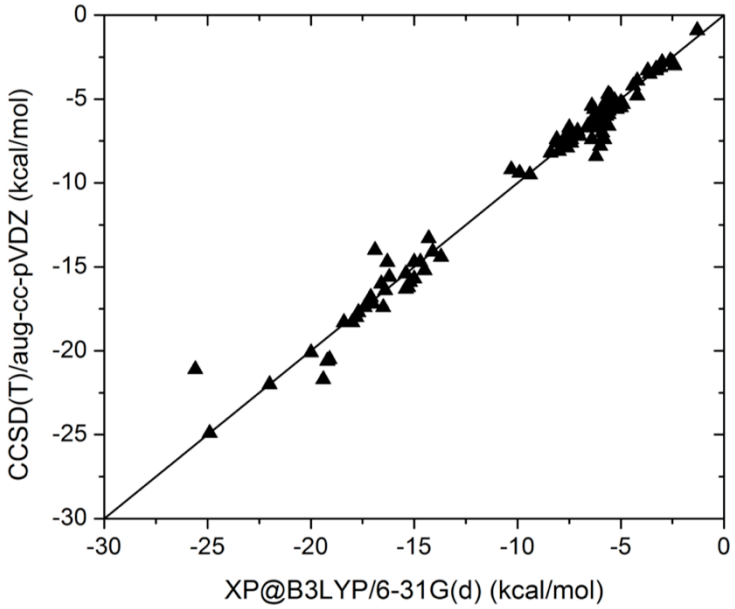


Figure 3-6. Comparison of the computed hydrogen bond interaction energies obtained using the XP@B3LYP/6-31G(d) and CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ methods. The geometries used in the X-Pol calculations were obtained at the same level of theory, whereas those used in the coupled cluster energy evaluations were optimized with B3LYP/aug-cc-pVDZ.

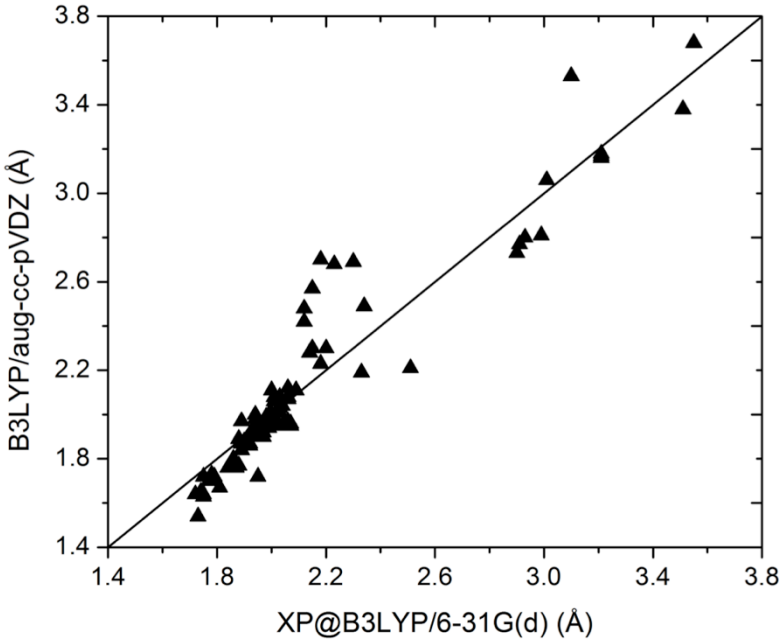


Figure 3-7. Comparison of the optimized hydrogen bond distances using the XP@B3LYP/6-31G(d) and B3LYP/aug-cc-pVDZ methods.

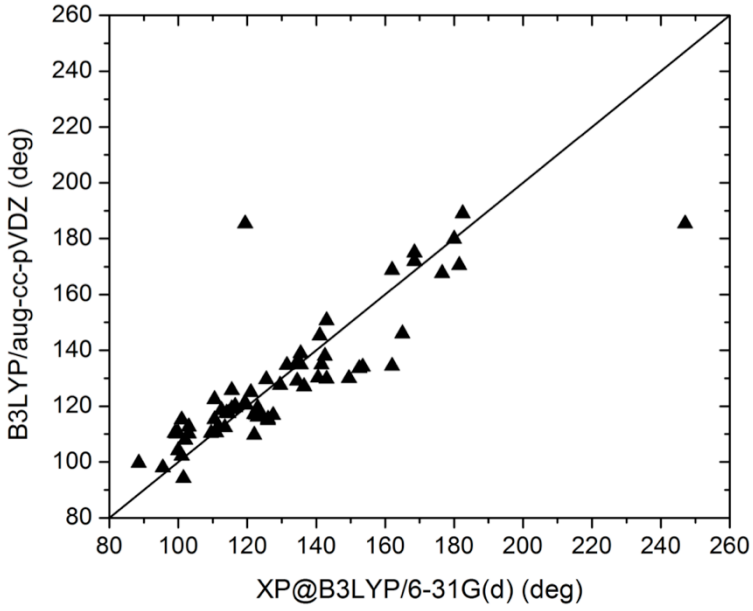


Figure 3-8. Comparison of the optimized hydrogen bond angles using the XP@B3LYP/6-31G(d) and B3LYP/aug-cc-pVDZ methods.

3.5. Concluding Remarks

The X-Pol potential uses an electronic structure method to model the mutual polarization effects between interacting fragments. It employs an effective Hamiltonian to model energy contributions from interactions between fragments beyond the electrostatic ones that are included self-consistently in the quantum mechanical fragment calculations. These additional interactions are dominated by exchange repulsion and dispersion energies,[8,17] which are approximated by the Lennard-Jones model.[4] The parameters of the Lennard-Jones function can be optimized to reproduce experimental data or accurate results from high-level calculations; in the present study, the CCSD(T)//aug-cc-pVDZ-binding energies and the optimized hydrogen bond distances and angles using B3LYP/aug-cc-pVDZ are chosen as the optimization target. Implicitly, the X-Pol results obtained with the optimized Lennard-Jones parameters for different atom types account for the energy component due to charge transfer in a way analogous to molecular mechanics force fields. If the specific description of charge transfer effects is important for a given problem, the fragment partitions need to be assigned in such a way that the electron donor and acceptor groups are both included in the same fragment. Alternatively, resonance charge delocalization effects can be modeled by the multiconfigurational, generalized X-Pol (GX-Pol) theory highlighted recently[47,53] or by using ensemble DFT.[16] Here, however, we tested the simpler approach in which charge transfer is only implicit. The goal is to develop and test a computational approach involving a modest computational cost that can be applied for fast calculations on large systems.

The explicit polarization (X-Pol) method is based on block localization of molecular orbitals within each fragment. The fragments can be assigned, for example, as individual molecules such as solvent molecules or as amino acid residues in a protein. If desired, important portions of the system can be treated as single large fragments, for example, one may take an entire active site of an enzyme as a single fragment. A key feature of the X-Pol method is that the block-localized orbitals of each fragment are optimized in the presence of the instantaneous electric field due to

all other fragments, and the mutual polarization among all fragments is determined using self-consistent field methods.[6,7] When the electronic integrals between different fragments are approximated by an external potential expansion, the computational costs can be greatly reduced.[4,8] Furthermore, the fragment block-localization scheme naturally leads to linear scaling in electronic structural calculations on large systems. The X-Pol approach provides a theoretical framework for developing next-generation force fields for macromolecular simulations using an explicitly quantum mechanical electronic structure theory.[54]

The present calculations employ the B3LYP hybrid density functional, and the Lennard-Jones parameters are optimized to higher-level reference data for a dataset containing 105 bimolecular, trimolecular, tetramolecular, pentamolecular, and heptamolecular complexes between one or more water molecules and an organic or inorganic compound or ion, representing the functional groups of amino acids and nucleobases. We found that the average deviation between the binding energies calculated by the XP@B3LYP/6-31G(d) and the CCSD(T)/aug-cc-pVDZ//B3LYP/aug-cc-pVDZ methods is about 0.8 kcal/mol, whereas the deviation in hydrogen bond distance is about 0.1 Å. It will be interesting to further test this kind of model through condensed-phase simulations, including the computation of liquid properties and solvation free energies.

Chapter 4

Quantum Mechanical Force Field for Water with Explicit Electronic Polarization[a]

4.1. Introduction

Critical to the success of dynamical simulations of chemical and biological systems is the potential energy function used to describe intermolecular interactions.[1,2] Because of the importance of aqueous solution and its unique roles in biomolecular interactions, water has been a subject of extensive and continuous investigation (a review in 2002 included a partial list of 46 water models,[3] while at least two dozen new models have appeared since that time).[4,5] An accurate and efficient model for liquid water also serves as an anchor for developing force fields for proteins, nucleic acids, and carbohydrates. Traditionally, the Lifson-type of effective, pairwise potentials have been used,[1,2,6] and much effort has also been devoted to incorporating many-body polarization effects into such force fields.[7] However, unlike the development of pairwise potentials, there is a great deal of uncertainty in the treatment of polarization effects, both in the choice of functional form and in the associated parameters. This is reflected in the fact that simple point charge models such as SPC,[8] TIP3P and TIP4P[9] quickly emerged as the standards in the 1980s for biomolecular force fields, but no standard polarizable force fields have emerged although dozens of polarizable potentials for water have been proposed.[3,4,10] We have developed a quantum mechanical framework in which each individual molecular fragment is treated by electronic structure theory.[11–14] Since polarization effects are naturally included in the self-consistent field (SCF) optimization of molecular wave functions, we call this method the explicit

[a] Reproduced with permission from Han, J.; Mazack, M. J. M.; Zhang, P.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2013**, *139*, 054503. Copyright 2013, AIP Publishing LLC.

polarization (X-Pol) theory.[11,14,15] Recent studies demonstrated the feasibility of X-Pol as a next generation force field for biomolecular simulations,[13] and encouraging results have been obtained using standard semiempirical Hamiltonians.[12,16] In the present paper we report a novel model for water, called XP3P, based on X-Pol theory and a three-point charge representation of the electrostatic potential, as a first step in our effort to develop a full quantum mechanical X-Pol force field for biomolecular and materials simulations.

The present quantum mechanical force field (QMFF) may be compared with phenomenological representations of electronic polarization in three commonly used methods in molecular mechanics, namely induced-dipole, Drude-oscillator, and fluctuating-charge models. In the induced-dipole approach,[17–20] atomic polarizabilities are assigned to the interaction sites, typically located on, but not limited to, atomic centers, from which induced point dipoles, representing the total electric field of the system, are obtained.[21] A commonly used method to assign atomic polarizabilities is the dipole interaction model (DIM) popularized by Applequist et al.[22] and extended by Thole[23] to incorporate short-range damping functions. Remarkably, the values optimized in DIM are quite transferrable,[24] requiring typically one parameter per element. The Drude-oscillator model may be considered as a point-charge equivalent of the induceddipole method.[25,26] Here, one or a set of point charges are harmonically linked to a polarizable site, in which the directions and distances of the Drude oscillators give rise to the corresponding induced dipole moments. The fluctuatingcharge[27–30] approach employs a chemical potential equalization scheme, in which the instantaneous partial charges minimize the energy of the system. The fundamental parameters used in the fluctuating-charge model correspond to the atomic electronegativity and hardness that are rigorously defined in density functional theory.[31]

Each of these classical methods has its advantages and shortcomings in practice. In the fluctuating-charge model, unphysical charge transfer effects between distant monomers can occur. Thus, charge constraints are required. On the other hand, the induced-dipole and the

Drude-oscillator model are difficult to use for representing molecular polarization involving a significant charge delocalization such as that across a conjugated polyene chain and the polarization of push-pull compounds (e.g., the crystal of p-nitroaniline). The Drude oscillator model has the advantage of simplicity in practice since any dynamics simulation code can be conveniently adapted to treat polarization effects by that method.

The X-Pol method relies on the partition of a large, condensed-phase system into molecular or submolecular fragments (or blocks),[11,12,15] which can be single solvent molecules like water, amino acid residues or nucleotide bases, small ions or enzyme cofactors, or a collection of these small units. The wave function of each molecular fragment is described by a Slater determinant of block-localized molecular orbitals that are expanded over basis functions located on atoms of the fragment. The total molecular wave function is approximated as a Hartree product of these fragmental, determinant functions. Consequently, Coulombic interactions between different fragments are naturally incorporated into the Hamiltonian, but short-range exchange repulsion, charge delocalization (also called charge transfer) and long-range dispersion interactions are not explicitly treated in the quantum chemical formalism.[32–34] These effects are included and optimized empirically to strive for accuracy (and efficiency) in X-Pol in the same spirit as that in force field development. The determinantal wave function for each monomer fragment can be approximated by wave function theory (WFT) at either an ab initio or a semiempirical level,[12,35] the density may be approximated by density functional theory (DFT),[35,36] or one can combine levels of theory,[37] but in this paper we use only semiempirical wave function theory. Although the present work involves only water, we note that the X-Pol theory can be used to model electronic polarization involving conjugated systems and significant charge delocalization contributions,[38] and the X-Pol model is also a reactive force field for modeling systems involving bond-forming and bond-breaking processes.

Semiempirical methods employing neglect diatomic differential overlap (NDDO)[39] are especially suited for QMFF development because of their computational efficiency. However, most such semiempirical models were not optimized to describe intermolecular interactions that are essential for modeling condensed-phase systems.[40–43] Part of the problems has been remedied through the incorporation of empirically damped dispersion functions.[44–49] Another important deficiency of many semiempirical models for treating nonbonded interactions is that molecular polarization is systematically underestimated. Recently, we have introduced a polarized molecular orbital (PMO) alternative,[49–51] in which a set of *p*-orbitals are added to each hydrogen atom.[52] It was found that the computed molecular polarizabilities for a range of compounds containing hydrogen, carbon, and oxygen are significantly improved.[49,51] Employing this strategy, we report here a parametrization of the PMO model for water (PMOw), which can be used in X-Pol for liquid simulations.

In the following, Section 4.2 summarizes the PMO parameterization for water and the development of the XP3P model liquid water. Computational details are given in Section 4.3. In Section 4.4, we present results and discussion. Section 4.5 concludes the paper with a summary of major findings.

4.2. Method

The X-Pol quantum mechanical force field is designed to model condensed phase systems with or without bondforming and bond-breaking processes. Thus, the X-Pol method can be used as a general-purpose force field in dynamics simulations of solvated proteins or as a reactive force field to model chemical reactions in solutions and in enzymes. In this section, we first describe the quantum chemical model designated as PMOw for water and compounds containing oxygen and hydrogen atoms. The acronym PMO is used to describe the general semiempirical model in which, in addition to a minimal basis set, a set of *p*-orbitals is added to hydrogen

atoms.[49,51] Then, we highlight its incorporation in X-Pol, called the XP3P model, for simulation of liquid water.

Table 4-1. Semiempirical Parameters for H and O Atoms in the PMOw model[a]

	H	O
U_{ss} (eV)	-11.15043	-111.86028
U_{pp} (eV)	-7.35459	-78.64105
β_s (eV)	-6.88125	-25.57063
β_p (eV)	-3.52628	-31.90404
ζ_s (Bohr^{-1})	1.17236	3.05303
ζ_p (Bohr^{-1})	1.05333	3.12265
α (Å^{-1})	3.05440	3.76880
g_{ss} (eV)	12.73667	17.36659
g_{sp} (eV)	8.04688	13.37288
g_{pp} (eV)	6.98401	14.78196
$g_{pp'}$ (eV)	10.65161	13.49319
h_{sp} (eV)	1.92149	4.42643

[a]The derived parameter, h_{pp}, is determined from g_{pp} and $g_{pp'}$, and has been set to a minimum value of 0.1 eV as implemented in the MOPAC program, $h_{pp} = \max\{0.1 \text{ eV}, (g_{pp} - g_{pp'})/2\}$.

4.2.1. Polarized Molecular Orbital Model for Water

The PMOw model is a new parameterization of the PMO method,[49] which is based on the MNDO formalism[53] with three key enhancements. First, a set of diffuse *p*-type basis functions is added on the hydrogen atoms.[50] This greatly improves the quality of the computed molecular polarizabilities and hence the treatment of hydrogen bonding interactions. Second, a damped

dispersion function, following the work of Tang and Toennies in wave function theory[54] and Grimme in density functional theory,[55] is included as a post-SCF correction to the electronic energy. In the present implementation, we have adopted the method and parameters proposed by Hillier and co-workers in the PM3-D method.[44–46] The inclusion of the damped dispersion terms further improves the description of intermolecular interactions and the performance of PMO on small molecular clusters.[44–47,49,51] Third, the PMOw model is parameterized for general applications to a specific class of compounds (see Section 4.4.1 for the set of parametrization data), and the optimization targets include molecular polarizabilities and non-bonded interactions as well as other properties used in the traditional semiempirical parameterization.[49] The parameters presented here are optimized for compounds containing oxygen and hydrogen atoms, especially for studying liquids, aqueous solutions, and proton transport. We note here that, in the same way that atoms are assigned types in molecular mechanics, the parameters for oxygen and hydrogen atoms in functional groups other than water (e.g., peptide bonds) need not be restricted to the same as used for such atoms in water. This departs from the philosophy that has usually been used in semiempirical methods,[56,57] in which general atomic parameters are used for all functionalities.

In theMNDO formalism,[53,58] there are 12 atomic parameters for each element, and the PMOw values for water and other compounds containing oxygen and hydrogen are listed in Table 4.1. These values are similar in many respects to the PMOv1 model introduced previously,[49] but they result from a new parametrization presented below. Three exceptions were made to the MNDO functional forms because of the addition of diffuse *p* basis functions on hydrogen atoms,[49] and they are listed as follows:

1. For the resonance integral involving *p* orbitals on hydrogen, the following conventions are used:

$$\beta_{lp}^{\mathrm{HH}} = 0 \tag{4.1a}$$

$$\beta_{lp}^{\mathrm{OH}} = \frac{\beta_l^{\mathrm{O}} + \beta_p^{\mathrm{H}}}{2} S_{lp} A_{lp} e^{\kappa_{lp} R_{\mathrm{OH}}} \tag{4.1b}$$

where l is the angular momentum quantum number, having the values of 0 (s orbital) and 1 (p orbital), and the subscript p denotes a p-orbital on hydrogen. Notice that Eq. (4.1b) is slightly different from the expression used in Ref. 49, in which the exponential function is absent. In Eq. (4.1b), β_l^{O} and β_p^{H} are standard MNDO-type parameters, A_{lp} and κ_{lp} are additional parameters introduced in PMO, and R_{OH} is the distance between oxygen and hydrogen atoms. S_{lp} in Eq. (4.1b) is an overlap integral ($\langle O_l | H_p \rangle$) between oxygen and hydrogen Slater-type orbitals using the parameters listed in Table 4-1, but specific exponents, ζ^{OO} and ζ^{HH}, are used for H–H and O–O pairs, respectively, in PMOw.

2. In standard MNDO,[53,58] the nucleus-electron attraction integral, $H_{\mu\nu}^{A}$, between electronic charge density on atom A and nucleus B is evaluated on the basis of the two-electron repulsion integral, $\langle \mu_A \nu_A | s_A s_B \rangle$.[59] In PMOw, if both A and B are hydrogen atoms, for a distribution of p orbitals (pp'), this is screened as follows:

$$H_{pp'}^{\mathrm{H}} = \left[1 - Be^{-\lambda R_{HH'}^2}\right] \left(H_{pp'}^{\mathrm{H}}\right)_{\mathrm{MNDO}} \tag{4.2}$$

3. For the homonuclear core-core repulsion integrals,[49,53,58] the standard values for α^O and α^H are replaced by $\hat{\alpha}^O$ and $\hat{\alpha}^H$. Note that α^O and α^H are used as in standard MNDO for core-core repulsion integrals between oxygen and hydrogen atoms.

The parameters in the standard MNDO formalism[53] (Table 4-1) and the additional parameters (Table 4-2) described above were adjusted by iterative optimization using a genetic algorithm for some of the systems and properties listed in Table A-1 in Appendix A. In

comparison with the results in Ref. 49, the present parameter set further improves the calculated molecular polarizability and dipole moment of water in the gas phase as well as the binding energy and dipole moment of water dimer (Table 4-3).

Table 4-2. Additional Semiempirical Parameters for Oxygen and Hydrogen in the Polarized Molecular Orbital Model and the Lennard-Jones Parameters in Explicit Polarization Model for Liquid Water

Parameter	Value
A_{sp}	0.03000
A_{pp}	0.15000
B	1.00000
k_{sp} (Å^{-1})	0.47069
k_{pp} (Å^{-1})	0.47069
l (Å^{-2})	1.10000
$\hat{\alpha}^{H}$ (Å^{-1})	2.52552
$\hat{\alpha}^{O}$ (Å^{-1})	3.03253
ζ^{HH} (Bohr^{-1})	1.28000
ζ^{OO} (Bohr^{-1})	2.76400
σ_H (Å)	0.800
σ_O (Å)	3.225
ε_H (kcal/mol)	0.05
ε_O (kcal/mol)	0.15

4.2.2. Explicit Polarization Theory

In X-Pol,[11,12,15] the system is partitioned into molecular or submolecular fragments, in which the total wave function of the system is assumed to be a Hartree product of the determinant wave functions of the individual fragments. In the present case, each fragment is simply a single water molecule, and the overall wave function is

$$\Phi = \prod_{a=1}^{N} \Psi_a \tag{4.3}$$

where N is the number of fragments in the system, and Ψ_a is a Slater determinant of doubly-occupied molecular orbitals (MOs) block-localized on molecule (fragment) a. The approximation of Eq. 4.3 implies neglect of the short-range exchange repulsion[33] and long-range dispersion interactions[60] between different fragments, which are corrected empirically below.[11,12,15] Use of Eq. 4.3 reduces the computational costs, allowing molecular dynamics and Monte Carlo simulations to be carried out for large systems efficiently with sufficient sampling.[12,13]

The effective Hamiltonian of the system is given by

$$H = \sum_{a=1}^{N} H_a^0 + \frac{1}{2} \sum_{a=1}^{N} \sum_{b \neq a}^{N} H_{ab} \tag{4.4}$$

where H_a^0 is the electronic Hamiltonian of fragment a in the gas phase and H_{ab} represents the effective interactions between molecules a and b:

$$H_{ab}(\rho_b) = -\sum_{i=1}^{M} V_i(\rho_b) + \sum_{A=1}^{Q} Z_A^a V_A(\rho_b) + E_{ab}^{\mathrm{XD}} \tag{4.5}$$

where M is the number of electrons and Q is the number of atoms in fragment a, Z_A^a would be the nuclear charge of atom A of fragment a if all electrons were treated explicitly but here it is the core charge since 1s electrons of oxygen atoms are in the core, and E_{ab}^{XD} is the exchange-dispersion correlation energy. The electrostatic potential $V_x(\rho_b)$, either at the electronic ($x = i$) or at the nuclear ($x = A$) position, due to the instantaneous charge density of fragment b is given by

$$V_x(\rho_b) = -\int \frac{\rho_b(\mathbf{r})d\mathbf{r}}{|\mathbf{r}_x - \mathbf{r}|} + \sum_{B=1}^{Q} \frac{Z_B^b}{|\mathbf{r}_x - \mathbf{R}_B^b|} \tag{4.6}$$

Here, $\rho_b(\mathbf{r})$ is the electron density of fragment *b* derived from the corresponding wave function Ψ_b (or Kohn–Sham Slater determinant),[11,12] and $\mathbf{R}_B^b$ denotes the nuclear coordinates.

We define the total interaction energy of a condensed phase system by

$$E_{tot} = \langle \Phi | H | \Phi \rangle - \sum_{a=1}^{N} \langle \Psi_a^0 | H_a^0 | \Psi_a^0 \rangle \tag{4.7}$$

The energy defined in Eq. 4.7 corresponds to the total energy of the condensed-phase system relative to that of infinitively separated fragments. Since all molecules are identical in pure liquid water in the present study, the last summation term in Eq. 4.7 is simply NE_a^0 with $E_a^0 = \langle \Psi_a^0 | H_a^0 | \Psi_a^0 \rangle$ being the energy of an isolated monomer. It is often useful for interpretive purposes to consider the dimeric interaction energies between two fragments even for a potential that includes many-body polarization effects as in the present X-Pol potential. To this end, we define the interaction energy between fragments *a* and *b* by[12]

$$E_{ab} = \frac{1}{2}\left(\langle \Psi_a | H_{ab} | \Psi_a \rangle + \langle \Psi_b | H_{ba} | \Psi_b \rangle \right) \tag{4.8}$$

The two terms in Eq. 4.8 corresponds to *a* embedding in *b* and *b* embedding in *a*, respectively, both in the presence of the rest of the system, and they are not always numerically equivalent in practice[11] even though they describe the same intermolecular interactions. The definition of Eq. 4.8 ensures that $E_{ab} = E_{ba}$.

