2012-12-05

Model Development and Application of Molecular Simulations for the Study of Proton Transport in Bulk Water and for the Prediction of Dipole Moments of Organic Compounds

Abhishek Asthana
Brigham Young University - Provo

Model Development and Application of Molecular Simulations for the Study of Proton Transport in Bulk Water and for the Prediction of Dipole Moments of Organic Compounds

Abhishek Asthana

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Dean R. Wheeler, Chair
W. Vincent Wilding
Thomas H. Fletcher
Thomas A. Knotts
John Hedengren

Department of Chemical Engineering

Brigham Young University

November 2012

ABSTRACT

Model Development and Application of Molecular Simulations for
the Study of Proton Transfer in Bulk Water and for the
Prediction of Dipole Moments of Organic Compounds

Abhishek Asthana
Department of Chemical Engineering, BYU
Doctor of Philosophy

The present work demonstrates the application of molecular simulations (MD) in two different areas: proton transport in bulk water and estimation of the dipole moment of polar organic compounds. In both areas, relatively few successful and robust methodologies exist.

In the first part, a new polarizable water model is developed for MD simulations of the proton transport process. The model was parametrized from a combination of quantum chemical calculations and experimental water properties. The model was implemented in MD simulation studies of liquid water at room temperature, as well as with excess protons. For pure water the model gave good agreement with experimental properties. The proton transport rate for a single excess proton also gave a good match with the experimental value. The water model was further extended to include chloride ions. At 0.2 M concentration the resulting density and structure agreed well with experiment, and the proton transport rate was found to be slightly reduced. The model was further extended to include multiple excess protons.

For the second part of the project, an open source *ab initio* MD program, SIESTA, was used to perform simulations of several organic compounds which potentially have multiple stable conformations, to determine their average dipole moments. A series of methods was developed. The most robust method involved modifications to the SIESTA code and statistical analysis of the resulting configurations, in order to more accurately predict the average dipole moment. The resulting dipole moments were in good agreement with the experimental values for cases in which experimental values were reliable. Based on this study, a general method to estimate the average dipole moment of any compound is proposed.

Keywords: molecular dynamics, water model, proton transfer, ab initio, dipole moment, conformational isomers

ACKNOWLEDGMENTS

I gratefully acknowledge the help and guidance of my advisor Dr. Dean R. Wheeler. He taught me how to deal with complicated problems with a methodical approach. To Dr. W. Vincent Winding, Dr. Richard L. Rowley, and Dr. Thomas A. Knotts for providing helpful resources and discussions related to this work. I also acknowledge the support of my parents and my wife Megha. To my friends, especially Dr. Sonal Patel, for their support during the tough times of my PhD. This work was supported by National Science Foundation and the DIPPR Laboratory.

Contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Motivation

Molecular dynamics (MD) simulations have been used in the last few decades to study multiple processes of industrial importance. What makes them so useful is that they enable an atomic-level analysis of a system and allow for the calculation of thermodynamic, transport, and dielectric properties based on how the dynamics of the entities of the system evolve in time. The present work illustrates the application of MD simulations in two areas. Firstly, MD simulations have been used to perform an in-depth investigation into the proton transport process in bulk water to better understand the solvation and dynamics of the excess protons during the process. Secondly, MD simulations have been applied to improve existing predictions and estimations of dipole moments of polar organic compounds by sampling multiple accessible conformations.

1.1.1 Proton transport in bulk water

Proton transfer (PT) in aqueous media plays an important role in fundamental biological and chemical processes including bio-energetics, cell signaling, and acid base chemistry. The PT process also accompanies numerous complex chemical reactions of industrial importance. These reactions are generally heterogeneous and involve multiple electrons transferred between the solution and the catalyst surface. An elementary step involves one or no electron being transferred, so there are multiple elementary steps and numerous intermediates for a

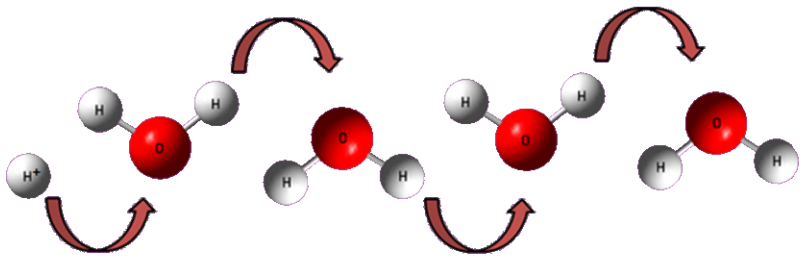
typical electrochemical reaction. This makes the experimental observation of the reaction mechanism difficult. In addition, there are other factors in an electrochemical environment, like the presence of an electric double layer that influences a surface reaction and further hinders an accurate experimental analysis. Theoretical methods to study PT and the subsequent complicated reactions provide useful insight into studying these phenomena.

Our particular interest in studying PT in aqueous solutions comes from the fact that a realistic depiction of PT is a necessary precursor to the analysis of the oxygen reduction reaction (ORR) [1] that takes place on the cathode of low-temperature fuel cells. The slow kinetics of the ORR is responsible for the low efficiency of fuel cells. The ORR is a 4-electron reduction which suggests complex surface phenomenon with multiple elementary reactions. Gattrell and MacDougal [2] suggested that the first step of the ORR is the rate limiting. In spite of the immense importance of this reaction, it is still not understood well.

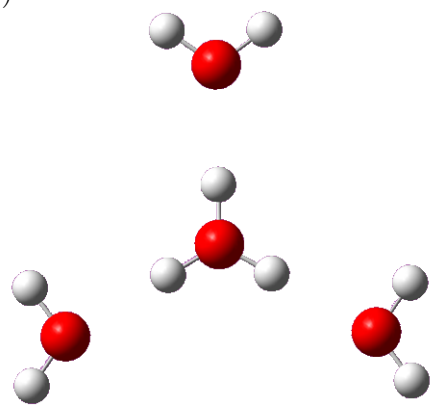
Diffusion of protons through bulk water occurs via two mechanisms: vehicular diffusion and structural diffusion. Vehicular diffusion involves the diffusion of a hydronium ion, as a whole, through liquid water. Structural diffusion on the other hand, as shown in Fig. 1.1(a), involves cleavage of covalent bonds, as a proton is transferred through a network of water molecules. This mechanism of PT is referred to as the Grotthuss mechanism. Most of the diffusivity of PT in bulk water is due to the Grotthuss mechanism and is much faster than vehicular diffusion. Thus, when compared to the diffusivity of other small cations (like K^+, Na^+, and NH_4^+), a proton has abnormally high diffusivity.

In liquid water, the solvation of an excess proton is idealized by two forms: Eigen ([$H_3O(H_2O)_3$]$^+$) [3] and Zundel ([$H(H_2O)_2$]$^+$) [4] cations. As shown in Fig. 1.1(b) and (c), in an Eigen cation, a hydronium ion is coordinated with 3 water molecules and in a Zundel cation, a hydronium ion is coordinated with one water molecule. In the presence of excess protons in bulk water, both Eigen and Zundel states are observed, and the Grotthuss mechanism involves an interconversion between these two solvation structures. The finer details of the hopping and transport mechanism is still debated, however [5].

(a)



(b)



(c)

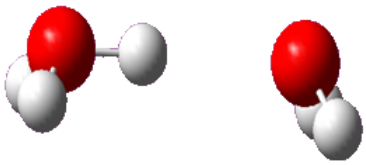


Figure 1.1: (a) PT via Grotthuss mechanism (b) The Eigen cation (c) The Zundel cation

Multiple experimental studies have been performed in order to better understand the solvation of protons and the mechanism of PT. After the discovery of Eigen and Zundel cations, mono-protonated water clusters ($[H(H_2O)_n]^+$) , were studied using vibrational predissociation spectroscopy [6–9], in an attempt to determine the relative occurrence of the Zundel and Eigen ions. NMR spectroscopy studies [10–18] were employed in order to determine the activation energy and rate constants for the PT reaction. Interconversion between Zundel and Eigen structures has also been observed experimentally [19]. Despite these extensive

experimental studies, doubt still remains on the relative population of the Zundel and Eigen cations [20, 21]. Also, the rate-limiting step of PT needs further investigation [22–24]. A detailed simulation study of PT, that accurately models bulk water behavior and adequately represents the PT reaction, can provide an additional insight into the PT mechanism, and help us understand the factors affecting the process better.

Studying electrochemical phenomena is aided by the development of a robust force-field for water that captures the electrostatics of the system accurately and responds realistically to the electrochemical environment. In addition, the force-field must also allow for bond breaking and forming. Existing force-fields for the atomic-level computer simulation of reactive systems either do not capture the bulk behavior accurately, or do not handle multiple excess protons, or are not transferable enough to be extended to heterogeneous reactions. The purpose of this work is to develop a force-field to simulate large systems with multiple excess protons, which captures the essential physics of the electrochemical system, and can be easily extended to heterogeneous reactions.

1.1.2 Estimation of dipole moments

The dipole moment of a molecule is a measure of separation of positive and negative charges and hence, a measure of the system polarity and the strength of the electric field a molecule may exert on its neighbors. Even non-polar molecules can have an induced dipole when placed in an electric field. Dipole moments result from non-uniform distributions of charges on the various atoms, which manifests as electron density being shared unequally between the constituent atoms. In the presence of an external field, these dipoles tend to align in order to attain a minimum energy configuration.

The dipole moment is one of the defining characteristics of a molecule. It has a large effect on the properties of fluids, as interactions between dipoles on different molecules contribute significantly to the intermolecular interactions in dense systems. Also, the dipole moment lies at the heart of multiple experimental methodologies to study system characteristics like

vibrational dynamics and conformational equilibrium [25]. The dipole moment is also used in prediction of dielectric constant [26], which is an important property for understanding solvation of ions and response to external electric fields. The dipole moment has also been used in several studies as an important parameter to determine molecular structure, bond angles, and resonance [27, 28]. Thus, the dipole moment of a molecule is an important property to know and hence it is desirable to accurately calculate or predict it for various compounds of interest.

The existing experimental techniques for determination of dipole moments can be divided into two categories. The first category of techniques is based on estimating dipole moments using measured dielectric constants. These techniques include the method postulated by Debye [29, 30] with modifications and improvements by Onsager [31] and others [32–36]. This category of technique is used to determine dipole moments in vapors and dilute solutions of polar substances in non-polar solvents. Although this method has been successfully used to predict approximate values of dipole moments of numerous molecules, it is less accurate for highly polar compounds and for compounds that have strong intermolecular interactions, like hydrogen bonds in the case of alcohols and water.

The second category of experimental techniques for determination of dipole moments is based on microwave spectroscopy and molecular beams [37, 38]. The accuracy of the microwave spectroscopy method, which is based on the Stark effect [39], depends upon the measurement of electric field strength, which entails the measurement of effective spacing between the electrodes. The spacing cannot be determined easily and hence the capability of this method are limited. The molecular beam method, which is similar in principle to the microwave spectroscopy method, offers the highest inherent accuracy of any known method of dipole moment determination. Unfortunately, the range of molecules that this method can be applied to is very limited - mostly, due to inherent complexities in more structured molecules, this method can only be applied to simple molecules, like diatomics and a few triatomics.

Due to the above mentioned limitations of experimental techniques, we use theoretical methods for calculation of dipole moments. The theoretical method involves calculation of accurate electronic structure around the atoms of a molecule, which gives an accurate picture of the charge distribution in the molecule, thus enabling us to calculate an accurate dipole moment.

Traditionally, the theoretical estimation of dipole moments of compounds has been based on optimized geometries of molecules, meaning a single conformation believed to be the ground state. However, at room temperature, many molecules will access multiple conformations that each have different dipole moments. Thus, we investigated the possibility of improving these estimations by considering time-averaged geometries instead of just the optimized geometry. In order to do this, we used first principle MD or *ab initio* MD simulations (AIMD), which enable the calculation of accurate electronic densities around moieties of a molecule, which are then used to calculate the dipole moment. Various polar organic compounds with different functional groups were included in the study for a comprehensive analysis.

1.2 Scope of Work

As stated previously, this work demonstrates the application of MD simulation in two fields: PT studies and dipole moment prediction. A comprehensive analysis of the PT process constitutes the bulk of the present work, and the remainder focuses on application of AIMD simulations for dipole moment calculations.

1.2.1 PT studies

A novel force field methodology has been developed that allows for the simulation of large systems with multiple excess protons. The philosophy behind the present work was to start from a polarizable charge density description on molecules that ensures reasonably accurate

electrostatics in diverse environments, and let the reactions take place naturally under the influence of appropriate stimuli.

The force field for water proposed here treats hydrogen and oxygen as separate entities that can associate to form molecules or dissociate in the presence of a stimulus (like an excess proton nearby). This approach enables us to describe the molecule in terms of flexible pair potentials, and thus the molecule can undergo changes in its geometry and dipole moment in the presence of solutes, surfaces, and variations in electric field.

In our model, diffuse and point charges are combined (see Fig. 2.2) with an Ewald (lattice) sum for an accurate description of short and long-range electrostatic forces. Short-range electrostatics are more accurately described by the use of diffuse charges, which lead to damped Coulombic interactions. This is especially important for intramolecular interactions on the order of 0.1 nm. The diffuse charge density, as described in the next section, is polarizable and contains charge, dipole, and quadrupole components. Bond making and breaking, including electron transport, occurs naturally under the framework of a consistent set of rules.

The model was implemented into the MD framework and was validated via reproduction of bulk water properties. A thorough analysis was then performed for a system representing liquid water with a single excess proton. The methodology was then extended to study bulk water systems with multiple excess protons, and the dynamics and solvation of the protons were studied. The model was then further extended to include counterions in our simulations, and the effect of their presence on the PT process was studied. The Model can be easily extended to study surface reactions too, but these analyses were beyond the scope of the project.

1.2.2 Dipole moment predictions

A new methodology was developed in order to predict dipole moments of various compounds. SIESTA [40], a free open source AIMD program, was used to perform these calculations. The

aim of the endeavor was to find the dipole moment based on time-averaged geometries, or, to calculate dipole via inclusion of various conformations of the same compound.

As a first step, certain polar organic molecules were selected. Using the SIESTA AIMD code, the evolution of geometry of these molecules was studied and these geometries were used to calculate dipole moment based on partial charges assigned in accordance with a charge assignment scheme. This study was performed for tributylphosphate (TBP). As the next step, the SIESTA code was modified to calculate dipole moment based on the actual electronic structure on the atoms of the molecules. In addition to TBP, this study is performed for four more molecules: 1,2-dichloroethane (DCE), 1,2-dinitrioethane (DNE), 1,2-ethanediol (EDL), and 1,1,1,2-tetrafluoroethane (TFE). The method was then validated for select conformations generated by our simulations by comparing the dipole moment values calculated for the same using computational chemistry software. Where available, the comparison was also made to experimental values of dipole moments.

1.3 Outline

The remainder of this document is organized as follows.

Background. Chapter 2 is a brief description of the existing methods used to study reactive systems in aqueous media. Also, a description of recent advances in development of new reactive force-field for water has been included.

*Properties of bulk water and Preliminary PT studies.*Chapter 3 describes the development of a force-field model for simulation of bulk water. Different aspects of the water model and its implementation to perform MD simulation are discussed in detail and the properties obtained from the simulations are reported.

PT with multiple protons and counterion. Chapter 4 explains the extension of the water model to studying PT process in bulk water, with one excess proton. The extension of model

to include multiple excess protons and counterions is also discussed and the results of the simulation are reported.

Dipole Moment studies. Chapter 5 gives a detailed description of the methodology used to estimate the average dipole moment of various polar organic compounds. The simulation tools and the statistical-mechanical analyses used to develop the method are also described in detail.

Conclusions. Chapter 6 presents the conclusions drawn from this work as well as some possible future extensions and applications of the model and methods developed in the study.

Chapter 2
Background

2.1 Molecular Dynamics Simulation

Molecular Dynamics (MD) is a category of computer simulation in which constituent particles of the system are allowed to interact for a particular time under Newtonian laws of motion. This method has been used for decades to predict thermophysical properties of gases and condensed matter, to understand and analyze atomic level phenomena and processes such as the solvation of an ion in aqueous solution and diffusion of ions near a surface.

It is beyond the scope of the present work to explain all aspects of the MD simulation technique required to write a computer program to simulate bulk liquid systems. Nevertheless, several key concepts will be introduced in this chapter as a basis on which following chapters will elaborate. Specifically, the concepts introduced here briefly explain what the boundary conditions of the simulation are; the level of detail of the phenomenon being studied that is accessible to us via classical and AIMD simulations; how the forces and potentials in classical MD and AIMD simulations are calculated; and how these forces are used in the equations of motion used to advance the dynamics of system in time. Additional details can be obtained from standardized texts and other sources [41, 42]. In addition, Appendix A contains the code that was used for this research, which may be examined for programming details.

In the MD simulation technique, we simulate a collection of particles bouncing around in a box called the "unit cell". These particles may be atoms, molecules, ions, or clusters of

these, and the particles may be modeled as one entity or may be split into discrete centers, which are called "sites". The general algorithm of MD technique entails the definition of a functional form of a potential for the system which is used to interact with the sites of a system. The forces thus calculated from these interactions are used to obtain new positions and velocities of the constituent sites, in accordance with Newton's second law. Thus, the problem breaks down into the solution of $3N$ coupled second-order ordinary differential equations (ODEs), where N is the number of sites, and the ODEs that must be solved have the form:

$$\mathbf{f}_i = -\frac{\delta U}{\delta \mathbf{r}_i} = m_i \frac{\delta^2 \mathbf{r}_i}{\delta t^2} \tag{2.1}$$

where $\mathbf{f}_i, m_i, \mathbf{r}_i$ are respectively the force, mass, and position vector of the site i, t is the time, and U is the potential energy of the system, which is equal to the sum total of all the interaction energies of the constituent species.

2.1.1 Initial and boundary conditions

Unless the initial setting of the constituent sites of a system are in an improbably high energy configuration (which may cause a simulation to crash), MD is generally insensitive to initial conditions. Since there are $6N$ degrees of freedom ($3N$ position and $3N$ momenta), where N is of order of hundreds, the system is sufficiently random that a trajectory of several picoseconds is sufficient to equilibrate the system and make it forget the starting configuration. Thus, any configuration (provided unnatural overlap between the atoms is avoided), may be used as a starting configuration. The simulation properties, however, should only be recorded after the system is well equilibrated, that is, when the potential energy of the system converges to (or oscillates about) a long-term equilibrium value.

Periodic boundary conditions (PBCs) are used at the faces of the unit cell as the boundary condition. PBCs minimize wall effects and are preferable for obtaining bulk properties.

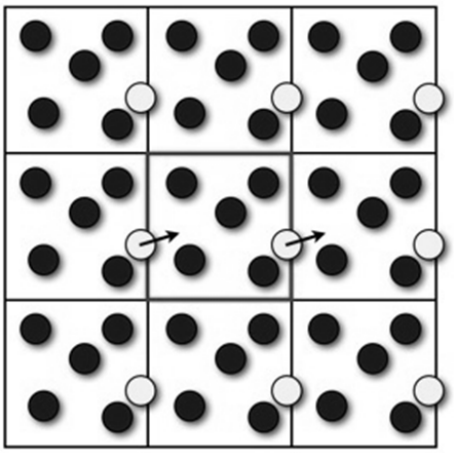


Figure 2.1: 2-dimensional representation of periodic boundary conditions [43]

Nevertheless, the cell length still needs to be large enough to capture the important short-range heterogeneties, typically around 2 nm for liquid simulations of small molecules.

PBCs are often used to simulate large systems by modeling a small part that is far from its edge. To simulate bulk behavior, each of the unit cell is surrounded with copies of itself, as shown in Fig. 2.1.

The centrally placed unit cell is surrounded by its exact images in all directions, so that the sites in that unit cell interact with sites in that unit cell, as well as the sites in the surrounding image cells. In addition, as indicated by arrows in Fig. 2.1, when a molecule wanders off outside a face of the unit cell, it is replaced by one of its images moving into the unit cell at the opposite face. In this way, a relatively small number of sites in a unit cell can imitate the bulk behavior.

2.1.2 Representation of system and the level of detail

The functional form of a force field model, which enables the calculation of U and $\mathbf{f}_i$(Eq. 2.1), encapsulates the physics of the system. The more accurately the model captures the physics of the system, the more realistically it represents the system at the atomic and the sub-atomic level. There are different ways to incorporate the physical details of a system into

the MD framework. These methods enable us either to represent the physical phenomena with an approximate formulation, or to include additional detail of the electronic structure around the atoms into the MD framework.

A realistic description of a system is very desirable but it can lead to an excessively computationally intensive force-field. For example, in AIMD simulations, the MD framework can be used in conjunction with quantum mechanics to find the accurate electronic cloud structure for all constituent sites of a system, at each step of MD. This kind of accuracy is desirable when one is studying chemical reactions because this method incorporates a better representation of electronic behavior and allows covalent bonds to be made and broken. Also, when a detailed analysis of electric or magnetic moments needs to be performed, an accurate description of electron cloud structure can help provide valuable insight. Such calculations, although very accurate, are computationally very expensive and do not allow us to access larger length-scales and longer time-scales. In fact, the AIMD studies are generally limited to simulating small systems (<100 atoms) for shorter periods of time (a few picoseconds).

As an alternative to AIMD, one can study a system using classical MD simulations, where one defines a pair-potential describing the interaction energy of the system. These pair potentials can be simplified or complicated, depending upon how much of the physics of the system we are able to incorporate. The classical MD simulations with simplified pair-potentials allow us to access larger systems and longer time-scales, but, traditionally, they do not allow for bonds to break or form, and contain little information about the electronic structure. For applications where electronic density is important, like when studying large reactive systems, we need to be able to include, in the force-field model, a realistic description of physical phenomenon occurring during the reactive event. For this purpose, the pair-potential expressions are modified, which results in a more complex expression for interaction of sites of the system.

2.1.3 Forces and potentials in classical MD

In classical MD, force on a site i equals the negative gradient of potential energy of the unit cell with respect to the position of particle i, and is calculated in accordance with Eq. 2.1. The potential of the unit cell in classical MD is taken to be the total of pairwise interactions between unique pairs of sites and images of the site:

$$U = \sum_{i} \sum_{j \neq i} u_{ij}(r_{ij}) \tag{2.2}$$

where i and j range from 1 to N. $u_{ij}(r)$ is a spherically symmetric potential function and r_{ij} is the *minimum image distance*, or the distance between site i and the nearest site or image j. The pairwise interactions are spherically truncated, at some value of r. For a cubic unit cell of length L, the pairwise interactions are truncated at $r_c = L/2$. An Ewald sum is used for long-range Coulombic interactions, as is discussed in Chapter 3.

Traditionally, a simplified pairwise potential, such as Lennard-Jones (LJ) potential [44], along with Coulombic point charges (due to the charges placed on the sites of system) are used to approximate the interaction between a pair of sites. The pairwise potential used in the present study has diffuse and point charges combined (see Fig. 2.2) with an Ewald (lattice) sum for an accurate description of short and long-range electrostatic forces. Short-range electrostatics are more accurately described by the use of diffuse charges, which lead to damped Coulombic interactions. This is especially important for intramolecular interactions on the order of 0.1 nm. The diffuse charge density, as described in Chapter 3, is polarizable and contains charge, dipole, and quadrupole components. Bond making and breaking, including electron transport, occurs naturally under the framework of a consistent set of rules.

Many methods are available in the literature for the incorporation of "many-body" effects in the formulation of inter-atomic potentials [45–48]. These methods attempt to approximate

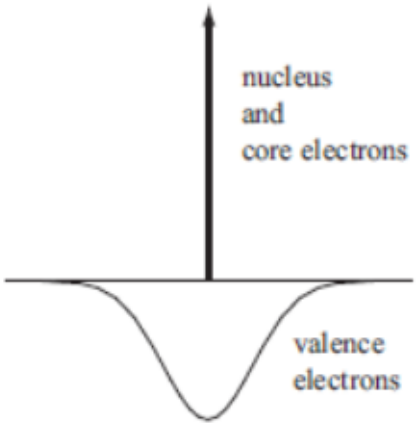


Figure 2.2: Illustration of generalized charge distribution around each atomic site.

the many-body effects by including more than two atoms/sites in the formulation of potentials, like the Axilrod-Teller three-body potential [49]. It is to be noted that Coulomb's law is always true and the interaction energy between charges in any system can still be described by pairwise interactions between sites. However, the presence of other atoms/molecules changes the distribution of electronic charge. So, if a model is able to correctly mimic the changes in charge distribution, then the model, in effect, can approximate the effect of the presence of other atoms/molecules. In the present work, the pairwise model allows for variation in shape, size, and magnitude of charge density, thus allowing for changes in charge distribution due to the presence of other atoms/molecules. Hence the model should capture the most significant aspects of the many-body effects.

The pairwise potentials are parametrized in advance by *ab initio* calculations, empirical adjustments, or a combination of the two. Parametrization of the components of the pairwise potentials is discussed in detail in Chapter 3.

2.1.4 Forces and potentials in AIMD

In classical simulations, a single potential energy surface (PES) is represented in the force field. This PES corresponds to the ground state configuration of the system. In excited states,

chemical reactions, or when studying electric or magnetic dipole or multipole moments, a more accurate representation of electron density is needed. In such cases, the electronic structure can be obtained from first principles by using a quantum chemistry method, such as Density Functional Theory (DFT). This class of simulations, where potential energy of a system is calculated on the fly (that is, accurate electronic structure is determined at every MD step to calculate the potential energy), is called the AIMD simulation. The following development defines relevant equations and terms, and illustrates various techniques used to obtain the correct electronic structure around the sites of system, in an AIMD simulation.

2.1.4.1 The Schrödinger equation

The determination of accurate electronic structure entails the solution of the time independent Schrödinger equation:

$$\hat{H}\Psi = E\Psi \tag{2.3}$$

Here, $\hat{H}$ is the system Hamiltonian, E is the total energy of the system, and Ψ is the wavefunction describing the stationary electronic state, implying that the Ψ is independent of time. The wavefunction Ψ that satisfies Eq. 2.3 is a probability amplitude in quantum mechanics describing the quantum state of a particle and how it behaves. The quantity $|\Psi|^2$ corresponds to the probability density of finding a particle in a given place, if the particle's position is measured. The Hamiltonian operator $\hat{H}$ can be split into kinetic and potential energy operators:

$$\hat{H} = \hat{T} + \hat{V} = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \tag{2.4}$$

where m is the particle's mass, V is its potential energy, ∇^2 is the Laplacian, and $\hbar$ is the reduced Plank's constant. Eq. 2.4 is true for a single particle. If there are N particles in a system, the corresponding equation for system Hamiltonian becomes:

$$\hat{H} = \sum_{i=1}^{N} -\frac{\hbar^2}{2m_i}\nabla_i^2 + V(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) \tag{2.5}$$

where ∇_i^2 is the Laplacian operator specific to the particle i, m_i is the mass of particle i, and the potential energy V is now a function of spatial positions of all N particles. It is to be noted here that under the Born–Oppenheimer approximation [50], the nuclei of the molecules/atoms are seen as fixed, generating a static external potential in which the electrons are moving. This approximation is justified because nuclei are much heavier than electrons, and hence move much slower. Thus, the Eq. 2.3 is only solved for the electrons of the system.

2.1.4.2 Solution of the Schrödinger equation

In an N electron system, the wavefunction Ψ that satisfies Eq. 2.3 (with the Hamiltonian given by Eq. 2.5), is the true wavefunction of the electrons and E the energy of the system, which may be in ground state or excited state. The potential energy V in Eq. 2.5 includes the potential energy of electrons moving in a constant external field of nuclei and the electron-electron interaction energy. While the interaction between nuclei and electrons is separable into simpler single particle equations, the electron-electron coulomb interaction is not separable, which makes the solution of the Schrödinger equation complicated for a system with more than one electron.

The electron-electron Coulombic interaction may be split into two parts: electron exchange and electron correlation. The electron exchange is due to the Pauli's exclusion principle, which states that no two electrons with the same spin may occupy the same quantum state. To understand the correlation term, imagine that for simplicity of computation, the instantaneous electron-electron repulsion was replaced by the repulsion of each electron with an average electron charge cloud. This approach for solving the Schrödinger equation was proposed in 1927 and is known as the Hartree approach. Although this approach ensures that individual position vectors of each electron do not need to be considered, it introduces

an error in the calculated wavefunction and energy. The error is energy, that is, the difference between the energy using Hartree approach and the exact energy of the system is called the total correlation energy.

The Hartree approach did not consider any electron exchange or correlation. The method was improved by Fock, who proposed using Slater determinants [51] for the expansion of the wavefunctions, enabling the inclusion of antisymmetry (Pauli's exclusion principle). The Hartree-Fock (HF) method was thus formulated and it included electron exchange but no electron correlation was considered. Numerous post-Hartree-Fock methods, like configuration interaction (CI) [52, 53], coupled cluster (CC) [54], Møller–Plesset perturbation theory (MP2, MP3, MP4) [55, 56], were developed in an attempt to improve on the HF method by including electron correlation, but the problem with these methods is that they require huge amount of computational effort, which makes it virtually impossible to apply them efficiently to larger, more complex systems. This is where DFT provides an appealing alternative as it enables the inclusion of electron exchange and correlation inexpensively, when compared to the existing methods. The basic idea behind DFT is that the electron-electron coulomb interaction is divided into the Hartree term (where electron-electron repulsion is replaced by electron - electron cloud repulsion) and an exchange-correlation functional, which is a function of spatially dependent electron density and includes the many particle interactions. This reduces the many-body problem, which depends on individual position vectors of electrons, into a single-body problem which depends on electron density.

2.1.4.3 The DFT method

The DFT method came into existence after the Hohenberg-Kohn (H-K) theorems [57] were postulated. The first H–K theorem demonstrates that the ground state properties of a many-electron system are uniquely determined by an electron density that depends on only 3 spatial coordinates. This theorem makes possible the reduction of many-body functions of $3N$ variables into a functions of only 3 variables. The second H–K theorem defines an energy

functional for the system and proves that the correct ground state electron density minimizes this energy functional. These two theorems lead to the development of the Kohn-Sham DFT (KS-DFT) framework.

Under the KS-DFT framework, a system of N interacting particles is converted into a fictitious system of non-interacting particles that generate the same electron density as any given system of interacting particles. The particles are assumed to move in a local effective potential, which is called the Kohn-Sham potential. This approach enables us to break the many-particle Schrödinger equation into N equations for the non-interacting particles:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + v_{\text{eff}}(\mathbf{r})\right)\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}) \tag{2.6}$$

Here, $v_{eff}(\mathbf{r})$ is the Kohn-Sham potential, $\phi_i(\mathbf{r})$ is the wavefunction of the non-interacting electron i, and ε_i is the corresponding energy. Total energy of the system of N interacting particles is the sum of the Eigen values ε_i, and the electron density for the system is given by:

$$\rho(r) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2 \tag{2.7}$$

The Kohn-Sham potential $v_{eff}(\mathbf{r})$ in Eq. 2.6 is formulated as:

$$v_{\text{eff}}(\mathbf{r}) = e^2 \int \frac{\rho(\mathbf{r}')}{\rho(\mathbf{r}-\mathbf{r}')}d\mathbf{r}' + V_{XC}(\mathbf{r}) \tag{2.8}$$

The $V_{XC}(\mathbf{r})$ term, called the exchange-correlation potential, attempts to account for the many-particle interactions, and is the only unknown of this method. In different versions of DFT, this term is approximated as a simple or complex function of local density, $\rho(\mathbf{r})$. For example, in local density approximation (LDA), $V_{XC}(\mathbf{r})$ is only a function of $\rho(\mathbf{r})$, while in generalized gradient approximation (GGA), $V_{XC}(\mathbf{r})$ is a function of $\rho(\mathbf{r})$ local density as well as the gradient of $\rho(\mathbf{r})$ at $\mathbf{r}$.

