MOLECULAR MODELING OF AQUEOUS SOLUTIONS: THE EFFECTS OF MOLECULAR POLARIZATION ON CLASSICAL FORCEFIELD DEVELOPMENT

by

Peter James Dyer

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical Engineering

December, 2008
Nashville, Tennessee

Approved:

Professor Peter T. Cummings

Professor Kenneth A. Debelak

Professor M. Douglas LeVan

Professor Clare M. McCabe

Professor Jens Meiler

to my Mum and Dad

ACKNOWLEDGEMENT

Over the course of this dissertation, there has been a number of people who has helped me along the way which has made this a more enjoyable and fulfilling experience, both educationally and personally.

Firstly I would like to thank my advisor, Dr Peter T Cummings. Thank you for the valuable and plentiful advice and knowledge you have shared over the entire graduate study process.

I would like to thank my fellow lab members, especially Patrick Redmill and Hugh Docherty for useful discussing and for comic relief over the years. The daily coffee break was the highlight of the workday, and has made the final year much more manageable.

I would like to thank Department of Energy for financial support throughout the degree, through grant number DE-FG02-03ER15385, without your funding this work would not have been possible.

And finally, I would like to thank my parents, for their encouragement and support during my time overseas.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

BACKGROUND

1.1 Outline

This thesis is about the formulation of new molecular dynamics (MD) forcefields to describe the interactions of water and solutes in aqueous environments, particularly the role of polarizability in forcefield transferability. A forcefield is a mathematical function which describes the physical interaction between two or more molecular entities. In general, the better the forcefield describes the true physical interaction, the more successfully it will be able to reproduce the actual behavior of a fluid composed of molecules. The development of MD forcefields, where the interactions between molecules and evolution of the molecular motion is based on classical mechanics, are required for studying molecular systems which are either too large or complex to be practically simulated via *ab initio* methods or too expensive or impossible to do via an experimental investigation.

In this work, the systems described by the new forcefields will be compared to both experimental observations and *ab initio* calculations. We will examine how the inclusion of polarization and the use of Gaussian distributions for electrostatic charges enhance the ability of the new forcefields to reproduce the physical properties of the systems, and how they differ from simpler models.

The layout of this thesis begins as follow, it begins with an introduction to water and aqueous solutions and the importance of performing molecular simulations for the understanding of these systems. Included is a summary of previous simulation studies

from the wider academic community of water and aqueous solutions. The aim of Chapter II is to describe an assortment of techniques used in classical simulations and *ab initio* calculations which have been used in this thesis. It is hoped that this introduction will show how these methods complement each other, and how assumptions are used to move from highly computationally expensive and accurate calculations to less computationally expensive calculations, while still retaining sufficient molecular interactions to produce scientifically meaningful results.

Chapter III and Chapter IV examine the effect of adding point polarizability to non-polar but polarizable solutes, and how this affects the solubility of simple solutes in bulk water and aqueous solutions. Specifically, Chapter III focuses on a range of popular simple rigid models for water in bulk conditions. Multiple water models are simulated in order to assess the ability of each water model to reproduce the bulk water behavior, and to investigate how the errors in the reproduction of the water solvent affect the solubility of non-polar solutes in aqueous solutions. The excess chemical potential is calculated for a range of temperatures at 1 Bar. Chapter IV investigates the effect of charged co-solvents (ions) on the solubility of non-polar solutes. The effect of the ions on the electric field and how it changes the solubility of the solutes, as the ion concentration increases at standard pressure and temperature are investigated.

Chapter V is a comparison between a recently developed classically-based water model (GCPM), which uses Gaussian distributed electrostatic charges and explicit polarizability, and *ab initio* Car-Parrinello molecular dynamics under a range of state conditions. The objective is to examine the ability of the classical water model to reproduce the mean and distribution of the induced dipole moment of bulk water.

Chapter VI shows the development of ion-water models to use with GCPM water. This shows that by using a polarizable ion forcefield, together with a polarizable model of water it is possible to reproduce the induced dipole moments for the ion compared to Car-Parrinello simulations results, and the correct kosmotropic and chaotropes behavior of the ions in ion-water clusters. We have also show the behavior of the ion and water structure at infinite dilution.