Table 4-3. Computed Equilibrium Properties for Water Monomer and Dimer from Different Polarizable Water Models and *ab initio* MP2/(CBS) with CCSD(T) Corrections along with Experimental Data

		PMOw	XP3P	AMOEBA[a]	POL5/TZ[a]	*ab initio*	Expt.[b]
H_2O	AE (kcal/mol)	233.0	233.0			229.3[c]	232.2
	IP (eV)	13.20	13.20			12.42	12.68
	r (Å)	0.955	0.957	0.957	0.957	0.9589[d]	0.9572
	θ (deg)	104.6	104.5	108.5	104.5	104.16[d]	104.52
	α (Å^3)	1.27	1.27	1.41	1.29	1.45[e]	1.45
	q^H (e)	0.16	0.34	0.26		0.35	N/A
	q^O (e)	-0.31	-0.67	-0.52		-0.70	N/A
	μ (D)	1.88	1.88	1.77	1.85	1.84[e]	1.86[f]

Continued on next page

Table 4-3. Computed Equilibrium Properties for Water Monomer and Dimer from Different Polarizable Water Models and *ab initio* MP2/(CBS) with CCSD(T) Corrections along with Experimental Data – *Continued*

		PMOw	XP3P	AMOEBA[a]	POL5/TZ[a]	*ab initio*	Expt.[b]
$(H_2O)_2$	ΔE_b (kcal/mol)	-5.1	-5.2	-4.96	-4.96	-5.0[g]	-5.44
	R_{OO}[h] (Å)	2.89	2.90	2.89	2.90	2.92	2.98
	α[h] (deg)	6.2	1.3	4.2	4.7	4.8	-1±10
	ϕ[h] (deg)	115	165	123	117	125	123 ± 10
	$\langle \mu_{mol} \rangle$ (D)	2.10	2.16				
	μ (D)	2.39	3.85	2.54	2.44	2.65	2.64

[a]Ref. 10. [b]Ref. 61. [c]Ref. 62. [d]Ref. 63. [e]Ref. 64. [f]Ref. 65. [g]Ref. 66. [h]Definitions of geometric parameters of $(H_2O)_2$ are illustrated in the inlet of Figure 4.1.

The exchange-dispersion correlation energy can be incorporated with an explicit density dependent term and added to the Fock operator as described in the work of York and co-workers.[32,67] Alternatively, the damped dispersion term that is an intrinsic part of the PMOw model can be used with the addition of a repulsive potential. Here, in the spirit of simplicity for a force field, we adopt a Lennard-Jones potential to approximate the remaining energy contributions[11,12,15] not included in the PMOw electronic structure method.[49] (Thus there are two R^{-6} terms, one in PMOw for intrafragment interactions and one in the Lennard-Jones term associated with interfragment interactions.) The Lennard-Jones term introduces two empirical parameters per atom type:

$$E_{ab}^{XD} = \sum_{A}^{Q}\sum_{B}^{Q} 4\varepsilon_{AB}\left[\left(\frac{\sigma_{AB}}{R_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{R_{AB}}\right)^{6}\right] \tag{4.9}$$

where ε_{AB} and σ_{AB} are obtained from the geometric mean of atomic parameters such that $\varepsilon_{AB} = (\varepsilon_A\varepsilon_B)^{1/2}$ and $\sigma_{AB} = (\sigma_A\sigma_B)^{1/2}$. These parameters are also listed in Table 4.2.

4.2.3. The XP3P Model for Liquid Water

The electrostatic potential (ESP) in Eq. 4.6 can be determined explicitly by evaluating the associated one and two-electron integrals in SCF calculations. However, this would have not saved much computational time, and would have missed the point of developing a fragment-based technique in electronic structure calculations. As we have proposed previously,[11,15,35] it is desirable to employ a more computationally efficient method to approximate the external potential $V_x(\Psi_b)$. In the present application to liquid water, we use a simple, three-point-charge approximation to $V_x(\Psi_b)$. Consequently, we call this X-Pol potential with three-point charges for water the XP3P model.

Several methods based on atomic partial charges for approximating the quantum external potential were described originally for the X-Pol potential,[11,12] and some of them were adopted later in other fragment-based molecular orbital models.[68] Although the use of atomic charges

obtained from fitting the quantum mechanical $V_x(\Psi_b)$ has been successfully used in several molecular mechanics force fields,[69,70] it is known that the ESP-fitting method sometimes yields unreasonably large partial charges on structurally buried atoms.[71] In addition, large variations could occur as a result of structural fluctuations to expose buried atoms during a dynamics simulation. A general approach is the multi-center multipole expansion of the quantum mechanical ESP,[72] and this method has been used in the effective fragment potential model;[73] multi-center multipolar representations could also be used with X-Pol.[74] A conceptually simple alternative is to use atomic charges derived from a population analysis such as the Mulliken or Löwdin population method.[75] When used with small, well balanced basis sets, the Mulliken or Löwdin charges can provide a good representation of the relative atomic electronegativity and they are computationally efficient. Scaled Mulliken population charges have been used and shown to be effective in statistical mechanical Monte Carlo simulations of liquid water using an explicit QMFF.[12]

Another way of approximating the external potential for intermolecular interactions is to employ partial atomic charges that are mapped from the density matrix to reproduce experimental dipole moments (in contrast to ESP fitting). This has been called a class IV charge model, and it can be parametrized to show good consistency for a variety of electronic structure methods and basis sets.[76,77] Alternatively, partial atomic charges can be derived to rigorously reproduce the molecular moments to any order of accuracy from a Lagrangian multiplier procedure. Following the method proposed by Thole and van Duijnen[78] and extended by Swart and van Duijnen,[79] we applied the Lagrangian multiplier approach to semiempirical methods,[80] which are known to yield excellent molecular dipole moments in comparison with experiments. In this approach, both the total molecular dipole moment and the local atomic hybridization contributions of the approximate NDDO wave function are reproduced exactly. In the present implementation, we preserve the total and local molecular dipole moments. In addition, we included in the procedure

the capability to reproduce experimental molecular polarizability and its atomic decomposition according to the dipole interaction model.[80] We called this method the dipole preserving and polarization consistent (DPPC) charge model.[80]

Specifically, the DPPC charge has two contributions, the Mulliken population charge, q_A^{MP} and the residual charges Δq_A^B due to preservation of atomic *s* and *p* hybridization dipole moments:[80]

$$q_A^{\mathrm{DPPC}} = q_A^{\mathrm{MP}} + \sum_{B=1}^{Q} \Delta q_A^B \tag{4.10}$$

where the residual charge Δq_A^B on atom *A* due to the constraint that the residual moment is identical to the atomic hybridization contribution from atom *B*:

$$\mu_B^{\mathrm{hyb}} = -\left(\mathbf{P}_{sp}\right)_B \cdot \mathbf{D}_B = \sum_{A=1}^{Q} \Delta q_A^B \mathbf{R}_A \tag{4.11}$$

where $(\mathbf{P}_{sp})_B$ is a diagonal matrix with the densities $P_{sp_x}^B$, $P_{sp_y}^B$, and $P_{sp_z}^B$, on atom *B*, $\mathbf{D}_B$ is the corresponding dipole integral, and $\mathbf{R}_A$ denotes the coordinates of atom *A*. The residual charges Δq_A^B that reproduce the hybridization component of molecular dipole moment, μ_B^{hyb}, are predominantly localized on atoms closest to atom *B*. Since the molecular dipole moment is determined from

$$\mu_{QM} = \sum_{A=1}^{Q} q_A^{\mathrm{MP}} R_A + \sum_{B=1}^{Q} \mu_B^{\mathrm{hyb}} \tag{4.12}$$

in semiempirical methods employing the NDDO approximation,[81] it is clear that the atomic charges given in Eq. 4.10 reproduce exactly the full quantum mechanical dipole moment and the local, atomic hybridization contributions:

$$\sum_{A=1}^{Q} q_A^{\mathrm{DPPC}} \mathbf{R}_A = \mu_{QM} \tag{4.13}$$

The residual charges depend on geometry and atomic electronegativity, and an expression for them was given in Ref. 80. The advantage of using the DPPC charges over the ESP-fitted ones is that local properties of the dipole integrals are explicitly accounted for and fully utilized to generate the partial atomic charges. The method to generate DPPC charges is applicable both to neutral and ionic molecules, independent of the origin of coordinates.[80]

4.3. Computational Details

The parameterization of the PMOw model was carried out by iterative optimization using a genetic algorithm that has been detailed in Ref. 49. The PMOv1 set of parameters overestimated the dipole moment of water (2.19 D) and underestimated the interaction energy for the water dimer (4.7 kcal/mol) in comparison with the target values of 1.85 D from experiment[65] and 5.0 kcal/mol from CCSD(T) and MP2/(CBS) calculations.[66] The PMOw parametrization improves these quantities for application to water and its ions.

Statistical mechanical Monte Carlo simulations were performed on a system consisting of 267 water molecules in a cubic box, employing the PMOw Hamiltonian. Based on procedures described previously,[12,16] periodic boundary conditions were used along with the isothermal-isobaric ensemble (NPT) at 1 atm and temperature ranging from −40 to 100 °C. As in the development of other empirical potentials including the successful SPC,[8] TIP3P, and TIP4P models[9] and the polarizable AMOEBA,[10] SWM4-NDP[26] and POL5/TZ[82] potentials for water (and many other water models not explicitly compared in this paper), the parameterization was performed only at 25 °C. The XP3P model based on the PMOw Hamiltonian has four Lennard-Jones parameters, ε_O, ε_H, σ_O, and σ_H. We have kept the ε_H and σ_H values used in a previous X-Pol simulation of liquid water with the AM1 Hamiltonian (called the MODEL potential for water), and we made small adjustments of the other two values (3.225 Å and 0.15 kcal/mol)[12] to reproduce the liquid density and heat of vaporization within 1% of the experimental values at

25 °C. In the parameterization stage, spherical cutoff with a switching function between 8.5 Å and 9.0 Å based on oxygen-oxygen separations was employed, and a long-range correction to the Lennard-Jones potential was included. (The SPC and TIP3P/TIP4P models[8,9] and later the TIP5P model[83] were also developed using cutoff distances, which were as small as 7.5 Å with a box of 125 or 216 water molecules.) Although it is possible to use Ewald sums to treat long-range electrostatic interactions,[84] we have not used the particle-mesh Ewald implementation in the present Monte Carlo calculation. In Monte Carlo simulations, new configurations were generated by randomly translating and rotating a randomly selected water molecule within ranges of ±0.13 Å and ±13°. In addition, the volume of the system was changed randomly within the limit of ±150 $Å^3$ on every 550th attempted move, and the coordinates of oxygen atoms were scaled accordingly. (Note that in the Monte Carlo calculations, the waters are rigid, so the hydrogen positions also adjust when the oxygen positions are adjusted.) These options were slightly adjusted to maintain an acceptance rate of about 45% at each temperature in the Metropolis sampling. In each simulation, at least 5×10^6 configurations were discarded for equilibration, which was followed by an additional 1×10^7 to 1.1×10^8 configurations for averaging. About 6×10^6 configurations can be executed per day on a 6-core Intel Xeon X7542 Westmere processor at 2.66 GHz.

The XP3P model was further examined in molecular dynamics simulations for 500 ps in the NVT ensemble, using the Lowe-Andersen thermostat[85,86] and a volume fixed at the average value from the Monte Carlo simulation; the number of water molecules in the dynamics simulations was also 267. The monomer geometries were enforced by the SHAKE/RATTLE procedure.[87] Although long-range electrostatic interactions can be computed using the particle-mesh Ewald summation that has been extended for the X-Pol potential,[84,88] we have used 9.0 Å cutoff in the present study. The velocity Verlet integration algorithm was used with a 1fs time step.

The total energy of the system was obtained from fully converged wave functions for each water molecule for each microscopic configuration, although different procedures were utilized in the Monte Carlo sampling and in molecular dynamics simulations. In Monte Carlo, an initial set of DPPC charges, derived from an initial guess of the X-Pol wave function, e.g., that from the previous configuration (with random perturbation to some randomly selected elements in the density matrix), are incorporated into the Fock matrix in terms of one-electron integrals (as in combined QM/MM schemes) in the subsequent iteration step during the self-consistent field (SCF) optimization. Then, a new set of orbital coefficients is obtained to generate updated DPPC charges for the next iteration until the electronic energy is converged to 5×10^{-5} eV for each monomer and to 10^{-5} for the partial atomic charges (in atomic units) between consecutive iterations. In Monte Carlo simulations, the Fock operator is constructed analogously to a combined QM/MM scheme,[89] which is not fully variational with respect to the change of the charge density; the external potential does incorporate the complete electrostatic effects in a self-consistent manner.[11,12] The procedure is efficient in Monte Carlo simulations since the electronic integrals are not required from all other molecular fragments, and it does not pose problems because gradients are not needed. This is the method proposed in the original development of the method for Monte Carlo calculations,[11,12] and it was used a few years later in the fragment molecular orbital model of Kitaura and co-workers.[90] For molecular dynamics simulations, a fully variational Fock operator for each monomer was used in which the external potential consists of contributions both from the DPPC charges and the explicit electron densities of all other fragments.[14,35] Here, analytic gradients can be directly obtained from the optimized X-Pol wave function. In molecular dynamics simulations, the criteria for energy and density conversion were set as 10^{-9} eV for energy and 10^{-6} for density matrix elements. The average energy difference from the two approaches in Monte Carlo and molecular dynamics is less than 1.5% in the computed heat of vaporization.

The Monte Carlo simulations were performed using the MCSOL program for X-Pol simulations,[91] while molecular dynamics simulations were carried out using a newly developed X-Pol program[92] written in C++ which has been interfaced both with CHARMM[93] and NAMD.[94] All *ab initio* electronic structure calculations were performed using GAUSSIAN 09.[95] All calculations were run on a constellation of clusters at the Minnesota Supercomputing Institute.

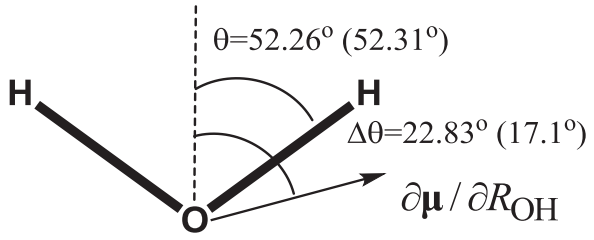
4.4. Results and Discussion

4.4.1. Gas-Phase Properties

Properties for the optimized water monomer and dimer using the PMOw and XP3P models are listed in Table 4-3 along with experimental data and the results from two empirical polarizable potentials that have been examined by Ren and Ponder.[10] The PMOw parameters were optimized against experimental or high-level *ab initio* data for a series of small molecules containing hydrogen and oxygen atoms (Table A-1 in Appendix A), including the properties listed in Table 4-3. In particular, the computed atomization energy (233.0 kcal/mol) and dipole moment (1.88 D) for water from PMOw agree with the corresponding experimental data that have been summarized in Ref. 49 (232.6 kcal/mol and 1.85 D, respectively). The Mulliken population charges from the PMOw wave function and the DPPC charges used in the XP3P potential are also listed in Table 4-3; the latter yields exactly the same molecular dipole moment as that from the QM calculation. An important quantity critical to describing hydrogen-bonding interactions is the molecular polarizability, which also shows good agreement with experiment (a deviation of 14%). This represents a major improvement over all previous NDDO-based models, which typically have errors more than 60% for water. Nevertheless, a question arises on whether or not the somewhat smaller polarizability would affect liquid properties. To address this issue, it is interesting to consider polarizable potential functions for water, in which the experimental gas-

phase electrostatic properties are not always enforced. This is illustrated by the use of smaller molecular polarizabilities in these empirical force fields, and this was justified as to reflect the relatively larger electric field than the mean field of the bulk due to the highly inhomogeneous environment in the first solvation shell;[96] for example, polarizabilities are set to 1.41, 1.29, and 0.98 $Å^3$ in the AMOEBA,[10] POL5/TZ[10,82] and SWM4-NDP[26] models, respectively, all of which yield similar heats of vaporization and similar densities of liquid water at ambient conditions.

The optimized bond length and bond angle for water are 0.9552 Å and 104.61° using PMOw; these values are in excellent agreement with the experimental values of 0.9572 Å and 104.54°.[97] Thus, either the optimized or the experimental monomer geometry can be used in the XP3P potential for liquid simulations discussed below. The change of the molecular dipole moment with geometry variation for the water monomer has an intriguing nonlinear dependence, which is not correctly reproduced in nearly all polarizable and non-polarizable potentials for water, except the TTM2-F model[98] that was specifically fitted with a function to reproduce an accurate *ab initio* dipole moment surface.[99] This is illustrated in Scheme 4.1, which shows that the dipole derivative with respect to an O–H stretch, $\partial\boldsymbol{\mu}/\partial R_{OH}$, lies significantly outside of the two O–H bonds of water. An angle of $\Delta\theta = 22.8°$ was obtained based on the vibrational absorption intensities.[98,100,101] For comparison, the present PMOw model yields a value of $\Delta\theta = 17.1°$, in reasonable agreement with experiment. This is encouraging since this information was not included in the PMOw parametrization process; it is purely a result of the qualitatively correct treatment of chemical bonding interactions in the present quantum mechanical model.



Scheme 4-1. Illustration of the angle between the molecular dipole moment derivative and the O-H bond vector in water monomer. Experimental values are given first, followed by the PMOw results in parenthese.

The potential energy profile for the water dimer along the O–O separation is illustrated in Figure 4.1, and the computed binding energies from PMOw and the XP3P potential are −5.1 and −5.2 kcal/mol, respectively, slightly greater than the best estimate of −5.0 kcal/mol from *ab initio* calculations using MP2/(CBS) + ΔCCSD(T) with the 6-311++G(d,p) basis set,[66] but somewhat smaller than an estimated value (−5.4 kcal/mol) based on measured molecular vibrations.[102] For comparison, both the POL5/TZ[82] model and the AMOEBA model yield a binding energy of −5.0 kcal/mol.[10] The equilibrium structures optimized using the full PMOw Hamiltonian and the fragmental XP3P potential are listed in Table 4-3. The O–O distances from the PMOw and XP3P models agree well with those from POL5/TZ and AMOEBA, which yield 2.89 Å and with the *ab initio* value of 2.91 Å.[66] Ren and Ponder found that the flap angle θ (the flap angle is defined as the angle between the C_2 axis of the hydrogen bond acceptor monomer and the O–O distance vector, depicted in the inset of Figure 4.1) is dependent on the monomer quadrupole moment, and that it was necessary to use explicit quadrupole terms in the AMOEBA model to yield a flap angle in agreement with the *ab initio* results. The results on the flap angle in the water dimer from the PMOw and XP3P models are also good, and the small tilt angle, α, from the hydrogen bond donor is also predicted. However, the large flap angle is not preserved in the

XP3P model. The structures and energies on other stationary points of water dimer are given in Table A-2 in Appendix A.

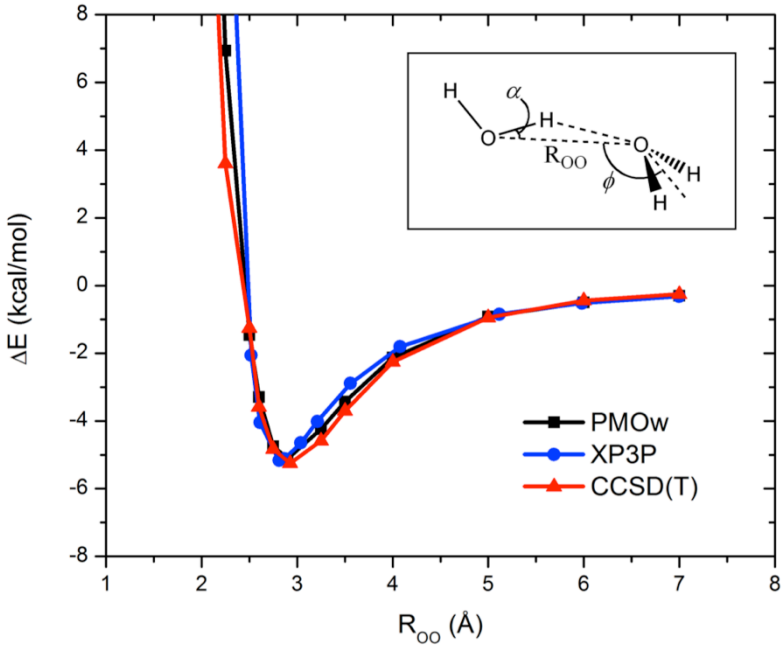


Figure 4-1. Potential energy profiles for a water dimer at the hydrogen bonding configuration from the PMOw (black) and the XP3P (blue) models for water along with CCSD(T) results (red). Definition of the geometrical parameters listed in Table 4-3 are given in the structure shown as inset in the upper right-hand corner. The CCSD(T) results are obtained with the aug-cc-pVDZ basis set on fully optimized geometries at various fixed O–O distances. Studies have shown that extrapolation to the complete basis set limit from the current size does not affect the computed energies by more than 0.2 kcal/mol.[103] All other geometric parameters are optimized.

Small water clusters (Figures A-1 and A-2 in Appendix A), including the cyclic configurations of the trimer, tetramer, and pentamer, four configurations of the hexamer, and the cubic D_{2h} arrangement of the octamer have been examined (Table 4.4). All clusters were fully optimized with PMOw using the conjugated gradient method with NAMD.[88,90] A configuration

was considered optimized when its gradient norm fell below 0.5 kcal mol^{-1} $Å^{-1}$. The best theoretical estimates for these systems are from the work of Bryantsev et al., who performed single-point MP2/(CBS) along with a CCSD(T) correction (simply called CCSD(T) results in this discussion) at the B3LYP/6-311++G(2*d*,2*p*) optimized structures.[66] As in the work of Ren and Ponder,[10] we list in Table 4.4 the total binding energies, the average O–O distances (R_{OO}), average O…H–O hydrogen bond angles ($\langle\phi\rangle$), and the total (μ) and average monomer ($\langle\mu_{mol}\rangle$) dipole moments. Of all water clusters, the average monomer dipole moments from the POL5/TZ and AMOEBA models[10] fall between the values computed using the PMOw and the XP3P method, and the trends are in accord with that estimated by Gregory et al.[104] using a portioning scheme for the electron density. Overall, the computed binding energies rom PMOw and XP3P methods are in good agreement with the CCSD(T) results, with root-mean-square (RMS) deviations of 1.2 and 2.4 kcal/mol, respectively. The performance of the AMOEBA force field is excellent, whereas the POL5/TZ model slightly underestimates the binding energies.[10,82] For the hexamers, the ordering of relative stability is cage > book > prism > cyclic from CCSD(T), and cage = prism > book > cyclic from PMOw. For comparison, the ordering from the MP2/CBS + ΔCCSD(T) calculations with 6-311++G(d,p) basis[66] and AMOEBA optimizations is prism > cage > book > cyclic.[10] In any event, the three non-cyclic structures of the water hexamer are energetically similar in binding, whereas the cyclic configuration is noticeably less stable than the other three.