Once we have selected a particular form for $V_{XC}(\mathbf{r})$, since it is a function of $\rho(\mathbf{r})$, we start with an initial guess for $\rho(\mathbf{r})$, calculate $v_{eff}(\mathbf{r})$, and then solve the Eq. 2.6 for $\phi_i(\mathbf{r})$, for the N electrons, and calculate the new $\rho(\mathbf{r})$ using Eq. 2.7, and loop over until we find a self-consistent solution.

After the solution of the Schrödinger equation under the KS-DFT framework, the potential energy of the electrons of the system is obtained. This is then added to the potential energy for the nuclei-nuclei interactions (which are treated classically) to obtain the net potential energy of the system. It is to be noted here that although AIMD methodology uses several approximations, these are mostly based on theoretical considerations. The classical MD potentials, that are empirically fitted to *ab initio* data, contain some information about the approximate electronic structure, but classical MD simulations in no way have the access to the level of detail that AIMD has, such as density of electronic states or other electronic properties.

2.1.5 Equations of motion

The second-order ODE Eq. 2.1 is solved using the fourth-order Gear predictor-corrector scheme. Gear's predictor-corrector method is a high-order integrator. The basic idea behind the scheme is to use information about the position and its first n derivatives ($n = 4$ here) at time t to arrive at a prediction for position and its first n derivatives at time $t + \Delta t$. Forces (or accelerations) are then calculated at the predicted positions, and then the predicted accelerations are adjusted in accordance with the actual accelerations calculated. In addition, we use the observed discrepancy between the predicted and the actual values to improve future predictions for position and its derivatives. Several coefficients are used in order to obtain corrected values of position and its derivatives. These constants are such that they yield an optimal compromise between the accuracy and stability of the algorithm and they are different for first and second order ODEs.

The above procedure is for the solution of Newton's equations of motion in the micro-canonical (NVE) ensemble. In practice, there is usually the need to perform MD simulations under specified conditions of temperature and/or pressure. Thus, in the literature there exist a variety of methodologies for performing MD simulations under isochoric-isothermal (NVT) or isobaric-isothermal (NPT) conditions. Simulation in other ensembles entails a reformulation of the Hamiltonian equations of motion to include the constraints of constant T and/or P. Among them the most widely used is the Nosé-Hoover method [58, 59].

In the NVT ensemble, the average temperature is regulated by a fluctuating heat bath, called the 'thermostat'. When simulating in NPT ensemble, we need additional methodology to regulate the pressure to keep the average pressure at the desired value. The mechanism to regulate pressure is called the 'barostat'. For these purposes, the new equations used to describe the dynamics of the system are:

$$\ddot{\mathbf{r}}_i = \frac{\mathbf{f}_i}{m_i} - \zeta \dot{\mathbf{r}}_i \tag{2.9}$$

$$\dot{\epsilon} = \frac{P - \langle P \rangle}{Q_P} \tag{2.10}$$

$$\dot{V} = 3\epsilon V \tag{2.11}$$

$$\dot{\zeta} = \frac{T - \langle T \rangle}{Q_T} \tag{2.12}$$

Here V is the unit cell volume, P and T are the instantaneous pressure and temperature, ϵ is the rate of strain or dilation of the unit cell, and ζ is a relative heat removal. Both ϵ and ζ have the units of time^{-1}. Q_P and Q_T are feedback constants for the barostat and the thermostat, respectively, and these were obtained by trial and error. The values of these feedback constants is given in Chapter 3, along with the other simulation details.

A fourth-order Gear predictor-corrector scheme is used to a integrate the equations of motion listed above.

2.2 Method Selection

The present work exemplifies the utility of MD simulation in studying complex systems at the atomic and sub-atomic levels. While a classical force-field has been used to study the PT process, an AIMD methodology is utilized to perform the dipole moment analysis.

A classical force-field for studying the PT process was used because in order to mimic bulk liquid behavior, we needed to study larger systems, and for better sampling, we needed to access larger timescales. In addition, for future application of the method to the study the ORR, effects of the electrochemical double layer, solvent, catalyst defects, and control of electrode potential need to be incorporated. While these effects can be approximated roughly in an AIMD simulation, more molecules and a longer duration are required to have our simulation more representative of the actual system.

While calculating the dipole moments of polar organic molecules based on time-averaged geometries, an accurate representation of the electronic structure was needed for an accurate calculation. This is the case because dipoles are especially sensitive to geometry variations and the distribution of electronic charge. Hence, we need to sample multiple conformations of the molecule whose dipole moment is to be studied. AIMD is used for this purpose. Since there is no time component involved and only multiple conformations are to be sampled, other sampling methods, like the Metropolis algorithm (Monte Carlo) could work as well. But for this study we use AIMD because the scheme ensures that molecules automatically access multiple conformations as the system evolves in time. Hence AIMD is used over other sampling methods more for convenience than for any other reason. Another advantage of using AIMD versus a scheme like Monte-carlo is that the wavefunctions for electrons do not

need to be calculated from scratch for every step because the wavefunctions from previous step can be used as initial guesses.

Section 2.3 describes the prior research pertaining to PT in the condensed phase. Section 2.4 gives a brief background on calculation/prediction of dipole moments.

2.3 Proton Transfer Studies

Simulation of the proton transfer (PT) reaction is not a trivial task because there are numerous electrochemical environmental factors that have an effect on the PT reaction, some of which are mentioned here. First, the solvation structure of an excess proton affects its transport properties, and simulations should adequately reproduce the structure for an accurate analysis. Second, polarizability must be captured precisely, since the variations in charge density have a pronounced effect on transport of excess protons. Third, for an accurate description of the interactions of a proton with the solvent and the counterion, a sufficient number of particles should be included in the simulation. Fourth, due to small energy barriers and small reaction energies, the reaction rate constants are very sensitive to the energy differences. The analysis is further complicated by the ability of proton to display quantum effects like tunneling, and some way of accounting for it must be included in the model.

2.3.1 AIMD studies of PT

Tuckerman and coworkers [60] have studied PT using Car-Parinello MD (CPMD) [61] to simulate 32 water molecules with one excess proton. From this study, they propose that proton diffusion occurs through the interchange between two species: the Eigen cation $H_9O_4^+$ and the Zundel cation $H_5O_2^+$ and the rate limiting step is the cleavage of the second solvation shell of H_3O^+ species. These results were confirmed by Marx et al. [62], who used a combined path-integral MD (PIMD) [63] and CPMD approach to study the same system. They also

report a considerable delocalization tendency of excess charge over a few hydrogen bonds, owing to quantum fluctuations.

Several other *ab initio* methods have been used [60, 64–68] to study the PT mechanism, but all these studies are limited to simulating small clusters of water ($\sim$ 32 molecules) for a short duration ($\sim$ 10 ps), because of computational limitations. While they are informative, it is unclear whether the simulations have adequately captured bulk water behavior and solvation effects. We need to be able to access larger time scales and much bigger systems in order to get reliable statistics on kinetic, transport, and interfacial properties.

2.3.2 Classical MD to simulate PT

To treat larger systems and to access longer timescales, a methodology is required to enable classical MD simulations to study the PT process. Several groups have addressed this problem and the rest of this sub-section focuses on various force-field models that different groups have proposed.

2.3.2.1 Voth's MS-EVB model

The multistate empirical valence bond (MS-EVB) model was developed by Voth and coworkers [69] and is based on the EVB method proposed by Warshel et al. [70] The underlying principle of the EVB method is that the energy of the system is modeled using a superposition of covalent-bonding states as shown in Fig. 2.3. In each of the states, the nuclei are in the same positions and the network of covalent bonds are changed. By varying the weighting coefficients of the states, covalent bonds can be made and broken in a continuous energy-conserving way.

The original version of the model, MS-EVB1, gave a good match to gas and bulk water properties, but in order to match these properties, the authors introduced a charge scaling parameter which reduced the net charge on the hydronium ion by about 25%. For an accurate representation of the system, this parameter was removed in the later version of the method

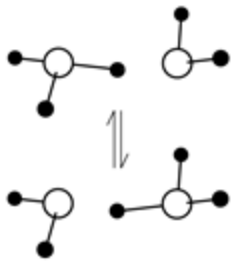


Figure 2.3: Illustration of superimposed states for the EVB method for Zundel ion

(MS-EVB2) by making some changes to the functional form of the potential. The MS-EVB algorithm first identifies the possible EVB states near the excess protons, and then a system Hamiltonian matrix is formed in accordance with particular formulations of pairwise interaction potential. The model parameters were derived by reproducing *ab initio* data for gas phase cluster geometries and formation energies [69]. The model reproduced some bulk water properties accurately and evidence of the Grotthuss mechanism was observed. The diffusivity of the excess proton was too low by about 50%.

MS-EVB3 [71], an improvement on MS-EVB2, was proposed by Wu et al. in 2008. In this new scheme, the underlying water model used in MS-EVB1 and MS-EVB2 was replaced by the SPC/Fw model, which gave a better match to liquid water properties, and also provided a more accurate description of the solvation structure and diffusion dynamics of the excess protons, although the reported diffusivity of the excess proton was still too low by about 50%.

The more recent simulations based on the MS-EVB methodology include up to 250 water molecules and up to 8 excess protons. This model has been demonstrated to work in numerous biological and chemical systems [72–76] and it provides a good representation of bulk transport of protons but it is not clear how this model would be extended to new chemistries including surface reactions. Specifically, it is not known how the superposition-of-states principle will be extended when a solid catalyst surface is present.

2.3.2.2 ReaxFF model

ReaxFF is a reactive force field developed by Goddard and coworkers [77]. It is based on bond-order potentials and is capable of simulating bond breaking and formation within the classical MD framework. Originally, the ReaxFF model was developed for hydrocarbons, but it has been applied successfully to a large number of chemical systems, such as organic reactions [78], thermal decompositions of polymers [79], reactions of energetic materials [80], catalytic formation of carbon nanotubes [81], dissociation of H_2 on Pt surfaces [82], and catalysis and ion transport in fuel cells [83].

The method is based on bond-order-bond-energy correlation. The bond orders are calculated on the basis of inter-atomic distance. The correct bond order is enforced using energy penalties for over-coordination and under-coordination of atoms. Charge is allowed to transfer between the sites using the Electronegativity Equalization method [84]. Charges and bond orders are updated every time step. The total energy of the system is the sum total of contributions of energies for bond distance, lone pair, over-coordination, under-coordination, valence interactions, energy penalty, conjugation, torsion, hydrogen bonding, van der Waals, and coulomb interaction, with each term having its own individual formulation and parameters. The model is parametrized by fitting it to experimental and quantum mechanical data.

Rahaman et al. [85] used the ReaxFF model to simulate PT in the glycine-water system. The force field parameters were optimized to reproduce the quantum mechanically derived energies of a lone glycine molecule, glycine-water dimer, and glycine-$(water)_2$ complexes.Their simulations included 213 water molecules and 1 excess proton. A crude but inexpensive method of "sliding restraint" was used to compare energy barriers for direct (water-free) and water-assisted (with one or two water molecules) proton transfer. The activation energy barrier for the direct mechanism was found to be a few kcal/mol higher than the water-mediated mechanisms.Water assisted PT event was found to have a lower energy barrier and hence a more likely event. Thus the results suggested that the proton transfer

reaction is most likely mediated by a single water molecule. The transport properties could not be calculated because the PT event was crudely forced to occur (without correction for the perturbation), and with insufficient sampling of the PT events.

Although the ReaxFF method has been applied to study numerous reactive systems, its inherent complexity and numerous parameters limit its transferability. In addition, a huge amount of *ab initio* data is required to parametrize the model.

2.3.2.3 Keffer's model

Keffer and coworkers [86] recently proposed a novel reactive algorithm for simulation of PT in aqueous systems. In their work, a transition state for PT obtained from *ab initio* data is mapped onto a set of geometric and energetic triggers to describe the structural diffusion of the proton. The geometric triggers include the distance between donor and acceptor oxygen being less than a particular cut-off, distance between the donor oxygen and transferring proton being greater than equilibrium O-H bond distance, the angle formed between donor and acceptor oxygen and the transferring proton being nearly linear, and the angle formed between transferring proton, acceptor oxygen and other hydrogens of water being near equilibrium angles of a hydronium ion. Another geometric trigger requires that the hydronium ion be coordinated with three other water molecules, thus forming an Eigen cation. The sole energetic trigger requires the proton to have sufficient energy to overcome the barrier associated with the PT reaction. The functional form of these triggers was selected and the parameters were adjusted to reproduce experimental values of the rate constant and activation energy for the PT reaction. In the course of their simulations, whenever these triggers are satisfied, an instantaneous reaction reaction takes place, and the proton hops over to the acceptor oxygen.

After the PT event, the system was equilibrated so that energetically and structurally, the system remains unaffected. To elaborate, it is needed to adjust molecular positions because the O-O radial distribution function (RDF) for $H_2O - H_2O$ is different than that for

$H_3O^+ - H_2O$. For example, the first peak in the $O(H_2O)$-$O(H_2O)$ RDF is at 2.8 Å, whereas the first peak in the $O(H_3O^+)$-$O(H_2O)$ is at 2.55 Å. Thus a local equilibration was needed to be performed. An objective function, as a function of interatomic distances and the kind of pair (O-O, O-H, O-H*), was defined, and the fix was performed for the one hydronium and one water molecule involved in the PT reaction, and the first solvation shell of water molecules around the reacting pair. The model was applied to study PT in bulk water and PT in dilute acid (HCl) solutions. For PT in bulk water, the model provided a good fit for the values for rate constant, activation energy, and heat of the PT reaction, and they confirmed that most of the proton diffusivity was due to structural diffusion (Grotthuss mechanism). In presence of the Cl^-counterion, they found that the environment of an Eigen ion was disrupted both structurally and energetically, reducing the probability of the occurrence of PT event.

Although this approach was used successfully to study PT in bulk and dilute acid solutions, it is not immediately clear how the concept of triggers would need to be modified when other species, like a solid catalyst, are present in the system.

2.3.2.4 The central force model

Bresme and coworkers proposed a force field [87], which is based on classical interatomic potentials following the pioneering ideas of Stillinger et. al. [88]. In this approach, one considers the oxygen and hydrogen species to be unbonded sites, held together in the water geometry with soft potentials. For the three pairs in bulk water system (O-O, O-H, H-H), one defines three pairwise potentials, whose parameters are tuned in order to reproduce the right water geometry and dipole moments, and the right frequencies of bond vibrations. Bresme and coworkers demonstrated the functionality for bulk water, by comparing to experimental transport, structural, and thermodynamic properties. Their simulations include 256 water molecules and the run time was 1.5 ns.

A similar concept, in combination with recursive fitting of radial distribution functions, was recently used by Hoffmann and coworkers to simulate bulk water and PT through hydrated Nafion membrane [89]. They observed multiple PT events in their simulation but the self diffusion coefficient of proton was too low by about 40%. Their simulations include 216 water molecules and the run time was 0.8 ns.

Both Bresme and Hoffmann observed the significant presence of anomalous autoionized water (H_3O^+ and OH^- ions) in their simulations. In particular, Bresme and coworkers observed only 85% neat water, and Hoffman and coworkers observed only 90% neat water. This is an undesirable effect and for a more accurate representation of system, the population of these extraneous species must be negligible. In the present work, like the central force model, there are no rigid bonds forming the hydrogen and oxygen sites into water molecules. Like Bresme's and Hoffman's model, the model used in this work uses pairwise potentials to hold the oxygen and hydrogen sites together, but it was ensured (with the help of a three-body term in the interatomic potential) that the water molecules would not break up spontaneously in a bulk water simulation.

2.3.2.5 Summary

The existing models to simulate reactions in bulk phase are either too expensive, non-transferable, or physically inaccurate. A robust and polarizable model that captures the essential physics of the system accurately is needed. A force field that can describe a few key test cases with a minimum number of adjustable parameters is more likely to meet these requirements because such a force field ensures that we are not empirically fitting data and overtuning it to a particular system. In Chapter 3 we propose a model intended to simulate multiple reactive systems with different chemical species.

2.4 Dipole Moment Studies

Listed in this section are various theories and methods that have been used for the estimation of molecular dipole moments.

2.4.1 Microwave spectroscopy

When a polar molecule is placed in an external electric field of magnitude E, the rotational energy levels of the molecule are split through the Stark effect. The Stark splitting can be adequately defined by second-order perturbation theory and the correction ΔW to the rotational energy is given by:

$$\Delta W = (A + BM^2)E^2 \tag{2.13}$$

Here M is the magnetic quantum number, which ranges from $-L$ to L, where L is related to the total angular momentum of an orbital in absence of an external field. The coefficients, A and B depend upon the components of dipole moments and can be determined experimentally, since the energy level separations can be easily calculated from the analysis of the microwave spectrum. It is therefore possible to extract the dipole components from experimentally determined A and B coefficients. When degenerate levels exist, a first order Stark effect occurs which is of the form:

$$\Delta W = \pm CME \tag{2.14}$$

The coefficient C is proportional to the dipole components and can be readily calculated. The first and second-order Stark perturbation terms, in many cases, permit the determination of dipole moment with a reasonable accuracy. Higher order terms may be considered if more accuracy is desired. A sample used for this method does not need to be pure and when two or more rotational isomers are present, their moments may be measured individually.

Also, since separate components of the dipole moment vector are determined, information is available on the orientation of the moment relative to the molecular framework.

In spite of all the advantages of the microwave method, there are several difficulties associated with it. In many spectra, a competing perturbation due to nuclear quadrupole hyperfine structure [39] may be present which may be comparable to the Stark effect and makes the analysis of data complicated. Another difficulty arises when the Stark effect of all accessible transitions are not sensitive to one of the components of dipole moments, making the determination of that component, and hence the magnitude of total dipole moment, inaccurate. Also, the accuracy of the the determined dipole moment depends upon the measurement of electric field, which entails the measurement of the effective spacing between electrodes, which may not be so easily determined.

2.4.2 Molecular beam electric resonance

This method is similar in principle to the microwave method because it also makes use of the Stark effect of rotational transitions, but in this method, the measurement is carried out on a molecular beam. In a molecular beam, the spectral line-widths are much smaller and the accuracy with which the Stark splittings can be measured is greatly improved. Also, the apparatus for this method can be designed so that the electric field is more uniform and can be measured accurately. Due to these reasons, the molecular beam method offers the highest inherent accuracy of any known method of dipole moment determination.

The main limitation of this method is that due to the multiple states involved in the case of more structured molecules, the molecular beam method can be applied to simple molecules only, mostly monatomic and diatomic (and a very few triatomic) molecules.

2.4.3 Determination of dipole moment from dielectric constant

The foundation of this method was laid down by Debye [30], when he proposed a general equation interrelating dipole moment and the dielectric constant:

$$\frac{\varepsilon - 1}{\varepsilon + 2}\frac{M}{\rho_m} = \frac{4\pi N_A}{3}\left(\alpha_0 + \frac{\mu^2}{3k_BT}\right) \tag{2.15}$$

where the left hand side of the equation is the molar polarization, ρ_m is the mass density, M is the molecular weight, N_A is the Avogadro's number, k_B is the Boltzmann's constant, T is the temperature, and α_0 is the polarizability by distortion [90] and $\mu^2/(3k_BT)$ is the polarizability by orientation.

Although Eq. 2.15 holds true for a wide variety of gases and vapors, it is less accurate for pure liquids and for gases and vapors in which strong intermolecular forces occur. In order to improve the accuracy of this method, modifications to the Debye equations were proposed by Onsager [31], which were later generalized by Kirkwood et al. [33]. These corrections enable good agreement between the dipole moment values calculated for normal liquids with the values obtained from vapor phase measurements [91]. Several other correlations for determining dipole moment from dielectric measurements are available in the literature [92–94].

The methods mentioned above provide some improvement on predictions of dipole moment for polar fluids, but the overall reliability of these methods is low. One reason for these poor predictions can be that the orientation-polarization effects are not being fully accounted for in these formulations. These orientation-polarization effects are particularly prominent in polar systems, as is the case with strongly hydrogen-bonded liquids like water and alcohols. Interestingly, even for a well-studied molecule like water, there is not a reliable experimental value of the liquid dipole moment.

2.4.4 *Ab initio* determination of dipole moments

It is possible to calculate the exact gas-phase dipole moment for a particular conformation of a molecule using computational chemistry software. For any given conformation, first the electronic structure around the sites of a molecule is determined and then based on this

electronic structure, the dipole moment of the molecule is estimated. Multiple [95–97] studies of dipole moment estimation have been conducted using *ab initio* methods at different levels of theory [98] and basis sets [98].

In general, when calculating the dipole moment, the geometry of the molecule being studied is optimized and for the obtained optimized geometry, the electronic structure calculations are performed. Thus, only a single conformation is usually considered in such studies. However, at room temperature, many molecules will access multiple conformations that each have different dipole moments. With MD, multiple conformations can be accessed in a single simulation. Thus, in the present work, AIMD simulations have been used in order to calculate the time-averaged dipole moment of various molecules. The details of the AIMD simulations, including molecules studied, tools used, type of simulations performed along with the results obtained are presented in Chapter 5.

Chapter 3
Bulk Water Simulations

3.1 Introduction

In this chapter, we develop the force field methodology that has been used in the present work for simulation of the PT process. We also present results for liquid water simulations using the developed force field. A detailed description of application of the model to the simulation of excess protons in bulk water is presented in Chapter 4.

When attempting to simulate a reactive system, there are certain physical phenomena that must be captured or reproduced. Firstly, the short- and long-range electrostatic interactions must be modeled accurately. In dense systems, the Hamiltonian used to represent the system must adequately represent the short-range Coulombic interactions between the sites of the molecules. These short-range interactions need to be spherically truncated at some reasonable cut-off distance in the interest of saving computational effort. However, if we simply truncate the interactions and completely ignore the long-range interactions, there can be pronounced errors in structure, thermodynamics, and particularly electrostatic properties [99]. Thus, while modeling bulk liquids, the long-range interactions should be factored in accurately and preferably in a way that the computational time is not increased significantly. In our model, diffuse and point charges are combined with an Ewald (lattice) sum for an accurate description of short- and long-range electrostatic forces. Short-range electrostatics are more accurately described by the use of diffuse charges, which lead to damped Coulombic

interactions. This is especially important for intramolecular interactions on the order of 0.1 nm.

Another desired aspect of a model for bulk liquid simulations is polarizability. Polarizability is the measure of change in a molecule's electronic density (specifically, the valence charge density) in response to an applied electric field, or changes in the electrostatic environment. The determination of accurate electronic structure is beyond the scope of classical MD. However, the model used for MD simulation of a dense polar system must be capable of reproducing the changes in overall charge density in response to external stimuli. In the present work, this is accomplished by allowing the diffuse charge density to vary in magnitude and shape whenever there is a change in the electrochemical environment around a particular site.

The ultimate purpose of the force-field developed in this work is to study reactions and for doing so, the model must allow for bond breaking and formation. A couple of decades ago, force-fields used in classical MD simulations did not allow for cleavage and re-formation of covalent bonds, and one had to turn to AIMD simulations to study reactions. Over the last two decades, several force-fields have been proposed to extend classical MD to reactive systems. A detailed account of these force-fields was presented in Chapter 2. In the current work, as mentioned previously, a methodology similar to the CFF (see Chapter 2) has been used. In our simulations, hydrogen and oxygen as separate entities that can associate to form molecules or dissociate in presence of a stimulus (like an excess proton nearby). This approach enables us to describe the molecule in terms of flexible pair potentials, and thus the molecule can undergo changes in its geometry and dipole moment in the presence of solutes, surfaces, and variations in electric field. The Bond making and breaking, including electron transport, occurs naturally under the framework of a consistent set of rules.

Transferability is a desired quality in any force-field, as it ensures that the model will still work when some new species are introduced in the simulation and/or the model is applied to studying a new system. A lot of force-fields, including ReaxFF [77], use numerous

parameters in different parts of the potential terms (see Chapter 2). This makes the model not only hard to parametrize, but also limits its applicability to a new system. It is desirable to have minimal adjustable parameters because this ensures the model is not overtuned to a particular system and is general enough to be applied to other systems. In our model, there are only two major adjustable parameters.

Although the model was developed in context of bulk water, our intention was to develop a general model applicable to many systems including solid catalyst surfaces and polymer membranes. The original aim of this project was to study the oxygen reduction reaction that occurs on the cathode of fuel cells, which would have entailed the inclusion of a platinum catalyst surface and a nafion polymer exchange membrane [100]. Due to the inherent complications involved in implementing the model into the MD framework, we could not reach the point where we can test the model in systems involving catalyst and membranes. Nonetheless, we were able to test its functionality in systems involving multiple excess protons and counterions, as is demonstrated in Chapter 4.

The following sections describe the model and its application to MD simulations of bulk water. The results for bulk water simulations are discussed in Section 3.4.

3.2 Model Description

Described here is a classical model that includes submodels for Coulombic potential, central-force potential, and van der Waals potential. The net potential energy is the sum total of these three components, and the interatomic forces are the negative gradient of the total potential energy. The submodels are intended to use a minimum of adjustable interaction parameters per site, which can be regressed from *ab initio* calculations as discussed in the section 3.4.

3.2.1 Coulombic submodel

As shown in Fig. 2.2, each atomic site i is modeled with a positive Coulombic point charge (q_i^c) and a negative diffuse charge (q_i^v), so that the total charge is $q_i = q_i^c + q_i^v$. Table 3.1 lists the values of q_i^c and q_i^v used in the water model. Note that the q_i^v for hydrogen is set to zero. This was done to reduce computational cost, since putting diffuse charge on hydrogen sites did not make a significant difference in our results. For simplicity, the model employs Gaussian-based charge densities. The combined core and valence charge distribution is described as:

$$\rho_i(\mathbf{r}) = q_i^c \delta(\mathbf{r}) + q_i^v \frac{(1+\sqrt{2}\gamma_i \mathbf{b}_i \cdot \mathbf{r})^2}{1+b_i^2} \gamma_i^3 \pi^{-3/2} \exp(-\gamma_i^2 r^2) \tag{3.1}$$

where $\mathbf{r}$ is a vector from the site center, and γ_i is a charge inverse-width parameter. Here, q_i^c is a fixed quantity, while q_i^v varies during the course of a simulation. Dimensionless adjustable vector b_i determines the amount of "p-character" or dipole [101] in the valence charge and is determined on-the-fly during a simulation, using dynamical equations.

Using this charge density expression, the Coulomb overlap integrals were evaluated for different interactions, like core to core (c-c), core to valence (c-v), and valence to valence (v-v). The Coulombic potential energy of the system includes all these interactions, and is formulated as:

$$U = \frac{1}{2} \sum_{i,j} q_i^v C_{ij}^{vv} q_j^v + \sum_{i,j} q_i^v C_{ij}^{cv} q_j^c + \frac{1}{2} \sum_{i,j} q_i^c C_{ij}^{cc} q_j^c + \sum_i q_i^v \phi_i^0 + U_{\text{ext}} \tag{3.2}$$

where C_{ij}^{vv} contains interaction integrals for v-v, C_{ij}^{cv} for c-v, and C_{ij}^{cc} for c-c. These interaction integrals are formulated as follows

$$
C_{ij} = \gamma_{ij} \begin{cases} \frac{2}{\sqrt{\pi}} \frac{1+\frac{7}{3}b_i^2+\frac{49}{60}b_i^4}{(1+b_i^2)^2} & \text{if } i=j \\ f_0(s_{ij}) + \frac{1}{(1+b_i^2)(1+b_j^2)} \Big\{ \big[2\Delta_{ij}(1+b_j^2) & \text{if } i \neq j \\ +2\Delta_{ji}(1+b_i^2) + 2D_{ij} - \frac{1}{2}(b_i^2+b_j^2) - b_i^2 b_j^2 \big]\ f_1(s_{ij}) & \\ + \big[\Delta_{ij}^2(1+b_j^2) + \Delta_{ji}^2(1+b_i^2) + 4\Delta_{ij}\Delta_{ji} + \Delta_{ij}(2D_{ij}-b_j^2) & \\ +\Delta_{ji}(2D_{ij}-b_i^2) + \frac{1}{2}D_{ij}^2 + \frac{1}{4}b_i^2 b_j^2 \big]\ f_2(s_{ij}) & \\ + \big[2\Delta_{ij}\Delta_{ji}(\Delta_{ij}+\Delta_{ji}+D_{ij}) - \frac{1}{2}(\Delta_{ij}^2 b_j^2 + \Delta_{ji}^2 b_i^2) \big]\ f_3(s_{ij}) & \\ +\Delta_{ij}^2\Delta_{ji}^2\, f_4(s_{ij}) \Big\} & \end{cases} \tag{3.3}
$$

$$
C_{ij}^* = \gamma_i \begin{cases} \frac{2}{\sqrt{\pi}} \frac{1+\frac{5}{6}b_i^2}{1+b_i^2} & \text{if } i=j \\ f_0(s_{ij}^*) + \frac{1}{1+b_i^2}\left[(2\Delta_{ij}^* - \frac{1}{2}b_i^2)\, f_1(s_{ij}^*) + (\Delta_{ij}^*)^2 f_2(s_{ij}^*) \right] & \text{if i} \neq \text{j} \end{cases} \tag{3.4}
$$

$$
C_{ij}^{**} = \begin{cases} 0 & \text{if } i=j \\ r_{ij}^{-1} & \text{if } i \neq j \end{cases} . \tag{3.5}
$$

The above expressions for $i \neq j$ make use of a number of auxiliary variables: $\Delta_{ij} = g_{ij}\mathbf{b}_i \cdot \mathbf{s}_{ij}$, $\Delta_{ji} = g_{ji}\mathbf{b}_j \cdot \mathbf{s}_{ji} = -g_{ji}\mathbf{b}_j \cdot \mathbf{s}_{ij}$, $\Delta_{ij}^* = \sqrt{2}\mathbf{b}_i \cdot \mathbf{s}_{ij}^*$, and $D_{ij} = g_{ij}g_{ji}\mathbf{b}_i \cdot \mathbf{b}_j$. Furthermore, $g_{ij} = \sqrt{2}(\gamma_{ij}/\gamma_i)$, $\mathbf{s}_{ij} = \gamma_{ij}\mathbf{r}_{ij}$, and $\mathbf{s}_{ij}^* = \gamma_i \mathbf{r}_{ij}$. Vector $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ is the displacement vector between sites i and j and has magnitude r_{ij}. The mixed-charge inverse width is given by$\gamma_{ij} = (\gamma_i^{-2} + \gamma_j^{-2})^{-1/2}$ .Variable $\mathbf{s}_{ij}$ is a dimensionless site-site distance that is scaled by the size of the diffuse charge widths and is used in additional formulas below. A series of functions is used in Eq. 3.10 and Eq. 3.11 where:

$$
f_0(s) = \mathrm{erf}(s)/s \tag{3.6}
$$

This function is commonly used in Coulombic simulations with an Ewald lattice sum. The higher-order functions ($k > 0$) are obtained recursively,

$$f_k(s) = -\frac{df_{k-1}}{d(s^2)} = \frac{1}{2s^2}\left[(2k-1)\, f_{k-1}(s) - \frac{2}{\sqrt{\pi}}\, \exp(-s^2)\right] \tag{3.7}$$

In Eq. 3.2 ϕ_i^0 is the site base potential, a parameter that allows the Coulombic model to be corrected for short-range and quantum effects that are otherwise neglected. It is given by

$$\phi_i^0 = \frac{|q_e|\gamma_i}{1+b_i^2}\left[-\frac{1}{4}a_0\gamma_i(3+5b_i^2) + \frac{1}{\sqrt{2\pi}}\,\frac{1+\frac{7}{3}b_i^2+\frac{49}{60}b_i^4}{1+b_i^2} + \lambda_i(\mathbf{b}_i^* \cdot \mathbf{b}_i - \tfrac{1}{2}b_i^2)\right]. \tag{3.8}$$

where a_0 is the Bohr radius and $|q_e|$ is the magnitude of the electron charge. The first term in the square brackets is a quantum kinetic energy term; the second term is a correction for the previous inclusion in U of valence electrons interacting with themselves (through C_{ii}^{vv}). The last term in the square brackets is an empirical correction to bias the ground state site dipole pi in the direction of $\mathbf{b}_i^*$, with λ_i being an adjustable dipole-strength parameter. Vector $\mathbf{b}_i^*$ is oriented relative to the locations of the nearest neighbors (in this case, nearest hydrogens about each oxygen):

$$\mathbf{b}_i^* = \sum_{j\neq i}\begin{cases} s_{ij}^{-1}\mathbf{s}_{ij} & \text{if } s_{ij} \leq 1.74 \\ s_{ij}^{-1}\mathbf{s}_{ij}\exp\left(11.4(s_{ij}-1.74)^2\right) & \text{if } s_{ij} > 1.74 \end{cases} \tag{3.9}$$

This piecewise expression is designed to smoothly adjust the dipole whenever an excess proton is approaching or leaving an oxygen site.