In Chapter VII, the intramolecular forcefield of the Thiol bond, between benzendithiolate (BDT) and gold clusters is investigated using various DFT functionals, are used to develop intramolecular bonding parameters. The effect of the variance in the bonding parameters is assessed, by examining the resulting packing structure of a self-assembled BDT monolayer bound to a gold surface. Finally, in chapter VIII, the overall conclusions of this work and possible future directions will be presented.

1.2 Water

Water is one of the most common substances on the Earth's surface, and is the most important fluid for our existence. All three phases are present in the natural environment, with liquid water covering approximately 70.8 percent of the earth[1]. Water's abnormal properties, are necessary for the existence of life, especially the ability of intracellular water to dissolve cellular constituents[2]. In biological life forms, water is the solvent which allows transport of nutrients, the ability to break down proteins and carbohydrates for digestion and is important for the regulation of cellular volume[3]. Water's behavior around dissolved biomolecules contributes to the 'hydrophobic effect', which contributes to protein folding, and plays an important role in the specificity and affinity of protein and DNA interactions[4]. Water molecules buried inside protein

molecules are so structurally important to the folding of proteins that is has been suggested that they should be included into protein structures[5]. As water is so important in biological systems, the understanding of the behavior of water, its aqueous solutions and the behavior of small solutes dissolved in the aqueous solution is vital for understanding the interaction between aqueous solutions and other larger biological molecules[6,7].

Water is a deceptively simple molecule comprised of two hydrogen atoms joined to a single oxygen atom. The hydrogen atoms are proton donors and the 'lone electron pairs' on the oxygen are proton acceptors, resulting in the ability of each water molecule to have four strong hydrogen bonds in a tetrahedral orientation. It is the formation of the hydrogen bonds that gives water its unique and unexpected properties. Some of these properties include a density maximum at 4^0C, as well as relatively high melting and boiling points for its molecular weight[8]. It has a high permanent dipole created from the electronegativity of the oxygen atom and the unsymmetrical nature of the angle between the hydrogen-oxygen bonds. The high dipole and dielectric constant gives water the ability to stabilize and dissolve ionic and polar species, the strong non-ideality of aqueous electrolyte solutions, and the ability of water to cluster around solutes has been related to complex water-water and water-solute intermolecular interactions.

The ability to simulate liquids at room temperature was a relatively new tool in the field of condensed matter physics, when, in 1974, Stillinger and Rahman[9] published simulations of the first widely used computer model for liquid water. Since then countless water models have been developed in order to better represent water in classical molecular simulations. Guillot[10] provides a good review of models developed prior to 2002. Classical water models have three main components; molecular structure, short

range interactions or van der Waals interactions, electrostatic interactions (partial charges) and, in addition to this, some models include explicit polarization which accounts for many body effects.

Most water models have a rigid structure, in which the bond length and angles between the atoms are fixed (*Figure 1-1*). Rigid water models are normally set to the isolated gas phase geometry with a HOH angle of 104.52^0, or to the angle between the arms of a tetrahedral (two H arms and two arms for the lone electron pairs), with a HOH angle of 109.47^0. Flexible models[11-13] allow the atoms to act like springs, permitting intramolecular vibrations, added in order to reproduce the vibrational modes of the isolated molecule.

Short-range interaction potentials are used to account for the van der Waals forces acting on the molecule, consisting of dispersion attraction and short range repulsion. In practice, for water models, the short range interactions are often only added to the oxygen atom, and include implicitly the van der Waals forces of the hydrogen atoms.

The charges on the water model are usually modeled as point charges, (corresponding to Dirac delta functions in charge density), with positive charges located at the hydrogen (H) atoms and a negative charge located on the oxygen (O) atom or at an offset a small distance along the HOH bisector. This offset has been shown to be vital in reproducing the ice I structure upon freezing[14]. The total charge of each molecule is zero, as water is an overall neutral molecule. The magnitude of the charge is sometimes set to reproduce the isolated dipole of water[15-21]. This is especially true for some polarizable models, since an induced dipole component is added to incorporate the effect of neighboring molecules. For non-polarizable models, the fixed dipole moment is increased beyond the isolated molecular value to implicitly take into account induction

effects thus enhancing the accuracy of the model at the state conditions where the water model is designed to be applied, which is usually at ambient conditions, i.e. 298 K and 1.01325 Bar. The number of electrostatic charges used in a water model normally varies between three and five, but increasing the number of sites has not consistently meant that accuracy is improved[22]. Recently there has been an increased awareness of the importance of higher multipoles, with the work of Abacal and Vega[23], showed that there is a trend between the melting point of hexagonal ice (Ih or everyday ice) and the quadrupole of the water model.