Table 4-4. Computed and Experimental Properties for Water Clusters

		PMOw	XP3P	POL5/TZ[a]	AMOEBA[a]	*Ab initio*[b]	Expt.[a]
Trimer cyclic	ΔE_b (kcal/mol)	-14.8	-15.7	-13.4	-15.3	-15.8	
	$\langle R_{OO} \rangle$ (Å)	2.87	2.77	2.90	2.81	2.81	2.845
	$\langle \phi \rangle$ (deg)	105.1	125.6		151.5	110.4	152
	$\langle \mu_{mol} \rangle$ (D)	2.14	2.46	2.22	2.29	2.3	
	μ (D)	1.19	0.01	1.21	1.09	1.07	
Tetramer cyclic	ΔE_b (kcal/mol)	-27.5	-28.9	-25.5	-27.7	-27.4	
	$\langle R_{OO} \rangle$ (Å)	2.74	2.68	2.769	2.76	2.75	2.79
	$\langle \phi \rangle$ (deg)	116.5	145.9		168.0	121.6	
	$\langle \mu_{mol} \rangle$ (D)	2.22	2.71	2.47	2.55	2.6	
	μ (D)	0.0	0.0	0.0	0.0	0.0	

Continued on next page

Table 4-4. Computed and Experimental Properties for Water Clusters – *Continued*

		PMOw	XP3P	POL5/TZ[a]	AMOEBA[a]	*Ab initio*[b]	Expt.[a]
Pentamer cyclic	ΔE_b (kcal/mol)	-35.7	-39.7	-34.1	-36.5	-35.9	
	$\langle R_{OO} \rangle$ (Å)	2.73	2.66	2.74	2.76	2.73	2.76
	$\langle \phi \rangle$ (deg)	126	159		176	132	
	$\langle \mu_{mol} \rangle$ (D)	2.26	2.82	2.57	2.64	2.7	
	μ (D)	1.17	0.02	1.19	0.92	0.93	
Hexamer cyclic	ΔE_b (kcal/mol)	-43.3	-49.0	-41.8	-44.8	-44.3	
	$\langle R_{OO} \rangle$ (Å)	2.72	2.65	2.74	2.75	2.72	2.76
	$\langle \phi \rangle$ (deg)	130	167		179	139	
	$\langle \mu_{mol} \rangle$ (D)	2.28	2.86	2.62	2.70	2.7	
	μ (D)	0.0	0.0	0.02	0.0		

Continued on next page

Table 4-4. Computed and Experimental Properties for Water Clusters – *Continued*

		PMOw	XP3P	POL5/TZ[a]	AMOEBA[a]	*Ab initio*[b]	Expt.[a]
Hexamer prism	ΔE_b (kcal/mol)	-47.8	-44.4	-41.9	-45.9	-45.3	
	$\langle R_{OO} \rangle$ (Å)	2.84	2.76	2.79	2.80	2.86	
	$\langle \phi \rangle$ (deg)	121.0	128.7			123.1	
	$\langle \mu_{mol} \rangle$ (D)	2.24	2.72	2.52	2.60		
	μ (D)	2.40	3.29	2.91	2.57	2.70	
Hexamer cage	ΔE_b (kcal/mol)	-47.8	-45.2	-41.8	-45.9	-46.0	
	$\langle R_{OO} \rangle$ (Å)	2.80	2.76	2.78	2.80	2.83	2.82
	$\langle \phi \rangle$ (deg)	118	126			121	
	$\langle \mu_{mol} \rangle$ (D)	2.22	2.72	2.49	2.58	2.6	
	μ (D)	2.05	2.01	2.44	2.16	1.90	1.82 – 2.07[c]

Continued on next page

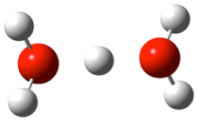
Table 4-4. Computed and Experimental Properties for Water Clusters – *Continued*

		PMOw	XP3P	POL5/TZ[a]	AMOEBA[a]	*Ab initio*[b]	Expt.[a]
Hexamer book	ΔE_b (kcal/mol)	-46.2	-48.3	-42.5	-45.8	-45.8	
	$\langle R_{OO} \rangle$ (Å)	2.75	2.70	2.79	2.78	2.78	
	$\langle \phi \rangle$ (deg)	121	144			127	
	$\langle \mu_{mol} \rangle$ (D)	2.24	2.79	2.55	2.63		
	μ (D)	2.40	2.22	2.45	2.29		
Octomer	ΔE_b (kcal/mol)	-77.7	-69.5			-72.6[d]	
	$\langle R_{OO} \rangle$ (Å)	2.74	2.72			2.81	
	$\langle \phi \rangle$ (deg)	163	164			163	
	$\langle \mu_{mol} \rangle$ (D)	2.20	2.86				
	μ (D)	0.0	0.0			0.0	

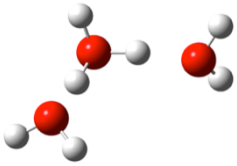
[a]Ref. 10. [b]Refs. 66 and 104. [c]Ref. 104. [d]MP2/(CBS) limit, see Ref. 101.

We have also examined several configurations of micro-solvated proton $H^+(H_2O)_n$, where n = 2, 3, 4, and 6 (Scheme 4.2). Depicted in Figure 4.2 are the potential energy profile for a proton migration between two water molecules at fixed O–O distances of the global minimum R_{min} (OO), R_{min} (OO) + 0.2 Å, and R_{min} (OO) + 0.4 Å from PMOw, and MP2/aug-cc-pVDZ optimizations. The equilibrium structure has an R_{min} (OO) separation of 2.46 and 2.40 Å, respectively, from these theoretical models. With a basis set comparable to aug-cc-pVDZ, the MP2 results on these proton clusters are very close to CCSD(T)-F12 results with jun-cc-pVTZ basis.[105] The PMOw O–O distance is about 0.05 Å longer than the MP2 result. In all cases, the proton is essentially symmetrically located between the two water molecules (Figure 4.2(a)). A small barrier appears when the O–O distance is stretched by 0.2 Å. The PMOw model yields a barrier of 1.9 kcal/mol, compared to 1.9 kcal/mol from MP2. Further stretching the O–O distance to R_{min} (OO) + 0.4 Å increases the barrier heights to 7.9, and 7.5 kcal/mol, respectively. There are numerous studies of proton-water clusters and proton transfer barriers with a variety of computational methods;[105–108] a thorough comparison with earlier studies is beyond the scope of the present work.

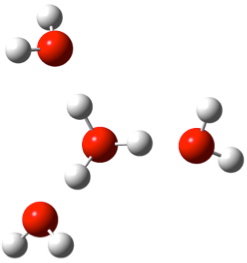
The binding energies between additional water molecules and $H_5O_2^+$ are listed in Table 4.5, along with the MP2/aug-cc-pVDZ results. Overall, the agreement is good, with a mean-signed deviation of 1.6 kcal/mol. Note that unconstrained optimization of the structure $H^+(H_2O)_6$ (IV) using PMOw collapses to isomer (III). Thus, the value in Table 4.5 was obtained by fixing the relative torsion angles of the hydrogen atoms of the central $H_5O_2^+$ unit to the MP2 values. Overall, the results from the PMOw model are in good accord with MP2 calculations and other theoretical models.



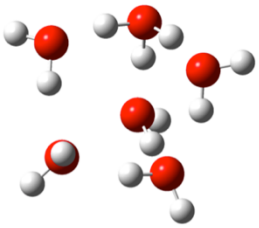
$H^+(H_2O)_2$



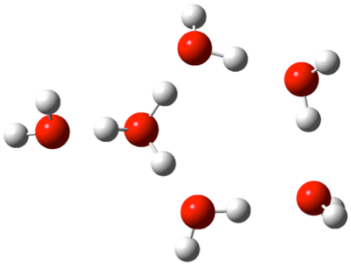
$H^+(H_2O)_3$



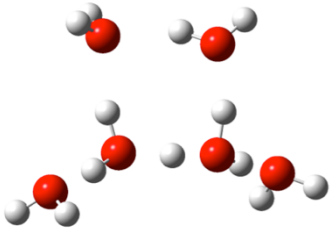
$H^+(H_2O)_4$



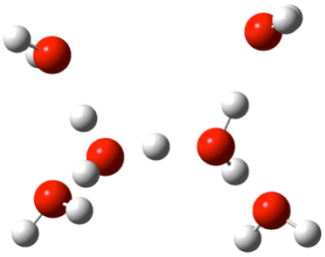
$H^+(H_2O)_6$ (I)



H+$(H_2O)_6$ (II)



$H^+(H_2O)_6$ (III)



$H^+(H_2O)_6$ (IV)

Scheme 4-2. Optimized geometries of $H^+(H_2O)_n$ clusters from PMOw.

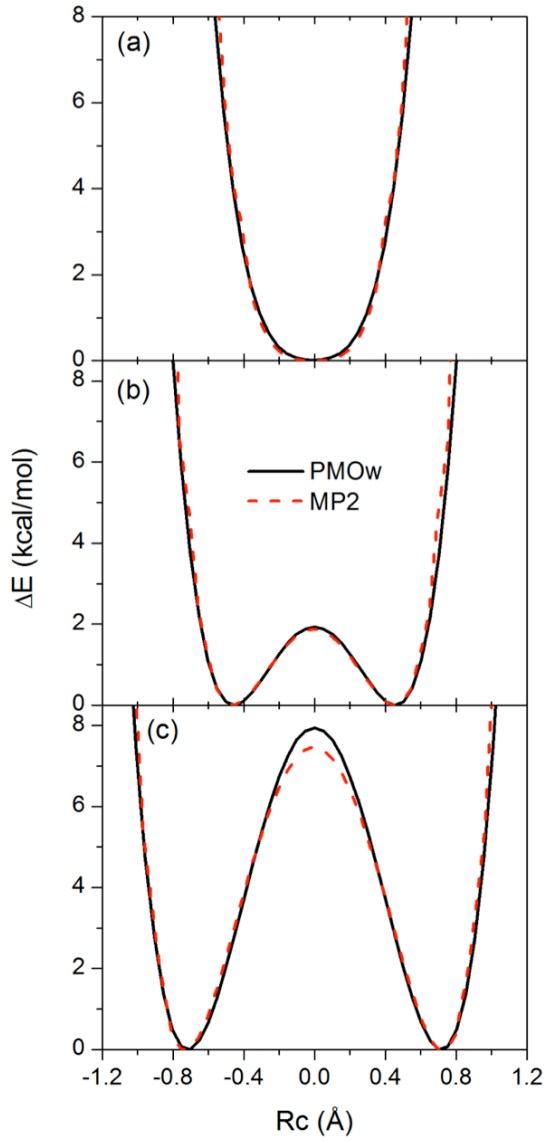


Figure 4-2. Potential energy profile for $H^+(H_2O)_2$ in the gas phase as a function of the proton transfer coordinate, defined as the distance from the mid-point between the two oxygen atoms, (a) at the minimum geometry, R_{min}(OO), (b) at a fixed O-O separation of R_{min}(OO) + 0.2 Å, and (c) at a fixed O-O distance of R_{min}(OO) + 0.4 Å from PMOw (black), and MP2/aug-cc-pVDZ (red, dashed) calculations. Geometries were optimized with fixed O-O distances.

Table 4-5. Computed Interaction Energies for $H^+(H_2O)_n$ complexes from the PMOw and MP2 methods[a]

Complex	PMOw	MP2
$H_5O_2^+ \ldots H_2O$	-21.4	-23.8
$H_5O_2^+ \ldots (H_2O)2$	-39.8	-43.8
$H_5O_2^+ \ldots (H_2O)4$ (Isomer I)	-68.9	-71.8
$H_5O_2^+ \ldots (H_2O)4$ (Isomer II)	-67.3	-71.8
$H_5O_2^+ \ldots (H_2O)4$ (Isomer III)	-66.6	-71.0
$H_5O_2^+ \ldots (H_2O)4$ (Isomer IV)	-60.5	-69.7

[a]Interaction energies are calculated by $\Delta E = E(\text{cluster}) - \left\{E\left(H_5O_2^+\right) + nE\left(H_2O\right)\right\}$, where n is the number of water molecules.

4.4.2. Liquid Properties

4.4.2.1. Properties at 25 °C

The computed and experimental thermodynamic and dynamic properties of liquid water at 25 °C and 1 atm are listed in Table 4.6, along with the results from TIP3P,[9,83,109] AMOEBA,[10] and SWM4-NDP.[26] The standard errors (±1σ) were obtained from fluctuations of separate averages over blocks of (2–4) × 10^5 configurations. A correction, by integrating the Lennard-Jones potential beyond the cutoff distance, for the Lennard-Jones potential neglected by the cutoff has been included, and this contributes to the total computed heat of vaporization by about 1%. Long-range electrostatic interactions were not corrected in the Monte Carlo simulations. Previous studies using empirical force fields indicate that there is little size dependency of the computed properties for liquid water, and these effects will be investigated in a future study. (The TIP3P and TIP4P potential functions were developed with 125 water molecules with a cutoff of 7.5 Å without long-range corrections.[9,83,109,110])

Table 4-6. Liquid Properties of the XP3P Model for Water along with Those from Experiments, and the TIP3P, AMOEBA, and SWM4-NDP Models

	XP3P	TIP3P[a]	AMOEBA[b]	SWM4-NDP[c]	Expt.[d]
$E(l)$ (kcal/mol)	-9.83 ± 0.01[e]	-9.82	-9.89	-9.92	-9.98
ΔH_{vap} (kcal/mol)	10.42 ± 0.01[e]	10.35	10.42	10.45	10.51
ρ (g/cm^3)	0.996 ± 0.001	1.002	1.000	1.000	0.997
C_p (cal mol^{-1} K^{-1})	21.8 ± 1.0	20.0	20.9		18.0
$10^6\ \kappa$ (atm^{-1})	25 ± 2	60			46
$10^5\ \alpha$ (K^{-1})	37 ± 3	75			26
μ_{gas} (D)	1.88	2.31	1.77	1.85	1.85
μ_{liq} (D)	2.524 ± 0.002	2.31	2.78	2.33	2.3-2.6
$10^5\ D$ (cm^2/s)	2.7	5.1	2.02	2.3	2.3
ε	97	92	82	79 ± 3	78
t_D (ps)	8.8			11 ± 2	8.3
t_{NMR} (ps)	2.6			1.87 ± 0.03	2.1

[a]Refs. 83 and 109. [b]Ref. 10. [c]Ref. 26. [d]See text for details. [e]The average $E(l)$ from molecular dynamics simulations employing the variational Fock operator is -9.99 kcal/mol over 400 ps. This gives a heat of vaporization of 10.52 kcal/mol.

The average density of XP3P is 0.996 ± 0.001 g/cm^3 at 25 °C, which is within 1% of the experimental value and is similar to results obtained with other polarizable and non-polarizable force fields (Table 4.6).[10,26,83,109] The total energy per monomer of liquid water, $E_i(l)$, is related to the heat of vaporization by

$$\Delta H_{\text{vap}} = -E_i(l) + P\left(V_{\text{gas}} - V_{\text{liq}}\right) + \Delta Q - \left(H^0 - H\right) \tag{4.14}$$

where V_{gas} and V_{liq} are the molar volumes of water in the gas phase (ideal) and in the liquid, P is the pressure, ΔQ is the quantum corrections to inter and intramolecular degrees of freedom between the gas and liquid, and the last term, $(H^o - H)$, is the enthalpy departure function.[111] Although ΔQ and $(H^o - H)$ has been tabulated and can be explicitly included[9,12,110] and this amount to a total correction of −0.06 kcal/mol at 25 °C, they have typically been neglected.[10,26,83] In this case, ΔH_{vap} is simply approximated by $-E_i(l) + RT$, which is also adopted in the present study (Table 4.6). The calculated heat of vaporization from the XP3P model is 10.42 ± 0.01 kcal/mol using the non-variational approximation in Monte Carlo simulations,[11,12] and the value is increased to 10.58 kcal/mol using the variational Fock operator in molecular dynamics.[14,35,92] The variational X-Pol approach used in molecular dynamics simulations lowers the interaction energy of the liquid by about 1.5% relative to the non-variational approach used in Monte Carlo. Overall, the agreement with experiment[112,113] is good, although there is greater deviation in the non-variational approach. The quality of the XP3P quantum mechanical potential for these two critical thermodynamic properties is comparable to that of the widely used SPC, TIP3P and TIP4P models for water[8,9] and to that of the recent polarizable models.[10,26,82,114]

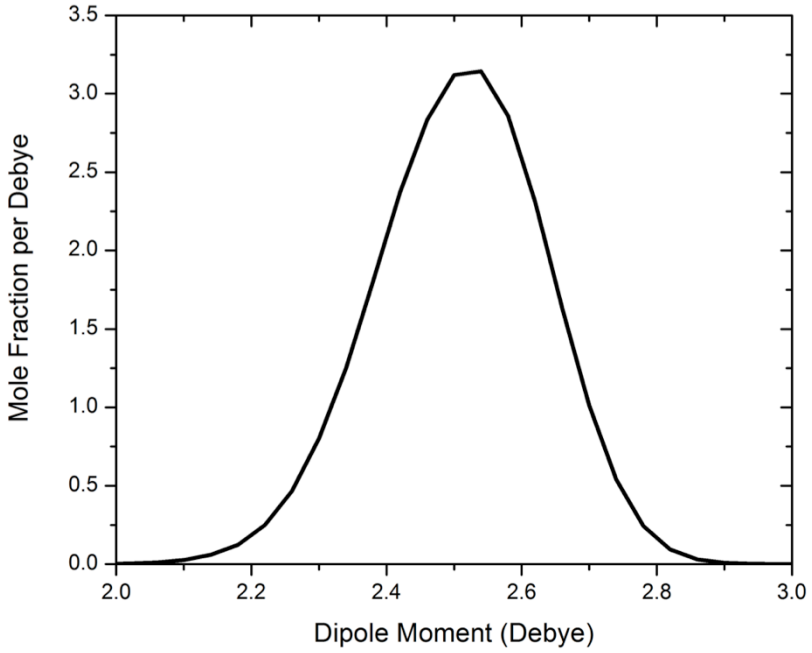


Figure 4-3. Distribution of the scalar molecular dipole moment in liquid water from Monte Carlo simulations with the XP3P potential at 25 °C and 1 atm. The units for the ordinate are mole percent per Debye.

The distribution of the magnitudes of monomer dipole moments from polarized wave functions in the liquid is shown in Figure 4.3; these dipole moments span a range from 2.1 to 2.9 D, and they yield an average $\langle \mu_{\mathrm{liq}} \rangle$ of 2.524 ± 0.002 D. The width at half maximum in the dipole distribution is 0.30 D (a half-width of 0.8 D was reported for the AMOEBA model,[10] which seems to be unrealistically large). Clearly, there is a major enhancement of the molecular dipole moment in the liquid, amounting to an increase over 35% relative to the gas phase value. For comparison, the AMOEBA model produced a much greater average, 2.78 D, or 50% greater than its gas phase value. The SWM4-NDP model yielded an average of 2.46 D,[26] similar to the present XP3P quantum mechanical model. Our previous investigation, employing the AM1 Hamiltonian to represent water monomers in X-Pol, resulted in an average dipole moment of 2.29 D;[12]

however, the smaller value is partly due to the much smaller molecular polarizability from AM1, and the weak polarization effect was corrected by scaling Mulliken population charges in that study. There is no experimental value for the dipole moment of liquid water (and in fact this quantity is not well defined), but values ranging from 2.3 to 2.6 D have been cited based on an estimate for ice Ih.[115,116] Finally, we note that *ab initio* molecular dynamics simulations yielded dipole moments ranging from 2.3 D to 3.8 D, depending on the method and functional used in DFT.[117] *Ab initio* molecular dynamics simulations seem to produce greater average dipole moments than polarizable force fields and the present XP3P model.

The dielectric constant of the liquid is related to the fluctuations of the total dipole moment of the simulation box and it is dependent on the boundary conditions used to treat long-range electrostatics.[118,119] We employed the reaction field approximation in the NVT ensemble at 25 °C and experimental density, where intermolecular interactions are truncated at R_{cut} = 9.0 Å. Under these conditions, a reaction field contribution is added to the electrostatic potential in Eq. 4.6:[118,120]

$$V_x^{RF}\left(\rho_b\right) = V_x\left(\rho_b\right)\left\{1 + \frac{2\left(\varepsilon_{RF}-1\right)}{2\varepsilon_{RF}+1}\left(\frac{\left|\mathbf{r}_x - \mathbf{R}_B^b\right|}{R_{cut}}\right)^3\right\} \tag{4.15}$$

where ε_{RF} is the dielectric constant of the continuum. The static dielectric constant ε is determined from Eq. 4.16.[120–122]

$$\frac{\left(\varepsilon-1\right)\left(2\varepsilon_{RF}+1\right)}{2\varepsilon_{RF}+\varepsilon} = \frac{4\pi}{3k_BT}\frac{\left\langle \mathbf{M}^2\right\rangle}{\left\langle V\right\rangle} \tag{4.16}$$

where **M** is the total dipole moment of the simulation box and $\langle V \rangle$ is the average volume per monomer. Ideally the reaction field dielectric ε_{RF} should be the same as that of the liquid in the cutoff sphere, although previous studies suggest that a choice of ε_{RF} in the range of $\varepsilon \le \varepsilon_{RF} \le \infty$ typically yields consistent results,[123] and a value of 160 has been used in the present study. The liquid dipole fluctuation converges slowly, and we have carried out 16 separate simulations, each

lasting about 15×10^6 configurations at 25 °C. An average value of 97 ± 8 was obtained by removing the two highest and two lowest values; the present average is greater than the experimental value of 78. Interestingly, Sprik argued that an average dipole moment of 2.5 – 2.6 D in liquid water would lead to the correct dielectric constant at room temperature,[124] and a similar observation was used in the parameter optimization process by Lamoureux et al.[96] In view of the average dipole moment from the XP3P liquid, which falls in the middle of this range, it is likely that a better agreement with experiment could be obtained if the simulations were further converged by extending the simulation to 100×10^6 configurations or more in each simulation. It is interesting to note that Ren and Ponder obtained a static dielectric constant of 82, in spite of a significantly larger dipole moment of 2.78 D of the liquid from the AMOEBA potential.[10] In that work, the authors argued that the correct average H–O–H angle was responsible for the good agreement between experimental and calculated liquid dielectric constant.[10,125,126]

Displayed in Figure 4.4 are the distributions of the binding energies per monomer in liquid water at a temperature range of −40 °C to 100 °C. The binding energies in Figure 4.4 correspond to the interaction energy of one monomer with the rest of the system. In a polarizable model, the total energy of the liquid also includes the energy cost needed to polarize the electronic wave function (also called self-energy, see below). Thus, in contrast to the use of a pairwise potential, the average energy, $E_i(l)$, per monomer in Table 4.6 is not exactly equal to half of the binding energy at 25 °C from Figure 4.4, but it is smaller by the amount of the self-energy. This is a reflection of the non-additive nature of a polarizable force field.[127] Note that such a self-energy term has been used to develop the SPC/E model.[128]

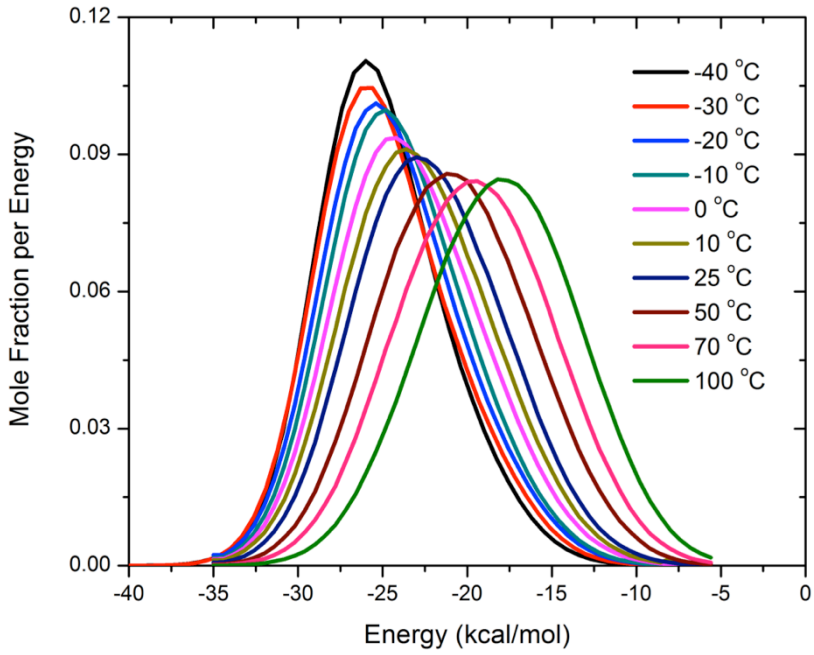


Figure 4-4. Distribution of the binding energies of water in the liquid at temperatures ranging from -40 °C to 100 °C. The binding energy corresponds to the total interaction energy of one water with the rest of the bulk solvent.

We have estimated several thermodynamic properties involving molecular fluctuations. The intermolecular contribution to the isobaric heat capacity C_p of water is defined below and can also be computed from the enthalpy fluctuations by

$$C_P = \left(\frac{\partial \langle H_i(l) \rangle}{\partial T} \right)_P + 3R = \frac{\langle H_i(l)^2 \rangle - \langle H_i(l) \rangle^2}{RT^2} + 3R \qquad (4.17)$$

where $H_i(l) = E_i(l) + PV_{liq}$ is the average enthalpy of the system per monomer. The total heat capacity of the liquid C_P for a rigid monomer model is determined by adding the classical kinetic energy contributions from translation and rotation of a water molecule ($3R$).[83] The average from the fluctuation formula in Eq. 4.17 is 22 ± 1 cal mol^{-1} K^{-1}, which is greater than the experimental value at 25 °C.[129,130] Path integral simulations by Vega et al. showed that inclusion of nuclear

quantum effects lowers the computed heat capacity by up to 6 cal/mol.[131] Quantities based on the fluctuation formula, including C_P (isobaric heat capacity), α (coefficient of thermal expansion), and κ (isothermal compressibility) are difficult to converge; they can also be estimated from the numerical derivatives of their definitions. The derivative estimate from liquid enthalpies vs. T yields a C_p of 19 cal mol^{-1} K^{-1} at 25 °C. The coefficient of thermal expansion (α) and the isothermal compressibility (κ) are determined from fluctuations of volume and enthalpy, with a computed value of 37×10^{-5} K^{-1} for α and 25×10^{-6} atm^{-1} for κ, respectively. These quantities show relatively large deviations from experiment ($\alpha = 25.6 \times 10^{-5}$ K^{-1} and $\kappa = 45.8 \times 10^{-6}$ atm^{-1})[129] due to their convergence.