Some of the above expressions needed to calculate the energy U are fairly complicated and nonlinear and contain terms up to fourth order in b_i, which would have led to an expensive force-field. So the following expressions, truncated at second order in b_i are used. This is reasonable because b values are small.

$$C_{ij}^{vv} = \gamma_{ij} \begin{cases} \frac{2}{\sqrt{\pi}} \left(1 + \frac{1}{3} b_i^2\right) & \text{if } i = j \\ f_0(s_{ij}) + \left[2\Delta_{ij} + 2\Delta_{ji} + 2D_{ij} - \frac{1}{2}(b_i^2 + b_j^2)\right] \, f_1(s_{ij}) & \text{if } i \neq j \\ + \left[\Delta_{ij}^2 + \Delta_{ji}^2 + 4\Delta_{ij}\Delta_{ji}\right] \, f_2(s_{ij}) & \end{cases} \tag{3.10}$$

$$C_{ij}^{cv} = \gamma_i \begin{cases} \frac{2}{\sqrt{\pi}} \left[1 - \frac{1}{6} b_i^2\right] & \text{if } i = j \\ f_0(s_{ij}^*) + (2\Delta_{ij}^* - \frac{1}{2} b_i^2) \, f_1(s_{ij}^*) + (\Delta_{ij}^*)^2 f_2(s_{ij}^*) & \text{if } i \neq j \end{cases} \tag{3.11}$$

$$C_{ij}^{cc} = \begin{cases} 0 & \text{if } i = j \\ r_{ij}^{-1} & \text{if } i \neq j \end{cases} \tag{3.12}$$

$$\phi_i^0 = |q_e| \gamma_i \left[-\frac{3}{4} a_0 \gamma_i (1 + \frac{2}{3} b_i^2) + \frac{1}{\sqrt{2\pi}} \left(1 + \frac{1}{3} b_i^2\right) + \lambda_i (\mathbf{b}_i^* \cdot \mathbf{b}_i - \frac{1}{2} b_i^2) \right] \tag{3.13}$$

Simplified long-range forms of C_{ij} and C_{ij}^* where $i \neq j$ are obtained by taking the limit $s_{ij} \to \infty$ in the above expressions. In general they may be used in place of the more expensive short-range expressions when $s_{ij} \geq 4.5$. The results are

$$\begin{aligned} C_{ij}^{LR} = & \frac{1}{r_{ij}} + \frac{\sqrt{2}}{r_{ij}^2} \left[\gamma_i^{-1} \mathbf{b}_i \cdot \hat{\mathbf{r}}_{ij} - \gamma_j^{-1} \mathbf{b}_j \cdot \hat{\mathbf{r}}_{ij} \right] - \frac{2}{r_{ij}^3} \gamma_i^{-1} \gamma_j^{-1} \left[3(\mathbf{b}_i \cdot \hat{\mathbf{r}}_{ij})(\mathbf{b}_j \cdot \hat{\mathbf{r}}_{ij}) - \mathbf{b}_i \cdot \mathbf{b}_j \right] \\ & + \frac{1}{2r_{ij}^3} \left\{ \gamma_i^{-2} \left[3(\mathbf{b}_i \cdot \hat{\mathbf{r}}_{ij})^2 - b_i^2 \right] + \gamma_j^{-2} \left[3(\mathbf{b}_j \cdot \hat{\mathbf{r}}_{ij})^2 - b_j^2 \right] \right\} \end{aligned} \tag{3.14}$$

and

$$C_{ij}^{*LR} = \frac{1}{r_{ij}} + \frac{\sqrt{2}}{r_{ij}^2} \gamma_i^{-1} \mathbf{b}_i \cdot \hat{\mathbf{r}}_{ij} + \frac{1}{2r_{ij}^3} \gamma_i^{-2} \left[3(\mathbf{b}_i \cdot \hat{\mathbf{r}}_{ij})^2 - b_i^2 \right] , \tag{3.15}$$

where $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/r_{ij}$ is a unit vector. These expressions naturally include all electrostatic interactions that scale as r^{-1} to r^{-3}, namely charge-charge (r^{-1}), charge-dipole (r^{-2}), dipole-dipole (r^{-3}), and charge-quadrupole (r^{-3}) interactions.

In a simulation we need to calculate forces on particles ($\mathbf{r}_i$), charge shapes ($\mathbf{b}_i$), and charges (q_i^v). These forces are then used in the equations of motion that govern the dynamics of the system. In the following discussion, the formulations of the three kinds of forces is presented.

3.2.1.1 Force on particles

The basic definition of the force on particle, $\mathbf{F}_i$, is

$$\mathbf{F}_i = -\frac{\partial U}{\partial \mathbf{r}_i} = \sum_{j \neq i} (\mathbf{F}_{ij}^{vv} + \mathbf{F}_{ij}^{vc} + \mathbf{F}_{ij}^{cv} + \mathbf{F}_{ij}^{cc} + \mathbf{F}_{ij}^{\phi 0}) + \mathbf{F}_i^{\text{ext}}\,, \tag{3.16}$$

which requires that we get the position derivatives (gradients) of C_{ij}, C_{ij}^*, C_{ij}^{**}, and ϕ_i^0. These are given below. The force between a valence charge and another valence charge is:

$$\begin{aligned}
\mathbf{F}_{ij}^{vv} = q_i^v \frac{\partial C_{ij}}{\partial \mathbf{r}_{ij}} q_j^v = -2\gamma_{ij}^2 q_i^v q_j^v \{ & (\mathbf{s}_{ij} + g_{ji}\mathbf{b}_j - g_{ij}\mathbf{b}_i)\, f_1(s_{ij}) \\
& + \left[(2\Delta_{ij} + 2\Delta_{ji} + 2D_{ij} - \tfrac{1}{2}b_i^2 - \tfrac{1}{2}b_j^2)\, \mathbf{s}_{ij} \right]\, f_2(s_{ij}) \\
& + \left[(\Delta_{ji} + 2\Delta_{ij})\, g_{ji}\mathbf{b}_j - (\Delta_{ij} + 2\Delta_{ji})\, g_{ij}\mathbf{b}_i \right]\, f_2(s_{ij}) \\
& + (\Delta_{ij}^2 + \Delta_{ji}^2 + 4\Delta_{ij}\Delta_{ji})\, \mathbf{s}_{ij}\, f_3(s_{ij}) \} \qquad (3.17)
\end{aligned}$$

The force between a valence charge and a core charge is given by:

$$\begin{aligned}
\mathbf{F}_{ij}^{vc} \;=\; & q_i^v \frac{\partial C_{ij}^*}{\partial \mathbf{r}_{ij}} q_j^c = -2\gamma_i^2 q_i^v q_j^c \left\{ (\mathbf{s}_{ij}^* - \sqrt{2}\mathbf{b}_i)\, f_1(s_{ij}^*) \right. \\
& + \left[(2\Delta_{ij}^* - \tfrac{1}{2}b_i^2)\, \mathbf{s}_{ij}^* - \Delta_{ij}^* \sqrt{2}\mathbf{b}_i \right]\, f_2(s_{ij}^*) \\
& \left. + (\Delta_{ij}^*)^2 \mathbf{s}_{ij}^*\, f_3(s_{ij}^*) \right\} \qquad (3.18)
\end{aligned}$$

and

$$\mathbf{F}_{ij}^{cv} = q_j^v \frac{\partial C_{ji}^*}{\partial \mathbf{r}_{ij}} q_i^c = 2\gamma_j^2 q_i^c q_j^v \left\{ (\mathbf{s}_{ji}^* - \sqrt{2}\mathbf{b}_j) f_1(s_{ji}^*) \right. \\ \left. \left[(2\Delta_{ji}^* - \tfrac{1}{2} b_j^2) \mathbf{s}_{ji}^* - \Delta_{ji}^* \sqrt{2} \mathbf{b}_j \right] f_2(s_{ji}^*) + (\Delta_{ji}^*)^2 \mathbf{s}_{ji}^* f_3(s_{ji}^*) \right\} \quad (3.19)$$

The force between a core charge and another core charge is given by:

$$\mathbf{F}_{ij}^{cc} = q_i^c \frac{\partial C_{ij}^{**}}{\partial \mathbf{r}_{ij}} q_j^c = -q_i^c q_j^c r_{ij}^{-3} \mathbf{r}_{ij} \,. \quad (3.20)$$

The Force due to the distortion of b-values is given by:

$$\mathbf{F}_{ij}^{\phi 0} = \frac{\partial U^{\phi 0}}{\partial \mathbf{r}_{ij}} = |q_e| \gamma_{ij} q_i^v \mathrm{e}^{-s_{ij}^2} s_{ij}^{-1} \left\{ \gamma_i \lambda_i \left[\mathbf{b}_i - (2 + s_{ij}^{-2})(\mathbf{b}_i \cdot \mathbf{s}_{ij}) \mathbf{s}_{ij} \right] \right. \\ \left. - \gamma_j \lambda_j \left[\mathbf{b}_j - (2 + s_{ij}^{-2})(\mathbf{b}_j \cdot \mathbf{s}_{ij}) \mathbf{s}_{ij} \right] \right\} . \quad (3.21)$$

and

$$\mathbf{F}_i^{\mathrm{ext}} = -\partial U_{\mathrm{ext}} / \partial \mathbf{r}_i \,. \quad (3.22)$$

Recall that $\Delta_{ij} = g_{ij} \mathbf{b}_i \cdot \mathbf{s}_{ij}$, $\Delta_{ji} = -g_{ji} \mathbf{b}_j \cdot \mathbf{s}_{ij}$, $\Delta_{ij}^* = \sqrt{2} \mathbf{b}_i \cdot \mathbf{s}_{ij}^*$, $\Delta_{ji}^* = \sqrt{2} \mathbf{b}_j \cdot \mathbf{s}_{ji}^*$, $D_{ij} = g_{ij} g_{ji} \mathbf{b}_i \cdot \mathbf{b}_j$, $g_{ij} = \sqrt{2}(\gamma_{ij}/\gamma_i)$, $\mathbf{s}_{ij} = \gamma_{ij} \mathbf{r}_{ij}$, $\mathbf{s}_{ij}^* = \gamma_i \mathbf{r}_{ij}$, $\mathbf{s}_{ji}^* = \gamma_j \mathbf{r}_{ji} = -\gamma_j \mathbf{r}_{ij}$, and $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$. As before, in the case where $s_{ij} \geq 4.5$ we can use long-range expressions in place of functions f_0, f_1, etc., which allows more compact (and easier to compute) interactions.

3.2.1.2 Force on charge-shapes

The force, $\mathbf{F}_i^b$ on charge shapes ($\mathbf{b}_i$) is given bythe following expressions. The force on b-values due to interactions between two valence charges is given by:

$$\begin{aligned}
\mathbf{S}_i = -\frac{1}{2}\sum_k\sum_j \frac{\partial C_{kj}}{\partial \mathbf{b}_i} q_k^v q_j^v &= -\frac{1}{2}\frac{\partial C_{ii}}{\partial \mathbf{b}_i}(q_i^v)^2 - \sum_{j\neq i}\frac{\partial C_{ij}}{\partial \mathbf{b}_i} q_i^v q_j^v \\
&= -\frac{2}{3\sqrt{\pi}}\gamma_{ii}(q_i^v)^2\,\mathbf{b}_i \\
&\quad - q_i^v \sum_{j\neq i} q_j^v \gamma_{ij} g_i \left[(2\mathbf{s}_{ij} + 2g_j\mathbf{b}_j - g_i\mathbf{b}_i)\, f_1(s_{ij}) + 2(\Delta_i + 2\Delta_j)\mathbf{s}_{ij}\, f_2(s_{ij})\right] \\
&= -2\sqrt{2} q_i^v \gamma_i^{-1} \sum_{j\neq i} q_j^v \gamma_{ij}^2 f_1(s_{ij})\,\mathbf{s}_{ij} \\
&\quad + \left\{ \left[-\frac{2}{3\sqrt{\pi}}\gamma_{ii}\,(q_i^v)^2 + 2q_i^v\gamma_i^{-2}\sum_{j\neq i} q_j^v \gamma_{ij}^3 f_1(s_{ij}) \right] \mathbf{I} - 4q_i^v\gamma_i^{-2}\sum_{j\neq i} q_j^v\gamma_{ij}^3 f_2(s_{ij})\,\mathbf{s}_{ij}\mathbf{s}_{ij}^{\mathrm{T}} \right\} \mathbf{b}_i \\
&\quad + 4q_i^v\gamma_i^{-1}\sum_{j\neq i} q_j^v\gamma_j^{-1}\gamma_{ij}^3 \left[-f_1(s_{ij})\,\mathbf{I} + 2f_2(s_{ij})\,\mathbf{s}_{ij}\mathbf{s}_{ij}^{\mathrm{T}}\right]\mathbf{b}_j \qquad (3.23)
\end{aligned}$$

The force on b-values due to interactions between a valence charge and a core charge is given by:

$$\begin{aligned}
\mathbf{S}_i^* = &-\sum_k\sum_j \frac{\partial C_{kj}^*}{\partial \mathbf{b}_i} q_k^v q_j^c = -\frac{\partial C_{ii}^*}{\partial \mathbf{b}_i} q_i^v q_i^c - \sum_{j\neq i}\frac{\partial C_{ij}^*}{\partial \mathbf{b}_i} q_i^v q_j^c \\
=&\frac{4}{3\sqrt{\pi}}\,\gamma_i q_i^v q_i^c\,\mathbf{b}_i \qquad (3.24) \\
&- 2\gamma_i q_i^v \sum_{j\neq i} q_j^c \left[\left(\sqrt{2}\mathbf{s}_{ij}^* - \tfrac{1}{2}\mathbf{b}_i\right) f_1(s_{ij}^*) + \sqrt{2}\Delta_i^*\mathbf{s}_{ij}^*\, f_2(s_{ij}^*)\right] \\
=&-2\sqrt{2}\gamma_i q_i^v \sum_{j\neq i} q_j^c f_1(s_{ij}^*)\,\mathbf{s}_{ij}^* \\
&+ \left\{ 2\gamma_i q_i^v \left[\frac{2}{3\sqrt{\pi}}\,q_i^c + \sum_{j\neq i} q_j^c f_1(s_{ij}^*)\right]\mathbf{I} - 4\gamma_i q_i^v \sum_{j\neq i} q_j^c f_2(s_{ij}^*)\,\mathbf{s}_{ij}^*\mathbf{s}_{ij}^{*\mathrm{T}} \right\}\mathbf{b}_i \qquad (3.25)
\end{aligned}$$

The force on b-values due to distortion in b-values is given by:

$$\mathbf{S}_i^{\phi} = -\sum_k \frac{\partial \phi_k^0}{\partial \mathbf{b}_i} q_k^v = -2q_i^v \mathbf{b}_i \frac{\partial \phi_i^0}{\partial b_i^2} \tag{3.26}$$
$$= q_i^v |q_e| \gamma_i \left[\left(a_0 \gamma_i - \frac{2}{3\sqrt{2\pi}} \right) \mathbf{b}_i - \lambda_i^1 \mathbf{b}_i^* \right]$$

and

$$\mathbf{F}_i^{b,\text{ext}} = -\partial U_{\text{ext}} / \partial \mathbf{b}_i \,. \tag{3.27}$$

3.2.1.3 Force on charge

The forces on charges (f_i^q) is given by:

$$f_i^q = -\phi_i = -\frac{\partial U}{\partial q_i^v} = -\left(\sum_j C_{ij} q_j^v + \sum_j C_{ij}^* q_j^c + \phi_i^0 + \phi_i^{\text{ext}} \right) \tag{3.28}$$

It is to be noted that the derivative of U with respect to valence charge constitutes the chemical potential (ϕ_i) for valence charge at each site. As will be discussed in the next chapter, the difference in chemical potential of valence charges on oxygen sites forms the basis of electron transfer that accompanies the PT process.

To summarize, there are three main adjustable Coulombic parameters per site: q_i^c, γ_i, and λ_i. Variables q_i^v and b_i are determined from charge conservation and energy minimization principles and so these are not independent model parameters. In the case of hydrogen sites, $q_i^v = b_i = 0$ and this simplifies the model somewhat. The choice of site parameters allows the charge distribution for each site to be tuned so that the model can mimic molecular charge distribution and polarizability. The flexibility of the model combined with a physical basis will allows reasonable transferability of parameters between similar molecules, and makes the model more robust in heterogeneous electrochemical environments. This has been demonstrated to some extent in Chapter 4.

3.2.2 Central-force submodel

Further quantum effects such as electron correlation and exchange are treated classically. The oxygen and hydrogen ions are held together with pairwise short-range interaction potentials [88] between OH and HH pairs, which are formulated as:

$$U_{\mathrm{OH}}^{\mathrm{CFM}}(r) = D1_{\mathrm{OH}} \exp\left(2a_{\mathrm{OH}}\left(r_{e\mathrm{OH}} - r\right)\right) - D2_{\mathrm{OH}} \exp\left(a_{\mathrm{OH}}\left(r_{e\mathrm{OH}} - r\right)\right) \tag{3.29}$$

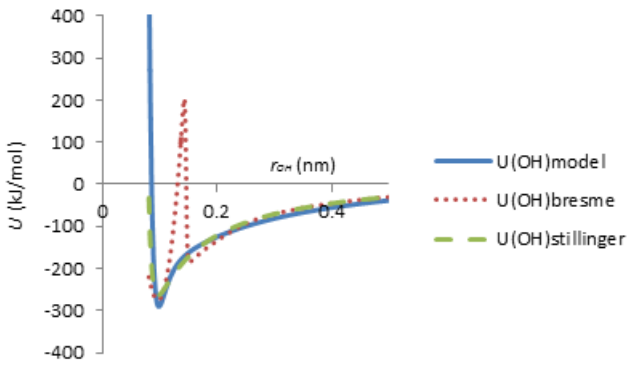
$$U_{\mathrm{HH}}^{\mathrm{CFM}}(r) = D_{\mathrm{HH}}\left(\left(1 - \exp\left(-a_{\mathrm{HH}}\left(r_{e\mathrm{HH}} - r\right)\right)\right)^2 - 1\right) \tag{3.30}$$

where, $r_{e\mathrm{OH}}$ and $r_{e\mathrm{HH}}$ are the respective bond lengths. Values of the fixed parameters used in Eq. 3.29 and Eq. 3.30 are listed in Table 3.1. These expressions and parameter values were chosen so as to reproduce the correct water geometry. Although it is of some value to get the right vibrational frequencies for the OH and HH bonds, we found that when this was attempted, we failed to get 100% neat water in our bulk simulations. This was one of the drawbacks of the models proposed by Bresme and coworkers, and Hoffmann and coworkers. The parameters were therefore adjusted so that water molecules would not break up, even though we did not match the experimental frequencies. The CFM potentials are plotted in Fig. 3.1. In addition to the CFM potentials, we also use a three-body potential in order to better control the HOH angle. This 3-body potential has the following functional form:

$$U_{\mathrm{HOH}}^{\mathrm{ang}} = A_{\mathrm{ang}} \exp\left(-\beta_{\mathrm{ang}} r_{\mathrm{OH1}}^2 - \beta_{\mathrm{ang}} r_{\mathrm{OH2}}^2\right)\left(\mathbf{r}_{\mathrm{OH1}} \cdot \mathbf{r}_{\mathrm{OH2}} - r_{\mathrm{OH1}} r_{\mathrm{OH2}} \cos\left(\theta\right)\right) \tag{3.31}$$

H1and H2 are the two nearest hydrogen neighbors to the oxygen. Values of the fixed parameters used are listed in Table **3.1**. These parameters (see Table **3.1**) were chosen so that the HOH angle for the water molecules in our simulations was around 109° and to get the right pair distribution functions, similar to the SPC/E water model [102].

(a)



(b)

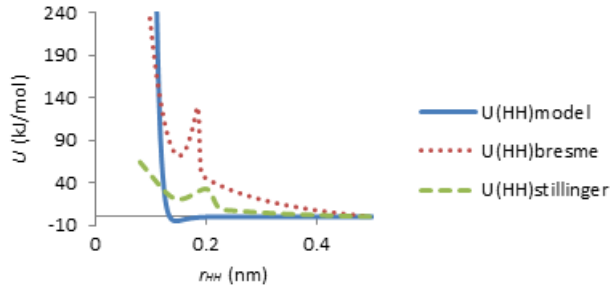


Figure 3.1: Comparison of pair potentials (a) OH pair (b) HH pair, used in model vs. the pairwise potentials used by Bresme and coworkers [87], and Stillinger et al. [88].

The net force on particles will be the sum total of forces due to pairwise OH and HH potentials and the HOH three body potential:

$$\mathbf{F}_{\mathrm{O}} = -\left(\frac{\partial U_{\mathrm{OH}}^{\mathrm{CFM}}}{\partial \mathbf{r}_{\mathrm{O}}} + \frac{\partial U_{\mathrm{HOH}}^{\mathrm{ang}}}{\partial \mathbf{r}_{\mathrm{O}}}\right) \tag{3.32}$$

$$\mathbf{F}_{\mathrm{H}} = -\left(\frac{\partial U_{\mathrm{OH}}^{\mathrm{CFM}}}{\partial \mathbf{r}_{\mathrm{H}}} + \frac{\partial U_{\mathrm{HH}}^{\mathrm{CFM}}}{\partial \mathbf{r}_{\mathrm{H}}} + \frac{\partial U_{\mathrm{HOH}}^{\mathrm{ang}}}{\partial \mathbf{r}_{\mathrm{H}}}\right) \tag{3.33}$$

3.2.3 Van der Waals potential

The extension to a many-molecule system is achieved by incorporation of a van der Waals (vdW) potential, in this case exponential-6 Buckingham, between each pair of oxygens:

$$U_{\mathrm{OO}}^{\mathrm{vdW}} = L_{\mathrm{OO}} \left[6\theta^6 \eta^{-7} \exp(\eta - \theta s_{\mathrm{OO}}) - (s_{\mathrm{OO}})^{-6} \right] \tag{3.34}$$

Such interactions have their origins in electron exchange and correlation and are not wholly independent from electrostatic interactions. From quantum mechanics, we know that the attractive correlation energy between two sites is approximated by the London formula [103]:

$$u_{12} = -\frac{3}{2} \alpha_1 \alpha_2 \frac{u_1 u_2}{u_1 + u_2} r_{12}^{-6} \tag{3.35}$$

where $\alpha_1$and α_2 are the site polarizabilities of sites 1 and 2 respectively, and u_1 and u_2 are the respective energies to promote an electron from an s to a p orbital. These values can be derived from our Coulombic submodel, and in particular q_i^c and γ_i parameters (e.g. Eq. 3.71 below), which means that independent parameters do not need to be derived at least for the attractive part of the vdW potential. L_{OO} is a fixed parameter that depends on oxygen site Coulombic parameters (given in Table 3.1). The shape of the repulsive part of the vdW potential is controlled by dimensionless parameters θ and η (also given in Table **3.1**). The above expression (Eq. 3.35) has been tested (Fig. 3.2) by comparing to *ab initio* results for molecular dimers.

If one wants to compare the exp-6 dispersion potential to an approximating Lennard-Jones potential, the corresponding energy and distance parameters are

$$\sigma_{\mathrm{OO}} = 2^{-1/6} \eta \theta^{-1} \gamma_{\mathrm{OO}}^{-1} \tag{3.36}$$

$$\varepsilon_{\mathrm{OO}} = (\eta - 6) \eta^{-7} \theta^6 L_{\mathrm{OO}} \tag{3.37}$$

The force on particle due to the vdW potential is given by the negative gradient of the vdW:

$$\mathbf{F}_{\mathrm{O}} = -\left(\frac{\partial U_{\mathrm{OO}}^{\mathrm{vdW}}}{\partial \mathbf{r}_{\mathrm{O}}} \right) \tag{3.38}$$

Table 3.1: List of model parameter values.

Parameter	Value
q^c_{O}	$1.02\,\lvert e\rvert$
γ_{O}	$14.47\,\mathrm{nm}^{-1}$
q^c_{H}	$0.4238\,\lvert e\rvert$
γ_{H}	$\infty\,\mathrm{nm}^{-1}$
λ_{O}	0.1584
$D1_{\mathrm{OH}}$	$300\,\mathrm{kJ/mol}$
$D2_{\mathrm{OH}}$	$315\,\mathrm{kJ/mol}$
a_{OH}	$88.0\,\mathrm{nm}^{-1}$
$r_{e\mathrm{OH}}$	$0.102\,\mathrm{nm}$
D_{HH}	$5\,\mathrm{kJ/mol}$
a_{HH}	$50.0\,\mathrm{nm}^{-1}$
$r_{e\mathrm{HH}}$	$0.161\,\mathrm{nm}$
A_{ang}	$1500\,\mathrm{kJ/mol}$
β_{ang}	$15.0\,\mathrm{nm}^{-1}$
θ_{set}	107°
L_{OO}	$1600\,\mathrm{kJ/mol}$
θ	4.25
η	16.0

3.3 Ewald Sum

The Ewald sum is an efficient way to handle long-range Coulombic interactions in a periodic system [104]. The basis for the Ewald sum is to temporarily replace each site charge-shape function ρ_i with one that has the same long-range interaction behavior. The cell potential due to the modified charge density is obtained from Poisson's equation, and is solved as a Fourier series in reciprocal space. If the replacement function ρ^r_i is smoother or more diffuse than the original ρ_i, it can more accurately be described by a Fourier series, leading to less computational expense. In a typical simulation, this means replacing point charges with Gaussian diffuse charges of the same magnitude. The Fourier series calculates smoothly varying long-range interactions accurately, but is less accurate for short-range interactions, due to finite truncation of the series as well as the fact that the original charge-shape function

has been modified. Therefore, a short-range (real-space) interaction between sites must be made to get the correct total potential of the cell.

In the present case, the traditional Ewald sum must be modified to account for the nonstandard original charge-shape functions. The criterion we use is that ρ_i^r and ρ_i should have the same total charge magnitude, dipole moment, and modified quadrupole moment. As in a conventional Ewald sum, ρ_i^r is based around a Gaussian charge distribution. We additionally add terms to generate dipoles and quadrupoles akin to ρ_i (Eq. 2.2). The result is:

$$\rho_i^r(\mathbf{r}) = \left[q_i^c + q_i^v \left(1 + 2\sqrt{2}\gamma_i^{-1}\alpha_E^2 \mathbf{b}_i \cdot \mathbf{r}\right)^2\right] (2\pi^{-1}\alpha_E^2)^{3/2} \exp(-2\alpha_E^2 r^2) \tag{3.39}$$

Here α_E is the Ewald convergence parameter. It is analogous to γ_i but smaller in magnitude.

In our Ewald framework the Coulombic potential of the system includes three pieces:

$$U_{\text{coul}} = U^s + U^c + U^r \tag{3.40}$$

where, U^s (the short-range potential) and U^c (the correction potential) are both calculated in real space. The short-range potential is based on the full Coulombic interaction between sites in which pair distance r_{ij} is less than cutoff r_{cut}:

$$U^s = \frac{1}{2} \sum_{i,j}^{r_{ij}<r_{\text{cut}}} q_i^v C_{ij}^{vv} q_j^v + \sum_{i,j}^{r_{ij}<r_{\text{cut}}} q_i^v C_{ij}^{cv} q_j^c + \frac{1}{2} \sum_{i,j}^{r_{ij}<r_{\text{cut}}} q_i^c C_{ij}^{cc} q_j^c + \sum_i q_i^v \phi_i^0 \tag{3.41}$$

The reciprocal potential is designed to account for all long-range interactions ($r_{ij} > r_{\text{cut}}$), including interactions between sites in the central cell and sites in the surrounding image cells. Using Fourier transforms, the interaction between two sites i and j can be expressed as:

$$u_{ij}^r(\mathbf{r}_{ij}) = \int\int \rho_i^r(\mathbf{r}_1)\,\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2 - \mathbf{r}_{ij}|}\,\rho_j^r(\mathbf{r}_2)\,d\mathbf{r}_1\,d\mathbf{r}_2$$
$$= \frac{1}{V}\sum_{\mathbf{h}} \hat{\rho}_i^r(\mathbf{h})\,\frac{4\pi}{h^2}\,\hat{\rho}_j^r(-\mathbf{h})\,\mathrm{e}^{\mathbf{i}\mathbf{h}\cdot\mathbf{r}_{ij}} \tag{3.42}$$

The first line of the equation is the real-space expression for the Coulomb energy between two distributed charges at centers r_i and r_j , where distribution functions ρ_i^r and ρ_j^r are given relative to the respective centers. The second line of the equation is the corresponding Fourier series representation of the pair energy, obtained by the convolution theorem and Fourier transforms of functions such as:

$$\hat{\rho}_i^r(\mathbf{h}) = \int \rho_i^r(\mathbf{r})\,\mathrm{e}^{-\mathbf{i}\mathbf{h}\cdot\mathbf{r}}d\mathbf{r} \tag{3.43}$$

Exponential quantity $\mathtt{i}$ is the imaginary number. Vector $\mathbf{h}$ is a reciprocal-lattice or wave vector and for a cubic unit cell is given by:

$$\mathbf{h} = 2\pi L^{-1}\mathbf{n} \tag{3.44}$$

where vector $\mathbf{n}$ is composed of three independent integers: n_1, n_2, and n_3. In Eq. 3.42 the sum over $\mathbf{h}$ excludes $\mathbf{h} = 0$, but otherwise includes all possible values of n_1, n_2, and n_3. In practice, one truncates the sum to $h < h_{\mathrm{cut}}$, based on the diminishing magnitude of terms as $h = |\mathbf{h}|$ increases.

The reciprocal potential of the system is obtained by a double sum over the potentials between all pairs of sites in the system. The double sum can be rearranged as follows

$$U^r = \frac{1}{2}\sum_i\sum_j u_{ij}^r = \frac{2\pi}{V}\sum_{\mathbf{h}\neq 0}\frac{1}{h^2}\left[\sum_i \hat{\rho}_i^r(\mathbf{h})\,\mathrm{e}^{-\mathbf{i}\mathbf{h}\cdot\mathbf{r}_i}\right]\left[\sum_j \hat{\rho}_j^r(-\mathbf{h})\,\mathrm{e}^{\mathbf{i}\mathbf{h}\cdot\mathbf{r}_j}\right] + U_0^r \tag{3.45}$$

where U_0^r accounts for the missing term ($\mathbf{h} = 0$) in the sum. Through a combination of system-charge neutrality and the so-called tinfoil boundary condition, we can take $U_0^r = 0$.