Polarizability is the distortion of the electron distribution from its normal shape due to the electric fields imposed by neighboring molecules. For water, the total dipole moment is increased from 1.855D[24] for the isolated molecule to experimental values, from high-energy x-ray measurements of the structure, of 2.9±0.6D[25] in liquid water at ambient conditions. The inability to precisely measure the total dipole moment of liquid water experimentally means that the actual value for the mean total dipole moment is subject to constant debate. Calculations the induced dipole moment using different *ab initio* methods unfortunately give different result. Coutinho *et al.*[26], report the total dipole moment of liquid water as 2.60 ± 0.14D, using a range of different quantum chemical calculations. Consistent with this result, is the work of Tu and Laaksonen[27] who report a value for the dipole moment of 2.65D, but it was higher than the work of Delle Site *et al.*[28] who obtain 2.47D. More importantly, Coutinho contended that the value of the total dipole moment of liquid water has an upper bound of 2.74D, due to the association between the with the water dipole moment and the dielectric constant[29]. Contrary to Coutinho's conclusions, Silvestrelli and Parrinello[30], used maximally localized Wannier functions[31] (Chapter V) to calculate the average total dipole moment of water of

2.95±0.05D, which is significantly higher than most other dipole moment calculation methods.

Nevertheless, this dramatic increase in the dipole moment between the isolated and condensed phases, needs to be accounted for if a model is going to be useful at both high and low water densities, and across a range of temperatures and heterogeneous solutions. One way to achieve this is to include explicit polarization. Unfortunately, polarization is a computationally intensive addition to water models, as it requires iteration over the entire system since the induced dipole moment on each molecule affects the induced dipole moment on all the other molecules. This iterative procedure continues until a converged solution for the electrostatic field is found in a self-consistent manner, called a self-consistent field (SCF). For this reason, explicit polarization is often excluded from simulations. A common *ad hoc* method to include polarization is to add a flexible point charge tethered to the oxygen atom with a harmonic potential. This is called a Drude particle with the technique called a Drude oscillator[32,33]. This allows a point charge to move around, changing the dipole and the related electrostatic interactions. This method removes the requirement of calculating a SCF, saving computational time. Though highly efficient, Drude particles have the disadvantage that the Drude particle is confined to the molecular plane, whereas experimentally polarizability of a water molecule is nearly isotropic[34].

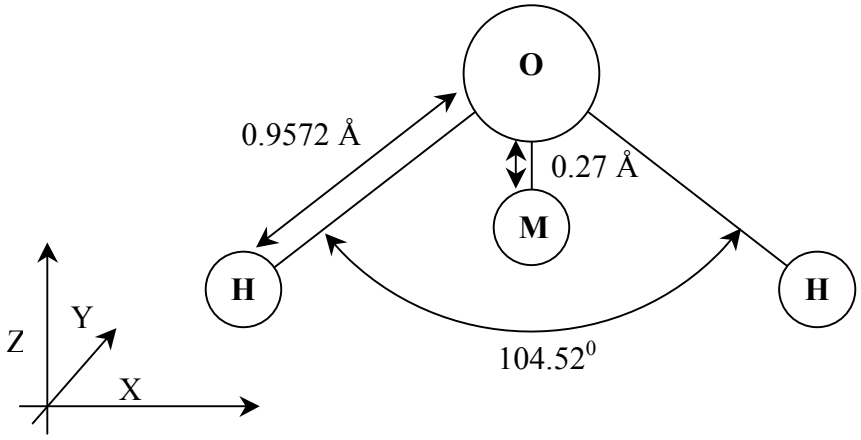


Figure 1-1: *Geometry of the GCPM molecule, a rigid molecule with an offset apex for the negative charge at point M. The positive partial charges are centered on the hydrogen atoms.*

1.3 Ions in aqueous solutions

Introduction

Water's importance comes from its ability as a solvent, and as a medium for chemical reactions. The aqueous solvation of ions plays an important role in chemical and biological systems. In biological application, ions are used to control the osmotic gradients related to the transport of biochemicals[35]. This is used to regulate the blood acidity, hydration and muscle functions. Ions also allow the transport of current through aqueous media, as in, for example, lead-acid batteries.