The self-diffusion coefficient of liquid water was determined using the Einstein formula[122] from molecular dynamics simulations with constant volume and temperature:

$$D = \lim_{t\to\infty} \frac{1}{6t} \left\langle \left| \mathbf{r}(t) - \mathbf{r}(0) \right|^2 \right\rangle \tag{4.18}$$

where $\mathbf{r}(t)$ is the position of the oxygen atom of water at time t. The diffusion coefficient was obtained as the slope from a linear fit of $\left\langle \left| \mathbf{r}(t) - \mathbf{r}(0) \right| \right\rangle / 6$ as a function of t, and we obtained a value of 2.7×10^{-5} cm^2 s^{-1}, which agrees with experiment.[132] It is known that non-polarizable potentials for water, such as SPC, TIP3P, and TIP4P, tend to overestimate the self-diffusion coefficient, while most polarizable force fields, including the present XP3P model, show significant improvement.[10,26,82,114] The computed diffusion coefficient is also affected by finite size of the simulation box, and extrapolation to infinity will further increase the value of the diffusion coefficient.[133]

The rotational correlation times, τ_2^α, of water with respect to the H–H and O–H axes are obtained from least-square fits of the orientational time-correlation function to a single exponential function, $C_2^\alpha(t) = Ae^{-t/\tau_2^\alpha}$, where α specifies the rotation axis. The orientation time-correlation function is defined as follows:[122]

$$C_2^{\alpha}(t) = \left\langle P_2\left[\mathbf{u}_i^{\alpha}(t)\mathbf{u}_i^{\alpha}(0)\right]\right\rangle \tag{4.19}$$

where P_2 is the second-order Legendre polynomial, and $\mathbf{u}_i^{\alpha}(t)$ is the unit vector along the α rotation axis of molecule i at time t. The time-integral of Eq. (4.19), $A\tau_2^{\mathrm{HH}}$, corresponds to the NMR rotational relaxation time of H_2O, τ_{NMR};[134] the present XP3P model yields a value of 2.6 ps, which may be compared with the experimental value (2.1 ps).[135] For comparison, the SWM4-NDP model predicts a τ_{NMR} value of 1.9 ps.[26] Similarly, the Debye dielectric relaxation time was determined from an exponential fit to the normalized autocorrelation function of the total dipole moment **M** of the system:[122]

$$C_{\mathrm{D}}(t) = \frac{\langle \mathbf{M}(t)\mathbf{M}(0)\rangle}{\langle \mathbf{M}^2(0)\rangle} \tag{4.20}$$

The Debye relaxation time characterizes the relaxation time of the hydrogen-bonding network in the liquid. The XP3P model shows that the Debye relaxation time is about 6% faster than the observed values (8.3 ps).[136] In comparison with other models, the present XP3P model performs well for these dynamic properties.[10,26,114]

The structure of liquid water is characterized by radial distribution functions (RDFs), $g_{xy}(r)$, which gives the probability of finding an atom of type y at a distance r from an atom of type x relative to the bulk. The RDFs computed at 25 °C from Monte Carlo simulations are shown in Figure 5 along with the neutron diffraction data. Overall, the agreement with experimental results is excellent. For the XP3P potential, the location of the maximum of the first peak of the O–O RDF is 2.78 ± 0.05 Å with a peak height of 3.0 (Figure 4.5(a)). For comparison, the corresponding experimental values are 2.73 Å and 2.8 from neutron diffraction[137] and X-ray absorption spectroscopy (XAS),[138] respectively. Integration of the O–O RDF to the first minimum at 3.30 Å yields an estimated coordination number of 4.5, which is in good agreement with the neutron diffraction result of 4.51 (integrated to 3.36 Å), but somewhat smaller than the XAS

result (4.7).[138] The oxygen-hydrogen and hydrogen-hydrogen radial distribution functions are also in accord with experiments.

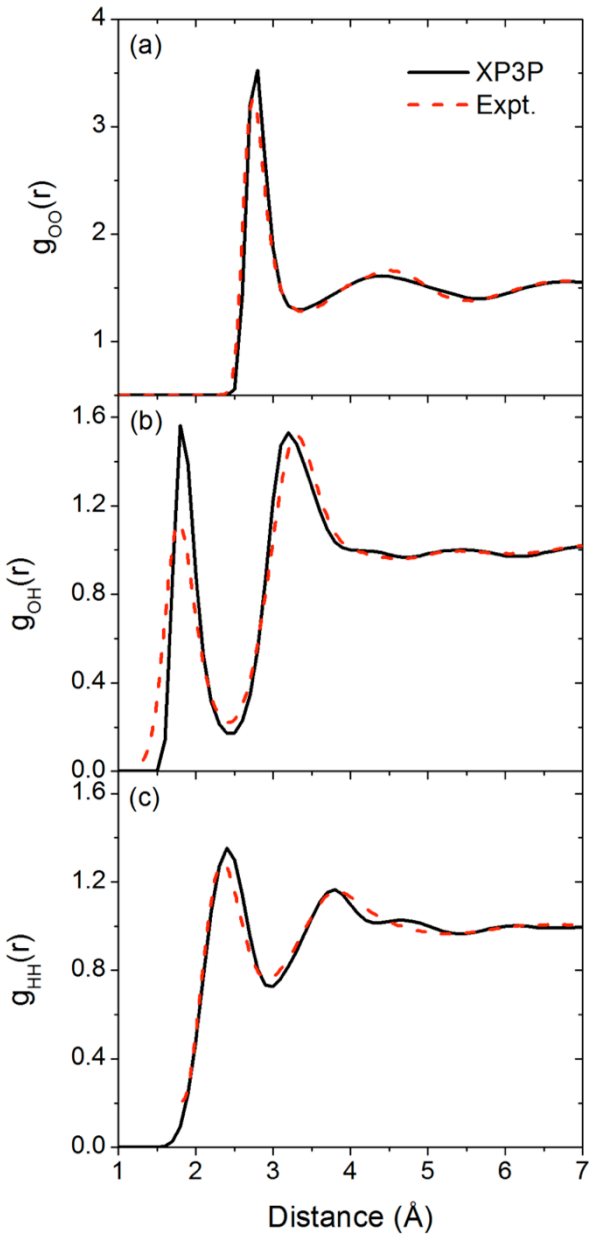


Figure 4-5. Computed (black) and experimental (red, dashed) oxygen-oxygen (a), oxygen-hydrogen (b), and hydrogen-hydrogen (c) radial distribution functions of liquid water at 25 °C and 1 atm.

4.4.2.2. Temperature-Dependent Liquid Properties

The computed liquid properties for ΔH_{vap}, C_P, ρ, α, and κ, at different temperatures ranging from −40 to 100 °C are listed in Table A-5 in Appendix A, and some of these are compared with experimental data in Figures 4.6–4.8. The formulas involving fluctuations of enthalpy and volume for C_P, α, and κ are known to have slow convergence even when Monte Carlo simulations were extended to over hundreds of millions of configurations. In the present simulations, C_P and α can also be determined directly from the enthalpy and volume derivatives with respect to temperature. For the isothermal compressibility, the fluctuation formula was used since the pressure was not changed in the present study.

The heats vaporization from −40 to 100 °C were obtained from the average energies plus RT for the PV term of an ideal gas; here, we have ignored the small corrections for the quantum vibrational energy difference and enthalpy departure function. For comparison, we have included the computed heats of vaporization in Figure 4.6 from the TIP5P model. The XP3P model agrees with the results from TIP5P quantitatively at temperature above 25 °C. Both XP3P and TIP5P overestimate ΔH_{vap} at temperature lower than 25 °C, but the TIP5P model yielded a greater deviation on supercooled water. Figure 4.6 shows that the change in ΔH_{vap} is nearly linear over the entire temperature range considered. This agrees with the experimental results on heat capacity, which is nearly constant at about 20 cal mol^{-1} K^{-1}.[129] The changes of heat capacity with temperature are given in Figure A-6 in Appendix A. The trends are in reasonable agreement with experiment at temperatures above 0 °C, although the sharp increase of C_P below 20 °C is not reproduced by the present simulations.

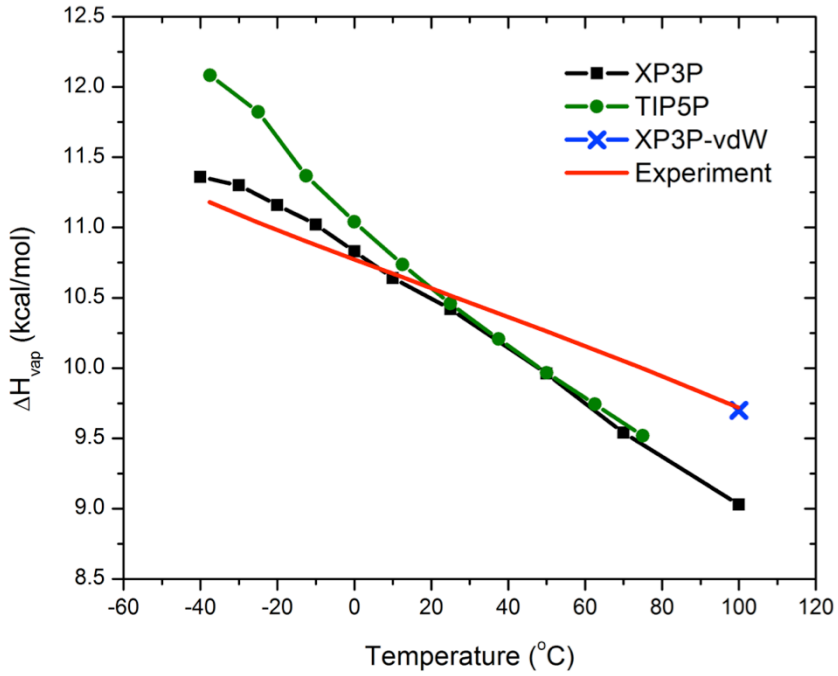


Figure 4-6. Computed (black) and experimental (red) heats of vaporization for liquid water. The results from the TIP5P model are illustrated in green.

The liquid density as a function of temperature is presented in Figure 4.7 along with the experimental density of liquid water. The XP3P model, which is optimized to reproduce the heat of vaporization and density at 25 °C, yields a maximum density at about −20 °C. Although the density maximum is significantly lower than the experimental value at 4 °C,[129] it is in fact remarkable in that there is a density maximum at all from the present model because other three-point-charge models do not possess this property with a reasonable temperature (except the SPC/E with much enhanced electrostatics). The computed density at temperature greater than 25 °C shows more rapid decline with increasing temperatures than experimental results.[129] This trend is similar to that found in the TIP*x*P series of models.[83] The densities for supercooled water are overestimated by 2% – 5% compared with the experimental data.[129] For comparison, among the non-polarizable models that do possess a density maximum, SPC/E[128] has a density maximum

at −38 °C,[139] TIP4P at −15 °C,[109] and TIP5P at about 0 °C; the TIP5P model was optimized to reproduce the temperature dependence of liquid density of water.[83] The AMOEBA model has a density maximum at 17 °C.[126]

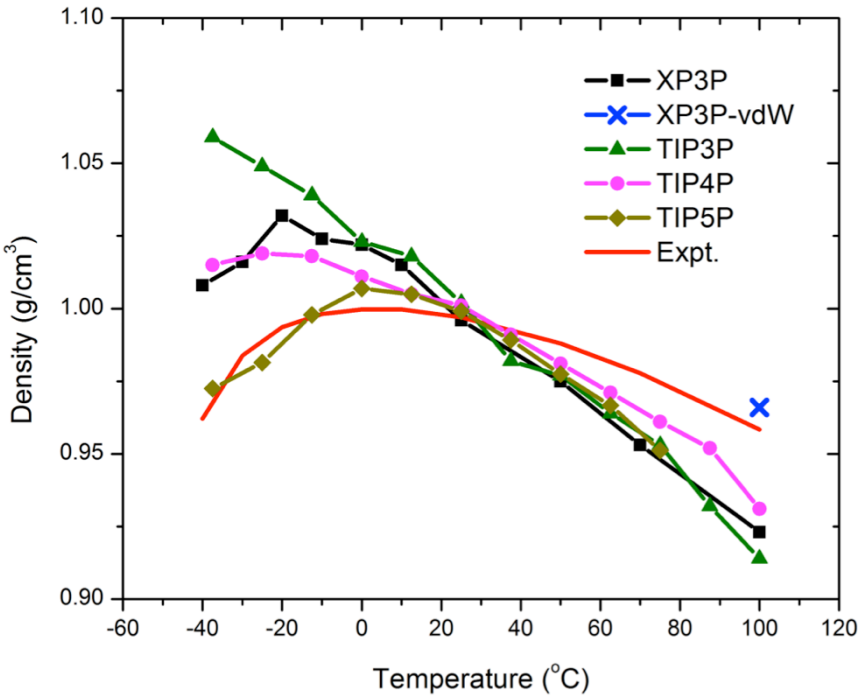


Figure 4-7. Computed (black) and experimental (red) densities for liquid water, along with those from the TIP3P (green), TIP4P (magenta), and TIP5P (dark yellow) models.

The temperature dependences of the computed density and ΔH_{vap} from the non-polarizable TIP*x*P series of models[79,109] and the polarizable AMOEBA potential[126] indicate that it is difficult, with fixed empirical parameters, to obtain good agreement (within 1%) with experiment for the entire temperature range from the supercooled liquid to the boiling point. This difficulty has been pointed out by Siepmann and co-workers, who used a charge-dependent van der Waals radius for oxygen in a fluctuating charge model for water.[140] Giese and York[32] developed a density-dependent van der Waals potential that can be directly incorporated into QM/MM style

simulations. We have further optimized σ_O at 100 °C to yield a better agreement with the experimental liquid density ρ. We found that a small change in σ_O from 3.225 to 3.205 Å is sufficient to produce a liquid density (0.962 g/cm^3) in good agreement with experiment (0.958 g/cm^3). This is shown by the blue cross point in Figure 4.7. Interestingly, the computed ΔH_{vap} (9.70 kcal/mol) was also found to be in excellent agreement with experiment (9.72 kcal/mol)[112] after this small adjustment (blue cross point in Figure 4.6). With this change, the average dipole moment is computed to be 2.470 ± 0.001 D, representing an increase of 0.042 D from 2.428 D computed with the original Lennard-Jones parameters in Table 2.2.

In view of the small change in the σ_O value, we suggest a simple temperature-dependent relationship for σ_O,

$$\sigma_O(T) = 3.225 - 2.667 \times 10^{-4}(T - 298.15) \tag{4.21}$$

in Å^3 where T is the absolute temperature. Alternatively, Eq. 4.21 may be rewritten in terms molecular dipole moment, which translates the expression to an aesthetically appealing, density-dependent one. In any event, it is straightforward to use Eq. 4.21 in Monte Carlo simulations, while it can be conveniently incorporated into a thermostat algorithm in molecular dynamics simulations.[86,141,142] However, a thorough examination of the performance of temperature-dependent van der Waals parameters is beyond the scope of the present work.

The computed coefficient of thermal expansion, α, follows the experimental trends nicely in Figure 4.8, and the negative values for supercooled water are consistent with the experimental values as a result of the existence of a density maximum vs. temperature.

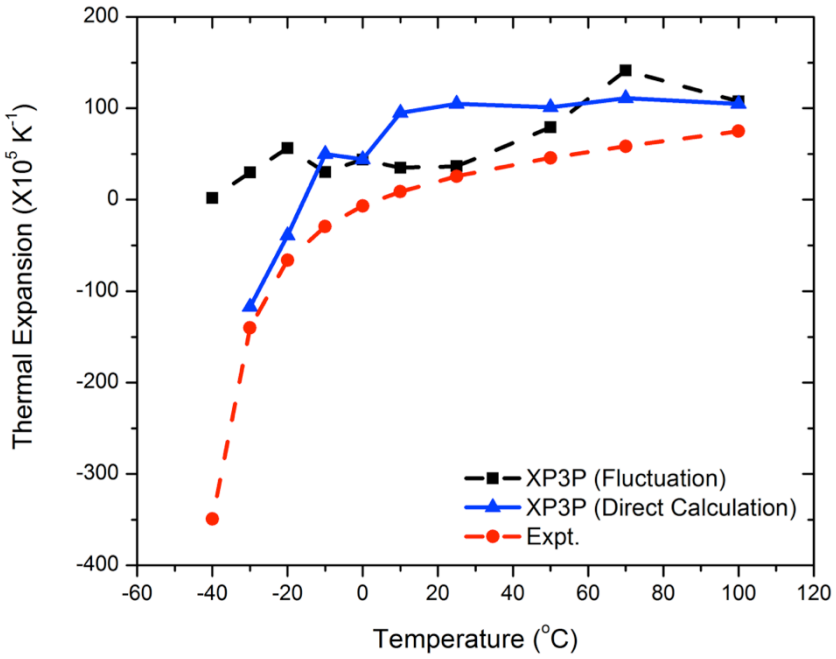


Figure 4-8. Computed (black) and experimental (red) coefficients of thermal expansion (α) for liquid water. The α values are also determined from numerical derivatives of liquid volume variations with temperature (blue).

The average dipole moment from the XP3P model decreases monotonically with increasing temperature (inset of Figure 4.9). The distributions of scalar dipole moment in the liquid at different temperatures are given in Figure 4.9. Consistent with Figure 4.4, the maximum positions are shifted towards smaller values as temperature increases, and this shift is accompanied by an increase in half width from about 0.26 D to about 0.32 D. The broader distribution of molecular dipole moment in liquid water at higher temperature reflects greater variations in the local hydrogen bonding networks and reduced average binding energies (Figure 4.4) and heats of vaporization (Figure 4.6). It is interesting to notice that the maximum dipole values in the distributions are not shifted at different temperatures (Figure 4.9); it is the population of the molecular dipole moment in the liquid that is broadened. This results in a shift of the maximum position towards smaller average values as the temperature increases. In a recent study, Raabe

and Sadus suggested that the introduction of bond and angle flexibility in a water model is responsible for the decrease in the dipole moment with increased temperature and for the good performance on computed dielectric constant and pressure-temperature-density behavior using a flexible water model.[143] However, the water geometry was severely distorted from the gas-phase structure and the average bond lengths and angles in the liquid states are both significantly larger than commonly accepted values of liquid water.[143,144] The results displayed in Figure 4.9 show that the change in electronic polarization at different thermodynamic state points also makes critical contributions to the variation of the molecular dipole moment.

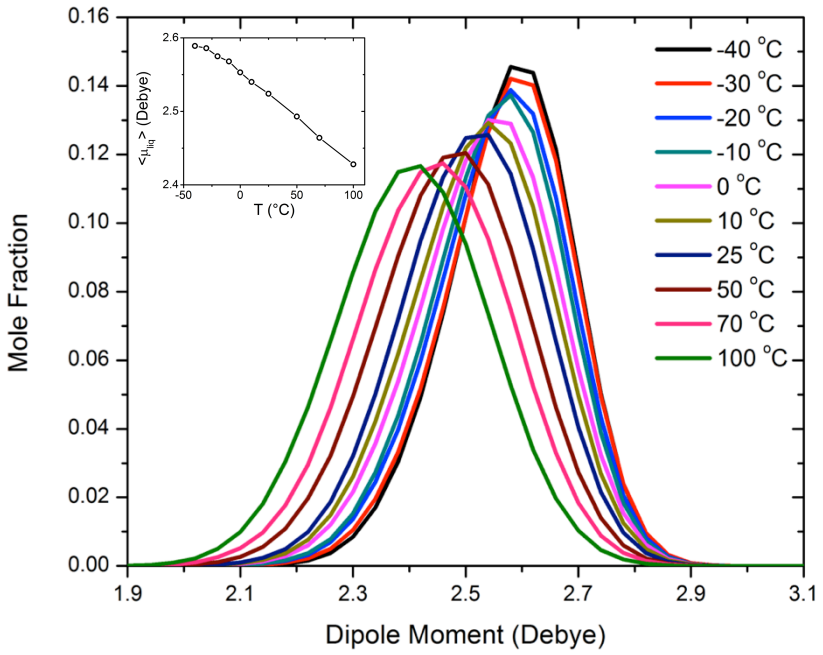


Figure 4-9. Distributions of scalar molecular dipole moments for liquid water at different temperatures along with the computed average molecular dipole moments in the inlet.

Computed radial distribution functions, which exhibit the expected trends as functions of temperature, are given in the supplementary material.[149] The loss of the liquid structure is observed with increasing temperature, and the height of the first peak in $g_{OO}(r)$ declines with

broadening of the peak as the first minimum disappears at high temperature (Figure A-8 in Appendix A). On the contrary, $g_{OO}(r)$ at low temperatures exhibits more structured RDFs. Similar trends are observed in both $g_{OH}(r)$ and $g_{HH}(r)$ as functions of temperature (Figures A-9 and A-10 in Appendix A).

4.4.2.3. Energy Decomposition Analysis of Liquid Water

The total binding energy, $E_i(l)$, from the XP3P water can be decomposed into specific contributing factors,[12,89,127] including vertical interaction energy and polarization energy. This analysis is useful for understanding the energy terms that are implicitly fitted in the development of polarizable or non-polarizable empirical potentials.

The vertical interaction energy represents the total energy of the liquid in which the wave function of each water molecule is not polarized, corresponding to that in the gas phase,

$$\Delta E_{\text{vert}} = \frac{1}{2}\sum_{a=1}^{N}\sum_{b\neq a}^{N}\left\langle \Psi_a^0 \left| H_{ab}^0\left(\rho_b^0\right)\right| \Psi_a^0 \right\rangle + E^{\text{XD}} \tag{4.22}$$

where $H_{ab}^0\left(\rho_b^0\right)$ is the interaction Hamiltonian between molecules *a* and *b*, in which the electrostatic potential defined in Eqs. 4.5 and 4.6 is obtained using the density of molecule *b* in the gas phase, ρ_b^0, and $E^{\text{XD}} = \sum_{a>b} E_{ab}^{\text{XD}}$ is the total van der Waals (i.e., the exchange-correlation term approximated by the Lennard-Jones potential in Eq. 4.9).

We emphasize that the term "vertical interaction energy" in energy decomposition analysis (EDA) is used to describe the interaction energy of the solvent molecules with their gas-phase, non-polarized electronic wave function relative to that of non-interacting molecules (Eq. 4.22).[89,127,145] This differs from the meaning of "vertical" that is associated with processes such as ionization and electronic excitation, where the geometries of the solute and the surrounding solvent are hypothetically kept in the un-ionized or the ground-state equilibrium configuration. In both cases the electronic wave function of the solute does change. In condensed-phase simulations, however, the energy accompanying the change of the electronic wave function

is called "polarization energy". Therefore, the term vertical is used to specify the interaction energy from an electronic state that is kept to remain in its gas-phase (electronic) configuration, prior to polarization.