The reciprocal-space sum in principal could describe all Coulombic interactions in the system if $\rho_i^r = \rho_i$; however, this would not be computationally efficient. Thus, the charge density used is instead as given in Eq. 3.40. The Fourier transform of the site charge density in Eq. 3.40 is:

$$\hat{\rho}_i^r(\mathbf{h}) = \left[q_i - \mathtt{i}\mathbf{h}\cdot\mathbf{p}_i + \tfrac{1}{6}\left(4\alpha_E^2\mathbf{I} - \mathbf{h}\,\mathbf{h}^{\mathrm{T}}\right) : \mathbf{Q}_i^*\right] \exp[-h^2/(8\alpha_E^2)] \tag{3.46}$$

where $\mathbf{p}_i$ is given by Eq. 3.68 , $\mathbf{I}$ is the identity matrix of order 3, and $\mathbf{Q}_i^*$ is given by

$$\mathbf{Q}_i^* = \frac{3q_i^v}{\gamma_i^2}\mathbf{b}_i\mathbf{b}_i^{\mathrm{T}} \tag{3.47}$$

This expression can then be substituted into Eq. 3.45, with the provision that we omit from the products the dipole-quadrupole and quadrupole-quadrupole interactions. This is done to be consistent with the simplified Coulombic overlap integrals in Eq. 3.10. This omission is fine as long as a corresponding change is made to U^c. Following some natural cancellation of imaginary terms and other algebra, we get

$$U^r = \frac{1}{2}\sum_{\mathbf{h}\neq 0}\Gamma(\mathbf{h})\,u_h(\mathbf{h}) \tag{3.48}$$

The Eq. 3.48 gives us the expression for reciprocal energy of the system. This equation uses some auxiliary functions that are listed below:

$$\Gamma(\mathbf{h}) = \frac{4\pi}{V}\, h^{-2} \exp[-h^2/(4\alpha^2)] \tag{3.49}$$

$$\begin{aligned} u_h(\mathbf{h}) = {} & \chi_{0c}\left[\chi_{0c} - 2\mathbf{h}\cdot\boldsymbol{\chi}_{1s} + (4\alpha^2\mathbf{I} - \mathbf{h}\,\mathbf{h}^{\mathrm{T}}) : \boldsymbol{\chi}_{2c}\right] + (\mathbf{h}\cdot\boldsymbol{\chi}_{1s})^2 \\ & + \chi_{0s}\left[\chi_{0s} + 2\mathbf{h}\cdot\boldsymbol{\chi}_{1c} + (4\alpha^2\mathbf{I} - \mathbf{h}\,\mathbf{h}^{\mathrm{T}}) : \boldsymbol{\chi}_{2s}\right] + (\mathbf{h}\cdot\boldsymbol{\chi}_{1c})^2 \end{aligned} \tag{3.50}$$

$$\chi_{0c}(\mathbf{h}) = \sum_j (q_j^v + q_j^c)\, \cos(\mathbf{h}\cdot\mathbf{r}_j) \tag{3.51}$$

$$\chi_{0s}(\mathbf{h}) = \sum_j (q_j^v + q_j^c)\, \sin(\mathbf{h}\cdot\mathbf{r}_j) \tag{3.52}$$

$$\boldsymbol{\chi}_{1c}(\mathbf{h}) = \sum_j \mathbf{p}_j \cos(\mathbf{h}\cdot\mathbf{r}_j) \tag{3.53}$$

$$\boldsymbol{\chi}_{1s}(\mathbf{h}) = \sum_j \mathbf{p}_j \sin(\mathbf{h}\cdot\mathbf{r}_j) \tag{3.54}$$

$$\boldsymbol{\chi}_{2c}(\mathbf{h}) = \tfrac{1}{3}\sum_j \mathbf{Q}_j^* \cos(\mathbf{h}\cdot\mathbf{r}_j) \tag{3.55}$$

$$\boldsymbol{\chi}_{2s}(\mathbf{h}) = \tfrac{1}{3}\sum_j \mathbf{Q}_j^* \sin(\mathbf{h}\cdot\mathbf{r}_j) \tag{3.56}$$

Note that u_h accounts for the charge-charge, charge-dipole, dipole-dipole, and charge-quadrupole interactions.

The correction potential in Eq. 3.40 is given by:

$$U^c = \frac{1}{2}\sum_{i,j}^{r_{ij}<r_{\mathrm{cut}}} q_i^v C_{ij}^{vv'} q_j^v + \sum_{i,j}^{r_{ij}<r_{\mathrm{cut}}} q_i^v C_{ij}^{cv'} q_j^c + \frac{1}{2}\sum_{i,j}^{r_{ij}<r_{\mathrm{cut}}} q_i^c C_{ij}^{cc'} q_j^c \tag{3.57}$$

The correction terms is needed because U^r (reciprocal potential) of necessity includes all site pairs in the system, whereas the short-range term U^s already includes the full interaction between a subset of the pairs, namely those where $r_{ij} < r_{\mathrm{cut}}$. U^c therefore removes this spurious partial double counting of pair interactions.

The correction terms ($C_{ij}^{vv'}$, $C_{ij}^{vc'}$, $C_{ij}^{cc'}$) of Eq. 3.57 can be derived from the first line of Eq. 3.42 and correspond to the real-space version of the interaction between modified charges

ρ^r. Because Eq. 3.39 is has a similar functional form to Eq. 3.1, the modified C_{ij} values in Eq. 3.57 can be derived from the original expressions of C_{ij}^{vv}, C_{ij}^{cv}, and C_{ij}^{cc} by sequentially replacing $\gamma_i \leftarrow \sqrt{2}\alpha_E$, $\gamma_{ij} \leftarrow \alpha_E$, and $b_i^{'} \leftarrow \sqrt{2}\gamma_i\alpha_i b_i$ in Eqs. 3.10–3.12, and are as follows:

$$C_{ij}^{vv'} = \alpha \begin{cases} \frac{2}{\sqrt{\pi}}\left(1 + \frac{1}{3}b_i'^2\right) & \text{if } i = j \\ f_0(s'_{ij}) + \left[2\Delta'_{ij} + 2\Delta'_{ji} + 2D'_{ij} - \frac{1}{2}(b_i'^2 + b_j'^2)\right]\, f_1(s'_{ij}) & \text{if } i \neq j \\ + \left[\Delta_{ij}'^2 + \Delta_{ji}'^2 + 4\Delta'_{ij}\Delta'_{ji}\right]\, f_2(s'_{ij}) & \end{cases} \tag{3.58}$$

$$C_{ij}^{cv'} = \alpha \begin{cases} \frac{2}{\sqrt{\pi}}\left(1 - \frac{1}{6}b_i'^2\right) & \text{if } i = j \\ f_0(s'_{ij}) + (2\Delta'_{ij} - \frac{1}{2}b_i'^2)\, f_1(s'_{ij}) + (\Delta_{ij}')^2\, f_2(s'_{ij}) & \text{if } i \neq j \end{cases} \tag{3.59}$$

$$C_{ij}^{cc'} = \alpha \begin{cases} \frac{2}{\sqrt{\pi}} & \text{if } i = j \\ f_0(s'_{ij}) & \text{if } i \neq j \end{cases} \tag{3.60}$$

The correction potential is applied in real space between the same set of pairs as the short-range potential U_s. Taken together, $U_r - U_c$ fully cancels out any interactions between ρ_r for $r_{ij} < r_{\text{cut}}$ and thus only includes long-range pair interactions and need not be calculated as frequently as U_s. This is the basis for a multiple-time step scheme used with this model, which is described in section (3.4.2).

To calculate the reciprocal forces on charge, particles, and charge shape, we use the following formulations. The reciprocal force on charges is given by:

$$\begin{aligned}\mathbf{F}^{\mathbf{q}r}_i &= \frac{\partial U^r}{\partial q_i^v} \\ &= \sum_{\mathbf{h}\neq 0} \Gamma(\mathbf{h}) \, \left[(\chi_i^{c1} + \chi_i^{c2}) \, \cos(\mathbf{h}\cdot\mathbf{r}_i) + (\chi_i^{s1} + \chi_i^{s2}) \, \sin(\mathbf{h}\cdot\mathbf{r}_i)\right]\end{aligned} \tag{3.61}$$

The reciprocal forces on position is given by:

$$\begin{aligned}\mathbf{F}^r_i &= -\frac{\partial U^r}{\partial \mathbf{r}_i} \\ &= \sum_{\mathbf{h}\neq 0} \Gamma(\mathbf{h}) \, \mathbf{h} \left\{[\chi_i^{c1}(q_i^v + q_i^c) + \chi_i^{c2} q_i^v] \, \sin(\mathbf{h}\cdot\mathbf{r}_i) - [\chi_i^{s1}(q_i^v + q_i^c) + \chi_i^{s2} q_i^v] \, \cos(\mathbf{h}\cdot\mathbf{r}_i)\right\}\end{aligned} \tag{3.62}$$

The reciprocal forces on b-values is given by:

$$\begin{aligned}\mathbf{F}^{\mathbf{b}r}_i &= -\frac{\partial U^r}{\partial \mathbf{b}_i} \\ &= -\sum_{\mathbf{h}\neq 0} \Gamma(\mathbf{h}) \, q_i^v \left\{ \left[(\mathbf{h}\cdot\chi_{1c} + \chi_{0s})\sqrt{2}\gamma_i^{-1}\,\mathbf{h} + \chi_{0c}\gamma_i^{-2}(4\alpha^2\mathbf{b}_i - (\mathbf{h}\cdot\mathbf{b}_i)\,\mathbf{h})\right] \, \cos(\mathbf{h}\cdot\mathbf{r}_i) \right. \\ &\qquad \left. + \left[(\mathbf{h}\cdot\chi_{1s} - \chi_{0c})\sqrt{2}\gamma_i^{-1}\,\mathbf{h} + \chi_{0s}\gamma_i^{-2}(4\alpha^2\mathbf{b}_i - (\mathbf{h}\cdot\mathbf{b}_i)\,\mathbf{h})\right] \, \sin(\mathbf{h}\cdot\mathbf{r}_i)\right\}\end{aligned} \tag{3.63}$$

where,

$$\chi_i^{c1}(\mathbf{h}) = \chi_{0c} - \mathbf{h}\cdot\boldsymbol{\chi}_{1s} + \tfrac{1}{2}(4\alpha^2\mathbf{I} - \mathbf{h}\,\mathbf{h}^{\mathrm{T}}) : \boldsymbol{\chi}_{2c} \tag{3.64}$$

$$\chi_i^{s1}(\mathbf{h}) = \chi_{0s} + \mathbf{h}\cdot\boldsymbol{\chi}_{1c} + \tfrac{1}{2}(4\alpha^2\mathbf{I} - \mathbf{h}\,\mathbf{h}^{\mathrm{T}}) : \boldsymbol{\chi}_{2s} \tag{3.65}$$

$$\chi_i^{c2}(\mathbf{h}) = (\mathbf{h}\cdot\boldsymbol{\chi}_{1c} + \chi_{0s})\sqrt{2}\gamma_i^{-1}\,\mathbf{h}\cdot\mathbf{b}_i + \tfrac{1}{2}\chi_{0c}\gamma_i^{-2}[4\alpha^2 b_i^2 - (\mathbf{h}\cdot\mathbf{b}_i)^2] \tag{3.66}$$

$$\chi_i^{s2}(\mathbf{h}) = (\mathbf{h}\cdot\boldsymbol{\chi}_{1s} - \chi_{0c})\sqrt{2}\gamma_i^{-1}\,\mathbf{h}\cdot\mathbf{b}_i + \tfrac{1}{2}\chi_{0s}\gamma_i^{-2}[4\alpha^2 b_i^2 - (\mathbf{h}\cdot\mathbf{b}_i)^2] \tag{3.67}$$

3.4 Results and Discussion

3.4.1 Pairwise model parametrization

The parameters of the model were adjusted to reproduce lone water polarizability, dipole moment, and the quadrupole moment, in addition to other properties.

The three mentioned properties can be calculated in terms of model parameters. The model dipole moment for each site due to the point dipole $\mathbf{b}_i$, is given by:

$$\mathbf{p}_i = \frac{\sqrt{2}q_i^v}{\gamma_i}\,\mathbf{b}_i \tag{3.68}$$

For each site,$\mathbf{p}_i$ is the dipole contribution from the sp-hybrid character of valence charge. To this is added the contribution from the gross separation of charge. Thus, the total dipole moment of water molecule is given by:

$$\mathbf{p}^{\text{tot}} = \sum_i \mathbf{p}_i + \sum_i (\mathbf{r}_i - \mathbf{r}_{\text{cm}})\,(q_i^v + q_i^c) \tag{3.69}$$

where $\mathbf{r}_{\text{cm}}$ is effectively the location of the central oxygen and the summations are over all atoms associated with a single water.

The quadrupole moment for our charge distribution model for a site i in terms of the model parameters, is given by:

$$\mathbf{Q}_i = \frac{3q_i^v}{\gamma_i^2}\left(\mathbf{b}_i\mathbf{b}_i^{\text{T}} + \frac{1}{2}\mathbf{I}\right) \tag{3.70}$$

Note this differs slightly from $\mathbf{Q}_i^*$ defined above. An analytical expression for the electronic polarizability for a single site, α_i can also be derived in terms of model parameters:

$$\alpha_i = \frac{-q_i^v}{\gamma^3 \left[-\frac{1}{3\sqrt{2\pi}} \left(q_i^v + |q_e|\right) + \frac{2}{3\sqrt{\pi}} q_i^c + \frac{1}{2} a_0 \gamma_i |q_e| \right]} \tag{3.71}$$

For a water molecule, $\mathbf{Q}_i^{\text{tot}} = \mathbf{Q}_{\text{O}}$ and $\alpha^{\text{tot}} = \alpha_{\text{O}}$ because $q_i^v = 0$ for the hydrogen sites. This expression for α also neglects polarizability due to bond stretching and angle variations.

Since this was an under-constrained problem, meaning that the number of properties being matched were less than the number of parameters, we looked at some additional properties. The present model has the ability to predict charge distribution and energies for the cationic and anionic forms of a molecule, which allows for the prediction of electron affinity and ionization energy. For parametrization, these properties were also matched. The model parameters are listed in Table 3.1. For the listed parameters, the corresponding electron affinity is -30 kJ/mol and the ionization potential was 1175 kJ/mol. The experimental values of these quantities are -16 kJ/mol and 1216 kJ/mol respectively [105, 106].

The potential energy scan (PES) of two approaching water molecules was also studied, and was also considered during the parametrization process. The two water molecules were allowed to approach towards each other in different orientations and the corresponding PES was plotted. Fig. 3.2 shows the PESs as predicted by model vs the *ab initio* PES. Fig. 3.2(a) shows the most important path of approach of two water molecules and the model underpredicts the attraction for this route. It is to be noted that the model is capable of exactly matching the *ab initio* data, but when we used parameters corresponding to this case in our MD simulations, the resulting bulk-water density was high and the heat of vaporization was also off. Since our aim was to develop a model that works well in all systems ranging from gas-phase to bulk liquid, we decided to put in additional repulsion, which led to minor mismatch between the model-predicted and *ab initio* PES, for this route.

The parameter space was thoroughly explored prior to settling on the given set of parameters. It was observed that two different types of sets of parameters gave good matches for the properties used for parametrizing the model. The characteristic difference between the two sets was the relative amount of contribution the two components (gross-charge separation and point dipole) make to the net dipole moment (which is the sum total of the two components). The relative amounts of contribution of each component can be varied by varying the parameters of the model. Among the two sets of parameters, one resulted in a higher contribution to dipole moment from the gross charge separation and the other resulted in a higher contribution to dipole moment from the point dipole. The latter case resulted in higher b-values. The underlying assumption of the model was that the b-values are small (see Section 3.2.1), which enabled us to neglect the higher order terms in Eq. 3.3 and Eq. 3.4. Hence, the latter case was an unfavorable one and this was confirmed when the model was used in the MD framework. The resulting structure of water obtained from the MD simulation was incorrect and the b-value did not seem to be converged. Thus, we selected the parameter set that resulted in a larger contribution to dipole moment from the gross charge separation.

3.4.2 Simulation details

A previously described MD simulation code [107] was modified to incorporate the new water model (see Appendix C). 256 oxygen pseudo ions and 512 hydrogen pseudo ions were simulated in a box at constant pressure simulations and at 298 K temperature. Periodic boundary conditions (PBCs) were used. The simulations were performed employing a multiple time step algorithm, with a time step of 0.1 fs, and secondary time step of 8 fs. The cutoffs used in this method for primary and secondary neighbors were, respectively, 0.65 nm and half the box length ($L = 2\,\mathrm{nm}$). The temperature of system was kept constant using a Noose-Hover thermostat, and the same was used for charge- and dipole-dynamics ($T = 5\,\mathrm{K}$). Long range

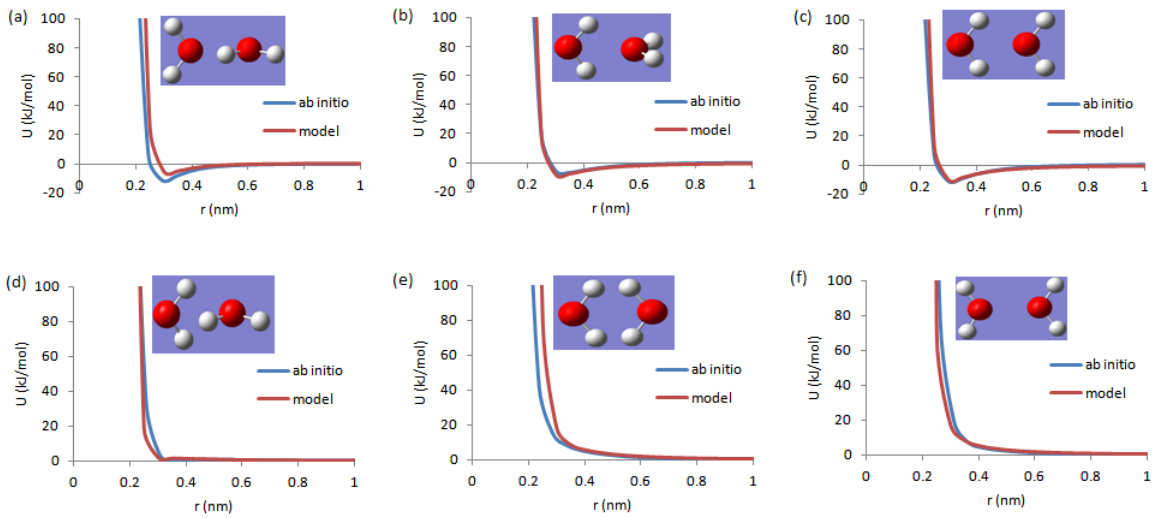


Figure 3.2: Water-Dimer potential energy scans. The water molecules are approached in different relative orientations and the model energy is matched with the corresponding *ab initio* energy. The *ab initio* calculations were performed at MP2/6-311(++)G (3df,2pd).

Coulombic interactions were handled using the Ewald summation method. A typical run involved $0.05 - 0.1$ ns for equilibration, and $0.5 - 1$ ns of production.

In our molecular dynamics simulations, since we do not have any rigid bonds, we needed a way to keep track of the hydrogen pseudo-ions that associate with the oxygen pseudo-ions to form water molecules. To this end, we employed a neighborhood list to keep track of the hydrogen neighbors that an oxygen ion is associated with. This list was updated every 0.2 fs. Another neighbor list exclusively kept track of the nearest oxygen neighbor of each species, and the information was used to exchange charge between species whenever needed (as described in section 4). This list was also updated every 0.2 fs.

In order to save time, our MD code uses a multiple-time step scheme, which is based on the fact that the long range forces do not need to be updated as frequently as the short range force. The code uses a multiple time scale method as follows: the shortest range forces (of order of intramolecular distance) are updated every time step (0.1 fs), other short range forces (< 0.6 nm distance) are updated every other time step (0.2 fs), and the long-range forces ($>$

0.6 nm) are updated every 8 fs. Also, to save time, the vdW and central force potentials and forces were completely tabulated, along with some parts of Coulombic potentials.

The feedback constants associated with the barostat and thermostat (see Section 2.1.5) are respectively: $Q_P = 0.7M/L$ and $Q_T = 0.01L^2M/Nk_B$, where M is the total mass of the particle, and L is the cell length and k_B is Boltzmann's constant. For solving the equations of motion for the valence charges (q_v) and the charge-shapes (b_i), we assign some fictitious masses to the respective degree of freedom. These masses were chosen to be as small as possible, so charges will rapidly equilibrate, without requiring an inordinately small time step to maintain stability. These masses were $m_q = 10^8 k_B \overline{T}_q (\delta t/|q_e|)^2$ and $m_b = 3.2$, where δt is the time step, and $\overline{T}_q = 5\,\mathrm{K}$ is the set point for the charge "temperature."

3.4.3 Structural and thermodynamic properties of bulk water

In order to assess the liquid structure resulting from our water model, first we examined the pair correlation functions or the radial distribution functions (RDFs) for all three atom pairs in our bulk water simulations (OO,OH, and HH). Fig. 3.3 shows the RDFs for the three pairs plotted against the experimentally obtained RDFs. For the OO pair, the first peak represents the first coordination shell of a water molecule. In good agreement with the experimental OO RDF first peak, it is located at 0.28 nm. The first OH peak is located at 0.1 nm, and it represents the hydrogen that is bonded to oxygen, while the first HH peak represents the other hydrogen that belongs to the same molecule, and is located at 0.16 nm. In the case of all three pairs, the model curves show modestly greater structuring than the experimental curves, as shown by the first-neighbor peak heights and the dips between the first and second neighbor shells. Furthermore, the model peaks corresponding to the second shell of water neighbors are slightly displaced to greater distances. This evidence suggests that the neighboring water molecules are "locked in" a little tighter than they should be. This causes the molecules in bulk to not move as freely as they would in a real system.

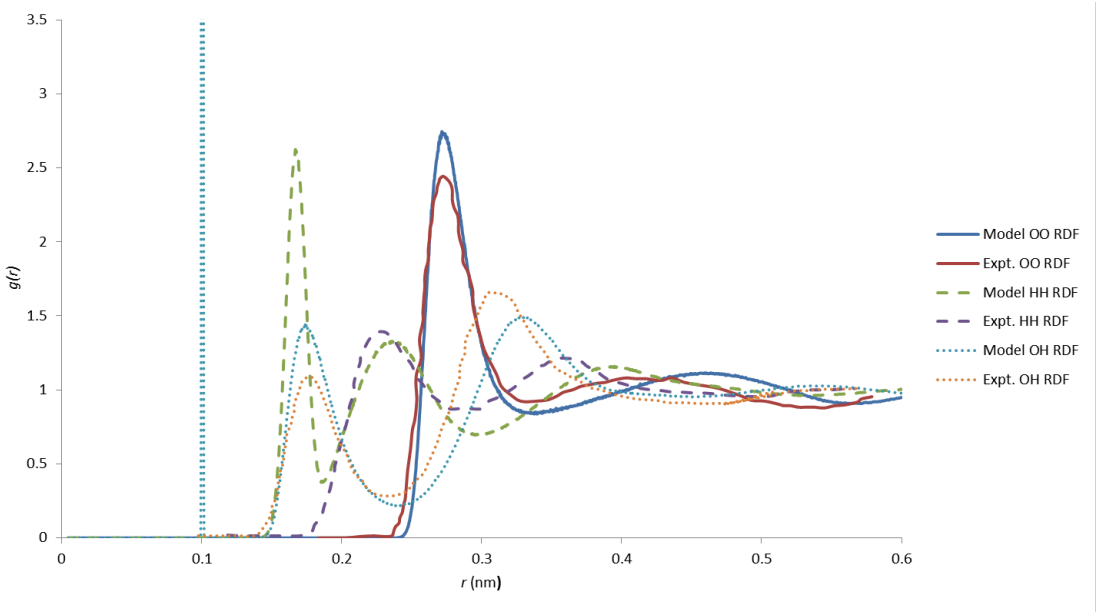


Figure 3.3: Model pair correlation function vs. experimental for Oxygen-Oxygen, Oxygen-Hydrogen, and Hydrogen-Hydrogen pairs

While the model RDFs are reasonably accurate, we nevertheless made multiple attempts to remedy this overstructuring effect by weakening the strength of interactions between atoms, including the CFM potential between oxygen and hydrogen. Such changes always caused problems with bulk water density or with anomalous autoionization of water, in which adjacent water molecules would spontaneously split into H_3O^+ and OH^- species. To our knowledge, all other models based on CFM have similar problems in matching the experimental RDFs. Because we considered the presence of autoionised water a significant misrepresentation of the actual liquid water system, we settled on the current set of interactions, even if they did not match the experimental RDFs as well as some other water models do.

Additional structural analysis was done by observing how the geometry of water molecules changes in course of a long simulation. As described previously, an oxygen atom, and its two nearest hydrogen neighbors were considered to constitute a water molecule. Over time, we kept track of the geometry of this water molecule, and found that the average bond length of the OH bond was 0.1002 nm and the average bond angle was 110.61°. In a gas-phase

simulation, the average bond length was around the same, while the average bond angle was around 114°. This increase in bond angle is consistent with experimental observation that the HOH angle in gas-phase is greater than in liquid phase [108]. The SPC/E model has a fixed bond length of 0.1 nm, and a fixed bond angle of 109.47°.

The Density of water calculated by our NPT MD simulations was 0.996 g/cm^3, which is in excellent agreement with the experimental value of 0.997 g/cm^3. The value obtained for bulk modulus (given by the correlation between pressure and volume fluctuation) from our simulations was 2.06×10^9 Pa , which was in good agreement with its experimental value of 2.2×10^9 Pa. Precise prediction of these properties, along with the fact that we have 100% neat water (no auto-ionization), RDFs that are in decent agreement to the experimental values, and close to the right geometry for the water molecules, give an indication of the quality of the simulated liquid structure.

3.4.4 Transport properties of bulk water

Since our aim is to study proton transport, it was important to us to study how well the model is doing in terms of prediction of dynamic properties of bulk water. The transport properties we looked into were self-diffusion coefficient and viscosity. Ionic conductivity of the system in presence of excess proton will be studied as a part of a detailed PT study, to be published at a future date.

For calculating these transport properties, we used the Green-Kubo approach, which relates the value of the time derivative of a mechanical variable at a specific time, designated as time zero, to its value at some later time. Fig. 3.4 shows the velocity correlation function (VACF) for oxygen sites, which is the integrand in the formula used to calculate the self-diffusion coefficient (D_{O}) in our MD simulation. The diffusivity is given by:

$$D_{\mathrm{O}} = \int_0^\infty \langle \mathbf{v}_{\mathrm{O}}(t) \cdot \mathbf{v}_{\mathrm{O}}(0) \rangle dt \tag{3.72}$$

The VACF tells us how strongly the velocity of particle is at time t, to what its value was at a previous time $t = 0$. Note that at $t = 0$, the VACF is a high because no time has elapsed and the velocities are identical. As time proceeds the value of the VACF in Fig. 3.4, expectedly, decreases as the correlation decays. This is because the particle undergoes collisions with other particles which change the direction and magnitude of its velocity from what it was originally. Also as would be expected for any dense fluid, the correlation becomes negative, because due to numerous and rapid collisions with other atoms/ions, velocity reversal takes place. After a long time, the VACF decays to zero as the velocities become uncorrelated.

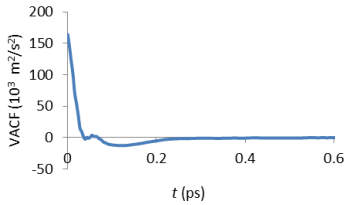
The self-diffusion coefficient is an important property to be matched, because it is solely responsible for vehicular diffusion. The vehicular diffusion of water is expected to be around the same as a hydronium ion, so if the self-diffusion coefficient is low, it leads to a lower PT rate prediction because the PT rate is the sum of vehicular and structural diffusion. As shown in Table 3.2, our model underpredicts the diffusion coefficient by almost 30%. This may be a consequence of an overly structured OO RDF, as described in previous section. The molecules do not have sufficient freedom to move around, which leads to a lower self-diffusion coefficient. Hoffmann and Bresme have also reported a lower self-diffusion coefficient and our model does a better job than these two models.

For viscosity (η), the shear stress auto correlation (SACF) function, shown in Fig. 3.4 was used. SACF is the integrand in the Green-Kubo formula for η :

$$\eta = \int_0^\infty \frac{V}{k_B T} \langle P_{xy}(t) \cdot P_{xy}(0) \rangle dt \tag{3.73}$$

Here V is the average cell volume, T is the temperature, and k_B is the Boltzmann constant. As is consistent with an over-structured OO RDF, and a low self-diffusion coefficient, the viscosity predicted by the simulation is over-predicted by about 11%. It is to be noted here that, within the current framework, it is possible to improve the prediction of transport properties if we relax the constraint of achieving the exactly right density. Specifically, if we

(a)



(b)

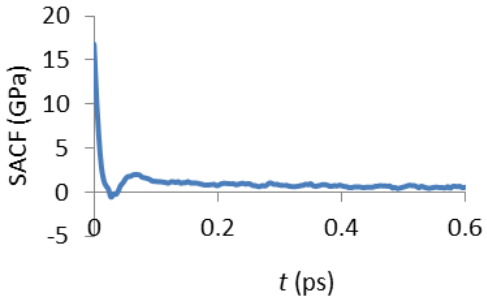


Figure 3.4: (a) The velocity auto-correlation function, used to calculate self-diffusion coefficient (b) The shear stress auto-correlation function, used to calculate viscosity

let the density be too low by 2.7%, we can increase the self-diffusion coefficient to around 1.9×10^{-5}cm^2/s and the viscosity to around 0.91 cP.

3.4.5 Dielectric properties

The relative dielectric constant ε of a material is the ratio of material's polarization response to that for a vacuum. It is an important measure of the accuracy of a model, but its calculation can be quite difficult due to its dependence on the fluctuations of the net system dipole moment ($\mathbf{M}$). ε is given by:

$$\varepsilon = \varepsilon_\infty + \frac{4\pi}{3k_BTV}\left[\langle\mathbf{M}^2\rangle - \langle\mathbf{M}\rangle^2\right] \tag{3.74}$$

This formula expresses the fact that the dielectric constant is proportional to the variance of the cell dipole moment. The value of ε generated this way converges slowly, and getting a reliable value generally requires much longer simulation times than for thermodynamic

properties. Here ε_∞ is the infinite-frequency dielectric constant. The experimental value of $\varepsilon_\infty = 1.79$ [109] was used.

Since there are no rigid bonds in our simulations, the use of PBCs complicates the calculation of dipole moment. That is to say, if one or both of the constituent hydrogens of a water molecule are on the opposite side of the box from the oxygen, a standard calculation of cell dipole moment would include an anomalously large dipole contribution from this water molecule. In order to correct for this, PBCs were modified so that the constituent hydrogens of a water molecule get transferred to the other side upon crossing a cell boundary only when the central oxygen is also transferred. Thus the hydrogens remain with their associated oxygen at all times, and the anomalous dipole moment is eliminated.

As listed in Table 3.2, the dielectric constant of water is under-predicted by about 7%. The average molecular dipole moment in bulk water for our model is also a good match with experiment, though there is considerable uncertainty on the experimental liquid dipole value [110]. Dielectric constant and molecular dipole moment are tightly coupled properties, and yet many popular water models are not able to simultaneously match both properties to the same degree as our model does. As a polarizable model, our model by design also matches the gas-phase value of water dipole, 1.85 Debye [110].