The effect of changing the ion concentration in aqueous solutions on the solubility of hydrophobic solutes in aqueous solution has been known for hundreds of years; with chemists using the addition of salt for the separation of chemicals. Salting-out (or Salting-in) effects is the decrease (or increase) in the solubility of hydrophobic molecules due in the increases in ion concentration. The mechanism for the salting-out effects is not well known, and there have been a number of mechanisms suggested in order to explain it.

The first mechanism is based on how the water molecules form a hydration layer around dissolved electrolyte and non-electrolyte solutes[36]. As water molecules prefer to surround an ion rather that a non-electrolyte solute, the more ions which are dissolved into the solution, the more water molecules will be involved in forming an ion hydration layer. This will reduce the number of water molecules available to surround the non-electrolyte solute. As a result, the solubility of the solute in the solution decreases, and precipitation of the solute will occur.

The second mechanism is based on the effect of the solute on the dielectric constant of the solvent[37,38]. If the saturated solution has a greater dielectric constant than that of bulk water, then salting-in occurs. Alternatively, if the saturated solution has a

lower dielectric constant than that of water, then salting-out occurs. This theory works well for small ions, but fails for larger ones, as the structural effects of the solution are not taken into account. The third is based on the internal pressure[39,40], or packing structure of the solution. The idea is based on the fact that a solute molecule occupies a volume in the solution, exerting a pressure on the solvent molecules. If the volume of the solution decreases after the addition of ions, this *pressure* changes the ion-solvent interactions to be more repulsive, causing the solute to precipitate out. Oppositely, if the volume of the solution increases, there is a salting-in effect as the ion-solvent interaction is more attractive. For a good review, see Grover and Ryall[41].

The ranking of how much effect ions have on the stability and solubility of proteins is known as the Hofmeister series[42]. Even though the Hofmeister series has been able to qualitatively explain the trends of protein in aqueous solutions, the ability to quantitatively describe the ion position in the Hofmeister series and its effect on a protein has not been obtained[43]. A small subsection of the Hofmeister series is shown below,

$$Ca^{2+} > Mg^{2+} > Na^{+} > K^{+} > NH_4^{+} \qquad [1\text{-}1a]$$

$$F^{-} > Cl^{-} > Br^{-} > I^{-} > ClO_4^{-} > SCN^{-} \qquad [1\text{-}1b]$$

where the ions on the left are kosmotropic (water structuring) and have a greater structuring effect on the solution and they strengthen the hydrophobic effect more than those on the right, which are more chaotropic (water disrupting).

Kosmotropes generally have small ionic radii, with multi-charged ions resulting with higher charge densities. The ion forms strong hydrogen bonds with the surrounding water, increasing the amount of structuring in the water around the ion, for example, Li^{+},

Ca^{2+}, Mg^{2+}, F^{-}[44]. Most of the kosmotropes are cations, as they generally have a smaller ionic radius.

Chaotropes, tend to have a larger ionic radius with a single charge, which will result in a lower charge density. This leads to weaker ion-water interactions relative to the cohesive water-water interactions. A chaotrope acts like a defect in the water structure compared to the pure water structure. Examples of chaotropes are SCN^{-}, ClO_4^{-}, I^{-}. For proteins, kosmotropes at a high concentration will stabilize proteins, whereas chaotropes at high concentrations destabilize proteins[45,46]. For a review of the Hofmeister effects see the work of Kunz *et al*[60].

Ion solvation in aqueous solution has been the focus of a wide range of experimental and theoretical research. In regard to computational chemistry and simulations, ionic solutions have been studied at both classical and *ab initio* levels. Reviews of aqueous solutions, including experimental characterization, are provided by Ohtaki and Radnai[47], Soper[48] and Rode *et al.*[49].