Table 4-7. Temperature-Dependent Energy Components (unit in kcal/mol)

T (°C)	$E_i(l)$	E_{vert}	E_{pol}	ΔE_{stab}	ΔE_{self}	E^{XD}	E_{ele}
-40	-10.89	-6.62	-4.27	-7.98	3.71	3.17	-14.06
-30	-10.81	-6.57	-4.24	-7.92	3.68	3.12	-13.93
-20	-10.66	-6.52	-4.14	-7.71	3.57	2.97	-13.63
-10	-10.50	-6.44	-4.06	-7.56	3.50	2.90	-13.40
0	-10.29	-6.35	-3.94	-7.30	3.36	2.74	-13.03
10	-10.08	-6.26	-3.82	-7.07	3.25	2.62	-12.70
25	-9.83	-6.17	-3.66	-6.76	3.10	2.49	-12.32
50	-9.32	-5.90	-3.42	-6.26	2.84	2.20	-11.52
70	-8.86	-5.69	-3.17	-5.78	2.61	1.97	-10.83
100	-8.28	-5.38	-2.90	-5.23	2.33	1.69	-9.97

The wave functions of the solvent molecules are polarized in the liquid, and the energy change induced by the mutual interactions with the rest of the system corresponds to the polarization interaction energy, which is defined by Eq. 4.23.[12,89,127]

$$\Delta E_{\text{pol}} = \left(\left\langle \Phi \middle| H \middle| \Phi \right\rangle - N E_a^0 \right) - \Delta E_{\text{vert}} = E_{\text{tot}} - \Delta E_{\text{vert}} \tag{4.23}$$

The polarization energy can be further separated into two physically significant terms, corresponding to the so-called self-energy, ΔE_{self}, which is an energy cost (also called energy penalty) needed to pay for distorting the molecular wave function, and a net stabilizing

contribution, ΔE_{stab}, which is responsible for polarizing the electronic wave function to lower the total energy of the system. These energy terms are given below,[12,38,89,127]

$$\Delta E_{\text{self}} = \sum_{a=1}^{N}\left[\left\langle \Psi_a \middle| H_a^0 \middle| \Psi_a \right\rangle - \left\langle \Psi_a^0 \middle| H_a^0 \middle| \Psi_a^0 \right\rangle\right] = \sum_{a=1}^{N} \Delta E_a \tag{4.24}$$

$$\Delta E_{\text{stab}} = \frac{1}{2}\sum_{a=1}^{N}\sum_{b\neq a}^{N}\left[\left\langle \Psi_a \middle| H_{ab}\left(\rho_b\right) \middle| \Psi_a \right\rangle - \left\langle \Psi_a^0 \middle| H_{ab}^0\left(\rho_b^0\right) \middle| \Psi_a^0 \right\rangle\right] = \frac{1}{2}\sum_{a=1}^{N}\sum_{b\neq a}^{N} \Delta\Delta E_{ab} \tag{4.25}$$

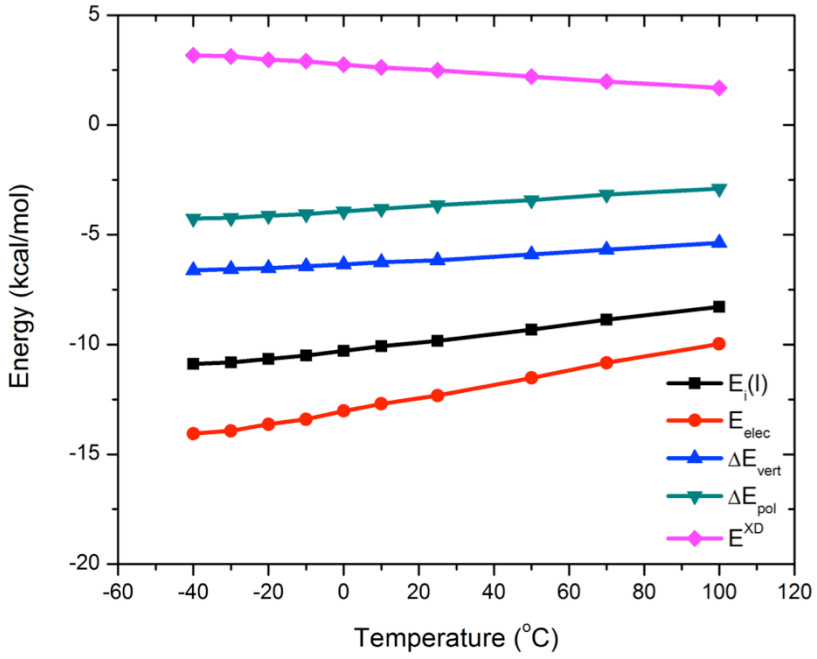


Figure 4-10. Average total interaction energies (black) per water in the liquid and their contributing components, including total electrostatic interaction energies (red), vertical interaction energies (blue), polarization energies (green), and exchange-dispersion correlation energies (magenta).

Shown in Table 4.7 and Figure 4.10 are the XP3P energy components at different temperatures. The vertical interaction energy contributes an almost constant percentage of the total binding energy, ranging from 60.8% at −40 °C to 65.0% at 100 °C. The increase of the

percentage with increasing temperature can be attributed to the increased volume of the system and reduced polarization effects at higher temperatures. At all temperatures used in the simulations, polarization effects are significant, contributing 35.0% – 39.2% of the total binding energies. At 25 °C, the average polarization energy is −3.66 kcal/mol (37.2% of $E_i(l)$). The van der Waals (or exchange-dispersion) term E^{XD} is dominated by the repulsive potential. The total electrostatic (non-van der Waals) component of the binding energy, $E_i(l)$, is the sum of the vertical and polarization interaction energies less the E^{XD} term, and it is about 20%–30% greater than the total binding energy in the 140 °C temperature range.

Table 4.7 shows that the average energy cost, i.e., self-energy (Eq. 4.24), needed to polarize the molecular wave function, is 3.10 ± 0.01 kcal/mol from the XP3P mode at 25 °C. This value is somewhat greater than the value estimated using the AM1 Hamiltonian (3.03 ± 0.01 kcal/mol).[12] If the classical expression for the self-energy,[128]

$$\Delta E_{\text{self}}^{\text{cl}} = \Delta\mu_{\text{ind}}^2 / 2\alpha \tag{4.26}$$

is used, where $\Delta\mu_{\text{ind}}$ is the induced dipole moment in the liquid, which is 0.64 D at 25 °C, and α is the molecular polarizability (1.27 Å^3) from the XP3P model, we obtain a self-energy of 2.35 kcal/mol, somewhat smaller than the quantum mechanical result (Eq. 4.23). The self-energy was used to correct the total energy of liquid water in the SPC/E model,[128] which has an effective dipole of 2.35 D ($\Delta\boldsymbol{\mu}_{\text{ind}}$ = 0.50 D). In that work, an estimate of $\Delta E_{\text{self}}^{\text{cl}}$ = 1.25 kcal/mol was used as an energy correction based on experimental polarizability of water. Table 4.7 shows that over the temperature range of −40 to 100 °C, ΔE_{self} varies from 3.69 kcal/mol to 2.33 kcal/mol, and the corresponding total polarization energies change from −4.25 to −2.90 kcal/mol.

4.5. Conclusions

A quantum mechanical force field (QMFF) for water with the explicit treatment of electronic polarization (X-Pol) has been described. Moving beyond the current Lifson-type, molecular

mechanics force fields (MMFF) that have been under continuous development in the past half century,[146–148] the present QMFF represents the condensed-phase system explicitly by an electronic structure method. Consequently, the internal energy terms in the traditional MMFF are replaced by a quantum mechanical formalism that naturally includes electronic polarization. An important aspect of the present procedure is the partition of a solution into molecular fragments such that the total wave function of the system is approximated as a Hartree product of antisymmetric, fragment wave functions. This approximation requires an empirical treatment of short-range intermolecular exchange repulsion and long-range dispersion interactions between different molecular fragments; however, one can model these effects using customary empirical formalisms. To this end, we have introduced a polarizable molecular orbital (PMO) model in the framework of the neglect diatomic differential overlap approximation. The present study represents a first step towards the goal of developing a full QMFF for the dynamic simulations of macromolecular systems as traditionally carried out with MMFF.

In this work, we introduce the first generation of a QMFF for water, making use of the PMO model specifically parameterized for compounds composed of hydrogen and oxygen, i.e., PMOw. The electrostatic potential responsible for the interactions among different fragments is model by a three-point charge representation that reproduces the total molecular dipole moment and the local hybridization contributions exactly. Consequently, the present QMFF for water, suitable for modeling gas-phase clusters, pure liquids, solid isomorphs, aqueous solutions, and the self-dissociation along with proton and anion transport, is called the XP3P model. The paper highlights the performance of the PMOw model for small water and proton clusters and simple proton transfer reactions, and the properties of liquid water using XP3P from a conglomeration of about 900×10^6 self-consistent-field calculations on a periodic system consisting of 267 water molecules. It is no exaggeration to say that this is the longest quantum mechanical simulation performed to date. More significantly, the unusual dipole derivative behavior of water, which is

incorrectly modeled in molecular mechanics, but is critical for a flexible water model, is naturally reproduced as a result of an electronic structural treatment of chemical bonding by XP3P. Much remains to be tested and investigated in future studies with the combined use of large clusters treated by PMOw embedded the XP3P liquid water. We anticipate that the present model is useful for studying proton transport in solution and solid phases as well as across biological membranes through ion channels.

Bibliography

Chapter 1

[1] Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.* **2003**, *66*, 27.

[2] MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584.

[3] Levitt, M. *Nat. Struct. Biol.* **2001**, *8*, 392.

[4] Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.

[5] Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236.

[6] Rick, S. W.; Stuart, S. J. *Rev. Comput. Chem.* **2002**, *18*, 89.

[7] Yu, H.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2005**, *172*, 69.

[8] Warshel, A.; Kato, M.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034.

[9] Applequist, J.; Carl, J. R.; Fung, K. –K. *J. Am. Chem. Soc.* **1972**, *94*, 2952.

[10] Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.

[11] Høye, J. S.; Stell, G. *J. Chem. Phys.* **1980**, *73*, 461.

[12] Rappé, A. K.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358.

[13] Sanderson, R. T. *Science* **1951**, *114*, 670.

[14] Jorgensen, W. L. *J. Chem. Theory Comput.* **2007**, *3*, 1877.

[15] Warshel, A.; Karplus, M. *J. Am. Chem. Soc.* **1972**, *94*, 5612.

[16] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

[17] The Nobel Prize Home Page. http://www.nobelprize.org (accessed Oct 9, 2013).

[18] Gao, J. *Rev. Comput. Chem.* **1995**, *7*, 119.

[19] Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298.

[20] Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632.

[21] Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.

[22] Gao, J. *J. Chem. Phys.* **1998**, *109*, 2346.

[23] Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890.

Chapter 2

[1] Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631.

[2] Hansson, T.; Oostenbrink, C.; van Gunsteren, W. F. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190.

[3] Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646.

[4] Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.* **2003**, *66*, 27.

[5] MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584.

[6] Yu, H.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2005**, *172*, 69.

[7] Foresman, J. B; Brooks, C. L., III. *J. Chem. Phys.* **1987**, *87*, 5892.

[8] Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236.

[9] Rick, S. W.; Stuart, S. J. *Rev. Comput. Chem.* **2002**, *18*, 89.

[10] Warshel, A.; Kato, M.; Pisilakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034.

[11] Gao, J.; Habibollazadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460.

[12] Gao, J.; Pavelites, J. J.; Habibollazadeh, D. *J. Phys. Chem.* **1996**, *100*, 2689.

[13] Xie, W.; Pu, J.; MacKerell, A. D.,Jr.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1878.

[14] Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 6379.

[15] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

[16] Forester, T. R.; McDonald, I. R.; Klein, M. L. *Chem. Phys.* **1989**, *129*, 225.

[17] Kristóf, T.; Liszi, J. *J. Phys. Chem. B* **1997**, *101*, 5480.

[18] Delhommelle, J.; Tschirwitz, C.; Ungerer, P.; Granucci, G.; Millié, P.; Pattou, D.; Fuchs, A. H. *J. Phys. Chem. B* **2000**, *104*, 4745.

[19] Delhommelle, J.; Millié, P.; Fuchs, A. H. *Mol. Phys.* **2000**, *98*, 1895.

[20] Nath, S. K. *J. Phys. Chem. B* **2003**, *107*, 9498.

[21] Kamath, G.; Lubna, N.; Potoff, J. J. *J. Chem. Phys.* **2005**, *123*, 124505.

[22] Lubna, N.; Kamath, G.; Potoff, J. J.; Rai, N.; Siepmann, J. I. *J. Phys. Chem. B* **2005**, *109*, 24100.

[23] Riahi, S.; Rowley, C. N. *J. Phys. Chem. B* **2013**, *117*, 5222.

[24] Reiffenstein, R. J.; Hulbert, W. C.; Roth, S. H. *Annu. Rev. Pharmacol. Toxicol.* **1992**, *32*, 109.

[25] Li, L.; Rose, P.; Moore, P. K. *Annu. Rev. Pharmacol. Toxicol.* **2011**, *51*, 169.

[26] Applequist, J.; Carl, J. R.; Fung, K. –K. *J. Am. Chem. Soc.* **1972**, *94*, 2952.

[27] Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.

[28] Ren, P.; Ponder, J. W. *J. Comput. Chem.* **2002**, *23*, 1497.

[29] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.

[30] Xie, W.; Pu, J.; Gao, J. *J. Phys. Chem. A* **2009**, *113*, 2109.

[31] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03*, Revision E.01; Gaussian, Inc.: Wallingford, CT, 2004.

[32] Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.

[33] Nosé, S. *Mol. Phys.* **1984**, *52*, 255.

[34] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

[35] Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *102*, 4613.

[36] Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638.

[37] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

[38] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

[39] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

[40] van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399.

[41] Olney, T. N.; Cann, N. M.; Cooper, G.; Brion, C. E. *Chem. Phys.* **1997**, *223*, 59.

[42] *CRC Handbook of Chemistry and Physics*, 88th ed.; Lide, D. R. Ed.; CRC Press: Boca Raton, FL, 2007: Section 9.

[43] *NIST Chemistry WebBook*; Linstrom, P. J.; Mallard, W.G., Eds. NIST Standard Reference Database Number 69, National Institute of Standards and Technology: Gaithersburg MD, 20899, http://webbook.nist.gov, (retrieved December 26, 2013).

[44] *CRC Handbook of Chemistry and Physics*, 88th ed.; Lide, D. R. Ed.; CRC Press: Boca Raton, FL, 2007: Section 6.

[45] Lemmon, E.W.; McLinden, M.O.; Friend, D.G. Thermophysical Properties of Fluid Systems. In *NIST Chemistry WebBook*; Linstrom, P.J.; Mallard, W.G., Eds. NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg MD, 20899, http://webbook.nist.gov, (retrieved October 2, 2008).

[46] Dupré, F.; Piaggesi, D.; Ricci, F. P. *Phys. Lett. A* **1980**, *80*, 178-180.

[47] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1989.

[48] Neumann, M.; Steinhauser, O. *Chem. Phys. Lett.* **1984**, *106*, 563.

[49] Andreani, C.; Petrillo, C.; Rocca, D. *Europhys. Lett.* **1988**, *5*, 145.

[50] Andreani, C.; Merlo, V.; Ricci, M. A.; Soper, A. K. *Mol. Phys.* **1991**, *73*, 407.

[51] Santoli, G.; Bruni, F.; Ricci, F. P.; Ricci, M. A.; Soper, K. A. *Mol. Phys.* **1999**, *97*, 777.

Chapter 3

[1] Levitt, M. *Nat. Struct. Biol.* **2001**, *8*, 392.

[2] Jorgensen, W. L. *J. Chem. Theory. Comput.* **2007**, *3*, 1877.

[3] Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298.

[4] Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.

[5] Gao, J. *J. Chem. Phys.* **1998**, *109*, 2346.

[6] Xie, W.; Gao, J. *J. Chem. Theory. Comput.* **2007**, *3*, 1890.

[7] Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108.

[8] Song, L.; Han, J.; Lin, Y. –L.; Xie, W.; Gao, J. *J. Phys. Chem. A* **2009**, *113*, 11656.

[9] Mo, Y.; Gao, J.; Peyerimhoff, S. D. *J. Chem. Phys.* **2000**, *112*, 5530.

[10] Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968.

[11] Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. *Ann. Rep. Comput. Chem.* **2007**, *3*, 177.

[12] Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. –P. *J. Chem. Theory. Comput.* **2007**, *3*, 1960.

[13] Gao, J.; Xia, X. *Science* **1992**, *258*, 631.

[14] Lin, H.; Truhlar, D. G. *Theor Chem Acc* **2007**, *117*, 185.

[15] Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198.

[16] Isegawa, M.; Gao, J.; Truhlar, D. G. *J. Chem. Phys.* **2011**, *135*, 084107.

[17] Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Theory. Comput.* **2010**, *6*, 2469.

[18] Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632.

[19] Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J. Phys. Chem. B* **2003**, *107*, 12039.

[20] Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. *J. Comput. Chem.* **2004**, *25*, 1431.

[21] Duan, L. L.; Mei, Y.; Zhang, D. W.; Zhang, Q. G.; Zhang, J. Z. H. *J. Am. Chem. Soc.* **2010**, *132*, 11159.

[22] Tong, Y.; Mei, Y.; Li, Y. L.; Ji, C. G.; Zhang, J. Z. H. *J. Am. Chem. Soc.* **2010**, *132*, 5137.

[23] Stoll, H.; Preuβ, H. *Theor. Chim. Acta* **1977**, *46*, 11.

[24] Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. *J. Chem. Phys.* **1970**, *53*, 4544.

[25] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.

[26] Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904.

[27] Truhlar, D. G.; Dahlke, E. E. *J. Chem. Theory. Comput.* **2007**, *3*, 1342.

[28] Truhlar, D. G.; Dahlke, E. E.; Leverentz, H. R. *J. Chem. Theory. Comput.* **2008**, *4*, 33.

[29] Tempkin, J. O. B.; Leverentz, H. R.; Wang, B.; Truhlar, D. G. *J. Phys. Chem. Lett.* **2011**, *2*, 2141.

[30] Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory. Comput.* **2009**, *5*, 1573.

[31] Mo, Y.; Peyerimhoff, S. D. *J. Chem. Phys.* **1998**, *109*, 1687.

[32] Mo, Y.; Gao, J. *J. Comput. Chem.* **2000**, *21*, 1458.

[33] Jacobson, L. D.; Herbert, J. M. *J. Chem. Phys.* **2011**, *134*, 094118.

[34] Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887

[35] Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*, 214103.

[36] Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.

[37] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

[38] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

[39] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

[40] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. *Gaussian 03*, Revision E.01; Gaussian, Inc.: Wallingford, CT, 2004.

[41] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. S. *J. Comput. Chem.* **1993**, *14*, 1347.

[42] Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard, W. A., III. *J. Chem. Theory Comput.* **2009**, *5*, 1016.

[43] Zhang, P.; Fiedler, L.; Leverentz, H. R.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2011**, *7*, 857.

[44] Gao, J.; Xia, X.; George, T. F. *J. Phys. Chem.* **1993**, *97*, 9241.

[45] Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008.

[46] Mo, Y.; Schleyer, P. v. R.; Wu, W.; Lin, M.; Zhang, Q.; Gao, J. *J. Phys. Chem. A* **2003**, *107*, 10011.

[47] Mo, Y.; Bao, P.; Gao, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6760.

[48] Gao, J.; Garner, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1986**, *108*, 4784.

[49] Gao, J. *ACS Symp. Ser.* **1994**, *569*, 8.

[50] Freindorf, M.; Gao, J. *J. Comput. Chem.* **1996**, *17*, 386.

[51] Riccardi, D.; Li, G.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467.

[52] Freindorf, M.; Shao, Y.; Furlani, T. R.; Kong, J. *J. Comput. Chem.* **2005**, *26*, 1270.

[53] Gao, J.; Cembran, A.; Mo, Y. *J. Chem. Theory Comput.* **2010**, *6*, 2402.

[54] Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459.

Chapter 4

[1] MacKerell, A. D., Jr. *J. Comput. Chem.* **2004**, *25*, 1584.

[2] Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.

[3] Guillot, B. *J. Mol. Liq.* **2002**, *101*, 219.

[4] Schropp, B.; Tavan, P. *J. Phys. Chem. B* **2008**, *112*, 6233.

[5] Vega, C.; Abascal, J. L. F.; Debenedetti, P. G. *Phys. Chem. Chem. Phys.* **2011**, *13,* 19660.

[6] Bernal, J. D.; Fowler, R. H. *J. Chem. Phys.* **1933**, *1*, 515.

[7] Jorgensen, W. L. *J. Chem. Theory Comput.* **2007**, *3*, 1877.

[8] Berendsen, H. J. C.; Postama, J. P. M.; van Gunsteren, W. F.; Hermans, J. in *Intermolecular Forces*; Pullmann, B. Ed.; D. Reidel Publising Company, Dordrecht, 1981; p. 331.

[9] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

[10] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.

[11] Gao, J. *J. Phys. Chem. B* **1997**, *101*, 657.

[12] Gao, J. *J. Chem. Phys.* **1998***, 109*, 2346.

[13] Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459.

[14] Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108.

[15] Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890.

[16] Wierzchowski, S. J.; Kofke, D. A.; Gao, J. *J. Chem. Phys.* **2003**, *119*, 7365.

[17] Vesely, F. J. *J. Comput. Phys.* **1977**, *24*, 361.

[18] Howard, A. E.; Singh, U. C.; Billeter, M.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 6984.

[19] Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1994**, *98*, 4180.

[20] Gao, J.; Habibollazadeh, D.; Shao, L. *J. Phys. Chem.* **1995**, *99*, 16460.

[21] Stout, J. M.; Dykstra, C. E. *J. Phys. Chem. A* **1998**, *102*, 1576.

[22] Applequist, J.; Carl, J. R.; Fung, K. –K. *J. Am. Chem. Soc.* **1972**, *94*, 2952.

[23] Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.

[24] van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399.

[25] Yu, H. B.; Hansson, T.; van Gunsteren, W. F. *J. Chem. Phys.* **2003**, *118*, 221.

[26] Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. *Chem. Phys. Lett.* **2006**, *418*, 245.

[27] Rappe, A. K.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358.

[28] Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141.

[29] Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.

[30] Valone, S. M. *J. Chem. Theory Comput.* **2011**, *7*, 2253.

[31] Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974.

[32] Giese, T. J.; York, D. M. *J. Chem. Phys.* **2007**, *127*, 194101.

[33] Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2010**, *6*, 2469.

[34] Gao, J.; Cembran, A.; Mo, Y. *J. Chem. Theory Comput.* **2010**, *6*, 2402.

[35] Song, L.; Han, J.; Lin, Y. –L.; Xie, W.; Gao, J. *J. Phys. Chem. A* **2009**, *113*, 11656.

[36] Han, J.; Truhlar, D. G.; Gao, J. *Theor. Chem. Acc.* **2012**, *131*, 1161.

[37] Wang, Y.; Sosa, C. P.; Cembran, A.; Truhlar, D. G.; Gao, J. *J. Phys. Chem. B* **2012**, *116*, 6781.

[38] Mo, Y.; Bao, P.; Gao, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6760.

[39] Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

[40] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

[41] Stewart, J. J. P. *J. Comput. Aid. Mol. Des.* **1990**, *4*, 1.

[42] Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J. Comput. Chem.* **2006**, *27*, 1101.

[43] Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173.

[44] Morgado, C. A.; McNamara, J. P.; Hillier, I. H.; Burton, N. A. *J. Chem. Theory Comput.* **2007**, *3*, 1656.

[45] McNamara, J. P.; Hillier, I. H. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362.

[46] McNamara, J. P.; Sharma, R.; Vincent, M. A.; Hillier, I. H.; Morgado, C. A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 128.

[47] Tuttle, T.; Thiel, W. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159.

[48] Korth, M.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 2929.

[49] Zhang, P.; Fiedler, L.; Leverentz, H. R.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2011**, *7*, 857.

[50] Fiedler, L.; Gao, J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 852.

[51] Isegawa, M.; Fiedler, L.; Leverentz, H. R.; Wang, Y.; Nachimuthu, S.; Gao, J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 33.

[52] Jug, K.; Geudtner, G. *J. Comput. Chem.* **1993**, *14*, 639.

[53] Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

[54] Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726.

[55] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

[56] Stewart, J. J. P. *Rev. Comput. Chem.* **1990**, *1*, 45.

[57] Zerner, M. C. *Rev. Comput. Chem.* **1991**, *2*, 313.

[58] Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907.

[59] Dewar, M. J. S.; Thiel, W. *Theor. Chim. Acta* **1977**, *46*, 89.

[60] Gao, J.; Wang, Y. *J. Chem. Phys.* **2012**, *136*, 071101.

[61] Liu, K.; Brown, M. G.; Saykally, R. J *J. Phys. Chem. A* **1997**, 101, 8995.

[62] Piris, M.; Matxain, J. M.; Lopez, X.; Ugalde, J. M. *J. Chem. Phys.* **2010**, *132*, 031103.

[63] Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Schaefer, H. F.; Quack, M. *J. Chem. Phys.* **2002**, *116*, 690.

[64] Maroulis, G. *J. Chem. Phys.* **2000**, *113*, 1813.

[65] Clough, S. A.; Beers, Y.; Klein, G. P.; Rothman, L. S. *J. Chem. Phys.* **1973**, *59*, 2254.

[66] Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard, W. A., III. *J. Chem. Theory Comput.* **2009**, *5*, 1016.

[67] Giese, T. J.; Chen, H. Y.; Dissanayake, T.; Giambasu, G. M.; Heldenbrand, H.; Huang, M.; Kuechler, E. R.; Lee, T. –S.; Panteva, M. T.; Radak, B. K.; York, D. M. *J. Chem. Theory Comput.* **2013**, *9*, 1417.