3.5 Conclusion

We have developed a water model that captures reasonably accurate electrostatic interactions, both at short and long distances. Molecule polarizability is implemented by the use of fluctuating charges, as well as fluctuating dipoles. This model was parametrized, as much as it was possible, from the *ab initio* data, and was then implemented into MD simulation. Novel time saving techniques like a modified Ewald sum specific to diffused charges and diffused dipoles and use of multiple neighborhood lists, were applied in MD simulation to make the code efficient. The model correctly avoids any significant amount of autoionization of

Table 3.2: Properties of liquid water models and experiments at ambient temperature

Model	T	ρ	$-U^{vap}$	P	D	ν	μ	ε
Exp[111–116]	298	0.997	9.92	0.00	2.30	0.89	2.6	78.3
Our Model	298	0.996	9.86	0.00	1.56	0.99	2.72	72.4
Hoffman CFM[89]	298	1.00			1.42			
Bresme CFM[87]	298	0.997	10.82	0.12	1.45		1.98	77
MSEVB[69]	298	1.00			2.9		2.7	
ST2[117]	298	0.997	8.69	0.62	2.9		2.35	69
SPC[118–120]	298	0.963	9.01	-0.00	3.3	0.40	2.27	68
SPC/E[102, 120]	298	0.998	9.89	-0.08	2.14	0.73	2.35	67
TIP4P[121, 122]	298	1.002	10.10	0.00	2.8	0.49	2.18	53

The properties listed in the table are: temperature T (K), density ρ (gm/cc), vaporization energy U^{vap}(KJ/mol), pressure P (Kbar), self-diffusion coefficient D ($\times 10^5\,\mathrm{cm}^2/\mathrm{s}$), viscosity ν (cP), dipole moment μ (Debye), and relative dielectric constant ε. Blank entries indicate no results provided in original sources.

water molecules, a problem found in other water models based on the central force potentials that allow bond formation and breaking.

Simulated bulk water properties were in reasonable agreement with the experimental values. The density predicted by the simulations matched almost exactly with the experimental value. The diffusion coefficient was low, which was in accordance with the over-structured RDF. The dielectric properties were also in good agreement with the experimental values.

The water model described here constitutes the basic framework of the methodology to study reactive systems. In the next chapter, the model's application to simulate PT is demonstrated.

Chapter 4
MD Simulations of Proton Transfer

4.1 Introduction

In this chapter is given a description of the implementation of proton transfer (PT) into the water model that was developed in the previous chapter. The basic concepts involved in modeling of the PT process are discussed in detail and the approach that is used in this work to realistically model the PT process is explained. For a more realistic description of an actual aqueous electrolyte, the model is then extended to include multiple excess protons coupled to chloride counterions.

To simulate PT in bulk water, it is essential that the force-field used to model the process must be able to accurately capture polarization, solvent effects, and the energy barriers for the PT reaction. In addition, since the model is to be used in a classical MD framework which does not allow for bond breaking and formation, the model must enable cleavage of covalent bonds under a consistent set of rules. All of these phenomena and characteristics of the bulk water system must be accurately represented under the additional constraint of getting the right thermodynamic and transport properties. Also, the quantum nature of protons must be incorporated in the model in some way.

Although incorporation of all these effects in a single force-field is a daunting task, there are ways to reproduce some the desired characteristics of the bulk water system using methods based on classical physics only. These methods entail a realistic representation of the system using a firm (and sometimes complicated) mathematical framework to accurately

depict various effects and phenomena. For example, polarizability may be captured reasonably well via a better description of charges and allowing for the charges to fluctuate. In the present work this was accomplished via the use of Gaussian-based charge densities with inherent dipole and quadrupole moments, and allowing the point dipoles placed on oxygen sites to fluctuate (see Chapter 3). Not only did our model capture the polarizability well, it also gave reasonably good matches for structural, thermodynamic, and transport properties of bulk water.

Solvent effects on the PT process may be incorporated intrinsically or extrinsically in the system Hamiltonian. In the present work, all interactions were dealt with intrinsically. Due to the small mass of a proton, some quantum effects also become prominent. Quantum tunneling (QT) is one such effect that may play an important role in the PT reaction. QT refers to the quantum mechanical phenomenon where a particle tunnels through a barrier that it classically could not surmount. In a classical MD simulation, QT may be approximated by using an effective energy barrier that is less than the true energy barrier for the PT reaction. A related quantum effect that may play an important role is delocalization of the proton. Delocalization refers to the wave-like character of the proton, such that the protonic charge is distributed over multiple locations. For instance, the thermal de Broglie wavelength for a proton at 298 K is 0.101 nm, which is significant. This could lead to a proton being simultaneously linked to multiple oxygens in the hydrogen bonded network during the PT event. There is no direct way to account for this in classical MD, but it may be approximated by reducing the PT reaction energy barrier and possibly making diffuse the proton core charge density. In the present chapter, we will demonstrate that our model represents the barrier heights of the PT reaction in the gas phase adequately well. Inclusion of a factor in the interaction energy between sites (see Eq 3.2) and minor changes to the model parameters to adjust the PT rate was our way of reducing the PT reaction energy barrier and incorporating the quantum effects.

The most significant challenge in modeling PT is allowing for bonds to break and form in classical MD simulations. Various groups have addressed this problem and have proposed different ways of dealing with breaking and formation of covalent bonds. A detailed account of such force fields was presented in Chapter 2. Only the core concepts involved in oxygen-hydrogen covalent bond cleavage and formation will be briefly compared and contrasted here.

Voth and coworkers [69], in their MS-EVB force-field, allow the breaking and formation of bonds based on a weighted superposition of most probable bonding states. The nuclei remain at the same position and the weights on different states are varied to simulate bond breaking and formation. Most probable bonding states were selected based on the distance between a hydronium ion and the water molecules and the number of water molecules in the first coordination shell.

Goddard and coworkers' ReaxFF model [77] uses interatomic distances to calculate bond order and correlates it to bond energy. Change in bond-orders is then used to simulate breaking and formation of covalent bonds, which is reflected in change in system energy via the correlation. Over- and under-coordination is penalized to keep the representation physically accurate.

Keffer and coworkers model for the PT reaction [86] entails the satisfaction of geometric and energetic triggers before a PT event can take place. Once the proton has transferred, the methodology requires local equilibration of the molecules in the vicinity of the donor and acceptor pair. The drawbacks of these methods, as discussed in Chapter 2, lie either in extension to new chemistries (inclusion of different species in the system, like a solid catalyst), parametrization (huge amount of *ab initio* data required), or computational expense.

The bond breaking and formation in the present work is accomplished by the use of a central force formalism, which eschews rigid bonds and holds the atoms of a molecule in the right geometry by the use of soft potentials. Such a framework allows the bonds to break

and form naturally. Such a description is a more realistic representation of an actual system and the approach ensures that the PT process occurs under a consistent set of rules.

A problem that arises while transferring a proton between oxygens is that the process should involve a net charge transfer of $+1|e|$ but most water models (current work included) do not have a total charge of $+1|e|$ on a proton or hydrogen site. So a mechanism is required that ensures the right amount of charge is exchanged between the accepting water molecule (which becomes a hydronium after accepting a proton) and the donating water molecule (which was a hydronium). Specifically, at the same time as the PT, we transfer electronic charge between the two oxygens involved, such that the net charge transferred is $1|e|$. However, in the MD framework we cannot suddenly and/or manually alter the charge on the oxygen sites, as it leads to instabilities. Thus, there must be a natural physical methodology for charge transfer. The methodology used in this work transfers charge based on differences in electrochemical potential and distance between the two oxygen sites.

In an actual bulk water system, the protons will be accompanied by the presence of an anion or a counterion (thus making the system a dilute acid solution). Thus a realistic simulation of the system must include a counterion. Some models choose to ignore it while others consider it as a non-polarizable negatively charged entity. In the current work, a counterion is included in the simulation and it is allowed to polarize to allow a more accurate description of the actual system.

In the following sections, first, a description of the charge transfer process accompanying the PT process is given. Next, we discuss how the model potential energy surface (PES) compares with *ab initio* PES for PT between two water molecules. Then, we discuss the implementation of the model into the MD framework in order to study PT in bulk water, with a single excess proton. The methodology is then extended to studying PT with a single excess proton and a counterion. Lastly, we discuss the application of model for studying PT with multiple excess protons.

4.2 Charge Transfer Accompanying the PT Reaction

As mentioned above, since there are no rigid bonds binding the O and the H sites, the PT occurs naturally whenever the electrochemical environment around the sites, in vicinity of excess protons, is favorable. As mentioned previously, the PT event needs to be accompanied by the electronic charge transfer process, because the PT event must involve a net charge transfer of $+1|e|$, and since $q_H^c = 0.4238|e|$ (see Table 4.2), a net electronic charge$-0.5762|e|$ must flow from the proton acceptor oxygen to the proton donor oxygen (see Fig. 4.1). To model this electronic charge transfer, we use electrostatic driving forces underlying our model. Each site has associated with it a chemical potential for charge, φ_i, which is simply the derivative of the potential energy U, with respect to the valence charge on the site, q_i^v:

$$\varphi_i = \frac{\delta U}{\delta q_i^v} = \sum_j C_{ij}^{vv} q_j^v + \sum_j C_{ij}^{cv} q_j^c + \phi_i^0 \tag{4.1}$$

This chemical potential is the spatially weighted average of the electric potential field about site i, with weight given by the Gaussian distribution in Eq. 3.1. This quantity encapsulates the electrostatic environment around the site i, and whenever there is an excess proton in the vicinity of a site, this is reflected in the φ_i value. Electrostatic potential φ_i is the appropriate driving force for charge transfer between any adjacent sites, in this case nearest oxygens involved in the PT. The charge flow rate ($\dot{q_i^v}$) is modulated by the following expression:

$$\dot{q_i^v} = -\frac{\varphi_i - \varphi_j}{A_R \exp\left(s_{\mathrm{OO}}\right)} + \frac{q_i^{v*} - q_i^v}{\tau_c} \tag{4.2}$$

The denominator of the first term on the right is a distance-dependent resistance for the flow of charge. A_R is a constant pre-exponential factor and $s_{\mathrm{OO}} = \gamma_{\mathrm{OO}} r_{\mathrm{OO}}$ is the dimensionless distance between the two O sites (i and j) involved in the valence charge transfer.

Although this approach, in a general way, ensures that the charge flows in the right direction when a PT event is taking place (or is about to take place), it does not ensure that a net charge of exactly $+1|e|$ will be transferred, which can create problems over long time

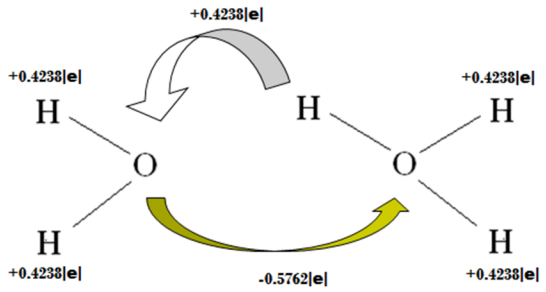


Figure 4.1: Proton transfer event accompanied by effective electron transfer from acceptor to donor oxygen site

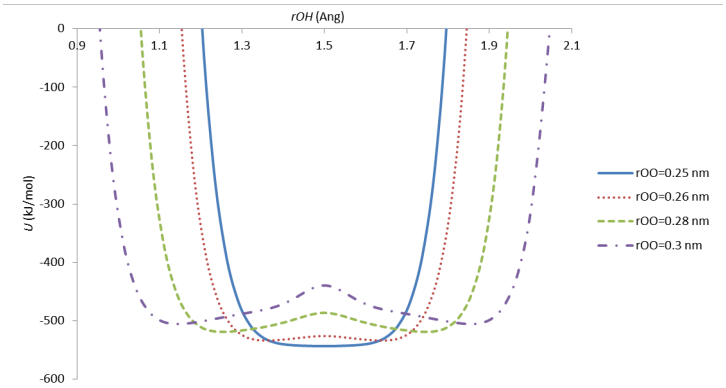
scales. In order to fix this problem, we included a second term in Eq. 4.2 that is an indirect mechanism to slowly correct the net charge transferred. In particular, the oxygen site valence charges asymptotically approach their design values of of$q_i^{v*} = -1.2914|e|$ for a hydronium ion and $q_i^{v*} = -1.8676|e|$ for a neutral water molecule. The parameters were adjusted empirically in order to get numerically stable and reasonable rates of charge transfer:$A_R = 30\frac{\text{J ps}}{\text{mol } |e|^2}$ and $\tau_c = 300$ ps.

4.3 PES of PT in Zundel ion

Fig. 4.2 shows the potential energy scan of transfer of a proton moving between two isolated water molecules, with the proton moving along the line joining the two oxygens. In the Figure, rOO is the distance at which the two oxygens and held at as proton moves from one to the other. In absence of any external field, we get a symmetric double-minimum curve. The barrier height of the symmetric PT event depends upon the distance between the two oxygen atoms. With our model, we have reproduced the barrier height for PT at different O-O distances and, as shown in Table **4.1**, it was in good agreement with the *ab initio* data.

When an electric field is applied along the PT axis, as shown in Fig 4.2, the curves distort and shift so as to make the state where proton is with the acceptor oxygen as more stable. This shows that the model responds to an external electric field, and during the PT event, the proton chooses to be with that water molecule around which the electrochemical

(a)



(b)

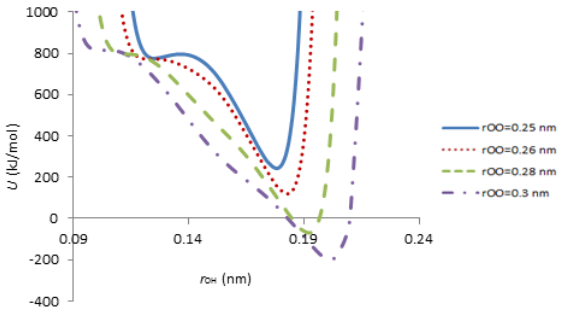


Figure 4.2: Potential energy surface for proton transfer along the axis joining the two oxygen sites (a) Symmetric case, (b) Electric field applied along the PT axis.

environment is more favorable. This characteristic of the model is important because in the bulk phase, the excess proton is under the influence of strong electric fields, and it will follow a PES that must be accurate for the PT mechanism.

4.4 MD Simulations with One Excess Proton

4.4.1 Preliminary studies

The PT scheme was implemented in an MD simulation. 256 water molecules were simulated in a 2 nm box with one excess proton. The simulations were set-up in a similar manner as the bulk water simulations, and were run for 16 ns of production time, after 1 ns of equilibration. The subsequent results from the MD simulations were encouraging. The density and RDFs

Table 4.1: Model predicted activation energy vs. *ab initio* activation

OO Distance (nm)	*Ab initio* Activation Energy (kJ/mol)	Model Activation Energy (kJ/mol)
0.25	0	0
0.26	5	6
0.28	28	30
0.30	67	71

for the three atom pairs, expectedly, did not change in the presence of an excess proton. Next, we examined the geometry of hydronium ions during the course of the simulation. The average OH bond lengths were found to be 0.101 nm, and the HOH bond angle was 113°. For comparison, in a gas phase simulation of an isolated zundel ion, these values were 0.101 nm and 116°, respectively. In the course of our MD simulation, multiple PT events including numerous barrier-recrossing events were observed. The preliminary estimate of PT rate in the bulk simulation was $0.15\,\mathrm{ps}^{-1}$, determined by dividing the number of PT events by the total simulation time. This rate is roughly a factor of 5 lower than the experimental value [123].

4.4.2 Further analysis of PT

4.4.2.1 Diffusion coefficient of excess protons

The method to calculate PT rate in bulk simulations in our preliminary studies was based on gross average of the number of PT events observed. A systematic method to study the diffusion of a proton, via both the structural and vehicular diffusion, was required for a detailed analysis. In order to accomplish this, we used a modified Green-Kubo (GK) approach. In the GK approach, as described in Section 3.4.4, a velocity correlation function (VCF) needs to be calculated in order to calculate the diffusion constant of a species. For this purpose, we need to keep track of the velocity of the excess protons at various time intervals. In an MD simulation for the PT process, it is difficult to keep track of the excess proton, because once a proton attaches to a water molecule to form a hydronium ion, all

three hydrogen atoms are now equivalent and any one of them can be transferred to an acceptor oxygen when a PT event occurs.

In order to get around this problem, instead of keeping track of the excess proton, we keep track of the oxygen that has the excess proton on it (that is, the oxygen of the hydronium ion). When a PT event occurs and the proton transfers over to the other oxygen, we start tracking the newly formed hydronium ion. There is a step change in the VCF because of the PT event, but the GK method is robust enough to handle discontinuous displacements because it integrates over the VCF curve. The following equations delineate the methodology:

$$VCF = \langle \mathbf{v}_{\mathrm{Pr}}^{\mathrm{rel}}(t) \cdot \mathbf{v}_{\mathrm{Pr}}^{\mathrm{rel}}(0) \rangle \tag{4.3}$$

where,

$$\mathbf{v}_{\mathrm{Pr}}^{\mathrm{rel}}(t) = \mathbf{v}_{\mathrm{O}}^{\mathrm{hydr}}(t) + \frac{\boldsymbol{\Delta}\mathbf{r}_{\mathrm{don-acc}}}{\boldsymbol{\Delta}t} \tag{4.4}$$

here, $\mathbf{v}_{\mathrm{Pr}}^{\mathrm{rel}}(t)$ is the corrected velocity of oxygen associated with the excess proton at time t, $\mathbf{v}_{\mathrm{O}}^{\mathrm{hydr}}(t)$ is the velocity of the oxygen which forms the hydronium ion at time t, and $\Delta\mathbf{r}_{\mathrm{don-acc}}$ is the distance between the donor and the acceptor oxygen atoms in a PT event. It is to be noted that $\Delta\mathbf{r}_{\mathrm{don-acc}}$is accumulated over time before it is used in averaging. Δt is the sampling interval.

If a proton transfers back to its old (previously donor) oxygen, the recrossing event will register as a negative of previous displacement, bringing the net sum of $\Delta\mathbf{r}_{\mathrm{don-acc}}$ contributions to zero. Hence, if a proton transfers back and forth between two oxygen, only the net displacement will be considered. With each recrossing event, a step change will be introduced in the VCF, and as stated previously, the GK method can handle step changes. Using the above VCF we calculate the diffusion constant in the same manner as we did in Section 3.4.4.

4.4.2.2 Results and discussion

With this modification to the methodology, we again simulated 256 water molecules with one excess proton. Reader is referred to Appendix C for the MD code. The simulations were set-up in a similar manner as previous simulations. After 16 ns of production time, the thermodynamic properties of the system were, as expectedly, found to be the same as the bulk water. The structural properties did not change either.

The VCF obtained from the simulation is shown in Fig 4.3. The PT rate based on the VCF data was found to be $8.07 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$, which was 13% lower than the experimentally observed PT rate of $9.3 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$. The contribution of vehicular diffusion (VD) to PT rate was $1.60 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$ and the contribution of structural diffusion (SD) was $6.47 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$. The VD contribution was approximately the same as the diffusion contant for water. This is expected because, since the mass of a water molecule and a hydronium ion is almost equal, they are expected to diffuse at almost the same rate. As mentioned in Section 3.4.4, this value is 30% too low, when compared to the experimental value of water diffusion constant. The VD contribution, although consitent with the bulk water simulation results, is thus low.

The experimental value of SD, obtained by substracting the total diffusion constant and the VD of water, is about $7.0 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$. The contribution of SD obtained from the simulations was only about 7.5 % too low, indicating that the simulations are able to mimic bond breaking and formation with adequate accuracy. Also, this indicates that the height of energy barriers for the PT reaction between a hydronium ion and the water molecule are accurately represented by the model.

The net rate of proton transfer obtained by compared to Voth and coworker's using the MS-EVB model was 50% too low. With only 13% error in the PT rate, our model matches the experimental value pretty well, although, the contribution of the VD is underpredicted.

Since the PT event is a rare phenomenon, we do not get adequate sampling to obtain a smoothly converging VCF curve. This is the reason behind the anamolous dip in the VCF curve at 1.4 ps, whereas one one normally expect a VCF curve to converge asymtotically

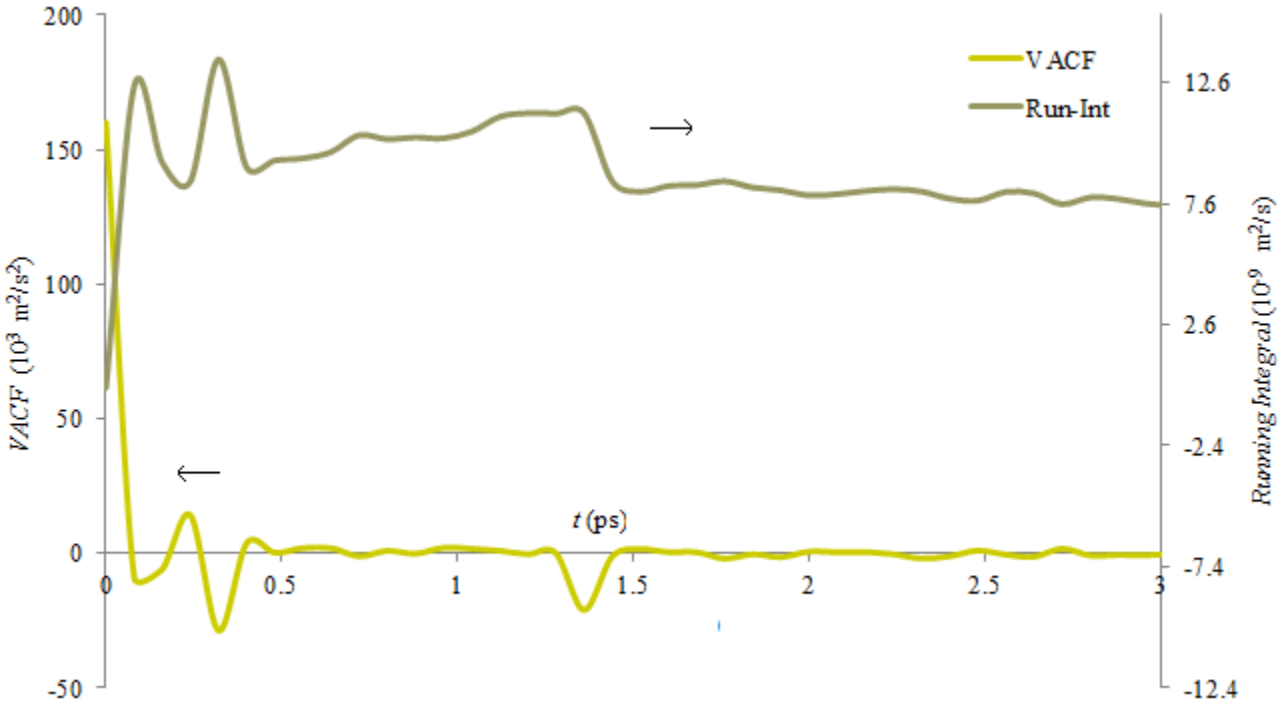


Figure 4.3: The velocity auto-correlation function, used to calculate the proton transfer rate

to zero at long times. Multiple proton transport events seem to be occuring near that time period. This sampling error is somewhat mitigated upon calculation of the running integral.

4.5 Inclusion of Counterion

One of the objectives of current study was to study the PT process in dilute acid solutions. Hydrochloric acid (HCl) was chosen for the current study. In accordance with the central-force methodology that has been used to describe the water molecules in our simulation, the HCl molecule also does not have a rigid bond. Hydrogen and chloride ion were treated like separate unbonded species and were held together with soft pairwise potentials.

An easy way to treat the chloride ion in a bulk MD simulation of PT is to consider it non-polarizable, but we decided that a polarizable chloride ion will be a more accurate representation of the actual system. So, a chloride ion in the simulation interacts interacts

with other species via Coulombic and vdW interactions, and also, it has the ability to polarize in response to the external stimuli. A detailed discussion of modeling of chloride ion is presented in the following discussion.

4.5.1 Model for chloride ion

In order to model a chloride ion, the description of charge density on the chloride ion site was kept the same as the other sites in the simulation. A positive core charge q^c_{Cl} and the inverse charge-width parameter γ_{Cl} were used as site parameters. The chloride site has a valence charge density of magnitude q^v_{Cl} and a point dipole b_{Cl}. The chloride site interacts with other sites in the system in accordance with the same Coulombic interaction potentials as described in Section 3.2.1.

A central force formulation is used to describe close-range interactions between $Cl^-$and H^+ionic species. For the sake of consistency, the functional form of these interaction potentials were kept the same as those between oxygen and hydrogen species. The potential for the $Cl^- - H^+$pair is given as:

$$U^{\mathrm{CFM}}_{\mathrm{ClH}}(r) = D1_{\mathrm{ClH}} \exp\left(2a_{\mathrm{ClH}}\left(r_{e\mathrm{ClH}} - r\right)\right) - D2_{\mathrm{ClH}} \exp\left(a_{\mathrm{ClH}}\left(r_{e\mathrm{ClH}} - r\right)\right) \tag{4.5}$$

4.5.2 Results and discussion

4.5.2.1 Chloride ion model parametrization

As one would expect, the interaction between chloride ions and other species, particularly hydrogen, must not be too weak nor too strong. In order to parametrize the model, we first fitted the q^c_{Cl} and γ_{Cl} parameters in order to reproduce 10% of experimental gaseous Cl^-ion polarizability . Ideally we would have liked to fit it to a higher fraction of gaseous ion polarizability, but when we did that, the chloride ion over-polarized when energy scans of the ion were performed with other species (H_2O and H_3O^+). Once we had the q^c_{Cl} and γ_{Cl}

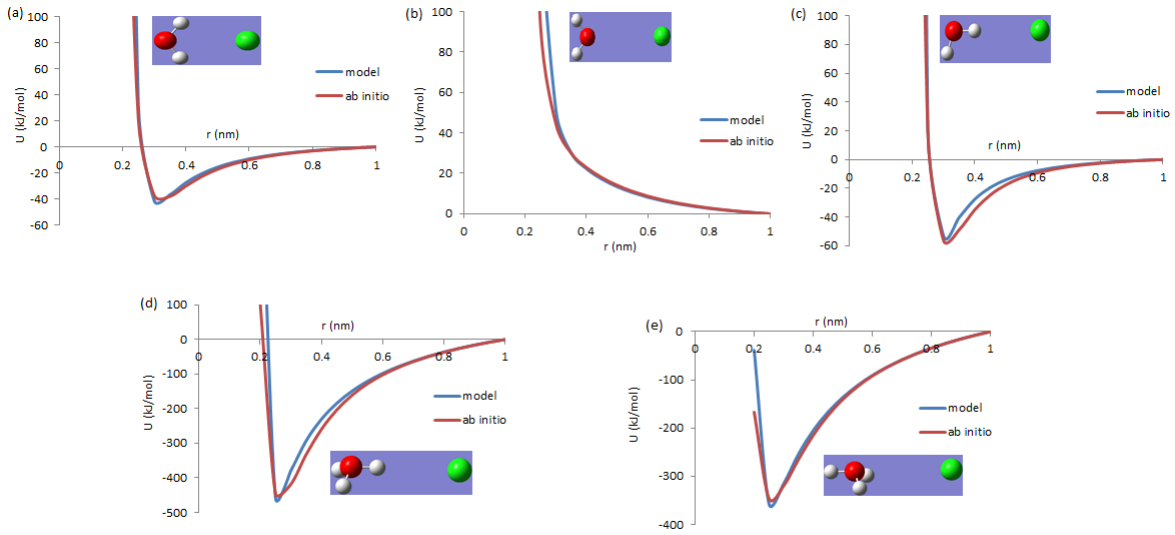


Figure 4.4: A comparison of *ab initio* PES and model PES for $Cl^- - H_2O$ (a-c) and $Cl^- - H_3O^+$ (d,e) pairs. The orientation of water, hydronium ion, and the lone chloride ion for different routes are shown next to the respective energy scan.

parameters, other parameters were adjusted to reproduce the energy scan data of $Cl^- - H_2O$ and $Cl^- - H_3O^+$ pairs. The *ab initio* data generated for comparison was at MP2/6-311+G (3df,2pd). Fig 4.4 compares the model PES and *ab initio* PES for the two scans. The selected set of parameters, listed in Table 4.2, gave good agreement between the model and *ab initio* data. Minor disagreements were allowed in order to match various other scan data as described in the following discussion.

In addition to comparing the PESs for the two pairs, we also compared the PESs for the cases when a proton is transferred between a hydronium ion and a chloride ion and also between water and a chloride ion. This was done in order to determine how likely it was for a chloride ion to accept/take a hydrogen from water and an excess proton from a hydronium ion. If the chloride ion is able to break water molecules and hydronium ions by accepting a proton, then it may have significantly undesirable effects on structural, thermodynamic and transport properties of the bulk system and also on the PT process.

Fig. 4.5 compares the PESs for PT scans for the two cases. Fig 4.5(b) shows that both the model and *ab initio* data predict that a chloride taking a hydrogen from water will be a

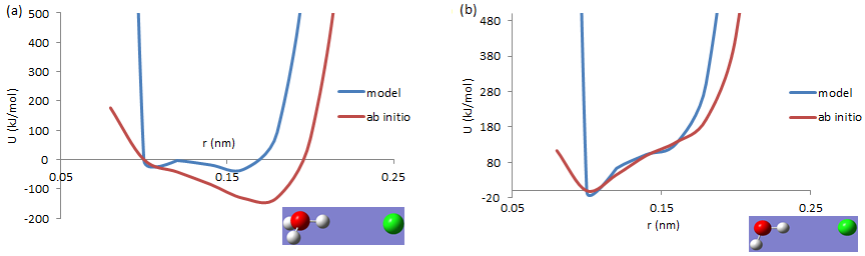


Figure 4.5: A comparison of *ab initio* PES and model PES when (a) the excess proton is pulled from hydronium ion to chloride ion (b) a hydrogen atom is pulled from the water molecule to the chloride ion. In both cases r is the distance between oxygen and transferring hydrogen. The orientations of the constituent species are indicated in the respective pictures.

very unlikely event. This is so because the corresponding configuration of a system where a chloride ion is associated with hydrogen is an extremely high energy one. For the selected set of parameters, the model and *ab initio* data for this case are in good agreement, except the short-range repulsion for oxygen and hydrogen and chloride ion (i.e. short range repulsion for the model is stiffer). Since the two pairs of species do not actually access the states where they are so close (<0.07 nm), the steep repulsion should not have any affect on bulk properties.

For the case of the PES for PT between a hydronium ion and a chloride ion, as shown in Fig 4.5(a), the *ab initio* data predict that when a proton is near a chloride ion, the system is more stable. This does not appear to be correct for aqueous systems because when HCl is added to water, it strongly dissociates. This discrepancy is likely due to the polarization of oxygen and chloride ion that takes place in the bulk liquid. Also, our preliminary MD simulations indicated that we needed to destabilize the configuration where hydrogen and chloride are strongly associated. Hence we decided to tune our model parameters so that the state where the proton in closer to chloride ion is less favorable (has higher energy) than what is predicted by the *ab initio* data. Although this choice results in a disagreement between the model and *ab initio* data (for this one case only), the corresponding model parameters enabled a better description of interactions in the bulk phase.

Table 4.2: List of parameter values for chloride ion.

Parameter	Value
q^c_{Cl}	$1.02\,\lvert e\rvert$
γ_{Cl}	$1.352\ \mathrm{nm}^{-1}$
λ_{Cl}	0
$D1_{\mathrm{ClH}}$	200 kJ/mol
$D2_{\mathrm{ClH}}$	0 kJ/mol
a_{ClH}	$35\ \mathrm{nm}^{-1}$
$r_{e\mathrm{ClH}}$	0.125 nm
L_{ClO}	1760 kJ/mol
θ_{ClO}	4.25
η_{ClO}	17.3

4.5.2.2 Simulation details

No new modifications to the MD code were required for the inclusion of the chloride ion. 256 oxygen pseudo ions, 513 hydrogen pseudo ions and a single chloride ion were simulated in a box at constant pressure simulations and at 298 K temperature. The time step used was the same as that used for bulk water simulations and simulations of PT in bulk with one excess proton (0.1 fs). Long range Coulombic interactions were handled using the Ewald summation method. A typical run involved 0.1 ns for equilibration, and 1 ns of production.