The solvation of ions in water is strongly dependant on the ion's size, and the sign and magnitude of the charge of the ion. Anions have a much more favorable free energy of hydration than that of a cation of the same size and charge magnitude[50-52]. The most frequently used method for modeling ions in aqueous solutions is to use a point charge model with a short range van der Waals interaction, often represented by a Lennard-Jones potential[53-56]. In more recent models, polarizability has been added to the ions, and it has been found to have a large effect on the nature of the ion-water interaction. Many-body effects, i.e. polarizability, in both the water and ion models are required for surface phenomena to be correctly reproduced in simulations[57,58].

Ions at the water/air interface

The behavior, stability and structure of ions at water interfaces are important given the role of interfaces between bulk water and biomolecules such as proteins, nucleic acids and membranes[61-63]. Surface solvation occurs when the ion resides preferably at the edge of the water/air interface. As a result, the second solvation shell preferentially starts to fill up before the first layer is completely full. Jungwirth and Tobias[64], showed that larger halide ions have a higher concentration of ions at the interface than in the bulk. This is in contrast what you would usually expect, as ions at the surface will increase the surface tension that will be unfavorable. However, the large halide ions breaking the hydrogen bonding structure in the bulk solution has a greater energy penalty than the increase in the surface tension from the ions being at the surface. As a result the larger ions preferentially move to the surface. For large clusters, the large halide ions do enter the bulk phase due to the entropic energy. Small cations in general are pulled in to the cluster, and undergo bulk solvation. Many simulations of the surface solvation of ions have been conducted[56,57,65-69]. Perara and Berkowitz[66] first predicted that a Cl^- anion solvated in a water cluster (up to 20 molecules) preferred surface solvation. There have been discrepancies between *ab initio,* classical simulations and experiments[68,70] concerning the number of water molecules required for bulk solvation to occur for halide ions. Simulations by Herce *et al.*[57], reported that the inclusion of polarizability is vital to reproduce the correct surface or bulk solvation behavior.

1.4 Experimental Methods

Experimental measurements are very important to molecular simulations as they can be used to verify the ability of the simulations to reproduce quantitatively the actual molecules behavior. Below is a summary of some molecular properties which are frequently calculated via molecular simulations and how they are measured experimentally.

Fluid structure

The structure of a fluid is the arrangement of the atoms around each other. The structure is often reported as a radial distribution function $g_{ij}(r)$, which gives the probability of two atom types *i* and type *j* being separated by a certain distance r, relative to the expect number of atoms at the same distance for a uniform concentration with the macroscopic density ρ. When a fluid has no structure, i.e. an ideal gas, the $g(r)$ has a value of 1 for all values of r. As a fluid becomes more structured or ordered there are larger deviations from the ideal gas (*Figure 1-2*). For water at standard conditions, there is structuring for the first 6-7Å, or two hydration shells, as seen in the oxygen-oxygen radial distribution function, $g_{oo}(r)$.

The structure of water was first studied by x-ray crystallography. X-rays are scattered by the electrons surrounding the nuclei[47], where the angle of the deflection of x-ray is compared to the straight trajectory. The degree of deflection gives information of the structure of the material. Using wide-angle X-ray scattering, the location of the nuclei separated by a distance similar to the X-ray wavelength (~1Å) can be determined, these distance are usually for determination intramolecular interactions. By changing the distance between the sample to the detector, small-angle X-ray scattering (SAXS) can be

preformed. Using the longer wavelengths, intermolecular separation distances can be measured. By using X-ray scattering, it was discovered that water was bound in a tetrahedral structure, but it could not account for the lack of long-range order in the liquid. X-ray scattering on water was always going to be problematic, as the hydrogen atom has a small electron density and as a result, it has small scattering power.

This problem was solved by the advent of the neutron diffraction, where beams of neutrons are diffracted by the interactions with the nuclei of the atoms. With a larger nucleus, there is a larger cross-area, resulting in more diffraction of the neutrons. So unlike x-ray diffraction, where the electron density is fixed by the atom type, the atomic nuclei can be increase through the substitution of one isotrope to another heavier isotrope. In the case of water, a hydrogen atoms (1 a.u.) can be replaced with deuterium (2 a.u.), increasing ability to extract hydrogen and oxygen infomation. With neutron diffraction, it is possible to determine the average orientation of water molecules around a central water molecule. Due to there being two hydrogen's for every oxygen in a water molecule, the $g_{HH}(r)$ peak is the most accurate, next being $g_{OH}(r)$ then the $g_{OO}(r)$. Conversely, water models are compared to the $g_{OO}(r)$, first and often only the oxygen-oxygen correlation is compared. A review of water structure from experiments and simulations is provide by Head-Gordon and Hura[71] and ion solvation via neutron scattering by Enderby[72].