[68] Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.

[69] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr., Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

[70] Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048.

[71] Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.

[72] Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, 1996.

[73] Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. *Annu. Rep. Comput. Chem.* **2007**, *3*, 177.

[74] Leverentz, H. R.; Gao, J.; Truhlar, D. G. *Theor. Chem. Acc.* **2011**, *129*, 3.

[75] Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986.

[76] Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820.

[77] Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 527.

[78] Thole, B. T.; van Duijnen, P. T. *Theor. Chim. Acta* **1983**, *63*, 209.

[79] Swart, M.; van Duijnen, P. T.; Snijders, J. G. *J. Comput. Chem.* **2001**, *22*, 79.

[80] Zhang, P.; Bao, P.; Gao, J. *J. Comput. Chem.* **2011**, *32*, 2127.

[81] Pople, J. A.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S136.

[82] Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **2001**, *115*, 2237.

[83] Mohoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.

[84] Zhang, P.; Truhlar, D. G.; Gao, J. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7821.

[85] Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.

[86] Koopman, E. A.; Lowe, C. P. *J. Chem. Phys.* **2006**, *124*, 204103.

[87] Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952.

[88] Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2.

[89] Gao, J.; Xia, X. *Science* **1992**, *258*, 631.

[90] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701.

[91] Gao, J.; Han, J.; Zhang, P. *MCSOL*, version 2012xp; Minneapolis, 2012.

[92] Mazack, M. J. M.; Gao, J. *X-Pol*, version 2013a1; University of Minnesota: Minneapolis, 2013.

[93] Brooks, B. R.; Brooks, C. L., III.; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.

[94] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.

[95] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.

[96] Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185.

[97] Benedict, W. S.; Gailar, N.; Plyler, E. K. *J. Chem. Phys.* **1956**, *24*, 1139.

[98] Burnham, C. J.; Xantheas, S. S. *J. Chem. Phys.* **2002**, *116*, 5115.

[99] Partridge, H.; Schwenke, D. W. *J. Chem. Phys.* **1997**, *106*, 4618.

[100] Whalley, E.; Klug, D. D. *J. Chem. Phys.* **1986**, *84*, 78.

[101] Rothman, L. S.; Rinsland, C. P.; Goldman, A.; Massie, S. T.; Edwards, D. P.; Flaud, J. M.; Perrin, A.; Camy-Peyret, C.; Dana, V.; Mandin, J. Y.; Schroeder, J.; McCann, A.; Gamache, R. R.; Wattson, R. B.; Yoshino, K.; Chance, K. V.; Jucks, K. W.; Brown, L. R.; Nemtchinov, V.; Varanasi, P. *J. Quant. Spectrosc. Radiat. Transf.* **2010**, *111*, 1568.

[102] Curtiss, L. A.; Frurip, D. J.; Blander, M. *J. Chem. Phys.* **1979**, *71*, 2703.

[103] Xantheas, S. S. *Struct. Bond.* **2006**, *116*, 119.

[104] Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M. G.; Saykally, R. J. *Science* **1997**, *275*, 814.

[105] Nachimuthu, S.; Gao, J.; Truhlar, D. G. *Chem. Phys.* **2012**, *400*, 8.

[106] Sadhukhan, S.; Munoz, D.; Adamo, C.; Scuseria, G. E. *Chem. Phys. Lett.* **1999**, *306*, 83.

[107] Kumar, R.; Christie, R. A.; Jordan, K. D. *J. Phys. Chem. B* **2009**, *113*, 4111.

[108] Goyal, P.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2011**, *115*, 6790.

[109] Jorgensen, W. L.; Jenson, C. *J. Comput. Chem.* **1998**, *19*, 1179.

[110] Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381.

[111] Kyle, B. G. *Chemical and Process Thermodynamics*; Prentice Hall PTR: New Jersey, 1999.

[112] Wagner, W.; Pruβ, A. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387.

[113] Haar, L.; Gallagher, E.; Kell, G. *NBS/NRC Steam Tables: Thermodynamic and Transport Properties and Computer Programs for Vapor and Liquid States of Water in SI Units*; Hemisphere Publishing Corporation: Washington, DC, 1984.

[114] Wang, J.; Cieplak, P.; Cai, Q.; Hsieh, M. J.; Wang, J. M.; Duan, Y.; Luo, R. *J. Phys. Chem. B* **2012**, *116*, 7999.

[115] Coulson, C. A.; Eisenberg, D. *Proc. R. Soc. London, Ser. A* **1966**, *291*, 445.

[116] Caldwell, J. W.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99*, 6208.

[117] Silvestrelli, P. L.; Parrinello, M. *Phys. Rev. Lett.* **1999**, *82*, 3308.

[118] Neumann, M. *J. Chem. Phys.* **1985**, *82*, 5663.

[119] Aragones, J. L.; MacDowell, L. G.; Vega, C. *J. Phys. Chem. A* **2011**, *115*, 5745.

[120] Alper, H. E.; Levy, R. M. *J. Chem. Phys.* **1989**, *91*, 1242.

[121] Barker, J. A.; Watts, R. O. *Mol. Phys.* **1973**, *26*, 789.

[122] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1989.

[123] Zhu, S. B.; Wong, C. F. *J. Chem. Phys.* **1993**, *98*, 8892.

[124] Sprik, M. *J. Chem. Phys.* **1991**, *95*, 6762.

[125] Hochtl, P.; Boresch, S.; Bitomsky, W.; Steinhauser, O. *J. Chem. Phys.* **1998**, *109*, 4927.

[126] Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2004**, *108*, 13427.

[127] Gao, J. *J. Comput. Chem.* **1997**, *18*, 1062.

[128] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

[129] Kell, G. S. *J. Chem. Eng. Data* **1975**, *20*, 97.

[130] Angell, C. A.; Oguni, M.; Sichina, W. J. *J. Phys. Chem.* **1982**, *86*, 998.

[131] Vega, C.; Conde, M. M.; McBride, C.; Abascal, J. L. F.; Noya, E. G.; Ramirez, R.; Sese, L. M. *J. Chem. Phys.* **2010**, *132*, 046101.

[132] Krynicki, K.; Green, C. D.; Sawyer, D. W. *Faraday Discuss.* **1978**, *66*, 199.

[133] Tazi, S.; Botan, A.; Salanne, M.; Marry, V.; Turq, P.; Rotenberg, B. *J. Phys. Cond. Matter* **2012**, *24*, 284117.

[134] Abragam, A. *The Principles of Nuclear Magnetism*; Clarendon Press: Oxford, U.K., 1961.

[135] Wilbur, D. J.; Defries, T.; Jonas, J. *J. Chem. Phys.* **1976**, *65*, 1783.

[136] Barthel, J.; Bachhuber, K.; Buchner, R.; Hetzenauer, H. *Chem. Phys. Lett.* **1990**, *165*, 369.

[137] Soper, A. K. *Chem. Phys.* **2000**, *258*, 121.

[138] Head-Gordon, T.; Johnson, M. E. *Proc. Nat. Acad. Sci.* **2006**, *103*, 7973.

[139] Baez, L. A.; Clancy, P. *J. Chem. Phys.* **1994**, *101*, 9837.

[140] Chen, B.; Xing, J.; Siepmann, J. I. *J. Phys. Chem. B* **2000**, *104*, 2391.

[141] Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511.

[142] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.

[143] Raabe, G.; Sadus, R. J. *J. Chem. Phys.* **2011**, *134*, 234501.

[144] Ichikawa, K.; Kameda, Y.; Yamaguchi, T.; Wakita, H.; Misawa, M. *Mol. Phys.* **1991**, *73*, 79.

[145] Mo, Y.; Gao, J. *J. Phys. Chem. B* **2006**, *110*, 2976.

[146] Lifson, S. *J. Chim. Phys. Physicochim. Biol.* **1968**, *65*, 40.

[147] Levitt, M.; Lifson, S. *J. Mol. Biol.* **1969**, *46*, 269.

[148] Levitt, M. *Nat. Struct. Biol.* **2001**, *8*, 392.

Appendix A

Supplementary Material for Chapter 4[a]

This supporting information includes tables containing the calculated properties for water and selected compounds containing hydrogen and oxygen atoms (the water monomer and dimer properties are used in the semiempirical polarized molecular orbital fitting) (Table A-1), optimized cartesian coordinates for the water dimer using PMOw, XP3P, and CCSD(T)/aug-cc-pVDZ (Table A-2), water clusters using PMOw and XP3P (Table A-3), proton-water clusters $H^+(H_2O)_n$, with n = 2, 3, 4, and 6, using PMOw and MP2/aug-cc-pVDZ (for n = 2, the optimized structures also include geometries at the B3LYP/aug-cc-pVTZ and M06-2X/aug-cc-pVTZ levels) (Table A-4), and thermodynamic properties from liquid simulations at temperatures ranging from -40 to 100 °C (Tables A-5 and A-6). Optimized structures for the water clusters with PMOw and XP3P are illustrated in Figures A-1 and A-2. Figure A-3 displays the root-mean squre displacement of water vs. time from molecular dynamics simulations of liquid water using XP3P at 25 °C. Figure A-4 shows the computed and single-exponential fitted reorientation time correlation function about the H-H vector. Figure A-5 depicts the computed and single-exponential fitted total dipole moment time correlation function. Figure A-6 gives the computed and experimental heat capacities at different temperatures, in which values computed using the fluctuation formula are given in black and those using the enthalpy derivatives in blue. Figure A-7 exhibits the computed and experimental isothermal compressibility at different temperatures. Figures A-8, A-9, and A-10 dispose the O-O, O-H, and H-H radial distribution functions of liquid water at temperatures ranging from -40 to 100 °C, respectively.

[a] Reproduced with permission from Han, J.; Mazack, M. J. M.; Zhang, P.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2013**, *139*, 054503. Copyright 2013, AIP Publishing LLC.

Table A-1. Targeted Molecular Properties Considered in PMOw Parameterization

Molecule	Property	PMOw	Target
H_2O	AE (kcal/mol)	233.0	232.2[a,b]
	IP (eV)	13.2	12.68[c]
	r (Å)	0.955	0.9575[d] (9-24)
	θ (deg)	104.6	104.52[d]
	α (Å^3)	1.27	1.45[e]
	q^O (e)	-0.31	
	q^H (e)	0.16	
	μ (Debye)	1.88	1.85[f]
$(H_2O)_2$	BE (kcal/mol)	-5.1	-5.0[g]
	$r_{14}{}^h$ (Å)	0.969	0.97[i]
	$r_{13}{}^h$ (Å)	0.952	0.96[i]
	$r_{24}{}^h$ (Å)	1.939	1.95[i]
	$r_{25}{}^h$ (Å)	0.960	0.96[i]
	$\theta_{413}{}^h$ (deg)	104.1	104.5[i]
	$\theta_{142}{}^h$ (deg)	170.9	172.9[i]
	$\theta_{126}{}^h$ (deg)	105.0	110.4[i]
	μ (Debye)	2.39	2.65[j]
OH^-	IP (eV)	2.76	1.83[m]
	r (Å)	0.841	0.964[n]
$OH^-(H_2O)$	BE (kcal/mol)	-27.7	-26.6[g]
	$r_{24}{}^h$ (Å)	1.217	1.12[o]
	$\theta_{213}{}^h$ (deg)	108.7	101.8[o]
H_2	AE (kcal/mol)	109.2	103.3[p]
	IP (eV)	13.7	15.4[p]
	r (Å)	0.687	0.741[d]
O_2	AE (kcal/mol)	89.8	118.0[a]
	IP (eV)	14.3	12.07[a]
	r (Å)	1.045	1.2074[c]
H_3O^+	r (Å)	1.020	0.976[q]
	θ (deg)	104.5	111.3[q]

Continued on next page

Table A-1. Targeted Molecular Properties Considered in PMOw Parameterization – *Continued*

Molecule	Property		PMOw	Target
H_2O_2	AE		191.6	252.3[a]
	IP		12.9	12.1[r]
	r_{OH} (Å)		0.928	0.950[s]
	r_{OO} (Å)		1.631	1.475[s]
	θ_{HOO} (deg)		87.3	94.8[s]
	φ_{HOOH} (deg)		145.2	119.8[s]
	μ (Debye)		0.893	1.57[t]
	RB (kcal/mol)	Cis	7.5	7.3[u]
		Trans	0.1	1.1[u]

[a]Ref. 1. [b]Ref. 2. [c]Ref. 3. [d]Ref. 4. [e]Ref. 5. [f]Ref. 6. [g]Ref. 7. [h]Geometrical description of $(H_2O)_2$ and $OH^-(H_2O)$ can be shown below:

[i]Ref. 8. [j]Ref. 9. [k]Ref. 10. [l]Ref. 11. [m]Ref. 12. [n]Ref. 13. [o]Ref. 14. [p]Ref. 15. [q]Ref. 16. [r]Ref. 17. [s]Ref. 18. [t]Ref. 19. [u]Ref. 20

Table A-2. Optimized Structures and Coordinates of Water Dimer Using PMOw, XP3P, and CCSD(T)/aug-cc-pVDZ

	PMOw	XP3P	CCSD(T)/aug-cc-pVDZ
r_{OO} (Å)	2.90	2.81	2.92
α (deg)	6.2	1.3	6.0
φ (deg)	115.0	165.3	120.8

Optimized coordinates using PMOw

O	-0.192670	-1.220867	0.796449
H	-0.149843	-1.029792	1.727852
H	-0.055213	-0.360730	0.371632
O	0.191267	1.206631	-0.742552
H	-0.575552	1.260349	-1.316839
H	0.921011	1.012409	-1.334542

Optimized coordinates using XP3P

O	-0.200309	-1.271823	0.879910
H	-0.124153	-0.876485	1.739193
H	-0.082431	-0.517446	0.272167
O	0.146000	0.938654	-0.821542
H	-0.546334	1.421094	-1.275040
H	0.946228	1.174006	-1.292688

Optimized coordinates using CCSD(T)/aug-cc-pVDZ

O	-1.519438	0.042235	-0.115367
H	-1.926117	-0.264377	0.704941
H	-0.563739	-0.015168	0.054153
O	1.395668	-0.034903	0.108692
H	1.739345	-0.617494	-0.582578
H	1.740665	0.838387	-0.123114

Table A-3. Optimized Coordinates of Water Clusters Using PMOw and XP3P

$(H_2O)_3$ cycle

PMOw

O	1.236380	1.117522	-0.020443
H	0.493789	1.034004	0.608665
H	2.011559	0.955442	0.511657
O	-1.173614	0.036723	1.173903
H	-1.173806	-0.569263	1.909942
H	-1.064999	-0.540050	0.391414
O	-0.113034	-1.088860	-1.221250
H	0.549449	-0.383130	-1.079765
H	-0.539723	-0.842389	-2.039121

XP3P

O	1.346839	0.864653	-0.020565
H	0.672030	0.658934	0.663464
H	1.944868	1.450562	0.433043
O	-0.836123	-0.114604	1.368817
H	-1.595123	-0.466014	1.823919
H	-0.910472	-0.507868	0.471376
O	-0.435533	-0.843175	-1.269997
H	0.313571	-0.244098	-1.056610
H	-0.274054	-1.078395	-2.178448

$(H_2O)_4$ cycle

PMOw

O	-1.435011	1.259749	-0.341777
H	-0.472933	1.460011	-0.215843
H	-1.861455	1.742987	0.363818
O	-1.282570	-1.455085	0.026105
H	-1.464996	-0.491939	-0.117522
H	-1.595071	-1.867331	-0.777408
O	1.444916	-1.267742	0.259805
H	0.476338	-1.462846	0.186657
H	1.677828	-1.597390	1.126016
O	1.272330	1.463079	0.055533
H	1.461256	0.494774	0.146367
H	1.778369	1.721734	-0.712751

XP3P

O	-1.485429	1.144962	-0.294133
H	-0.521496	1.384972	-0.199480
H	-1.899022	1.989916	-0.426091
O	-1.159458	-1.503161	0.003342
H	-1.387487	-0.539353	-0.118340
H	-2.006716	-1.927935	-0.060999
O	1.485090	-1.144825	0.295364
H	0.521296	-1.384945	0.199586
H	1.898314	-1.989486	0.430314
O	1.159463	1.503025	-0.004907
H	1.387354	0.539327	0.117900

H	2.007089	1.927506	0.056442

$(H_2O)_5$ cycle
PMOw

O	-1.233643	-1.511071	1.230758
H	-0.899074	-2.013689	1.972475
H	-1.049472	-0.570538	1.489617
O	0.101480	-2.057940	-1.077757
H	-0.574133	-2.440942	-1.632905
H	-0.377136	-1.863681	-0.230213
O	0.716128	2.200316	-0.093673
H	0.893369	1.508315	-0.782140
H	1.592638	2.476906	0.171480
O	-0.689323	1.059935	1.937634
H	-0.166777	1.481815	1.206295
H	-1.469957	1.608560	1.998175
O	1.219643	0.284854	-1.965678
H	0.791545	0.444239	-2.804376
H	0.835709	-0.578078	-1.663691

XP3P

O	-1.095095	-1.578172	1.191957
H	-1.505826	-1.979212	1.947509
H	-0.980424	-0.625349	1.490071
O	-0.065083	-1.899241	-1.237730
H	-0.277019	-2.795427	-1.466410
H	-0.463822	-1.811858	-0.319458
O	0.739711	2.130151	0.028380
H	0.874139	1.502031	-0.744360
H	1.252841	2.884735	-0.231850
O	-0.736187	0.957890	1.902776
H	-0.174872	1.430485	1.215991
H	-0.944118	1.656243	2.510539
O	1.055790	0.399443	-1.964373
H	1.366279	0.241763	-2.847027
H	0.644743	-0.484474	-1.719996

$(H_2O)_6$ cycle
PMOw

O	1.865633	1.973205	-0.068810
H	0.888004	2.136512	-0.034917
H	2.195096	2.467570	0.680604
O	0.776029	-2.602288	-0.068813
H	1.039430	-3.134794	0.680601
H	1.406273	-1.837288	-0.034919
O	2.641660	-0.629080	0.068814
H	2.294277	0.299218	0.034919
H	3.234525	-0.667223	-0.680600
O	-1.865633	-1.973205	0.068810
H	-2.195096	-2.467570	-0.680604
H	-0.888004	-2.136512	0.034917

O	-0.776029	2.602288	0.068813
H	-1.039430	3.134794	-0.680601
H	-1.406273	1.837288	0.034919
O	-2.641660	0.629080	-0.068814
H	-3.234525	0.667223	0.680600
H	-2.294277	-0.299218	-0.034919

XP3P

O	1.867096	1.855705	0.154416
H	0.902360	2.115066	0.022763
H	2.244353	2.657191	0.492941
O	0.673541	-2.544806	0.154401
H	1.179020	-3.272264	0.492919
H	1.380521	-1.839000	0.022749
O	2.540638	-0.689098	-0.154415
H	2.282879	0.276067	-0.022756
H	3.423373	-0.615068	-0.492937
O	-1.867096	-1.855707	-0.154416
H	-2.244354	-2.657190	-0.492940
H	-0.902360	-2.115066	-0.022763
O	-0.673542	2.544807	-0.154402
H	-1.179020	3.272265	-0.492921
H	-1.380522	1.838999	-0.022749
O	-2.540639	0.689098	0.154416
H	-3.423371	0.615068	0.492938
H	-2.282880	-0.276066	0.022757

$(H_2O)_6$ cage

PMOw

O	-2.045144	0.990395	1.791214
H	-1.890950	1.487588	2.589494
H	-1.557982	1.490419	1.094164
O	-0.582638	-1.319148	1.114250
H	-1.161875	-0.632575	1.514197
H	0.305260	-0.917455	1.195733
O	1.808073	0.170349	0.692978
H	2.018894	-0.207870	-0.200438
H	2.642732	0.192114	1.154859
O	-0.812439	-0.865689	-1.485884
H	-1.429858	-1.445410	-1.926723
H	-0.835801	-1.156506	-0.519697
O	-0.411077	1.808952	-0.290242
H	-0.676696	1.103512	-0.903787
H	0.431891	1.471346	0.058565
O	1.877599	-0.781072	-1.848322
H	2.021472	-0.091416	-2.495688
H	0.895538	-0.919536	-1.869674

XP3P

O	-1.905248	0.927242	2.070851
H	-2.379996	1.643571	2.477566
H	-1.456726	1.363810	1.305152

O	-0.593909	-1.252830	1.013020
H	-1.152913	-0.649885	1.551741
H	0.284010	-0.848243	1.126155
O	1.784689	0.234858	0.669141
H	1.975278	-0.086991	-0.256767
H	2.655529	0.341641	1.039397
O	-0.734282	-0.851029	-1.592110
H	-1.516327	-1.282383	-1.920390
H	-0.748092	-1.100263	-0.614354
O	-0.451112	1.695572	-0.092466
H	-0.645522	1.107186	-0.832372
H	0.406562	1.342019	0.221618
O	1.927001	-0.611074	-1.863182
H	2.208433	-0.866728	-2.735245
H	0.939638	-0.728473	-1.902750

$(H_2O)_6$ book
PMOw

O	0.092813	1.474458	0.925967
H	0.932072	1.486214	0.381349
H	0.248384	2.107202	1.625540
O	-0.006826	-1.381935	1.084625
H	-0.030675	-0.418330	1.232234
H	-0.825068	-1.530944	0.562263
O	-2.248297	1.400751	-0.535879
H	-1.411624	1.530123	-0.031258
H	-2.038859	1.717733	-1.412541
O	-2.370654	-1.372818	-0.437090
H	-2.450431	-0.394461	-0.516133
H	-3.165339	-1.634148	0.021974
O	2.329944	1.221743	-0.526528
H	2.371356	0.222886	-0.552854
H	2.270517	1.465061	-1.449495
O	2.164667	-1.463209	-0.497469
H	1.345910	-1.562978	0.069141
H	2.836112	-1.934348	-0.007846

XP3P

O	0.031045	1.605310	0.483625
H	0.951919	1.516586	0.075579
H	0.201330	2.180746	1.222204
O	0.031159	-1.234170	0.535308
H	-0.049344	-0.272625	0.608088
H	-0.911119	-1.496756	0.394304
O	-2.459433	1.193453	-0.429569
H	-1.536518	1.396905	-0.113507
H	-2.725643	2.007105	-0.841067
O	-2.639022	-1.453498	0.054675
H	-2.685856	-0.484545	-0.158183
H	-3.538377	-1.739734	-0.049752
O	2.474295	1.215930	-0.445048

H	2.610398	0.223843	-0.370846
H	3.278831	1.503247	-0.860999
O	2.572214	-1.427075	-0.164547
H	1.605142	-1.463276	0.113196
H	2.832975	-2.338444	-0.117477

$(H_2O)_6$ prism
PMOw

O	1.468184	-0.325617	1.380568
H	0.529145	-0.676397	1.510992
H	1.963534	-0.657547	2.124439
O	0.735122	2.019012	-0.053117
H	1.069467	1.429539	0.652244
H	1.100151	1.596921	-0.844229
O	1.689348	-0.555227	-1.359987
H	2.355420	-1.138554	-1.716835
H	1.779097	-0.643216	-0.382651
O	-1.875769	1.239276	-0.032376
H	-0.980903	1.663404	-0.056545
H	-2.488737	1.964903	0.045978
O	-1.086351	-1.266325	-1.413540
H	-0.145604	-1.061068	-1.556438
H	-1.493303	-0.390454	-1.342153
O	-1.045196	-1.090314	1.404745
H	-1.106555	-1.512188	0.518237
H	-1.504051	-0.245146	1.221670

XP3P

O	1.531427	-1.260917	1.513053
H	0.531430	-1.273326	1.591501
H	1.807499	-1.638580	2.340744
O	0.482996	2.190280	-0.747663
H	1.075258	2.853507	-1.084053
H	1.010907	1.349117	-0.847285
O	1.730624	-0.172548	-0.891627
H	2.485635	-0.586916	-1.296934
H	1.726964	-0.587968	0.027889
O	-1.780526	1.164533	0.206214
H	-0.965517	1.645064	-0.121395
H	-2.425404	1.861718	0.263722
O	-0.960090	-0.997506	-1.500813
H	-0.023447	-0.758038	-1.447667
H	-1.393926	-0.242253	-1.081383
O	-1.122062	-1.199266	1.416728
H	-1.306164	-1.691674	0.608519
H	-1.442587	-0.304237	1.151481

$(H_2O)_8$ D_{2h}
PMOw

O	1.396594	-1.395887	1.313489
H	1.499211	-1.498508	0.311714

H	2.016240	-2.015323	1.691983
O	-1.395937	-1.396590	-1.313118
H	-2.015434	-2.016287	-1.691560
H	-1.498580	-1.499118	-0.311332
O	-1.395877	-1.396207	1.318119
H	-0.433981	-1.514964	1.480256
H	-1.514954	-0.434321	1.480106
O	1.396435	1.396619	-1.313114
H	1.498922	1.499279	-0.311323
H	2.015976	2.016235	-1.691546
O	1.396489	-1.395829	-1.317735
H	1.515286	-0.433940	-1.479889
H	0.434621	-1.514904	-1.479871
O	1.395930	1.396571	1.318126
H	0.434005	1.515338	1.480117
H	1.514986	0.434699	1.480272
O	-1.396003	1.395846	-1.318147
H	-0.434140	1.514906	-1.480142
H	-1.514773	0.433900	-1.480172
O	-1.396593	1.396177	1.313082
H	-1.499023	1.498701	0.311289
H	-2.016403	2.015606	1.691396

XP3P

O	1.446808	-1.446128	1.445886
H	1.478908	-1.478205	0.438878
H	2.096796	-2.095889	1.693710
O	-1.446071	-1.446800	-1.445534
H	-2.095838	-2.096812	-1.693274
H	-1.478158	-1.478760	-0.438522
O	-1.300990	-1.301305	1.188708
H	-0.372224	-1.444814	1.462881
H	-1.444857	-0.372549	1.462726
O	1.446558	1.446782	-1.445488
H	1.478480	1.478871	-0.438475
H	2.096366	2.096762	-1.693205
O	1.301679	-1.301018	-1.188375
H	1.445233	-0.372249	-1.462512
H	0.372962	-1.444876	-1.462529
O	1.301044	1.301623	1.188748
H	0.372235	1.445168	1.462754
H	1.444867	0.372902	1.462908
O	-1.301233	1.301033	-1.188749
H	-0.372473	1.444857	-1.462773
H	-1.444743	0.372228	-1.462790
O	-1.446795	1.446479	1.445469
H	-1.478730	1.478428	0.438452
H	-2.096824	2.096272	1.693104

Table A-4. Optimized Proton-Water Cluster Geometries, $H^+(H_2O)_n$, with n = 2, 3, 4, and 6, Using PMOw and MP2/aug-cc-pVDZ, For n = 2, the optimized structures also include geometries at the B3LYP/aug-cc-pVTZ and M06-2X/aug-cc-pVTZ levels.