4.5.2.3 Structural and thermodynamic properties

In the presence of a single counterion, the radial distribution functions of O-O, O-H, and H-H pairs were unaltered and were found to be same as shown in Fig. 3.3. The density of the system was also unchanged (0.996 g/cc). The O-Cl and H-Cl RDF is shown in Fig. 4.6. The first and the second peaks of the O-Cl RDF are at around 3 nm and 5 nm, respectively, which indicates a good match to the experimental RDF data. Also, there is no overstructuring which indicates the interactions between oxygen and chloride ion are being represented correctly. The model H-Cl RDF, on the other hand, although has the first and the second peaks at the right distances, the curve is more structured than the experiment RDF.

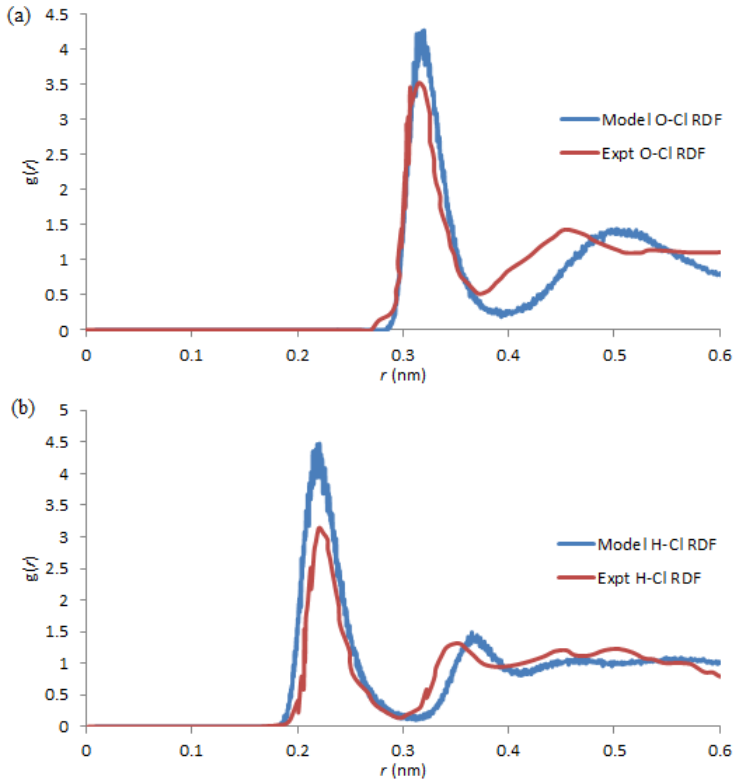


Figure 4.6: Model pair correlation function vs. experimental for (a) oxygen-chloride ion, and (b) hydrogen-chloride ion pairs.

This indicates increased attraction with the chloride ion. We have weakened the attraction between hydrogen and chloride ion from *ab initio* results, but the two species still seem to be too strongly bound. This increased attraction resulted in a low diffusion coefficient of the excess proton and hence led to a decreased PT rate in presence of the chloride ion. The PT rate in this case was found to be $6.02 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$, with VD contribution being $1.6 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$ and the SD contribution being $4.5 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$.

4.6 Extension to Multiple Excess Protons

By design, our force-field is capable of handling multiple excess protons. No change in formulation/methodology was required to extend the model to incorporate multiple excess protons. Up to 6 excess protons were included in this study, corresponding to maximum

concentration of about 1 M. The time step, equilibration time, and production time were all kept the same as before.

For this system, the structure pertaining to O-O, O-H, and H-H pairs was unchanged and was found to be same as shown in Fig. 3.3. The density of the system was also unchanged (0.996 g/cc). The proton conductivity from the $VACF$ data was found to be $7.94 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$, with VD contribution being $1.6 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$ and the SD contribution being $6.4 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$.

4.7 Conclusion

The water model was extended to include excess protons. In order to simulate PT, first a robust charge transfer scheme was developed to ensure that the right amount of charge is exchanged between the donor and the acceptor oxygens during the PT event. Second, the *ab initio* PES for proton transport between two water molecules was reproduced using our model and the two data, pertaining to the PT reaction energy barrier, were in good agreement. The model was then used in the MD framework and the PT rate found from preliminary studies, which was determined by simply counting the PT events, was found to be too low by a factor of 5.

More sophisticated techniques were used in order to better estimate the rate of proton transport. These involved using a Green-Kubo approach and determining the PT rate via diffusion of a proton. The PT rate from this approach was found to be $8.07 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$, which was 13% too low, though the contribution of structural diffusion calculated from simulations gave a good agreement with the experimental value. The model was then extended to include chloride ions for a better representation of an electrolytic system. The model parameters for the chloride ion were determined by reproducing *ab initio* data of energy scans for chloride-water and chloride-hydronium ion pairs and also by reproducing PES of proton transport between these two pairs. The functional form of the interaction potentials were

unchanged. The MD simulations for the system yielded the right density and structure of the system. The proton conductivity was found to be $6.02 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$. The method was then extended to include multiple excess protons in the system. The proton conductivity for this case was $7.94 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$.

Chapter 5
Dipole Moment Determination

5.1 Introduction

In Chapter 2, we discussed the experimental and theoretical techniques used to determine or estimate the dipole moment of various molecules. As stated earlier, previous theoretical methods of dipole moment determination involve the evaluation for only the ground state (lowest energy) configuration of the molecule only. Considering only the ground state electronic configuration can be inadequate when determining dipole moments, because even at room temperature a molecule may access multiple conformations that each have different dipole moments. Thus, considering multiple conformations and evaluating the average dipole moment from all of these conformations (that are accessible to a molecule) can improve the prediction. The MD method can be a useful tool when multiple configurations need to be accessed.

Multiple conformations can also be accessed by manually rotating the molecules around a dihedral angle or otherwise perturbing the geometry. Dipole moment calculations can be performed on each such geometry using computational chemistry software like Gaussian. The average dipole moment may then be calculated by a weighted average of dipole moments thus obtained, with the weights depending upon the energy of that particular configuration. For example, these states may be weighted using a Boltzmann weighting factor or probability:

$$p_i = \exp\left(-\frac{E_i}{k_B T}\right) \tag{5.1}$$

where E_i is the energy of a particular probed geometry and T is the temperature of interest. This approach has two drawbacks. First, for large molecules having branched structures and multiple dihedral angles, it becomes difficult and cumbersome to include all the possible structures in the analysis. Second, such calculations do involve the averages of dipole moment over multiple geometries but completely ignore the entropic contribution of each state.

Even though a particular state may be energetically favorable, it may not be as likely because it may not be favored entropically. In order to account for this entropic contribution, the probability, w_i, of occurrence of a state i of energy E_i is given by:

$$w_i = g(E_i) \exp\left(-\frac{E_i}{k_B\, T}\right) \tag{5.2}$$

where $g(E_i)$ is a temperature-independent quantity known as the *density of states* that corrects the weighting factors (probabilities) for entropic contributions.

The sampling in MD and MC simulations naturally accounts for the entropy effect. An advantage of using MC simulations over MD simulations is that in MC simulations, if there are energy barriers between conformations, they are easily crossed due to the inherent randomness involved in the MC sampling. But since the scheme is to be used with an *ab initio* methodology, which requires the calculation of electronic structure at each step, a downside to using an MC scheme is that it can be enormously expensive, because at every step of the simulation, the electronic structure is needed to be calculated from the beginning. As mentioned in Chapter 2, the MD scheme is advantageous because at every step, the electronic structure can be calculated by minor perturbations of the electronic structure from the previous step. In the interest of inclusion of entropy-based sampling and also saving the computational effort, we decided to use the MD scheme.

Attempting to study multiple conformations using MD simulations is not a new concept and is frequently used in studying proteins. Specifically, how proteins change their conformations over a period of time and how they fold have attracted special attention over the last

few decades [124]. These studies find application in high-performance drug discovery [125]. MD method provides a useful insight in understanding how the structure of proteins evolves in time and the MD simulations are used to examine the energetics and mechanisms of conformational change. In these simulations, individual trajectories of different time lengths are generated to sample conformational distribution.

The idea of including multiple conformations in prediction of the dipole moment, while ensuring that the electronic structure is correct for the respective geometry at each MD time step, is novel and to our knowledge this area of study is relatively unexplored. In the present work, AIMD simulations have been used in order to estimate the dipole moment of molecules that have conformational isomers. The AIMD methodology was explained in Chapter 2. Determination of accurate electronic structure at each step ensures that the dipole moment of the molecule is calculated correctly. The net dipole moment is calculated by averaging the dipole moments of the different configurations that a molecule attains as it evolves in time.

The main objective of the work developed in this chapter is to develop a general method for the estimation of average dipole moment of any compound. Towards this end, through preliminary studies, we developed an "approximate method" for dipole moment determination using an open source AIMD code, SIESTA (Spanish Initiative for Electronic Simulations with Thousands of Atoms). The approximate method was used to study tributylphosphate (TBP) molecule. TBP was selected because the experimental value of its dipole moment was known. It was also one of the compounds thoroughly studied in the theoretical dipole moment studies by Liu et al. [26]. Certain crude assumptions were made in the approximate method, which made the method less reliable. The results obtained from this preliminary study also indicated that we needed to refine the methodology for better dipole moment predictions.

The method was then refined by making certain modifications to the SIESTA code and the subsequent analysis to develop an "exact method" for average dipole moment determination.

Along with TBP, the compounds used for this study were: 1,2-dichloroethane (DCE), 1,2-dinitrioethane (DNE), 1,2-ethanediol (EDL), and 1,1,1,2-tetrafluoroethane (TFE). These compounds were selected because they have only one dihedral angle and are easy to analyze. After running simulations of these compounds with the modified SIESTA code, the resulting distributions of dipole were analyzed and a general procedure for calculation of average dipole moment was developed based on the findings.

In the present chapter, first a detailed discussion about the SIESTA AIMD code is presented. Next, the approximate method is discussed and the results obtained are presented. Next, the refinements on the approximate method are discussed and various aspects of the exact method are discussed in detail. A comprehensive analysis of the dipole moment distributions is also presented. Lastly, based on our studies, a general method to calculate the average dipole moment of any molecule is presented.

5.2 The SIESTA AIMD Code

The SIESTA AIMD code was used in this work to calculate dipole average dipole moment of various compounds. SIESTA is a software implementation for performing electronic structure calculations and *ab initio* molecular dynamics simulations of molecules. SIESTA is a fully self-consistent DFT method that predicts the physical properties of a collection of atoms. In the present study, as will be demonstrated in this chapter, the SIESTA code has been used to study the time evolution of dipole moments of various polar organic compounds. Before going over the implementation details, we will outline here some important aspects of the SIESTA program.

5.2.1 Pseudopotentials

The use of pseudopotentials [126] is an approximation to simplify the description of complicated systems. Specifically, a pseudopotential is an attempt to replace the complicated

effects of the motion of the core (i.e. non-valence) electrons of an atom and its nucleus with an effective potential, or pseudopotential, so that the Schrödinger equation contains a modified effective potential term instead of the normal Coulombic potential term for these core electrons. So, in this approach only the chemically active valence electrons are dealt with explicitly, while the core electrons are 'frozen', being considered together with the nuclei as rigid non-polarizable ion cores.

'Norm-conserving' pseudopotentials [127] enforce the condition that, outside of a cutoff radius, the norm (functional form) of each pseudo-wavefunction be identical to its corresponding all-electron wavefunction. Norm-conserving pseudopotentials are derived from an atomic reference state, requiring that the pseudo- and all-electron valence eigenstates have the same energies and amplitude (and thus density) outside a chosen core cutoff radius. SIESTA uses pseudopotentials to represent the electron-ion interaction.

The norm-conserving pseudopotentials used as input for the SIESTA program were generated using the ATOM program [128]. These generated pseudopotentials were verified by comparison to the pseudopotentials generated by the Octopus pseudopotential generator [129].

5.2.2 Basis set

A basis set is a set of functions that are used to create molecular orbitals. A molecular orbital is expressed as a linear combination of atomic orbitals, so, if ϕ_i is an atomic orbital, the molecular orbital ψ will be given by:

$$\psi = \sum_{i=1}^{K} c_i \phi_i \tag{5.3}$$

where c_i is the coefficient denoting the weight of each of the K atomic orbitals. The more atomic orbitals we use to describe the molecular orbital (that is larger the basis set size), the more accurately we are able to describe the molecular orbital, but the cost is greater

too. There are two types of formulations for the atomic orbitals that are used: Slater type orbitals (STOs) and Gaussian type orbitals (GTOs). STOs have radial and spherical components and they resemble the hydrogen orbitals. Originally, STOs were used as representations for atomic orbitals. GTOs, as the name suggests, resemble a Gaussian function in form. Choosing the GTO form of function leads to mathematical simplification and reduces computational effort, and thus is extensively used in almost all computational chemistry software.

SIESTA uses numerical atomic orbitals (NAOs) [130], which differ from traditional representations of atomic orbitals. They are represented numerically on radial grids rather than analytically. While this makes calculations more time consuming, many of the required radial integrations can be performed beforehand and tabulated. NAOs have the advantage that they can be designed with any shape without extra computational cost, and can easily be localized to avoid expensive long-range interactions, thus achieving better scaling.

5.2.3 Exchange and correlation

The details of exchange and correlation effects were presented in Chapter 2. As mentioned earlier, these effects are in general non-local and cannot be readily accounted for in independent-particle formulations, and in DFT they are approximated on the basis of an explicit exchange-correlation term (see Section 2.1.4.3). SIESTA allows for LDA and GGA approximations. In the present work, for simplicity, we have used the local density approximation (LDA).

5.2.4 SIESTA AIMD simulations

Isochoric-isothermal (NVT) simulations were run for a single molecule of the compounds being studied. The temperatures in the simulations varied from 300 K to 1500 K. Higher temperature runs were included in the study to enable access to unlikely configurations. Further justification of studying the molecules at higher temperatures will be given in the

following subsections. A Nosé-Hover thermostat was used to control the temperature. The time step used for the MD simulations was 0.3 fs. Such a small time step was chosen (as opposed to the recommended 1 fs time step) so that the fast motion of the hydrogen atoms in the system can be treated with accuracy. The run time varied from 30 ps to 100 ps.

The input configuration of the molecule of the compounds being studied were generated by optimizing the geometry of molecules at a low level of theory and small basis set. HF/STO-3G was used for all geometry optimization calculations. Such cheap calculations were performed because it was not desired to start the AIMD simulation with a configuration that was already near the absolute energy minima, because the intention was to explore multiple configurations in a single run. In some cases, where molecules are known to exist in two conformations (like cis and trans isomers of 1,2-dichloroethane), geometry optimization was performed for each of these conformations and separate simulations were run for each. SIESTA uses periodic boundary conditions, so each simulation cell is surrounded by its own replica in all directions.

SIESTA routinely outputs potential energy, total energy, temperature, pressure, atomic positions, and velocities at every MD time step. Also, the SIESTA code provides the information about the total charge density, electrostatic and total potentials, local density of states, and other information about the electronic structure. These data can be used for multiple analyses for any compound. We used the output data in two ways to perform the dipole moment studies on the selected compounds, which led to the development of the approximate method and the exact method.

5.3 Preliminary Study (Approximate Method)

The original SIESTA code does not output dipole moment at every time step. For the purposes of preliminary studies, no changes to the original SIESTA code were made. For the determination of dipole moment at each MD step, we used atomic positions (which

were readily available from the simulation output) and the charge density on various atoms. Without performing any detailed electronic charge density analysis or attempting to use any electronic structure data, we estimated the dipole moment by assigning partial charges to various atomic sites.The Merz-Singh-Kollman (MSK) scheme [131] was used to assign partial charges to the atoms of a molecule. In the MSK scheme, atomic charges are fitted to reproduce the molecular electrostatic potential at a number of points around the molecule. Unlike the charges predicted by the Mulliken scheme [132], the MSK charges do not depend upon the level of theory and basis set used to generate them, and hence the MSK scheme is more robust. A few studies have shown that in robustness, the MSK charge assignment scheme even surpasses other popular schemes like ChelpG [133]. The MSK charge assignment calculations were performed using Gaussian at HF level of theory and STO-3G basis set, only for the optimized geometry of the molecule. The inherent crude assumption of this preliminary analysis was that the charge densities on atoms do not change significantly as a molecules evolves in time. Thus, the partial atomic charges (calculated for only the optimized geometry) and the atomic positions from each MD step from SIESTA were used to calculate the approximate dipole moments.

Since the method development was still in its preliminary stages, we needed to test it on a compound for which the experimental dipole moment was known and which would help demonstrate the functionality of the methodology for a complex structure with numerous functional groups. Thus, the compound selected for this study was tributylphosphate (TBP). The experimental value of the dipole moment for the compound was known and for our study, TBP is an adequately complicated molecule with 3 butyl chains, along with 4 polar oxygen and phosphorus atoms. As mentioned previously, it is also one of the molecules that was thoroughly studied by Liu at al. [26]. The structure of TBP is shown in Fig 5.1.

Only a short amount of time (≈30 ps) can be simulated using AIMD simulations, thus, it is not feasible to include all configurations that a real molecule may have access to at room temperature. There are energy barriers that a simulated molecule at room temperature may

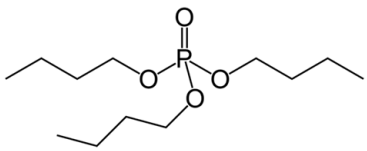


Figure 5.1: Structure of Tributylphosphate (TBP)

never cross for the time of the simulation and hence the states on the other side of the barrier may never get sampled. Thus, we ran the simulations of the TBP molecule at 5 different temperatures, ranging from 300 K to 1600 K, in order to facilitate barrier crossing and sampling multiple states. The results of simulations at all five temperatures for TBP are presented below.

5.3.1 Results and discussion

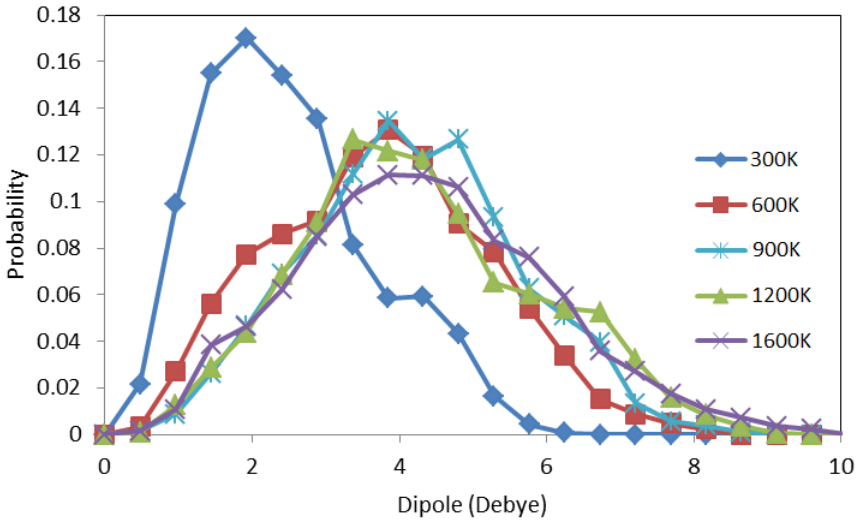
NVT simulations of TBP were run for 30 ps at 0.3 fs time step at 300 K, 600 K, 900 K, 1200 K, and 1600 K. The total run time for all five temperatures was 61 CPU-days. As stated earlier, the dipole moment at each step was calculated using geometries from the output of the SIESTA code and charges preassigned via MSK scheme. The average dipole moment of the molecule was calculated by averages of the approximate dipole moment calculated at each step. The resulting dipole distribution and average dipoles at different temperatures are plotted below.

At 300 K, the average dipole moment calculated from the simulation was around 2.5 D, which was 19% lower than the experimental value (3.08 D). The result was encouraging as we based our calculations on charges that were assigned based on the optimized geometry only. We hypothesized that the calculation of accurate charge density (electronic structure) at each MD step should give us a better match with the experimental values. The average dipole moments at higher temperatures, as shown in Fig 5.2(b), seem to converge to a constant value, instead of increasing monotonically, which is expected.

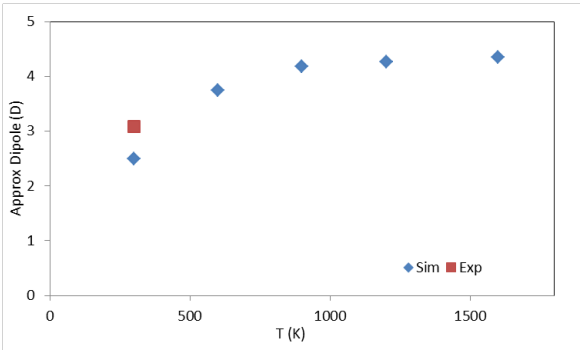
Fig 5.2(a) shows the distribution of dipole moments at various temperatures. For the 300 K simulations, the dipole distribution curve is biased significantly towards the low dipole moment values, while for higher temperature runs, the curve is shifted towards higher dipole moment values, indicating that higher temperature runs are able to access higher dipole moment configurations of the TBP molecule. This fact is also evident when we look at the potential energy vs. dipole moment curves (Fig 5.2(c)). Although there is a significant shift in the dipole moment distribution curves when we go from 300 K to 600 K, the distributions do not shift much beyond that for higher temperatures. But the higher temperature curves have longer tails and are very slightly biased towards higher dipole moment values. This, along with what we observe from Fig 5.2(c), indicates that although there is a significant overlap in the phase space sampled by the low and high temperature runs, higher temperature runs do occasionally access conformations that low temperature runs (especially 300 K temperature run) is unable to access. This proves that there is some value in studying high temperature simulations.

Based on what we learned from this study, we concluded that an improvement in methodology is required pertaining to both the way we calculate the dipole moments and the way we factor in the results of higher temperature runs in our estimation of average dipole moment at room temperature.

(a)



(b)



(c)

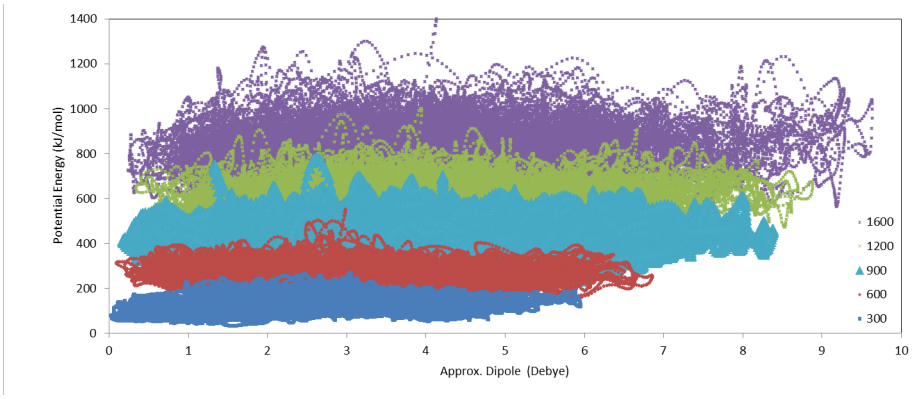


Figure 5.2: Results from preliminary work on TBP molecule. (a) Dipole distribution at various temperatures, (b) Average dipole vs. temperature (c) Potential energy vs. dipole trajectory at different temperatures. In every case, dipoles are estimates based on constant partial charges assigned to atoms.

5.4 Detailed Study (Exact Method)

Based on the results obtained from our preliminary studies, we concluded that refinements were needed in order to develop an improved method for the estimation of dipole moments. The improvements that we incorporated are summarized in the following list.

1. Calculation of exact dipole moment at every step: In the approximate method, we use less information than is available from the AIMD simulation. We only use the atomic positions as the molecule evolves in time under the effect of accurate interatomic forces. The SIESTA code, by default, does not have the option of calculating the dipole moment at every step. However, sufficient information is available to do that. Hence we modified the SIESTA code (written in Fortran-95) to calculate the dipole moment at every step using the accurate electronic charge density that is already available.

2. More freedom for motion of molecules: SIESTA, by default, allows a 10% spatial latitude in each of the three coordinate directions for allowing the molecules to rotate. That is to say, the code calculates the maximum distance between nuclei for each direction in the input file and multiplies the distance by 1.1 to get the box length in that direction. In order to make different conformations of a molecule more accessible, we increased the box-size to 1.5 times the box length that would just contain a molecule, in each direction.

3. More accurate representation of the gas phase: As discussed in detail in Chapter 2, periodic boundaries cause atoms in the simulation box to interact with their own images in order to simulate bulk behavior. SIESTA code also has PBCs which is problematic when we are attempting to study a gas-phase system. Increasing the box size reduces the interactions between the atoms and their images. Thus, increasing the box size (as described above) also helped us reduce the effect of PBCs and enabled us to study a molecule in the gas phase, without elaborate modifications to the SIESTA code.

4. Using higher-temperature runs more effectively: For some molecules, simulations at 300 K may not sample all accessible conformations of the molecule, since the energy barriers between various conformations may not be crossed in the simulation timescale. The higher temperature runs, on account of having more kinetic energy to cross over the energy barriers, have a higher likelihood of accessing different conformations of the molecule being studied. We thus explore a method to extrapolate higher-temperature results to room temperature. This theory and application of this statistical mechanics method is discussed in Sec 5.4.2.

5. More robust ways of validation: Dipole calculations for a few select configurations (that a molecule attains in the course of the AIMD simulation) were duplicated in order to validate the dipole moments we get as SIESTA output. In accordance with the analyses of computational chemistry methods to calculate dipole moment conducted by Liu et al. [26], the *ab initio* calculations at B3LYP level of theory and 6-311+G (3df,2p) basis set were performed to validate dipole moments.

As mentioned previously, the modified SIESTA code was then used to study five compounds: DCE, DNE, EDL, TFE, and TBP. For the smaller symmetric molecules (DCE, DNE, EDL) (Fig. 5.3(a)), simulations were run for both cis and trans isomers. For TFE, due to asymmetry, there are no cis and trans isomers. The two geometries used for the simulation of TFE differ in dihedral angle by 180 degrees (Fig 5.3(b)). NVT simulations of these five compounds were run at 300 K. The details of simulations, including the input geometries, computer run time, the time simulated, and the equilibration period are included in Table 5.1.

The distributions of dipole moments obtained from these simulations are analyzed and the overlap is determined based on a method we developed. This method is explained below in Section 5.4.1. If the overlap is not adequate because of poor sampling, we need to do higher temperature runs as well. The present work is exploratory, so in anticipation of this potential problem, we performed additional simulations at 500 K, 700 K, 900 K, and 1200 K. From each of these runs, dipole moment values are determined at the temperature of run and

Table 5.1: Simulation Details: A list of compounds studied with the exact method, along with the corresponding number of starting conformations (or geometries), computer run time in CPU days, simulated run time in ps, and the equilibration time in ps

Compounds	# starting geometries	Comp. time (CPU-days)	Sim. time (ps)	Eq. time (ps)
TBP	1 (Fig 5.1)	61	30	30
DCE	2 (Fig 5.3(a))	35	100	30
DNE	2 (Fig 5.3(a))	40	100	30
EDL	2 (Fig 5.3(a))	37	100	30
TFE	2 (Fig 5.3(b))	36	100	30

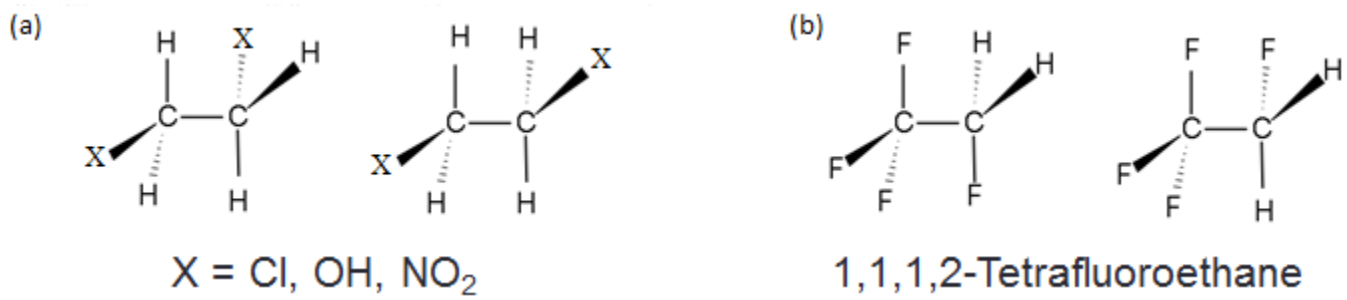


Figure 5.3: Structures of DCE ((a), X=Cl), EDL ((a), X=OH), DNE ((a), X=NO_2), and TFE (b)

then extrapolated back to 298 K using a method we developed. This extrapolation method is described in Section 5.4.2. The results of the SIESTA runs at different temperatures and subsequent analysis are presented in Section 5.4.3.

5.4.1 Quantifying the overlap

Although for the molecules studied, it was possible to visually determine if there is a good overlap between different simulations, to make the method robust and to eliminate guesswork, we used a modified version of a statistical test called the two-sample Kolmogorov-Smirnov (KS) test [134]. The KS test determines if two data-sets overlap significantly. One advantage of using this test is that it makes no assumption about the distribution of data, unlike the t -test, which assumes a normal distribution.

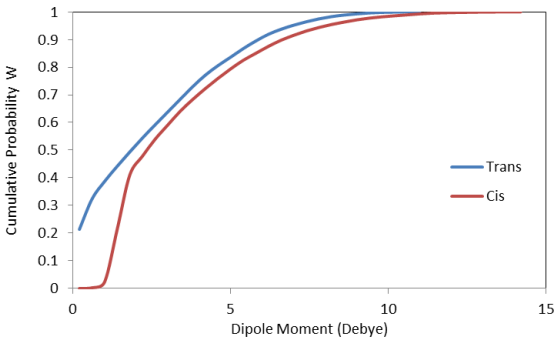


Figure 5.4: Cumulative dipole moment distributions for cis and trans simulations of DCE at 300 K

The first step in the KS method is to generate cumulative dipole distributions W_α for each simulation α.

$$W_\alpha(d_j) = \sum_{d_i \leq d_j} w_i^\alpha \,,$$

where d_i and d_j values span the same binned dipole values recorded in the simulations. Also, it is assumed that probabilities w_i^α for each simulation are normalized such that $0 \leq W_\alpha \leq 1$. In order to compare two simulations, distributions W_1 and W_2 are put on the same plot as shown for example in Fig. 5.4. Then we calculate the KS statistic D^{KS}, given by the maximum vertical difference between the distributions at any binned d_j value:

$$D^{KS} = max\left(|W_1(d_j) - W_2(d_j)|\right) \tag{5.4}$$

If we were to use the regular KS test, then for a 95% confidence interval, the two distributions are consistent if $D^{KS} \leq 0.001$. However, for our case we do not need the simulated distributions to be exactly identical with 95% confidence. Thus, based on the simulations run so far, we propose that adequate overlap between the two simulations corresponds to $D^{KS} \leq 0.25$. If this degree of agreement is not achieved, this is likely caused by poor sampling. To remedy poor sampling, simulations can be run for longer times, or if this is not likely to improve the results, then additional simulations at higher temperatures must be performed.