Dynamics

As molecular simulations describe dynamic systems, i.e. the motion of molecules over time, is of great importance. Dynamic properties include self-diffusion, which is the

displacement a molecule goes through over time, rotational diffusion, and the rate of replacement of the first hydration shell, which is an indicator for how strong a molecule or ion, holds onto its boundary water layer. A good review of the structure and dynamics of hydrated ions is provided by Ohtaki and Radnai[47]

Solubility

Accurate solubility measurements are important for the calculations of the salting effects of hydrophobic molecules in aqueous solution and determination of the excess chemical potential. As the solubility of hydrophobic molecules in aqueous solutions is very low, mass-spectrometry is used. Mass-spectrometry turns the sample into ionized fragments, which are separated according to their masses by passing through an electric or magnetic field. For ions of the same charge, the same amount of kinetic energy is transferred. The larger the mass of the ion, the slower the ion moves, and by measuring the speed, the type of ion can be determined. There are multiple variants of mass spectrometry, but in essence, the concentration is determined by the amount of each type of ions collected in each variant[73].

Vapor liquid equilibrium

Vapor liquid equilibrium (VLE) calculations require the use of accurate pressure, temperature and composition measurement techniques. There are four main methods used. In the first method, static type, a vessel of variable or constant volume is filled with each substance. The temperature and pressure is adjusted to bring about a phase separation. Stirrers are used to reach the equilibrium state, and samples are taken from each phase for gas chromatography (GC) or mass spectrometry analysis of the

composition. The second method, the recirculation method is similar to the static but each phase has an outlet which is pumped into the other phase. This allows for easy assess to the samples and rapid achievement of equilibrium. Thirdly, there is the open continuous flow method, in which the components are pumped into a separating cell, at equilibrium conditions. It is useful for mixtures with sensitive components, but requires a large amount of samples. Lastly, the synthetic method, is when a known mixture is added into a cell the temperature and pressure adjusted until a homogenous phase is formed. Then the pressure and temperature is varied until the mixture forms a new phase. This has the advantage that there is no sampling, but it is inaccurate as precise detection of phase transformation maybe difficult.