Optimized coordinates of $H^+(H_2O)_2$ using PMOw at various O-O separations.

$R(OO) = R_{min}$ (2.4576 Å), heat of formation = 45.61598 kcal/mol

O	1.230135	-0.000188	0.000377
O	-1.229794	-0.002765	-0.002109
H	0.000505	-0.000188	0.000377
H	1.543881	0.942210	0.000377
H	1.542787	-0.341451	-0.878470
H	-1.540968	-0.945913	0.007665
H	-1.546492	0.347828	0.871509

$R(OO) = R_{min} + 0.2$ (2.6576 Å), heat of formation = 46.74870 kcal/mol

O	0.272179	0.349354	-1.284836
O	-0.319385	-0.392854	1.189367
H	1.280192	0.349354	-1.284836
H	0.034132	1.328985	-1.284836
H	0.006796	0.013158	-0.271651
H	-1.259607	-0.378873	1.446509
H	-0.021758	-1.275991	1.475212

$R(OO) = R_{min} + 0.2$ (2.6576 Å), PT transition state, heat of formation = 48.68019 kcal/mol

O	0.414314	0.514299	-1.147448
O	-0.416353	-0.513012	1.149751
H	1.403175	0.514299	-1.147448
H	0.186758	1.476629	-1.147448
H	-0.000239	0.000953	0.001572
H	-1.405176	-0.519032	1.138657
H	-0.182800	-1.473934	1.152727

$R(OO) = R_{min} + 0.4$ (2.8576 Å), heat of formation = 49.75676 kcal/mol

O	0.252283	0.323444	-1.407005
O	-0.328384	-0.408817	1.285414
H	1.263601	0.323444	-1.407005
H	0.008338	1.304981	-1.407005
H	0.004232	0.006275	-0.410261
H	-1.223509	-0.336556	1.648286
H	0.011426	-1.226248	1.678383

$R(OO) = R_{min} + 0.4$ (2.8576 Å), PT transition state, heat of formation = 57.69576 kcal/mol

O	0.447186	0.557758	-1.232763
O	-0.448670	-0.556411	1.232712
H	1.432277	0.557758	-1.232763
H	0.218344	1.515848	-1.232763
H	-0.000073	-0.000429	-0.000281
H	-1.433692	-0.560811	1.226324
H	-0.215607	-1.513500	1.239528

Optimized coordinates of $H^+(H_2O)_2$ using MP2/aug-cc-pVDZ at various O-O separations.

$R(OO) = R_{min}$ (2.4030 Å), MP2/aug-cc-pVDZ = -152.8448665 hartree

O	-1.199982	-0.040499	-0.063244
O	1.199730	-0.039894	0.063041
H	-0.001959	-0.040154	-0.002051
H	-1.647063	0.781314	-0.326213
H	-1.685637	-0.438396	0.679391
H	1.689531	-0.441535	-0.674769
H	1.647137	0.781913	0.325266

$R(OO) = R_{min}$ + 0.2 Å (2.6030 Å), MP2/aug-cc-pVDZ = -152.841357 hartree

O	-1.301495	-0.060966	0.066212
O	1.298412	-0.022104	-0.055003
H	-0.229171	0.003832	0.026876
H	-1.677886	-0.315384	-0.799075
H	-1.712063	0.771553	0.370127
H	1.810650	-0.555086	0.572313
H	1.833134	0.759649	-0.259912

$R(OO) = R_{min}$ + 0.2 Å (2.6030 Å), PT transition state, MP2/aug-cc-pVDZ = -152.838357 hartree

O	-1.299185	-0.044533	0.078174
O	1.299132	-0.044247	-0.078056
H	-0.002284	0.000331	0.000447
H	-1.717022	-0.460491	-0.697377
H	-1.730257	0.816379	0.222370
H	1.718575	-0.461993	0.695556
H	1.731414	0.816011	-0.221946

$R(OO) = R_{min}$ + 0.4 Å (2.8030 Å), MP2/aug-cc-pVDZ = -152.83470 hartree

O	-1.400250	-0.069285	0.058650
O	1.400409	-0.009511	-0.039880
H	-0.376795	-0.017804	0.036035
H	-1.757746	-0.226889	-0.839202
H	-1.784178	0.748372	0.435579
H	1.949904	-0.616660	0.476388
H	1.967539	0.743349	-0.258959

$R(OO) = R_{min}$ + 0.4 Å (2.8030 Å), PT transition state, MP2/aug-cc-pVDZ = -152.822791 hartree

O	-1.398390	-0.039252	0.093461
O	1.398396	-0.039161	-0.093377
H	-0.001818	-0.009171	0.000251
H	-1.779524	-0.537411	-0.653697
H	-1.783347	0.855992	0.065970
H	1.780479	-0.537779	0.652908
H	1.784160	0.855675	-0.066097

Optimized coordinates of $H^+(H_2O)_2$ using B3LYP/aug-cc-pVTZ at various O-O separations.

$R(OO) = R_{min}$ (2.4052 Å), B3LYP/aug-cc-pVTZ = -153.260528 hartree

O	-1.201264	-0.036875	-0.055861
O	1.201293	-0.036944	0.055859
H	-1.658558	0.759783	-0.361432
H	-1.693939	-0.446218	0.670821
H	1.693887	-0.445608	-0.671253
H	1.658638	0.759504	0.361890

H	-0.000261	-0.036909	-0.000014

R(OO) = R_{min} + 0.2 (2.6052 Å), B3LYP/aug-cc-pVTZ = -153.25599 hartree

O	1.300047	-0.019796	0.047490
O	-1.302593	-0.055143	-0.062447
H	1.817671	0.753394	0.303791
H	1.831548	-0.553522	-0.556686
H	-1.719701	-0.338840	0.770211
H	-1.689456	0.778921	-0.380959
H	-0.219691	-0.040436	-0.016704

R(OO) = R_{min} + 0.2 (2.6052 Å), PT transition state, B3LYP/aug-cc-pVTZ = -153.253757 hartree

O	1.300465	-0.039953	0.075052
O	-1.300412	-0.039924	-0.074975
H	1.712521	0.825332	0.219442
H	1.744558	-0.485979	-0.662890
H	-1.745404	-0.485881	0.662466
H	-1.712123	0.825479	-0.219678
H	0.000026	-0.039938	0.000039

R(OO) = R_{min} + 0.4 (2.8052 Å), B3LYP/aug-cc-pVTZ = -153.24893 hartree

O	1.402252	-0.003709	0.028527
O	-1.401135	-0.067826	-0.049291
H	1.961055	0.738042	0.286689
H	1.960388	-0.642446	-0.429927
H	-1.777168	0.707483	-0.506562
H	-1.792999	-0.186782	0.836307
H	-0.360208	-0.044019	-0.020396

R(OO) = R_{min} + 0.4 (2.8052 Å), PT transition state, B3LYP/aug-cc-pVTZ = -153.23823 hartree

O	-1.399599	0.032195	0.091286
O	1.399654	0.032221	-0.091284
H	-1.779678	-0.858415	0.030936
H	-1.802685	0.584973	-0.597489
H	1.778900	-0.858753	-0.031078
H	1.802998	0.584662	0.597611
H	0.000027	0.032208	0.000001

Optimized coordinates of $H^+(H_2O)_2$ using M06-2X/aug-cc-pVTZ at various O-O separations.

R(OO) = R_{min} (2.3926 Å), M06-2X/aug-cc-pVTZ = -153.189893 hartree

O	-1.195139	-0.034291	-0.050677
O	1.195254	-0.034718	0.050761
H	-1.663005	0.741928	-0.386597
H	-1.696288	-0.449859	0.664618
H	1.695483	-0.446580	-0.667349
H	1.662484	0.741083	0.388599
H	0.000409	-0.034504	0.000057

R(OO) = R_{min} + 0.2 (2.5926 Å), M06-2X/aug-cc-pVTZ = -153.184697 hartree

O	-1.293267	-0.016186	-0.042125
O	1.297058	-0.054652	0.059422
H	-1.817468	0.743703	-0.318580
H	-1.831415	-0.568116	0.536984
H	1.717702	-0.336441	-0.770349
H	1.692792	0.766039	0.396835

H	0.208060	-0.038481	0.016731

R(OO) = R_{min} + 0.2 (2.5926 Å), PT transition state, M06-2X/aug-cc-pVTZ = -153.18268 hartree

O	1.294554	-0.041120	0.067264
O	-1.294553	-0.041099	-0.067287
H	1.718623	0.799510	0.289656
H	1.751719	-0.449910	-0.681914
H	-1.751432	-0.450168	0.681912
H	-1.718919	0.799435	-0.289460
H	0.000001	-0.041110	-0.000011

R(OO) = R_{min} + 0.4 (2.7926 Å), M06-2X/aug-cc-pVTZ = -153.177038 hartree

O	1.395961	-0.000684	0.020787
O	-1.394993	-0.062077	-0.052839
H	1.962540	0.704004	0.349771
H	1.957415	-0.634887	-0.436337
H	-1.789064	-0.269978	0.813954
H	-1.779931	0.742239	-0.445469
H	-0.358704	-0.039282	-0.025502

R(OO) = R_{min} + 0.4 (2.7926 Å), PT transition state, M06-2X/aug-cc-pVTZ = -153.166033 hartree

O	1.394009	-0.035287	0.083884
O	-1.393568	-0.034827	-0.083547
H	1.788415	0.848647	0.107918
H	1.813878	-0.548882	-0.622281
H	-1.815304	-0.551614	0.619105
H	-1.790737	0.847818	-0.107604
H	0.000220	-0.035057	0.000168

Optimized coordinates of $H^+(H_2O)_3$ using PMOw and MP2/aug-cc-pVDZ

PMOw, heat of formation = -44.58622 kcal/mol

O	0.704436	0.255118	0.831904
O	-1.852124	0.126354	0.718342
O	1.333263	-0.348794	-1.558136
H	-0.362606	0.255118	0.831904
H	0.947712	1.223506	0.831904
H	0.954628	-0.056767	-0.168886
H	-2.225832	-0.620844	1.220937
H	-2.393191	0.885007	1.003018
H	0.723241	-0.927056	-2.052678
H	2.199766	-0.786486	-1.659554

MP2/aug-cc-pVDZ = -229.144014 hartree

O	-0.000010	0.920932	-0.057161
O	2.082984	-0.464135	-0.011174
O	-2.082976	-0.464152	-0.011190
H	0.873249	0.366003	0.021242
H	0.000010	1.685049	0.541624
H	-0.873288	0.365985	0.021259
H	2.720711	-0.403287	-0.737518
H	2.572600	-0.775833	0.763476
H	-2.572455	-0.775984	0.763501
H	-2.720815	-0.403096	-0.737378

Optimized coordinates of $H^+(H_2O)_4$ using PMOw and MP2/aug-cc-pVDZ
PMOw, heat of formation = -131.85993 kcal/mol

O	1.707752	1.192487	1.024783
O	-0.682127	-2.004321	0.955938
O	-1.485267	1.144060	-1.371529
O	0.523292	-0.377357	-0.690917
H	2.678901	1.192487	1.024783
H	1.475337	2.135159	1.024783
H	-0.382190	-2.031781	1.878959
H	-0.727371	-2.939600	0.698506
H	-1.521000	1.437657	-2.296533
H	-2.386184	0.829357	-1.192852
H	-0.305049	0.185329	-0.999238
H	1.019379	0.287193	-0.050286
H	0.094575	-1.057792	-0.019299

MP2/aug-cc-pVDZ = -305.436809 hartree

O	0.000000	2.486412	-0.144841
O	-2.153296	-1.243206	-0.144841
O	2.153296	-1.243206	-0.144841
O	0.000000	0.000000	0.489114
H	-0.003174	3.158604	0.551629
H	0.445394	2.892818	-0.901585
H	-2.727951	-1.060686	-0.901585
H	-2.733845	-1.582051	0.551629
H	2.737018	-1.576554	0.551629
H	2.282557	-1.832132	-0.901585
H	0.818472	-0.524574	0.204381
H	0.045058	0.971104	0.204381
H	-0.863530	-0.446531	0.204381

Optimized coordinates of $H^+(H_2O)_6$ (Isomer I) using PMOw and MP2/aug-cc-pVDZ
PMOw, heat of formation = -298.73211 kcal/mol

O	1.747249	-0.842581	1.367204
O	-1.852919	-1.224226	0.377286
O	0.547003	1.752766	1.497460
O	-1.257728	1.298597	-0.608592
O	0.644360	-0.018843	-2.109487
O	0.430361	-2.054792	-0.528456
H	2.714230	-0.842581	1.367204
H	1.495949	0.106320	1.367204
H	-1.934004	-0.313478	0.009798
H	-2.683144	-1.661981	0.143840
H	1.092794	2.551665	1.497058
H	0.131610	1.749463	2.373350
H	-1.822174	2.050008	-0.824598
H	-0.689368	1.617394	0.131228
H	0.519700	-0.106646	-3.064783
H	-0.078045	0.586751	-1.815559
H	-0.532071	-1.845129	-0.159890
H	0.539571	-1.308068	-1.261619

H	1.027405	-1.666554	0.240627

MP2/aug-cc-pVDZ = -458.003233 hartree

O	-1.764653	-1.570429	-0.548998
O	2.050677	-1.060190	-0.787402
O	-1.718260	1.257265	-0.655151
O	1.159677	1.628266	-0.706185
O	0.045582	0.981928	1.749068
O	0.269764	-1.513005	0.959741
H	-2.563133	-2.112455	-0.494785
H	-2.039884	-0.648454	-0.713953
H	2.049624	-0.097903	-0.947981
H	2.959794	-1.375894	-0.875397
H	-1.737911	1.509051	0.282951
H	-2.391194	1.812057	-1.077526
H	1.583923	2.407364	-1.093913
H	0.246853	1.632549	-1.052606
H	0.237567	1.314923	2.637628
H	0.649939	1.450049	1.137073
H	1.050023	-1.433656	0.301325
H	0.195201	-0.615309	1.405345
H	-0.583102	-1.633008	0.403261

Optimized coordinates of $H^+(H_2O)_6$ (Isomer II) using PMOw and MP2/aug-cc-pVDZ

PMOw, heat of formation = -297.09042 kcal/mol

O	-0.956707	0.522573	-1.328682
O	-1.449830	-0.128960	1.145410
O	1.542419	0.561703	-1.293327
O	-1.564085	3.104647	-1.239952
O	2.241593	-1.504385	0.246658
O	0.154905	-2.262461	2.044920
H	0.138658	0.522573	-1.328682
H	-1.187097	1.528921	-1.328682
H	-1.187699	0.245539	-0.349259
H	-2.223477	3.444222	-1.861350
H	-0.844292	3.750227	-1.280015
H	1.914749	0.306971	-2.153838
H	1.845463	-0.184821	-0.686060
H	3.091427	-1.467838	0.699767
H	1.602885	-1.744104	0.947340
H	-0.973273	-0.932110	1.457501
H	-2.373659	-0.297640	1.373306
H	0.317154	-2.237662	2.998544
H	-0.094139	-3.181123	1.869315

MP2/aug-cc-pVDZ = -458.003245

O	-1.163339	-0.511254	-0.688764
O	1.040618	-0.453791	0.449888
O	-3.100929	-1.755757	0.545978
O	3.249204	-1.820582	-0.332338
O	1.174105	2.438547	0.660056
O	-1.283113	2.059738	-0.742564

H	-0.221171	-0.568194	-0.219500
H	-1.376975	0.472867	-0.735943
H	-1.870058	-0.997251	-0.170730
H	1.824818	-0.962378	0.138618
H	1.318802	0.475031	0.547083
H	-3.129087	-2.024556	1.474391
H	-3.633182	-2.408284	0.069878
H	3.731664	-2.406654	0.266186
H	3.482818	-2.120256	-1.221014
H	1.238427	2.785306	1.562656
H	1.847803	2.932927	0.169086
H	-0.561918	2.445751	-0.212624
H	-1.984303	2.720474	-0.806136

Optimized coordinates of $H^+(H_2O)_6$ (Isomer III) using PMOw and MP2/aug-cc-pVDZ
PMOw, heat of formation = -296.43425 kcal/mol

O	1.366113	-0.735842	0.012955
O	-1.108885	-0.652661	-0.001703
O	1.985924	-1.335386	-2.503333
O	-1.995064	-1.422114	2.438326
O	-1.344993	2.138807	0.168437
O	1.488505	1.895050	0.068438
H	0.242243	-0.735842	0.012955
H	1.566061	0.279017	0.012955
H	1.611021	-1.012250	-0.945068
H	-1.441795	-1.015307	0.873014
H	-1.326790	0.312665	0.071139
H	1.403086	-1.946727	-2.975116
H	2.873355	-1.685184	-2.666928
H	-2.834121	-1.895911	2.492712
H	-1.403693	-1.924781	3.011959
H	-1.887740	2.592274	-0.491374
H	-1.651511	2.494957	1.015288
H	0.559182	2.201064	0.014842
H	1.960917	2.430469	-0.580593

MP2/aug-cc-pVDZ = -458.002074

O	-1.163339	-0.511254	-0.688764
O	1.040618	-0.453791	0.449888
O	-3.100929	-1.755757	0.545978
O	3.249204	-1.820582	-0.332338
O	1.174105	2.438547	0.660056
O	-1.283113	2.059738	-0.742564
H	-0.221171	-0.568194	-0.219500
H	-1.376975	0.472867	-0.735943
H	-1.870058	-0.997251	-0.170730
H	1.824818	-0.962378	0.138618
H	1.318802	0.475031	0.547083
H	-3.129087	-2.024556	1.474391
H	-3.633182	-2.408284	0.069878
H	3.731664	-2.406654	0.266186

H	3.482818	-2.120256	-1.221014
H	1.238427	2.785306	1.562656
H	1.847803	2.932927	0.169086
H	-0.561918	2.445751	-0.212624
H	-1.984303	2.720474	-0.806136

Optimized coordinates of $H^+(H_2O)_6$ (Isomer IV) using PMOw and MP2/aug-cc-pVDZ
PMOw, heat of formation = -289.08270 kcal/mol

O	1.211858	-0.250202	0.187565
O	-1.203570	-0.478589	0.023270
O	2.421912	-1.696305	-1.772053
O	-2.346251	-1.336452	2.265958
O	-2.363299	1.721673	-0.917557
O	2.245341	2.260374	0.221294
H	-0.093606	-0.250202	0.187565
H	1.629956	0.658621	0.187565
H	1.631454	-0.809913	-0.526095
H	-1.625289	-0.823974	0.884936
H	-1.678653	0.357172	-0.321706
H	1.903474	-2.317923	-2.295114
H	3.210762	-2.193298	-1.525803
H	-2.950605	-2.087575	2.192099
H	-1.769121	-1.570361	3.005928
H	-2.643014	1.703240	-1.842681
H	-3.126258	2.082517	-0.445549
H	2.802686	2.538522	-0.514930
H	2.736855	2.527484	1.006647

MP2/aug-cc-pVDZ = -457.999948

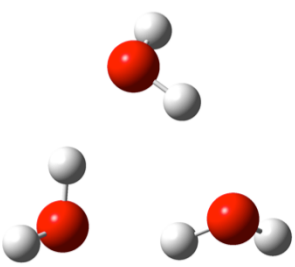
O	-1.114278	-0.390032	-0.440653
O	1.115217	-0.391980	0.442983
O	-2.994632	-1.605088	1.019228
O	2.990552	-1.606695	-1.021394
O	1.956110	1.983598	1.332140
O	-1.952638	1.986994	-1.331989
H	0.002029	-0.356290	0.001911
H	-1.477767	0.484323	-0.731341
H	-1.781962	-0.841218	0.134846
H	1.781341	-0.843410	-0.134471
H	1.479449	0.482486	0.733047
H	-3.024460	-1.709700	1.979699
H	-3.488743	-2.357414	0.666018
H	3.483563	-2.360682	-0.670161
H	3.016740	-1.710973	-1.982019
H	2.302739	2.063899	2.231609
H	2.452506	2.631822	0.813881
H	-2.448557	2.635915	-0.814146
H	-2.299532	2.066863	-2.231395

Table A-5. Computed average thermodynamic properties of water at different temperatures between -40 °C and 100 °C (values in parentheses for C_p and α are obtained from the direct derivative calculations)

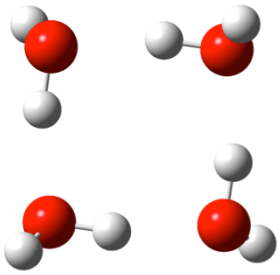
Temperature (°C)	ΔH_{vap} (kcal/mol)	C_p (cal mol^{-1} K^{-1})	ρ (g/cm^3)	$10^5\ \alpha$ (K^{-1})	$10^6\ \kappa$ (atm^{-1})
-40	11.36 ± 0.01	20.8 ± 0.7 (14)	1.008 ± 0.001	1.9 ± 3.5 (-77)	31.7 ± 2.2
-30	11.30 ± 0.01	17.4 ± 0.4 (18)	1.016 ± 0.001	29.8 ± 2.0 (-117)	16.5 ± 0.6
-20	11.16 ± 0.01	19.3 ± 0.5 (22)	1.030 ± 0.001	56.6 ± 3.6 (-39)	22.4 ± 1.2
-10	11.02 ± 0.01	22.2 ± 0.8 (24)	1.024 ± 0.001	30.2 ± 2.2 (50)	23.9 ± 1.3
0	10.83 ± 0.01	21.7 ± 0.7 (27)	1.022 ± 0.001	43.7 ± 2.6 (44)	39.3 ± 2.8
10	10.64 ± 0.01	21.2 ± 1.0 (25)	1.015 ± 0.001	35.0 ± 4.0 (95)	28.2 ± 2.0
25	10.42 ± 0.01	21.8 ± 1.0 (25)	0.996 ± 0.001	36.6 ± 3.0 (105)	25.0 ± 1.6
50	9.96 ± 0.01	25.5 ± 1.4 (28)	0.975 ± 0.001	79.3 ± 6.2 (101)	33.8 ± 2.3
70	9.54 ± 0.01	22.9 ± 1.1 (27)	0.953 ± 0.002	141.3 ± 13.8 (111)	78.3 ± 8.1
100	9.03 ± 0.01	21.8 ± 0.9 (25)	0.923 ± 0.002	107.6 ± 8.0 (105)	76.2 ± 6.8