5.4.2 Extrapolation method

The extrapolation method used in this study to extrapolate dipole moments calculated at high temperatures to lower temperature is similar to the expanded ensemble density of states (EXEDOS) method developed by Pablo and coworkers [135–137]. From the results obtained from the simulations, first we generate dipole moment and energy distributions. The probability of a state depends upon the time it spends in a particular energy state, which can be readily determined from simulation results. This probability of occurrence of a state depends upon energetic and entropic contributions in accordance with Eq. 5.2. The inherent assumption of this extrapolation method is that the density of states $g(E_i)$ is temperature-independent. Thus, when extrapolating the probability from higher temperature T_x to lower temperature T_0 (298 K in all cases here), the probability values, w_i^0, are given by:

$$w_i^0 = w_i^x \exp\left[\frac{E_i}{k_B}\left(\frac{1}{T_x} - \frac{1}{T_0}\right)\right] \tag{5.5}$$

where w_i^x are the probabilities of the state with energy E_i and dipole d_i at temperature T_x. Using these rescaled or extrapolated probabilities, we then calculate the average dipole moment at the new temperature using:

$$\langle d \rangle^0 = \frac{\sum_i w_i^0 d_i}{\sum_i w_i^0} \tag{5.6}$$

5.4.3 SIESTA results and analyses

The average dipole moments calculated form our simulations and the errors are listed in Table 5.2. For all cases except DCE, the simulations were run at 300 K and data were extrapolated back to 298 K. For DCE, the simulations were run at 500 K and extrapolated back to 298 K. The extrapolation method is explained in Section 5.4.2. Neglecting systematic errors of both methods and focusing only on the random or sampling errors, the experimental values and simulation values are within the range of calculated average dipole moment values

for DCE, EDL, and TBP. The aim of the current study was to develop a robust and self-consistent method for dipole moment prediction. To that end, the agreement between the predicted values from simulations and experimental values is acceptable. Nevertheless, the uncertainties in the simulations could be reduced by running the simulations longer.

The molecule-average dipoles in Table 5.2 come from 2 runs for each molecule except TBP. The average dipole moment for each run is extrapolated to the desired temperature (in this case 298 K) using Eq. 5.5. This value for each run is considered an independent sample. To get the average value for the molecule, we do a weighted average of the two (or more) runs as follows.

$$\langle d \rangle = \frac{\sum_\alpha \langle d \rangle_\alpha^0 \, N_\alpha}{\sum_\alpha N_\alpha} \tag{5.7}$$

where $\langle d \rangle_\alpha^0$ is the run-average dipole extrapolated to T_0 and N_α is the number of time steps in run α. The standard deviation of the mean, in the case of two samples, is

$$\sigma_{\langle d \rangle} = \frac{\sqrt{N_1 N_2}}{N_1 + N_2} \left| \langle d \rangle_1^0 - \langle d \rangle_2^0 \right| \tag{5.8}$$

To convert the standard deviation to an error value, it must be scaled by a t-table value, which in the case of two samples and 90% confidence is 3.06. The errors on the molecule average dipoles reported in Table 5.2 are based on two samples (cis and trans simulations) and 90% confidence interval.

Even though the agreement between simulation and experiment in Table 5.2 appears adequate, there are some unquantified uncertainties with the experimental values. In the case of the TFE molecule, three different values of dipole moment have been reported by three different groups (see Table 5.2). In addition, for DCE a value of 1.44 Debye was reported by McClellan et al. [138], which was later modified by them to 2.94 Debye (the value reported in Table 5.2). For the case of DNE, a gas-phase experimental value could not be found. The experimental value listed in the table was determined in benzene solvent. For some molecules, like nitroethane, the experimental value of dipole moment determined

Table 5.2: Comparison of experimental and calculated values of average dipole moment at temperatures around 300 K

Compounds	Expt. Avg. Dip. (Debye)	Calc. Avg. Dip. (Debye)
DCE	2.94 [138]	3.06 ± 1.14
EDL	2.52 [139]	2.94 ±1.00
DNE	4.45 [140]	3.56± 0.52
TFE	0.32 [141],0.62 [142], 2.06 [143]	1.31±0.47
TBP	3.08 [144]	3.19

in the gas phase (3.29 Debye) differs from that in liquid phase (3.60 Debye) [139]. This may also have contributed to the significant mismatch between the experimental and simulation values.

For the purposes of validation, the dipole moment values of initial geometries used as an input were calculated using SIESTA (run for just 1 step) and compared to the dipole moment values obtained from Gaussian at B3LYP/6-311+G (3df,2p). The results are listed in Table 5.3. For the sake of comparison, also shown in the table is the Gaussian values for the dipole moment values obtained at LSDA/STO-3G (Gaussian-II) and at LSDA/3-21+G (Gaussian-III). For reference, a water molecule was also included in this validation step. As can be seen from the table, for non-zero dipole moments, SIESTA underpredicts the dipole moment for the compounds studied, relative to Gaussian data (at B3LYP/6-311+G (3df,2p)). Fig 5.5 shows that the discrepancy between Gaussian- (at B3LYP/6-311+G (3df,2p)) and SIESTA-predicted dipole moments follows a definite trend and likely can be attributed to some systematic error. It is to be noted that if we try to fix the the average dipole moment values obtained from our simulations based on this systematic offset, it takes the values further away form experimental values. Also, in a single case (DCE cis) we decreased the energy cut-off (or error) for the orbitals by a factor of 10 from the default value. We observed that this increases the SIESTA dipole moment value for DCE cis from 1.79 D to 2.45 D, making it closer to the Gaussian-I value. This suggests another fruitful avenue for improvement would be to decrease the default energy cut-off in SIESTA.

Table 5.3: Comparison of SIESTA calculated dipole moments for respective conformations with Gaussian values for the two input geometries of DCE, EDL, DNE, TFE, and for TBP and water.

	Calculated Dipole Moment (Debye)			
Compound	Gaussian-I	Gaussian-II	Gaussian-III	SIESTA
DCE	2.76 (c), 0.00 (t)	3.22 (c), 0.00 (t)	3.31 (c), 0.00 (t)	1.79 (c), 0.00 (t)
EDL	2.55 (c),0.00 (t)	2.00 (c),0.00 (t)	3.32 (c),0.00 (t)	1.97 (c), 0.00 (t)
DNE	4.96 (c), 0.00 (t)	3.92 (c), 0.00 (t)	5.86 (c), 0.00 (t)	4.04 (c), 0.00 (t)
TFE	2.21 (1), 2.21 (2)	0.93 (1), 0.93 (2)	2.62 (1), 2.62 (2)	1.36 (1), 1.36 (2)
TBP	4.38	2.34	4.33	3.70
Water	1.85	1.70	2.67	1.40

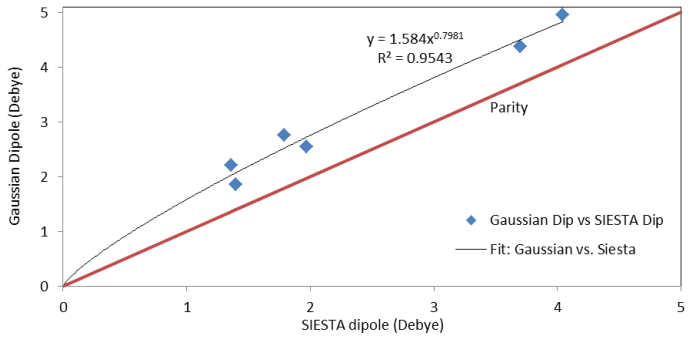


Figure 5.5: Systematic offset between Gaussian-I (at B3LYP/6-311+G (3df,2p)) and SIESTA dipole moment values

Another interesting observation from Tables 5.2 and 5.3 are that the SIESTA molecule-average dipole (Table 5.2) is not simply a weighted average of the dipoles of the starting cis and trans conformations (Table 5.3). It is obvious that the MD simulations are sampling from conformations with larger dipole moments than would be expected from ground-state-type cis and trans conformations. This is confirmed in Fig. 5.6 below.

The systematic error may be due to the underlying DFT method and basis set used to calculate the electronic structure, or the cutoff value used for truncation of orbitals. In the present simulations, the DFT method used with SIESTA simulations was LDA and the basis set consisted of NAOs (see Section 5.2.2), which when compared to the B3LYP level of theory and a large basis set used for the Gaussian data, will lead to a less accurate electronic

structure calculation. SIESTA does not do a B3LYP level of theory and traditional Gaussian based orbitals as these lead to a very expensive calculation. Also, large cutoff values for orbitals can be expensive. It is of some value to mix and match some of the available options in DFT methods, choice of orbital size and orbital cutoff values to find an optimum set of these three that will give better matches to the Gaussian dipole moment. However, such an analysis is beyond the scope of the present work.

The distributions of dipole moments for different molecules are shown in Fig. 5.6. It is evident from the Fig. 5.6(a-i) that the cis and trans simulations of the DCE molecule cover different regions of phase space. The cis simulation has a distinct peak near 1.8 Debye, which corresponds to additional sampling of the initial cis geometry (see Table 5.3). Likewise, the trans simulation samples mostly near its initial trans geometry (zero dipole). Thus, it seems that there is not a significant overlap between the dipole moment distributions from these two simulations, which, as mentioned previously, may give unreliable value of the average dipole moment. The D^{KS} value for these two distributions was calculated to quantify the overlap and, as listed in Table 5.4, it was greater than 0.25, which, according to the discussion in Section 5.4.1, also indicated a poor overlap. When we consider the energy distributions from the two simulations, we find that there is some overlap between the two distributions. This indicates that there are more than one low-energy state in the system, each having a different dipole and needing to be considered in the molecule-average dipole moment. However, DCE simulations at 300 K do not sample all these relevant states as there are significant energy barriers between them. In order to get around this problem, the dipole probability distribution of a higher-temperature run was chosen. Fig. 5.6(a-ii) shows the dipole moment distribution at 500 K and there is an improved overlap between the distributions from the two runs, as expected. The calculated D^{KS} value in this case was less than 0.25, which indicates reasonable overlap. Hence, this distribution was used to estimate the average dipole moment using a procedure described in the Subsection 5.4.2.

Table 5.4: D^{KS} values for DCE, DNE, EDL, and TFE dipole distributions at two temperatures. For DCE, the higher temperature run was required.

Compounds	D^{KS}at 300 K	D^{KS} at 500 K
DCE	0.3638	0.2274 (required)
EDL	0.1783	0.3670 (not required)
DNE	0.1764	0.1623 (not required)
TFE	0.08951	0.0756 (not required)

The DNE simulations (Fig 5.6(b)) with trans starting configuration distinctly samples in two states, depicted by the two peaks in its distribution. The cis simulation on the other hand, only samples one state but since there seems to be a significant overlap between the two in the region of the most likely conformation, the values are expected to be reliable. The calculated D^{KS} value less than 0.25, and hence the overlap was considered adequate. Fig. 5.6 (c) and (d) show dipole moment distributions for EDL, and TFE respectively, each at 300 K. The D^{KS} values for the two cases indicated adequate overlap between the distributions from two simulations, and thus the dipole moments calculated from 300 K runs are reliable.

5.6(e) shows the dipole moment distribution for TBP at 300 K. Although the TBP simulation samples dipole moments of around 2.0 Debye for about 40% of the time, it does sample higher dipole moments too, which is indicated by the distribution being biased towards higher dipole moment values. This results in a average dipole moment of about 3.2 Debye.

It is to be noted that at 500 K, for EDL, the overlap is worse than it was for 300 K (see Table 5.4). Thus, going to higher temperature does not significantly improve overlap between the two distributions in every case.

5.5 The General Procedure

Based on the analyses of the five compounds studied, a general procedure to calculate average dipole moments was developed. The procedure is shown in Fig 5.7. For any molecule to be studied, we start with the selection of two or more different geometries. If it is a symmetric

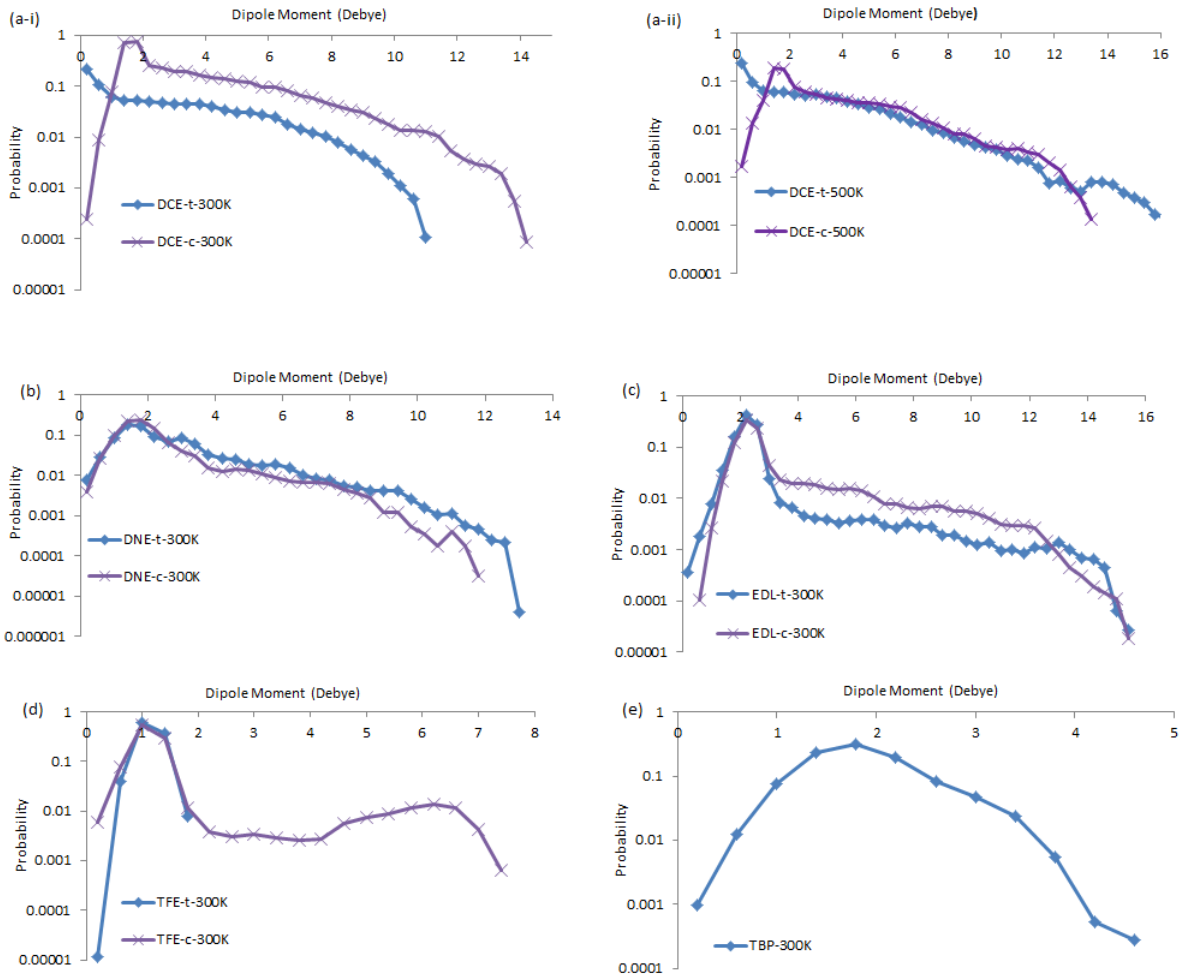


Figure 5.6: Dipole moment distributions for 300 K for trans (t) and cis (c) starting configurations of (a) DCE (b) DNE (c) EDL, two starting geometries (1 and 2) of (d) TFE, and (e) TBP

molecule with a single dihedral, we study the cis and trans conformations and if it is an asymmetric molecule, we perturb the geometry by rotating the groups along the dihedral. If it is a large molecule with multiple functional groups and multiple dihedral angles, then the positions of alkyl branches and functional groups may be slightly altered to produce new geometries. All the geometries must be optimized to generate a reasonable starting point (recommended level of theory and basis set: HF/STO-3G). These optimized geometries are then used as input to the SIESTA program.

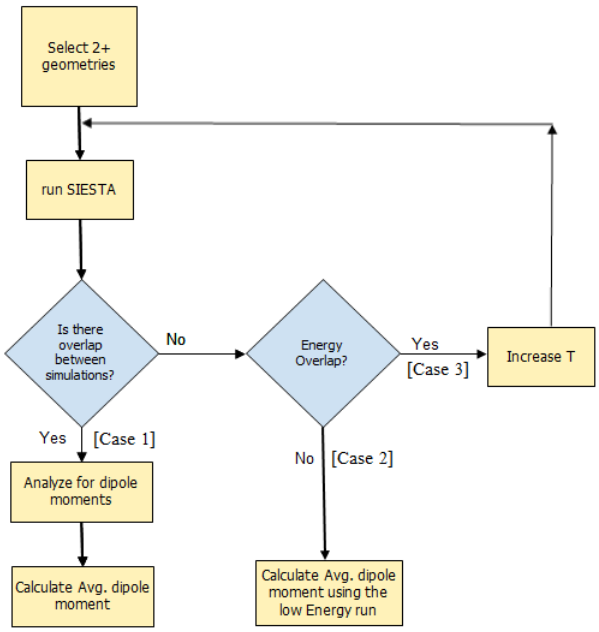


Figure 5.7: Procedure to determine the average dipole moment

Once the input geometry files, along with the pseudopotentials files (5.2.1), are ready for the molecules, NVT simulations of the selected geometries are performed with a Nosé-Hover thermostat. The time step should be kept small (recommended: 0.33 fs) and all simulations must be equilibrated (recommended: 20-30 ps). The production time for the simulations must be as large as possible. Due to computational limitations, we did not go beyond 100 fs run time, but for better results, it is advisable to run the simulations longer.

After running the simulations for the desired production time using the modified SIESTA code, dipole and energy probability distributions are determined from the SIESTA output files listing dipole and energy at each time step. Appendix A contains the modified subroutines and makefile of the SIESTA code required for this work. In addition, a special-purpose code was written for interpreting the dipole and energy output files and generating the probability distributions and temperature-extrapolated dipole moments.This code is included in the Appendix B. Once we have the distributions, we determine if there is a significant overlap between different simulations of the same molecule, by calculating the D^{KS} value for that

molecule and compare it to a standard KS-statistic value (as described in Subsection 5.4.1). If there is a reasonable overlap, then we simply calculate the average dipole moment and uncertainty in accordance with Eq. 5.7.

If the overlap between the dipole distributions is not significant, then it may be due to inadequate run time or there may be sampling issues. In order to differentiate between the two, we need to analyze the energy surfaces of the molecule. Fig. 5.8 gives a pictorial representation of the energy surface cases for a system with two states. If the different states of the molecule differ in energy only slightly, and the energy barrier is low, then the simulations are expected to sample those states adequately and a decent overlap is expected (see case 1 in Figs 5.8 and 5.7). In our study, DNE, EDL, and TFE simulations were of case 1 type.

The systems that are simulated may not overlap either because the other metastable states, being of high energy, are improbable at low temperature (see case 2 in Figs 5.8 and 5.7). Lack of overlap in this case indicates that the state is so improbable that it need not be considered in calculation of average dipole moment.

What qualifies as a "hard case" (case 3 in Figs. 5.7 and 5.8) is when two or more states of almost equal probability (that is, almost equal energy), are separated by a high energy barrier, making it unlikely for any 300 K simulation to adequately sample all such states. For such systems, one must run the molecule at higher temperatures so that it may have sufficient kinetic energy to cross barriers and sample all significant conformational states. To distinguish between cases 2 and 3, we generate energy probability distributions from the two simulations. The degree of overlap can be quantified by calculating a D_E^{KS} statistic for these energy probability distributions, just as we did previously for the dipole distributions. If there is a modest degree of overlap ($D_E^{KS} < 0.95$) then the molecule qualifies as a hard case and we must resort to higher temperature runs. In our study, DCE simulations were the one example of case 3 type. Once we have rerun the simulations as needed, the system

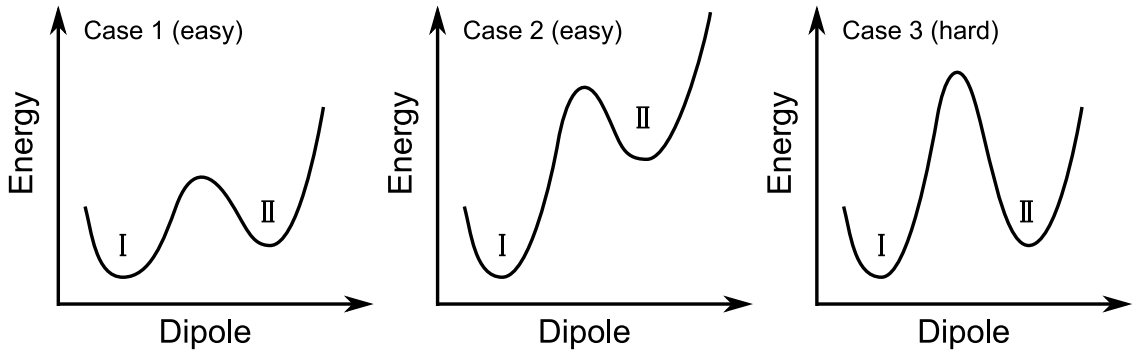


Figure 5.8: Diagrammatic representation of energy versus dipole (the reaction coordinate) for three cases for a molecule with two conformational states I and II.

should either have overlapping dipole distributions (case 1) or only one relevant distribution (case 2), and reliable molecule-average dipoles can be obtained.

5.6 Conclusion

The aim of the present study was to develop a general and reliable method to calculate average dipole moment of polar organic compounds with conformational degrees of freedom. In preliminary studies, SIESTA AIMD code was used without any modifications and the molecule dipole moment was estimated with the crude assumption that the local charge density on the sites remains unchanged during the course of simulation. The TBP molecule was studied using this method and the subsequent dipole moment distributions and the average dipole moment obtained gave encouraging results, but we judged that a modification in methodology would increase the accuracy of the method.

The SIESTA code was modified and was used to study TBP and four additional polar organic compounds. To get reliable values, simulations with different starting conformations of a given molecule must show some overlap in sampled states. This is indicated by a D statistic from KS theory and also by the calculated standard error of the average dipole moment for the combined simulations. If a low-temperature set of simulations do not have sufficient overlap, the reason may be that either the less-sampled states are improbable or

the energy barriers between almost degenerate states cannot be crossed. If latter is the case, then simulations at high temperature need to be performed to enable energy-barrier crossings and better sampling. To facilitate this, a method was also developed for extrapolating the dipole moment values from higher temperature runs to room temperature.

The dipole moments obtained from the simulations gave a good match with experimental values for three of the five compounds. Reasons for the mismatch between simulation and experimental dipole moments can be: the experimental values are inaccurately determined or biased, the simulations do not have adequate sampling of phase space, the simulations have biased dipole values due to limitations in quantum chemical methods, or a combination of these reasons. It is axiomatic that longer and more sophisticated DFT calculations and more accurate experiments will lead to better agreement, but it remains to be determined how much effort this will require. However, at the present level of effort, the dipole moments generated by theory and experiment differed, on average, by 8 % for DCE, EDL, and TBP molecules. The other two cases (TFE and DNE) likely have problems with the experimental results.

The method developed in the present study has two potential uses. First, it can be used in prediction of dipole moment for compounds whose experimental values are not available. Second, this method can be used as a tool for evaluating the quality of the experimental values. If the simulations appear to have adequate sampling for a molecule and there is still a disagreement between theory and experiment, then this suggests that the experiment could be problematic and should be revisited.

Chapter 6
Conclusion

6.1 Proton Transport Studies

6.1.1 Summary of results

Not many robust and polarizable force-field models are available in the literature that can simulate reactions. Here we have developed a polarizable water model, that includes multi-body effects, and that treats reactions naturally. The water model we developed has the potential of competing with more expensive (MS-EVB) or less physically-accurate Keffer's model. The model captures reasonably accurate electrostatic interactions, both at short and long interatomic distances, and allows for proton transfer to occur naturally when dictated by electrostatic forces, without an additional set of rules. Molecule polarizability is implemented by the use of fluctuating site dipoles, as well as fluctuating charges. This model was parametrized, as much as possible, from the *ab initio* data, and was then implemented into an MD simulation. Novel time-saving techniques like a modified Ewald sum specific to diffuse charges and diffuse dipoles and use of multiple neighborhood lists, were applied in MD simulation to make the code efficient. The model correctly avoids any significant amount of autoionization of water molecules, a problem found in other water models based on the central force potentials that allow bond formation and breaking.

Although the model was developed in the context of bulk water, we have attempted to make a model that can be easily extended to heterogeneous systems that include solid catalyst surfaces and polymer membranes. Also, since bond breaking and formation occurs

naturally in our simulations it is our hope that it can be used to study liquid-state chemical reactions. Not only the model can be used to study transport, thermodynamic, and structural properties of a reactive system, it can be used to calculate reaction rates and effects of other species in the system, such as a solvent. For instance, under the NEMD framework, the model could be used to study the change in conductance of an electrolyte with varying electric fields.

Several quantum effects, like tunneling and delocalization of protons may be important and are not accounted for directly in our simulations. These effects have been included indirectly in our potential energy function and by empirically lowering the PT reaction energy barrier. In the classical MD framework, a more detailed treatment was not possible. In spite of this approximation, the results were encouraging.

Simulated bulk water properties were in reasonable agreement with the experimental values and evidence of PT events were found in an MD simulation containing one excess proton. The preliminary PT rate, calculated with some crude assumptions, appeared to be too low by a factor of 5. Several modifications to the methodology were made to calculate the true PT rate. The refined method resulted in the PT rate that was $8.07 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$, which was off and too low by 13%. The method was then extended to include a counterion in our MD simulations. The structural and transport properties of the system gave good agreement with the experimental data, and the resulting PT rate was $6.02 \times 10^{-5}\,\mathrm{cm}^2/\mathrm{s}$. The method was also extended to include multiple excess protons.

6.1.2 Future work

6.1.2.1 Study of ORR

As mentioned previously, the ultimate aim of this study is to develop a robust and polarizable model for studying the oxygen reduction reaction (ORR) that occurs on the cathode of fuel cells. The water model developed in this work has been partially demonstrated to be transferable by verifying its functionality when different chemical species (counterions,

multiple protons) are added. The next step towards fulfilling the ultimate objective of the study will be to include a polymer membrane in the simulation. As a model for polymer membrane, a hydrated Nafion membrane may be selected. Multiple Nafion monomers may be included along with oxygen and hydrogen species. The molecular interactions between the sites of the Nafion, and the hydrogen and oxygen species of the simulation will still be given by the formulations in Section 3.2, and no new theoretical development is needed. The simulations with the Nafion membrane would be used for studying the relative rate of PT via three mechanisms: bulk transport (vehicular diffusion), Grotthuss mechanism (structural diffusion), and sulphonate-assisted PT (due to the polymer membrane). The results for proton conductivity may be compared with the experimental results of Doyle et al. [145, 146]. Once satisfactory results are obtained for PT in presence of Nafion membrane, the next step would be to add a platinum catalyst to the system. Pt(1,1,1) surface may be used for this purpose.

To study the ORR, first, in order to speed up the events of interest for better sampling, one would need to use a sampling technique, like transition path sampling (TPS) [147]. Second, candidate mechanisms for the ORR would need to be selected and the energy surfaces of these would be individually studied. Third, effect of solvation would be studied on the reaction pathways and energy barriers for the four elementary steps of ORR. The rate determining step would be identified by this study. Lastly, the effect of electrode potential and electrochemical double layer on the energy barriers would be studied. With these studies, an attempt to identify the factors that make the RDS of ORR slow may be determined. The results may be compared to an *ab intio* study of ORR by Sha et al. [148].

6.1.2.2 Other applications of the model

Besides demonstrating the simulation of proton transport in this work, several novel concepts have been introduced that are useful in the development of a force-field that intends to go beyond what traditional classical force-fields can do. First, the water model developed in this

work uses a small number of parameters to capture the essential physics of the system, which, as demonstrated in Chapter 4, helps making the model transferable. The transferability can be further tested by the application of model to reactive systems studied by Voth's MS-EVB and Goddard's ReaxFF (see Section 2.3.2). Second, the concept of central-force enables simulation of reactions while keeping the integrity of nonreacting molecules intact. Third, the weak vdW interactions between sites on different molecules are inter-related with the Coulombic interactions and do not need to be determined independently. Fourth, a novel Ewald sum, specific to a system of fluctuating charge and fluctuating point dipoles is implemented, along with some other time saving algorithms. These concepts have much broader application than just simulation of reactions. The application of these concepts enables one to study charged and reacting systems at the molecular level more efficiently, which can provide additional insight into simulated physical, chemical, or biological nanoscale phenomena.

6.2 Dipole Moment Studies

6.2.1 Summary of results

A method was developed to estimate the dipole moment of polar organic compounds. An open source AIMD code SIESTA was used in order to simulate several organic molecules. The molecule-average dipole moment was determined by analysis of the simulated dipole moment distributions, with careful attention to possible sources of error. A general method was then developed to estimate the dipole moment of any compound with conformational degrees of freedom. The major advantage of this method is that it involves the determination of reasonably accurate electronic structure via solution of the Schrödinger equation, which leads to better estimations of dipole moment.

One drawback of the method is that the determination of electronic structure at each MD step is a computationally demanding process and this makes the method very expensive.

Thus, for large molecules, the simulated run time cannot be more than a few hundred picoseconds, which may result in inadequate sampling. However, we have developed some robust methods to circumvent the problem of poor sampling. Careful analyses of the dipole distributions along with using data from higher temperatures provides a reasonably good estimate for the average dipole moment.

First an approximate method was developed which did not fully utilize all the available data from SIESTA output and was based on certain assumptions. The results obtained were encouraging but prompted improvements. The method was hence refined, pertaining to both the SIESTA program used to run simulations and subsequent analyses of dipole and energy distributions. The study was performed for for five polar organic compounds: TBP and four compounds with a single dihedral angle. The results for all compounds studied were within the range of experimental values except for DNE.

6.2.2 Future work

Due to time and computational limitations, in our studies the SIESTA AIMD code could not be run for more than about one week, which translated to around 100 ps simulated time for smaller molecules and only 30 ps for TBP. The sampling would improve if the simulations could be run longer and this should be the first thing to try in order to continue these studies. Next, better DFT methods and larger basis set for orbitals may be used to improve dipole moment predictions. With SIESTA, it is possible to use more sophisticated DFT methods, like PBE [149], RPBE [150], and LYP [151]. Also, it is possible to use multiple-zeta basis sets [152] with SIESTA, allowing larger orbitals and better representation of polarizability. Utilizing these options will increase the computation time, but can lead to a more accurate electronic structure calculation, and hence and an accurate dipole moment calculation. Of course, in some cases the easier method inadvertently benefits from cancellation of errors, and the more expensive method does not immediately lead to improved accuracy. These tests could be performed on the same five compounds that were used in the present study.

One of the major challenges faced in this study was that the SIESTA code could not be run in parallel, that is, simultaneously on multiple processors. If the SIESTA could be run on parallel processors, the efficiency would increase greatly and one would have access to higher time scales. This is also one of the prospective improvements that could be made to the study. To this end, a system administrator for the supercomputer center at BYU was consulted regarding this problem. He said there were some problems with the library that the SIESTA code depends on, and a solution was not obtained as part of the present project.

Once an optimum DFT-method and basis set size is selected, the method that is developed here could be used to estimate average dipole moment of any compound. The general procedure presented in Chapter 5, including the statistical tests, may need to be refined as one studies dipole moments of other compounds.

Bibliography

[1] T. Fuller, *http://www.che.gatech.edu/fuller/fc-workshop/lowtempfuelcell.pdf* (2005).

[2] M. Gattrell and B. MacDougall (John Wiley, 2003).

[3] M. Eigen, Angewandte Chemie, International Edition **3**, 1 (1964).

[4] G. Zundel and J. Fritsch, *The Chemical Physics of Solvation* (Elsevier, 1986).

[5] C. Wraight, Biochimica et Biophysica Acta **1757**, 886 (2006).

[6] M. Okumura, L. I. Yeh, J. D. Myers, and Y. T. Lee, Journal of Physical Chemistry **94**, 3416 (1990).

[7] J. M. Headrick, E. G. Diken, R. S. Walters, H. N. I., R. A. Christie, J. Cui, E. M. Myshakin, M. A. Duncan, M. A. Johnson, and K. D. Jordan, Science **308**, 1765 (2005).

[8] M. Miyazaki, A. Fujii, T. Ebata, and N. Mikami, Science **304**, 1134 (2004).

[9] H. A. Schwarz, Journal of Chemical Physics **67**, 5525 (1977).

[10] S. Meiboom, Journal of Chemical Physics **34**, 375 (1961).

[11] A. Loewenstein and A. Szoke, Journal of American Chemical Society **84**, 1151 (1962).

[12] Z. Luz and S. Meiboom, Journal of American Chemical Society **86**, 4768 (1964).

[13] R. R. Knispel and M. M. Pintar, Chemical Physics Letters **32**, 238 (1975).

[14] V. Graf, F. Noack, and G. J. Bene, Journal of Chemical Physics **72**, 861 (1980).

[15] D. L. Turner, Molecular Physics **40**, 949 (1980).

[16] W. J. Lamb, D. R. Brown, and J. Jonas, Journal of Physical Chemistry **85**, 3883 (1981).

[17] B. Halle and G. Karlstrom, Chemical Society, Faraday Trans. II **79**, 1031 (1983).

[18] R. Pfeifer and H. G. Hertz, Ber. Bunsen-Ges **94**, 1349 (1990).

[19] S. Woutersen and H. J. Bakker, Physics Review Letters **96**, 138305 (2006).

[20] N. Agmon, Israel Journal of Chemistry **39**, 493 (1999).

[21] D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, Nature **397**, 601 (1999).

[22] S. Cukierman, Biochem. Biophys. Acta: Bioenerg. **1757**, 876 (2006).

[23] H. Lapid, N. Agmon, M. K. Petersen, and G. A. Voth, Journal of Chemical Physics **122**, 14506 (2005).

[24] O. Markovitch and N. Agmon, Journal of Physical Chemistry A **111**, 2253 (2007).

[25] H. Piatkowski, L.;Bakker, Journal of Chemical Physics **136**, 164504 (2012).

[26] J. Liu, W. Wilding, and R. Rowley, Journal of Chemical & Engineering Data **56**, 2430 (2011).

[27] W. Moore, *Physical Chemistry* (Prentice-Hall,Englewood Cliffs, N.J., 1972).

[28] R. LeFevre, *Dipole Moments, Their Measurement and Application in Chemistry* (John Wiley and Sons, New York, 1953).

[29] P. Debye, Physik. Z. **13**, 97 (1912).

[30] Debye, Verhandl. Deut. Pkysik. Ges. **15**, 777 (1913).

[31] L. Onsager, Journal of American Chemical Society **58**, 1486 (1938).

[32] C. Maxwell, J., *Treatise on Electricity*, vol. II (Oxford, London, 1881).

[33] J. Kirkwood, Journal of Chemical Physics **7**, 911 (1939).

[34] J. Kirkwood, Trans. Faraday Soc. **42A**, 7 (1946).

[35] H. Frohlich, Trans **44**, 238 (1948).

[36] H. Frohlich, *Theory of Dielectrics* (Oxford University Press, London, 1958).

[37] J. Bogs and A. Deam, Journal of Chemical Physics **32**, 315 (1960).

[38] K. Higasi, Phys. Chem. Res. **28**, 284 (1936).

[39] R. Nelson, D. Linde, and A. Maryott, *Selected values of electric dipole moment for molecules in the gas phase* (1967), national Standard Reference Data Series - National Bureau of Standards 10.

[40] J. Soler, E. Artacho, J. Gale, A. Garcia, and J. Junquera, Journal of Physics: Condensed Matter **14**, 2745 (2002).

[41] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, 2002), 2nd ed.

[42] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, 2009).

[43] C. Iacovella, *Periodic boundary conditions* (2006), glotzer group. Depts of Chemical Engineering, Materials Science & Engineering, Macromolecular Science, and Physics, University of Michigan.

[44] J. Lennard-Jones, Proc. R. Soc. Lond. A **106**, 463 (1924).

[45] L. Perera and M. Berkowitz, Journal of Chemical Physics **95**, 1954 (1991).

[46] J. Trersoff, Phys. Rev. B **39**, 5566 (1989).

[47] M. Daw, S. Foiles, and M. Baskes, Mat. Sci. And Engr. Rep. **9**, 251 (1993).

[48] F. Cleri and V. Rosato, Phys. Rev. B **48**, 22 (1993).

[49] B. Axilrod and E. Teller, Journal of Chemical Physics **11**, 299 (1943).

[50] M. Born and J. Oppenheimer, Ann. Physik **84**, 457 (1927).

[51] J. Slater, Proc. Natl. Acad. Sci. **13**, 423 (1927).

[52] D. Maurice and M. Head-Gordon, Molecular Physics **96**, 1533 (1999).

[53] M. Head-Gordon, R. Rico, M. Oumi, and T. Lee, Chemical Physics Letters **219**, 21 (1994).

[54] T. Voorhis and M. Head-Gordon, Journal of Chemical Physics **115**, 5033 (2001).

[55] C. Moller and S. Plesset, M., Physical Review **46**, 618 (1934).

[56] K. Raghavachari and A. Pople, J., International Journal of Quantum Chemistry **14**, 91 (1978).

[57] P. Hohenberg and W. Kohn, Physical Review **136**, B864 (1964).

[58] S. Nose, Journal of Chemical Physics **81**, 511 (1984).

[59] W. Hoover, Phys. Rev. A **31**, 1695 (1985).

[60] M. Tuckerman, Laasonen, M. Sprik, and M. Parrinello, Journal of Physical Chemistry **99**, 5749 (1995).

[61] R. Car and M. Parinello, Phys. Rev. Lett. **55**, 2471 (1985).

[62] D. Marx, M. Tuckerman, J. Hutter, and M. Parinello, Nature **397**, 601 (1999).

[63] M. Tuckerman (1999), URL `http://homepages.nyu.edu/~mt33/jpc_feat/node11.html`.

[64] M. Tuckerman, K. Laasonen, M. Sprik, and M. Parrinello, Journal of Chemical Physics **103**, 150 (1995).

[65] M. Tuckerman, K. Laasonen, M. Sprik, and M. Parrinello, Journal of Physical Chemistry **99**, 5749 (1995).

[66] R. Car and P. M., Physics Review Letters **55**, 2471 (1985).

[67] D. K. Remler and P. A. Madden, Molecular Physics **70**, 921 (1990).

[68] Laasonen, M. Sprik, M. Parrinello, and R. Car, Journal of Physical Chemistry **99**, 9080 (1995).

[69] U. W. Schmitt and G. A. Voth, Journal of Physical Chemistry B **102**, 5547 (1998).

[70] A. Warshel and R. Weiss, Journal of American Chemical Society **102**, 6218 (1980).

[71] Y. Wu, H. Chen, F. Wang, F. Paesani, and V. G. A., Journal of Physical Chemistry B **112**, 467 (2008).

[72] Y. Wu and V. G. A., FEBS Lett. **552**, 23 (2003).

[73] Y. Wu and Vot, Biophysics Journal **85**, 864 (2003).

[74] A. Smondyrev and V. G. A., Biophysics **83**, 1987 (2002).

[75] Y. Wu, B. Ilan, and V. G. A., Biophysics J **92**, 61 (2007).

[76] H. Tepper and V. G. A., Journal of Physical Chemistry B **110**, 21327 (2006).

[77] V. Duin, Dasgupta, Lorant, and Goddard, Journal of Physical Chemistry A **105**, 9396 (2001).

[78] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. Goddard, Journal of Physical Chemistry A **105**, 9396 (2001).

[79] K. Chenoweth, S. Cheung, A. C. T. van Duin, and G. W.A., Journal of American Chemical Society **127**, 7192 (2005).

[80] S. A., A. C. T. van Duin, D. Chakraborty, S. Dasgupta, and G. W. A., Phys. Rev. Lett. **91**, 098301 (2003).

[81] K. Nielson, A. C. T. van Duin, J. Oxgaard, W. Deng, and W. Goddard, Journal of Physical Chemistry **109**, 493 (2005).

[82] J. Ludwig, D. Vlachos, A. C. T. van Duin, and W. Goddard, Journal of Physical Chemistry B **110**, 4274 (2006).

[83] G. W. A., A. C. T. van Duin, T. Jacob, and Y. Jang, Mol. Simul. **32**, 251 (2006).

[84] S. Rick and S. S.J., *Reviews in Computational Chemicstry* (Wiley, 2002).

[85] O. Rahaman, A. C. T. van Duin, W. A. Goddard, and D. Doren, Journal of Physical Chemistry B **115**, 249 (2011).

[86] M. E. Selvan, D. J. Keffer, S. Cui, and S. J. Paddison, Journal of Physics Chemistry C **114**, 11965 (2010).

[87] F. Bresme, Journal of Chemical Physics **115**, 16 (2001).

[88] Lemberg and Stillinger, Journal of Chemical Physics **62**, 1677 (1975).

[89] D. W. M. Hoffman, L. Kuleshova, and B. Duganno, Journal of Molecular Modeling **14**, 225 (2008).

[90] S. Kielich, ACTA PHYSICA POLONICA **XXVII**, 304 (1965).

[91] W. Heston, E. Hennelly, and C. Smith, Journal of American Chemical Society **72**, 2071 (1950).

[92] E. Guggenheim, Trans. Faraday Soc. **47**, 573 (1951).

[93] J. Smith, Trans. Fara **46**, 394 (1950).

[94] C. Smyth, *Dielectric Behaviour and Structure* (McGraw-Hill, NY, 1955).

[95] Y. Wang and J. Bowman, Journal of Chemical Physics **134**, 154510 (2011).

[96] C. Parkanyi, *Theoretical Organic Chemistry* (Elsevier, 1998).

[97] D. Rogers, *Computational Chemistry using the PC* (Wiley-Interscience, 2003).

[98] C. Cramer, *Essentials of Computational Chemistry: Theories and Models* (Wiley, 2002).

[99] S. Feller, R. Pastor, A. Rojnuckarin, S. Bogusz, and B. Brooks, Journal of Physical Chemistry **100**, 17011 (1996).

[100] G. W. A., B. Merinov, A. C. T. van Duin, T. Jacob, M. Blanco, V. Molinero, S. Jang, and Y. Jang, Molecular Simulation **32**, 251 (2006).

[101] W. J. Mortier, S. K. Ghosh, and S. Shankar, Journal of American Chemical Society **108**, 4315 (1986).

[102] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, Journal of Physical Chemistry **91**, 6269 (1987).

[103] F. London, Transactions of the Faraday Society **33**, 8 (1937).

[104] P. Ewald, Biogr. Mems Fell. R. Soc. **34**, 134 (1988).

[105] K. N. Joshipura, S. Gangopadhay, C. G. Limbachia, and M. Vinodkumar, Phys. : Conf. Ser. **80**, 012008 (2007).

[106] P. Cabral do cuto, S. G. Estaciao, and B. J. Costa, Journal of Chemical Physics **123**, 054510 (2005).

[107] J. Whleeler, D.R.; Newman, Journal of Physical Chemistry B **108(47)**, 18362 (2004).

[108] R. Ludwig, Angewandte Chemie, International Edition **40**, 1808 (2001).

[109] A. D. Buckingham, Proc. R. Soc. London, Ser. A **238**, 235 (1956).

[110] Y. Laaksonen, A.;Tu, Chemical Physics Letters **329**, 283 (2000).

[111] J. M. Haile, *Molecular Dynamics Simulation: Elementary Methods* (Wiley Professional: New York, 1997).

[112] D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water* (Oxford University, London, 1969).

[113] U. Kaatze and V. Z. Uhlendorf, Journal of Physical Chemistry **126**, 151 (1981).

[114] J. V. Sengers and J. T. R. Watson, Journal of Physical Chemistry **15**, 1291 (1986).

[115] K. A. Motakabbir and M. Berkowitz, Journal of Physical Chemistry **94**, 8359 (1990).

[116] A. D. Trokhymchuk, M. F. Holovko, and K. Heinzinger, Journal of Chemical Physics **99**, 2964 (1993).

[117] F. H. Stillinger and A. Rahman, Journal of Chemical Physics **60**, 1545 (1974).

[118] K. Watanabe and M. L. Klein, Journal of Chemical Physics **131**, 157 (1989).

[119] G. A. Wang, F.; Voth, Journal of Chemical Physics **122**, 144105 (2005).

[120] H. G. Wu, Y.;Tepper, Journal of Chemical Physics **124**, 024503 (2006).

[121] M. R. Reddy and M. Berkowitz, Journal of Chemical Physics **87**, 6682 (1987).

[122] Y. Mao, Y.;Zhang, Chemical Physics Letters **542**, 37 (2012).

[123] M. E. Marx, D.and Tuckerman, J. Hutter, and M. Parrinello, Nature **397**, 601 (1999).

[124] H. Scheraga, M. Khalili, and A. Liwo, Annu. Rev. Phys. Chem. **58**, 57 (2007).

[125] N. Okimoto, N. Futatsugi, H. Fuji, A. Suenaga, G. Morimoto, R. Yanai, Y. Ohno, T. Narumi, and M. Taiji, PLoS Comput Biol **5**, e1000528 (2009).

[126] P. Schwerdtfeger, ChemPhysChem **12**, 3143 (2011).

[127] G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[128] A. Garcia, URL `http://www.icmab.es/siesta/atom`.

[129] Octopus Pseudopotential Generator, URL `http://www.tddft.org/programs/octopus/pseudo.php`.

[130] J. Junquera, O. Paz, D. Sanchez-Portal, and E. Artacho, Phys. Rev. B **64**, 235111 (2001).

[131] U. Singh and P. Kollman, J. Comp. Chem. **5**, 129 (1984).

[132] J. Montgomery and M. Frisch, Journal of Chemical Physics **112**, 6532 (2000).

[133] F. Martin and H. Zipse, Computational Chemistry **26**, 97 (2005).

[134] W. Press, B. Flannery, S. Teukolsky, and V. Vetterling, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (Cambridge, 1992), 2nd ed.

[135] R. Shekhar, J. Whitmer, R. Malshe, J. Moreno-Razo, T. Roberts, and J. Pablo, Journal of Chemical Physics **136**, 234503 (2012).

[136] F. Wang and D. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[137] S. Singh, M. Chopra, and J. Pablo, Annu. Rev. Chem. Biomol. Eng. **3**, 369 (2012).

[138] A. McClellan, *Tables of Experimental Dipole Moments* (W.H. Freeman Pub., 1963).

[139] A. McClellan, *Tables of Experimental Dipole Moments* (Rahara Enterprises, 1989).

[140] Y. Lam, L. Koh, and H. Huang, Journal of Chemical Society Perkin Transactions **2**, 175 (1993).

[141] G. Gao, W. Wang, and X. C. Zeng, Fluid Phase Equilibria **137**, 87 (1997).

[142] T. Takagi, T. Sakura, T. Tsuji, and T. Hongo, Fluid Phase Equilibria **162**, 171 (1999).

[143] C. Meyer and G. Morrison, Journal of Physical Chemistry **95**, 3860 (1991).

[144] *Nist computational chemistry comparison and benchmark database*, URL `http://srdata.nist.gov/cccbdb/default.htm`.

[145] M. Doyle, L. Wang, Z. Yang, and S. Choi, Journal of Electochemical Society **150**, D185 (2003).

[146] M. Doyle, M. Lewittes, M. Roelofs, and S. Perusich, Journal of Physical Chemistry **105**, 9387 (2001).

[147] C. Dellago, P. Bolhuis, and P. Geissler, Advance Chemical Physics **123** (2002).

[148] Y. Sha, T. Yu, V. Merinov, and W. Goddard, Journal of Physical Chemistry Letters **1**, 856 (2010).

[149] J. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett **77**, 3865 (1996).

[150] B. Hammer, L. Hansen, and J. Norskov, Phys. Rev. B **59**, 7413 (1999).

[151] P. Gill, B. Johnson, and J. Pople, Journal of Chemical Physics **98**, 5612 (1993).

[152] C. Quinn, *Computational Quantum Chemistry: An Interactive Introduction to Basis Set Theory* (Academic Pre, 2002).

Appendix A
Modifications to the SIESTA Code

A.1 Modifications for calculation of dipole moment

- Following lines were added to the siesta_analysis.F code:

1. At the beginning of the code:

   ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮

   CONTAINS

   ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮

   use m_dipol

   use parallel, only: IOnode

2. At line number 260:

   ! Print electric dipole

   if (shape .ne. 'bulk') then

      if (IOnode) then

         write(6,'(/,a,3f12.6)') 'siesta: Electric dipole (a.u.) =', dipol

         write(6,'(a,3f12.6)') 'siesta: Electric dipole (Debye) =', (dipol(ix)/Debye,ix=1,3)

      endif

   if (cml_p) then

call cmlAddProperty(xf=mainXML, value=dipol/Debye, title='Electric dipole', dictref='siesta:dipol', units='siestaUnits:Debye')

endif !cml_p

endif

- Following line must be added to MAKEFILE to enable compilation: write_subs.o: m_dipol.o, parallel.o

A.2 Modifications for altering the box size

The automatic_cell.f code was modified. The new code that is used instead of the original version is:

```
C*****************automatic_cell.f*********************
subroutine automatic_cell(ucell,scell,na_u,xa,isa,charnet)
use precision, only: dp
use atmfuncs, only: rcut
use parallel, only: IONode
use units, only: Ang
implicit none
real(dp), dimension(3,3), intent(inout) :: ucell
real(dp), dimension(3,3), intent(inout) :: scell
integer, intent(in) :: na_u
real(dp), dimension(3,*), intent(in) :: xa
integer, dimension(*), intent(in) :: isa
real(dp), intent(in) :: charnet
integer :: ix, ia, is, iv
real(dp) :: rc, xmin, xmax
ucell(1:3,1:3) = 0.0_dp
```

```
scell(1:3,1:3) = 0.0_dp
do ix = 1,3
xmin = huge(1._dp)
xmax = -xmin
do ia = 1,na_u
is = isa(ia)
rc = rcut(is,0)
xmin = min( xmin, xa(ix,ia)-rc )
xmax = max( xmax, xa(ix,ia)+rc )
enddo
C Use a 50% margin for atomic movements
ucell(ix,ix) = 1.50_dp * (xmax - xmin)
scell(ix,ix) = ucell(ix,ix)
enddo
C build cubic cell if system is charged
if (charnet .ne. 0.0_dp) then
xmax = -huge(1._dp)
do ix = 1,3
if (ucell(ix,ix) .gt. xmax)
xmax = ucell(ix,ix)
enddo
do ix = 1,3
ucell(ix,ix) = xmax
scell(ix,ix) = xmax
enddo
endif
C
```

if (IOnode) then

write(6,'(/,a,3(/,a,3f12.6))') . 'siesta: Automatic unit cell vectors (Ang):', . ('siesta:', (ucell(ix,iv)/Ang,ix=1,3), iv =1,3)

endif

C

end subroutine automatic_cell

Appendix B
Analysis of Distributions

Included here is the computer program that is used for the analysis of the dipole moment and energy distributions and extrapolation of dipole moment from a higher temperature to lower temperature.

```
*234567890123456789012345678901234567890123456789012345678901234567890123456789012
c Code to make histograms of energy and dipole from SIESTA run
c
PROGRAM histogram
integer Ebinmax,Dbinmax
parameter (Ebinmax = 1000, Dbinmax = 50)
real*8 Trun,Tsamp,T298,Esamp,EKS,Eref,Dsamp,kB
real*8 Er,Elo,Ehi,Dlo,Dhi
integer samp,maxsamp,step
integer Ebin,Ebinlo,Ebinhi,Dbin,Dbinlo,Dbinhi
real*8 hist1(Ebinmax,Dbinmax),hist2(Ebinmax,Dbinmax)
real*8 hist3(Ebinmax,Dbinmax),hist4(Ebinmax,Dbinmax)
real*8 tot1E(Ebinmax),tot2E(Ebinmax),tot3E(Ebinmax)
real*8 tot4E(Ebinmax),tot1D(Dbinmax),tot2D(Dbinmax)
real*8 tot3D(Dbinmax),tot4D(Dbinmax)
real*8 Ecenter(Ebinmax),Dcenter(Dbinmax)
real*8 Tavg,Davg,Eavg,D298I(2),D298A(2)
```

```
character*25 Efile,Dfile,Ofile
c
c Run temperature should be average temperature of this run.
c Reference energy should be same for all runs of this molecule
c and should be approximately a weighted average energy.
Trun = 300.00 !K
Eref = -4205.70 !eV
T298 = 298.00 !K
c
c input files and output file
Efile = "energies-tbp.MDE"
Dfile = "dip-tbp.csv"
Ofile = "histogram-tbp.csv"
c
c histogram limits for rescaled energies = (E-Eref)/kB (in Kelvin)
Elo = -15000.
Ehi = 15000.
c
c histogram limits for dipoles (in Debye)
Dlo = 0.
Dhi = 20.
c
c boltzmann constant in eV/K
kB = 8.6173323E-5
c
c max possible number of samples (timesteps) from SIESTA runs
maxsamp = 1000000
```

```
c
c reset histograms and get bin centers for energy and dipole
do Ebin = 1, Ebinmax
Ecenter(Ebin) = Elo + (Ehi-Elo)*(Ebin-0.5)/Ebinmax
tot1E(Ebin) = 0.d0
tot2E(Ebin) = 0.d0
tot3E(Ebin) = 0.d0
tot4E(Ebin) = 0.d0
do Dbin = 1, Dbinmax
hist1(Ebin,Dbin) = 0.d0
hist2(Ebin,Dbin) = 0.d0
hist3(Ebin,Dbin) = 0.d0
hist4(Ebin,Dbin) = 0.d0
enddo
enddo
do Dbin = 1, Dbinmax
Dcenter(Dbin) = Dlo + (Dhi-Dlo)*(Dbin-0.5)/Dbinmax
tot1D(Dbin) = 0.d0
tot2D(Dbin) = 0.d0
tot3D(Dbin) = 0.d0
tot4D(Dbin) = 0.d0
enddo
Tavg = 0.d0
Eavg = 0.d0
Davg = 0.d0
D298I(1) = 0.d0
D298I(2) = 0.d0
```

```
D298A(1) = 0.d0
D298A(2) = 0.d0
c
c read in energies, temperatures, and dipoles from SIESTA output
c file and python-generated dipole file
c
open(15,file=Efile)
rewind(15)
read(15,*)
open(16,file=Dfile)
rewind(16)
ccc read(16,*)
c
do samp = 1, maxsamp
c
c read and validate sample from energy file
read(15,FMT=*,end=100) step, Tsamp, EKS, Esamp
if (step.ne.samp) write(6,*)
1 'problem with energy file data, step=',step
c
c read and validate sample from dipole file
read(16,FMT=*,end=100) step, Dsamp
if (step.ne.samp-1) write(6,*)
1 'problem with dipole file data, step=',step
c
c generate rescaled energies
Er = (Esamp-Eref)/kB
```

```
c convert energy values to bin number
Ebin = int((Er - Elo)/(Ehi - Elo)*(Ebinmax-1)) + 1
c check if energy bin index is exceeded
if (Ebin .lt. 1) then
write(6,*)'Elo too large, Er=',Er
Ebin = 1
elseif (Ebin .gt. Ebinmax) then
write(6,*)'Ehi too small, Er=',Er
Ebin = Ebinmax
endif
c convert dipole values to bin number
Dbin = int((Dsamp - Dlo)/(Dhi - Dlo)*(Dbinmax-1)) + 1
c check if dipole bin index is exceeded
if (Dbin .lt. 1) then
write(6,*)'Dlo too large, Dbin=',Dbin
Dbin = 1
elseif (Dbin .gt. Dbinmax) then
write(6,*)'Dhi too small, Dbin=',Dbin
Dbin = Dbinmax
endif
c increment bins, using different temperature options
hist1(Ebin,Dbin) = hist1(Ebin,Dbin) + 1.
hist2(Ebin,Dbin) = hist2(Ebin,Dbin) + exp(Er/Tsamp-Er/T298)
hist3(Ebin,Dbin) = hist3(Ebin,Dbin) + exp(Er/Trun-Er/T298)
hist4(Ebin,Dbin) = hist4(Ebin,Dbin) + exp(Er/Tsamp)
c run averages
Tavg = Tavg + Tsamp
```

```
Eavg = Eavg + Esamp
Davg = Davg + Dsamp
D298I(1) = D298I(1) + Dsamp*exp(Er/Tsamp-Er/T298)
D298I(2) = D298I(2) + exp(Er/Tsamp-Er/T298)
D298A(1) = D298A(1) + Dsamp*exp(Er/Trun-Er/T298)
D298A(2) = D298A(2) + exp(Er/Trun-Er/T298)
enddo
write(6,*) 'maxsamp not large enough'
goto 200
100 maxsamp = samp - 1
write(6,*) 'end of energy or dipole file, maxsamp=',maxsamp
200 close(15)
close(16)
c
c find nonzero ranges of histograms, and normalize
Ebinlo = Ebinmax
Ebinhi = 1
Dbinlo = Dbinmax
Dbinhi = 1
do Ebin = 1, Ebinmax
do Dbin = 1, Dbinmax
if (hist1(Ebin,Dbin).gt.0.d0) then
c adjust ranges up or down if nonzero number of samples
if (Ebinlo.gt.Ebin) Ebinlo = Ebin
if (Ebinhi.lt.Ebin) Ebinhi = Ebin
if (Dbinlo.gt.Dbin) Dbinlo = Dbin
if (Dbinhi.lt.Dbin) Dbinhi = Dbin
```

```
c normalize histograms
hist1(Ebin,Dbin) = hist1(Ebin,Dbin)/maxsamp
hist2(Ebin,Dbin) = hist2(Ebin,Dbin)/maxsamp
hist3(Ebin,Dbin) = hist3(Ebin,Dbin)/maxsamp
hist4(Ebin,Dbin) = hist4(Ebin,Dbin)/maxsamp
c accumulate 1D projections of histograms
tot1E(Ebin) = tot1E(Ebin) + hist1(Ebin,Dbin)
tot2E(Ebin) = tot2E(Ebin) + hist2(Ebin,Dbin)
tot3E(Ebin) = tot3E(Ebin) + hist3(Ebin,Dbin)
tot4E(Ebin) = tot4E(Ebin) + hist4(Ebin,Dbin)
tot1D(Dbin) = tot1D(Dbin) + hist1(Ebin,Dbin)
tot2D(Dbin) = tot2D(Dbin) + hist2(Ebin,Dbin)
tot3D(Dbin) = tot3D(Dbin) + hist3(Ebin,Dbin)
tot4D(Dbin) = tot4D(Dbin) + hist4(Ebin,Dbin)
endif
enddo
enddo
c accumulate run averages
Tavg = Tavg/maxsamp
Eavg = Eavg/maxsamp
Davg = Davg/maxsamp
D298I(1) = D298I(1)/D298I(2)
D298A(1) = D298A(1)/D298A(2)
c
c output file with probabilities over nonzero range
open(17,file=Ofile)
write(17,*) 'probability histograms from SIESTA run'
```

```
write(17,*) 'input files:„',Efile
write(17,*) '„',Dfile
write(17,*)
write(17,*) 'samples =,',maxsamp
write(17,*) 'Trun =,',Trun,',K, setpoint'
write(17,*) 'Tavg =,',Tavg,',K, average from run'
write(17,*) 'Eref =,',Eref,',eV, ref energy for molecule'
write(17,*) 'Eavg =,',Eavg,',eV, average from run'
write(17,*)
write(17,*) 'Davg =,',Davg,',D, average from run'
write(17,*) 'D298-I =,',D298I(1),',D, extrapolation to',
1 ' 298K using instant T'
write(17,*) 'D298-A =,',D298A(1),',D, extrapolation to',
1 ' 298K using average T'
write(17,*)
write(17,*)'All energies below rescaled by (E-Eref)/kB',
1 ' and have units Kelvin'
write(17,*) 'kB =,',kB,',eV/K'
write(17,*)
write(17,*)'Matrix 1: original phase-space probability'
write(17,*)',Dipole (Debye)'
write(17,'(A6,300(A1,ES14.7E3))')'Energy',
1 (',',Dcenter(Dbin),Dbin=Dbinlo,Dbinhi)
do Ebin = Ebinlo, Ebinhi
write(17,'(F8.0,300(A1,ES14.7E3))') Ecenter(Ebin),
1 (',',hist1(Ebin,Dbin),Dbin=Dbinlo,Dbinhi)
enddo
```

```
write(17,*)
write(17,*)
write(17,*)'Matrix 2: 298K extrapolated phase-space ',
1 'probability - intantaneous T'
write(17,*)',Dipole (Debye)'
write(17,'(A6,300(A1,F8.5))')'Energy',
1 (',',Dcenter(Dbin),Dbin=Dbinlo,Dbinhi)
do Ebin = Ebinlo, Ebinhi
write(17,'(F8.0,300(A1,ES14.7E3))') Ecenter(Ebin),
1 (',',hist2(Ebin,Dbin),Dbin=Dbinlo,Dbinhi)
enddo
write(17,*)
write(17,*)
write(17,*)'Matrix 3: 298K extrapolated phase-space ',
1 'probability - average T'
write(17,*)',Dipole (Debye)'
write(17,'(A6,300(A1,F8.5))')'Energy',
1 (',',Dcenter(Dbin),Dbin=Dbinlo,Dbinhi)
do Ebin = Ebinlo, Ebinhi
write(17,'(F8.0,300(A1,ES14.7E3))') Ecenter(Ebin),
1 (',',hist3(Ebin,Dbin),Dbin=Dbinlo,Dbinhi)
enddo
write(17,*)
write(17,*)
write(17,*)'Matrix 4: Density of states - instantaneous T'
write(17,*)',Dipole (Debye)'
write(17,'(A6,300(A1,F8.5))')'Energy',
```

```
1 (',',Dcenter(Dbin),Dbin=Dbinlo,Dbinhi)
do Ebin = Ebinlo, Ebinhi
write(17,'(F8.0,300(A1,ES14.7E3))') Ecenter(Ebin),
1 (',',hist4(Ebin,Dbin),Dbin=Dbinlo,Dbinhi)
enddo
write(17,*)
write(17,*)
write(17,*) '1-D projections / summations of above Matrices'
write(17,*)
write(17,*) ',Matrix1,Matrix2,Matrix3,Matrix4'
write(17,*) 'Energy,runT,insT298,avgT298,DOS'
do Ebin = Ebinlo, Ebinhi
write(17,'(F8.0,300(A1,ES14.7E3))') Ecenter(Ebin),
1 ',',tot1E(Ebin),',',tot2E(Ebin),
2 ',',tot3E(Ebin),',',tot4E(Ebin)
enddo
write(17,*)
write(17,*)
write(17,*) ',Matrix1,Matrix2,Matrix3,Matrix4'
write(17,*) 'Dipole,runT,insT298,avgT298,DOS'
do Dbin = Dbinlo, Dbinhi
write(17,'(F8.5,300(A1,ES14.7E3))') Dcenter(Dbin),
1 ',',tot1D(Dbin),',',tot2D(Dbin),
2 ',',tot3D(Dbin),',',tot4D(Dbin)
enddo
close(17)
c
```

```
stop
end
```

Appendix C
The MD Code for Proton Transfer

To obtain the MD code used in this work, please visit the following digital archive at the BYU Harold B. Lee Library:

http://hdl.lib.byu.edu/1877/2944

Two sets of files are included there. The first one includes the MD code and the input files for the bulk water simulations. The second set includes the MD code and input files for bulk water simulations with one excess proton and one counterion. In addition, the second set includes a python script that is used to calculate the proton diffusion constant from the information contained in two output files.