Table A-6. Computed average dipole moment and DPPC partial charges at different temperatures between -40 °C and 100 °C

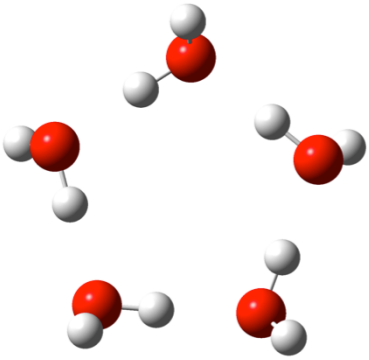
Temperature (°C)	μ (Debye)	DPPC partial charges (e)	
		O	H
-40	2.589 ± 0.003	-0.91892	0.45946
-30	2.586 ± 0.003	-0.91777	0.45888
-20	2.575 ± 0.003	-0.91382	0.45691
-10	2.568 ± 0.002	-0.91116	0.45558
0	2.553 ± 0.002	-0.90572	0.45286
10	2.540 ± 0.002	-0.90122	0.45061
25	2.524 ± 0.002	-0.89532	0.44766
50	2.493 ± 0.002	-0.88410	0.44205
70	2.464 ± 0.001	-0.87366	0.43683
100	2.428 ± 0.001	-0.86046	0.43023



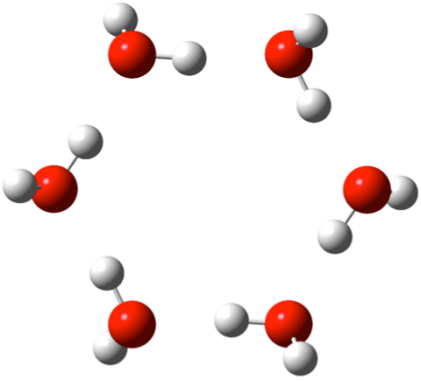
$(H_2O)_3$ cycle



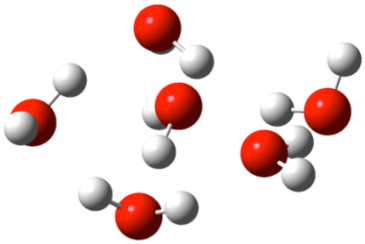
$(H_2O)_4$ cycle



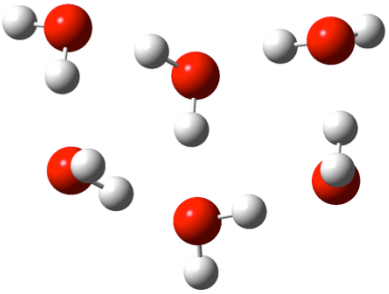
$(H_2O)_5$ cycle



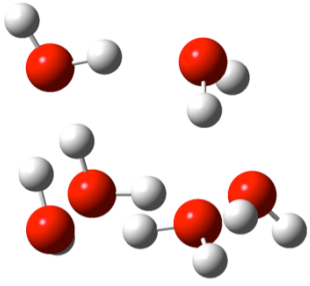
$(H_2O)_6$ cycle



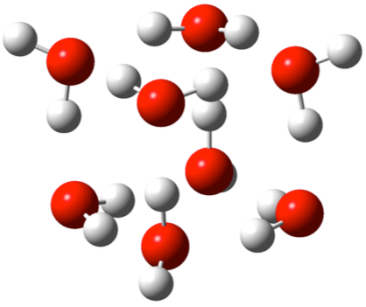
$(H_2O)_6$ cage



$(H_2O)_6$ book

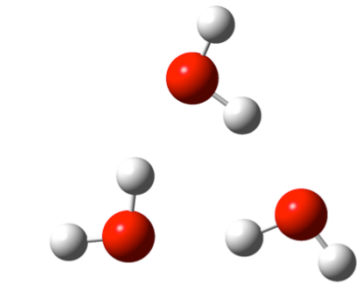


$(H_2O)_6$ prism

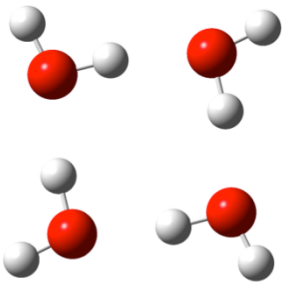


$(H_2O)_8$ D_{2h}

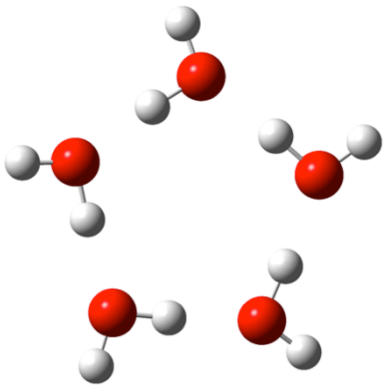
Figure A-1. Optimized water clusters with PMOw.



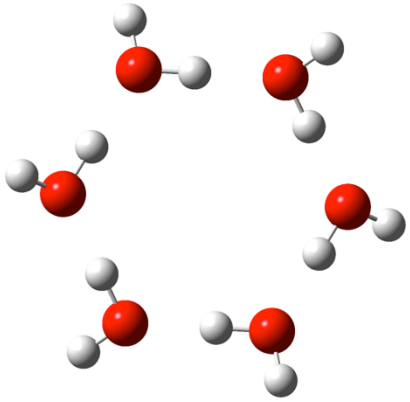
$(H_2O)_3$ cycle



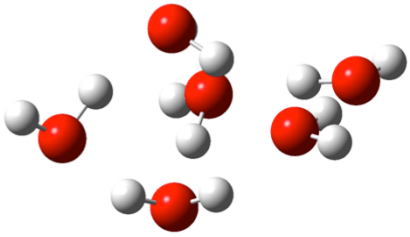
$(H_2O)_4$ cycle



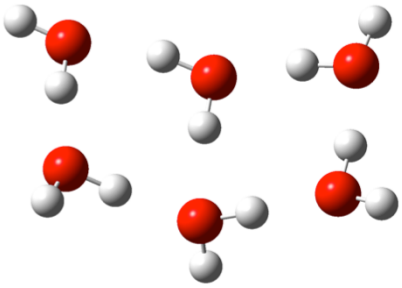
$(H_2O)_5$ cycle



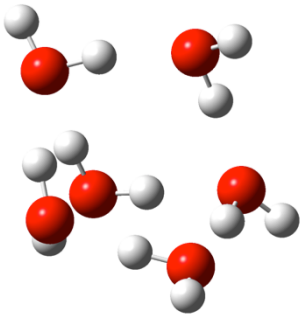
$(H_2O)_6$ cycle



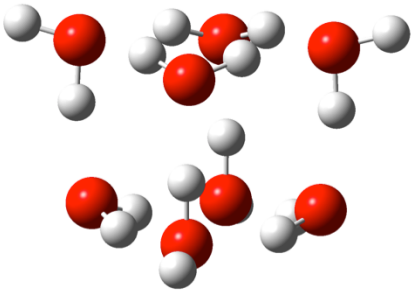
$(H_2O)_6$ cage



$(H_2O)_6$ book



$(H_2O)_6$ prism



$(H_2O)_8$ D_{2h}

Figure A-2. Optimized water clusters with XP3P.

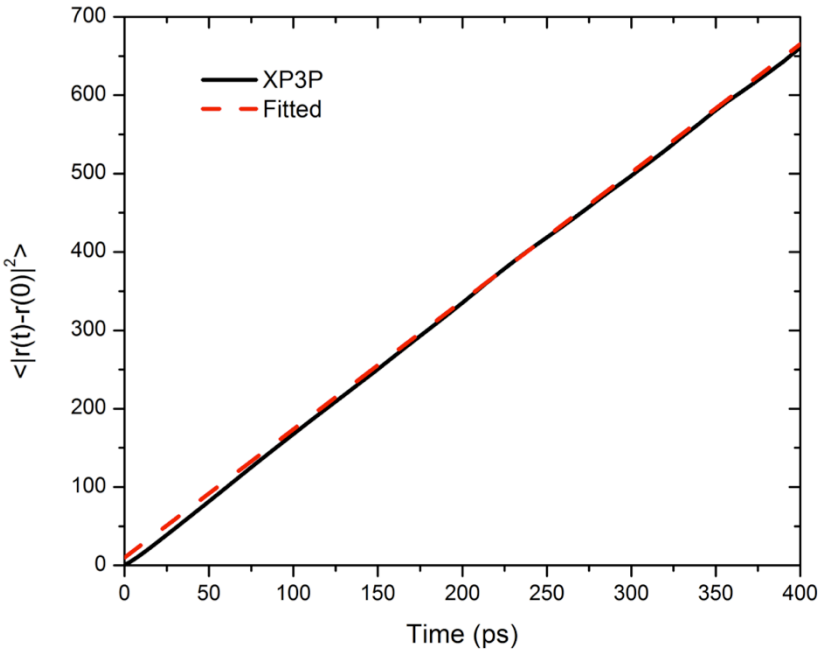


Figure A-3. Computed and fitted mean-squared deviation (MSD) plots with respect to time from XP3P.

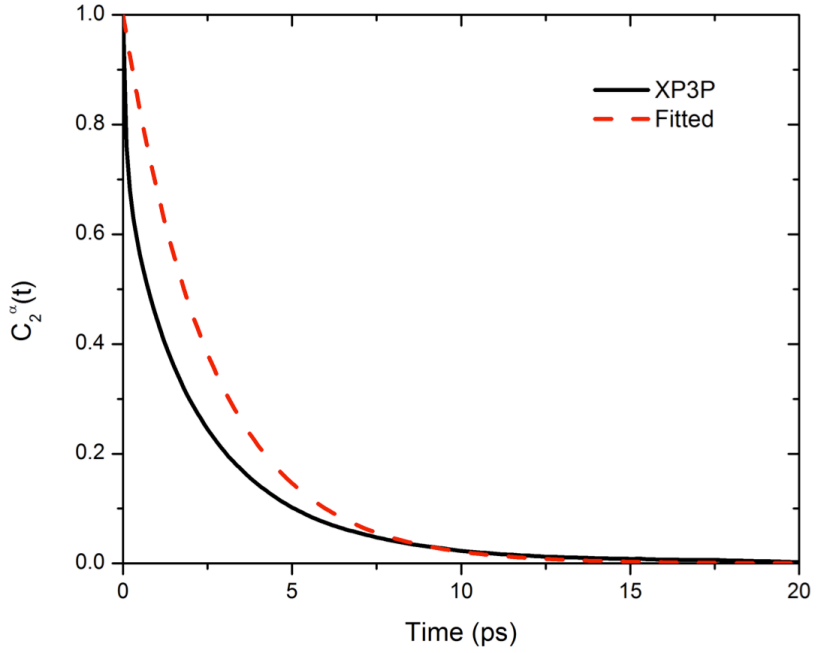


Figure A-4. Computed and fitted orientational time-correlation function from XP3P.

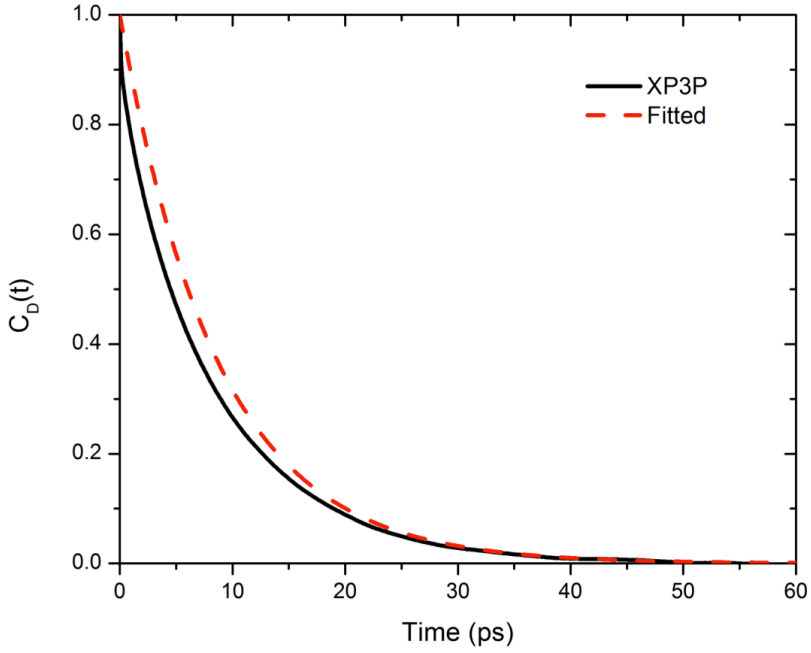


Figure A-5. Computed and fitted dipole autocorrelation function from XP3P.

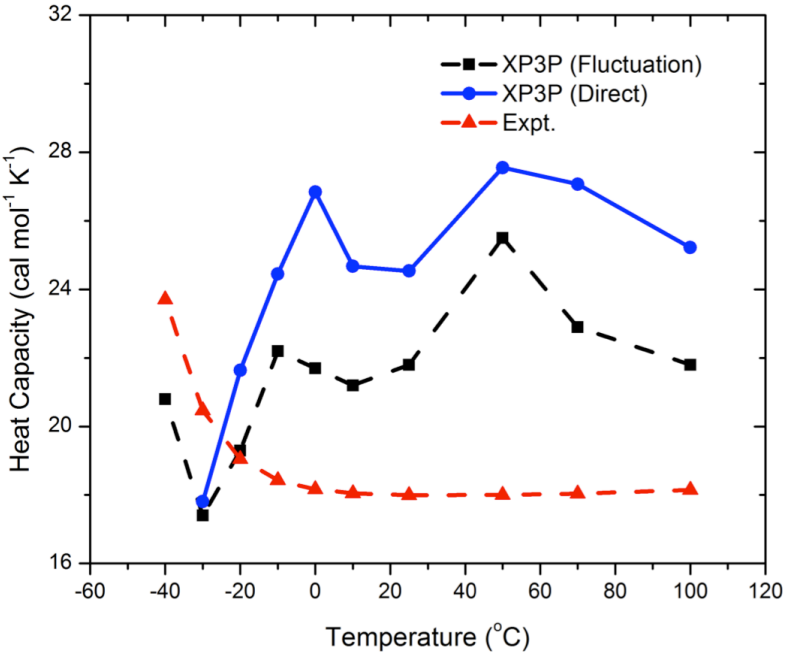


Figure A-6. Computed heat capacities from the fluctuation formula and direct numerical derivatives from XP3P at temperatures ranging from -40 °C to 100 °C, compared to those from experiment.

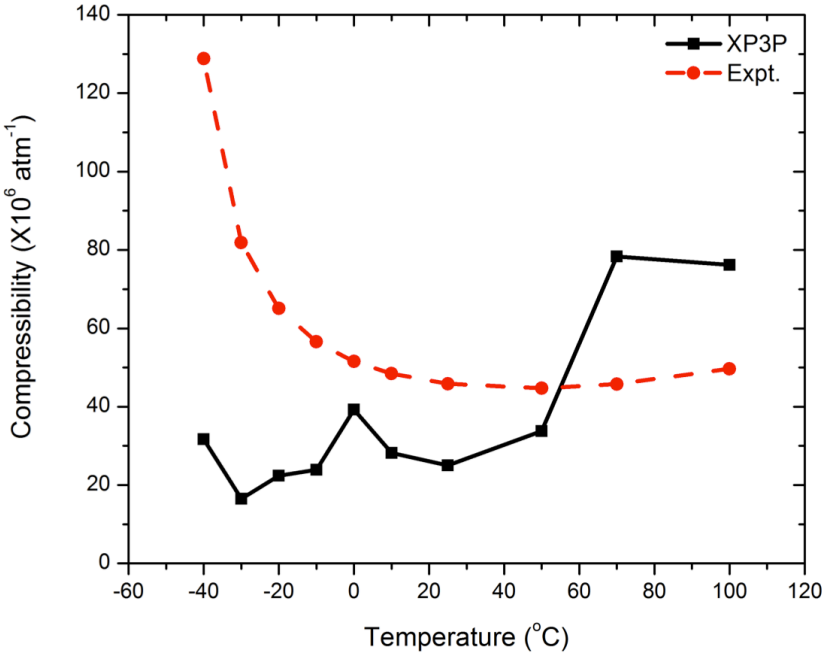


Figure A-7. Computed isothermal compressibilities from XP3P at temperatures ranging from -40 °C to 100 °C, compared to those from experiment.

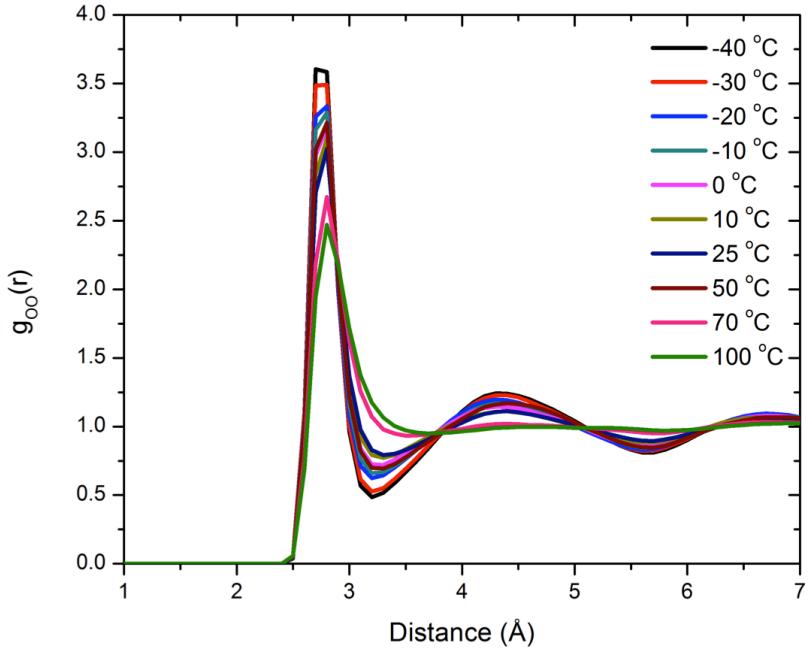


Figure A-8. Computed OO RDFs at temperatures ranging from -40 °C to 100 °C.

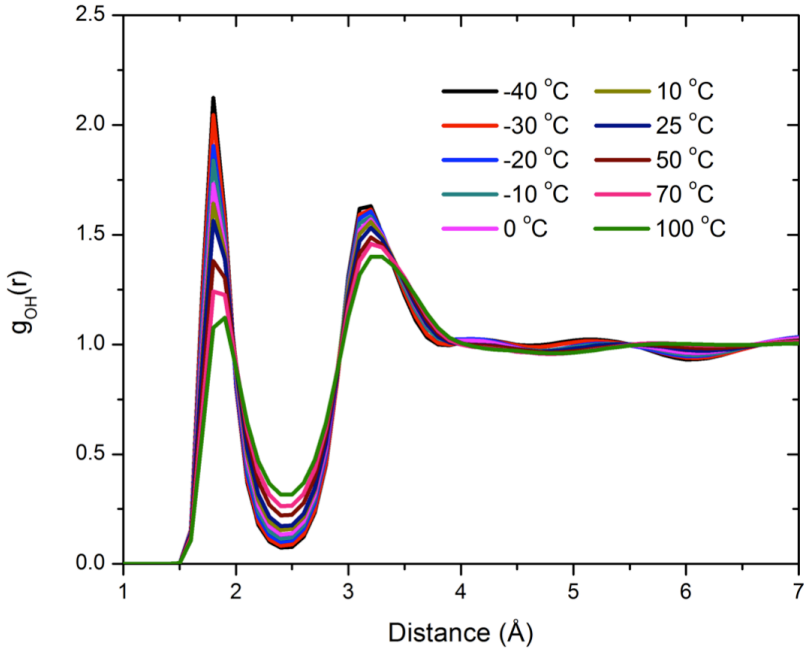


Figure A-9. Computed OH RDFs at temperatures ranging from -40 °C to 100 °C.

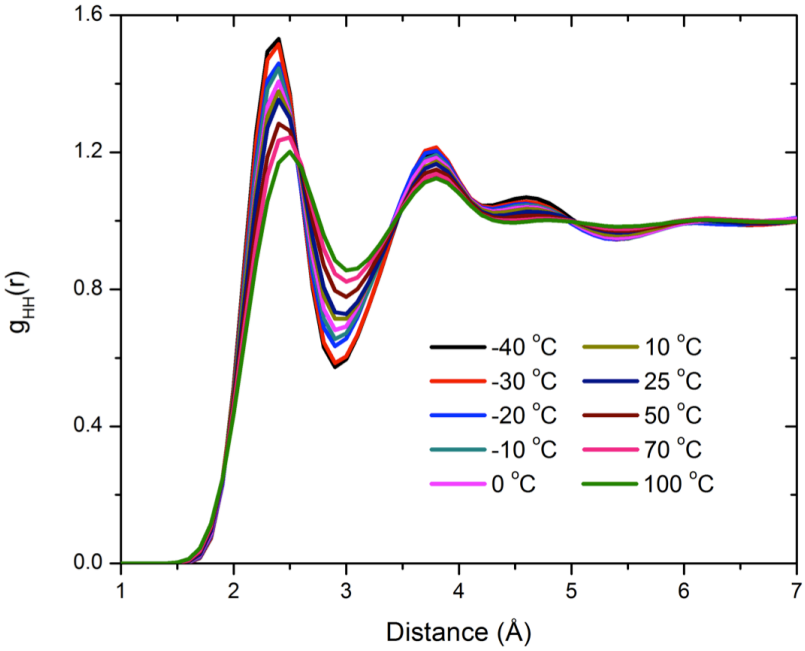


Figure A-10. Computed HH RDFs at temperatures ranging from -40 °C to 100 °C.

References

[1] Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622.

[2] Martin, J. M. L. *J. Chem. Phys.* **1992**, *97*, 5012.

[3] Page, R. H.; Larkin, R. J.; Shen, Y. R.; Lee, Y. T. *J. Chem. Phys.* **1988**, *88*, 2249.

[4] *CRC Handbook of Chemistry and Physics*, 93rd ed.; Haynes, W. M. Ed.; CRC Press: Boca Raton, FL, 2012: Section 9.

[5] *CRC Handbook of Chemistry and Physics*, 93rd ed.; Haynes, W. M. Ed.; CRC Press: Boca Raton, FL, 2012: Section 10.

[6] Dyke, T. R.; Muenter, J. S. J. Chem. Phys. 1973, 59, 3125.

[7] Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddad, W. A., III. *J. Chem. Theory Comput.* **2009**, *5*, 1016.

[8] Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Schaefer III, H. F.; Quack, M. *J. Chem. Phys.* **2002**, *116*, 690.

[9] Bouteiller, Y.; Desfrancois, C.; Abdoul-Carime, H.; Schermann, J. P. *J. Chem. Phys.* **1996**, *105*, 6420.

[10] Neumark, D. M.; Lykke, K. R.; Anderson, T.; Lineberger, W. C. *Phys. Rev. A* **1985**, *32*, 1890.

[11] Ruscic, B.; Wagner, A. F.; Harding, L. B.; Asher, R. L.; Feller, D.; Dixon, D. A.; Peterson, K. A.; Song, Y.; Qian, X.; Ng, C. –Y.; Liu, J.; Chen, W.; Schwenke, D. W. *J. Phys. Chem. A* **2002**, *106*, 2727.

[12] Celotta, R. J.; Bennett, R. A.; Hall, J. L. *J. Chem. Phys.* **1974**, *60*, 1740.

[13] Rosenbaum, N. H.; Owrutsky, J. C.; Tack, L. M.; Saykally, R. J. *J. Chem. Phys.* **1986**, *84*, 5308.

[14] Lee, H. M.; Tarkeshwar, P.; Kim, K. S. *J. Chem. Phys.* **2004**, *121*, 4657.

[15] Liu, J.; Salumbides, E. J.; Hollenstein, U.; Koelemeij, J. C. J.; Eikema, K. S. E.; Ubachs, W.; Merkt, F. *J. Chem. Phys.* **2009**, *130*, 174306.

[16] Sears, T. J.; Bunker, P. R.; Davies, P. B.; Johnson, S. A.; Spirko, V. *J. Chem. Phys.* **1985**, *83*, 2676.

[17] Robertson, A. J. B. *Trans. Faraday Soc.*, **1952**,*48*, 228.

[18] Redington, R. L.; Olson, W. B.; Cross, P. C. *J. Chem. Phys.* **1962**, *36*, 1311.

[19] Cohen, E. A.; Pickett, H. M. *J. Mol. Spetry.* **1981**, *87*, 582.

[20] Flaud, J. M.; Camy-Peyret, C.; Johns, J. W. C.; Carli, B. *J. Chem. Phys.* **1989**, *91*, 1504.