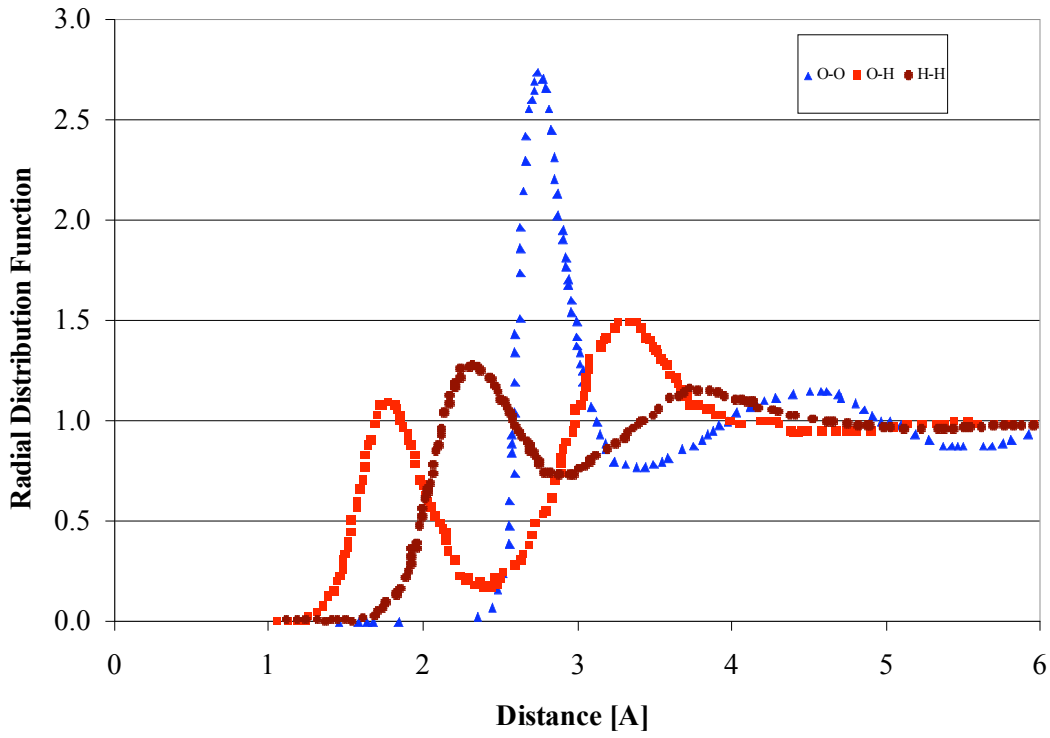


Figure 1-2: *Radial distribution functions of water at 298K and 1atm, using neutron scattering from Soper*[74]. *Blue triangles, is the* $g_{OO}(r)$, *Red squares is the* $g_{OH}(r)$, *Brown circles is* $g_{HH}(r)$,

References

[1] F. Franks, *Water, A comprehensive Treatise*. (Plenum Press, 1972).

[2] R. Cooke, I. D. Kuntz, Annual Review of Biophysics and Bioengineering **3**, 95 (1974).

[3] P. M. Wiggins, Microbiological Reviews **54** (4), 432 (1990).

[4] J. W. Schwabe, Current opinion in structural biology **7** (1), 126 (1997).

[5] U. Sreenivasan, P. J. Axelsen, Biochemistry **31**, 12785 (1992).

[6] M. J. Tait, F. Franks, Nature **230**, 91 (1971).

[7] P. E. Smith, B. M. Pettitt, Journal of Physical Chemistry **98** (39), 9700 (1994).

[8] B. Cabane and R. Vuilleumier, Comptes Rendus Geoscience **337** (1-2), 159 (2005).

[9] F. H. Stillinger, A. Rahman, Journal of Chemical Physics **60** (4), 1545 (1974).

[10] B. Guillot, Journal of Molecular Liquids **101**, 219 (2002).

[11] G. S. Fanourgakis and S. S. Xantheas, Journal of Physical Chemistry A **110** (11), 4100 (2006).

[12] S. B. Zhu, S. Yao, J. B. Zhu, S. Singh, G. W. Robinson, Journal of Physical Chemistry **95** (16), 6211 (1991).

[13] Y. J. Wu, H. L. Tepper, and G. A. Voth, Journal of Chemical Physics **124** (2) (2006).

[14] E. Sanz, C. Vega, J. L. F. Abascal, L. G. MacDowell, Physical Review Letters **92** (25), 255701 (2004).

[15] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L, Klein, Journal of Chemical Physics **79** (2), 926 (1983).

[16] P. Paricaud, M. Predota, A. A. Chialvo, P. T. Cummings, Journal of Chemical Physics **122** (24), 4511 (2005).

[17] G. Lamoureux, E Harder, L V. Vorobyov, B. Roux, A. D. MacKerell, Chemical Physics Letters **418**, 245 (2006).

[18] H. A. Stern, F. Rittner, B. J. Berne, and R. A. Friesner, Journal of Chemical Physics **115** (5), 2237 (2001).

[19] P. J. van Maaren and D. van der Spoel, Journal of Physical Chemistry B **105** (13), 2618 (2001).

[20] E. Rozners, J. Moulder, Journal of Molecular Structure: Theochem **712**, 167 (2004).

[21] H. Saint-Martin, J. Hernandez-Cobos, M. I. Bernal-Uruchurtu, I. Ortega-Blake, and H. J. C. Berendsen, Journal of Chemical Physics **113** (24), 10899 (2000).

[22] M. Lisal, I. Nezbeda, W. R. Smith, Journal of Physical Chemistry B **108**, 7412 (2004).

[23] J. L. F. Abascal, C. Vega, Physical Chemistry Chemical Physics **9**, 2775 (2007).

[24] S. A. Clough, Y. Beers, G. P. Klein, Journal of Chemical Physics **59** (5), 2254 (1973).

[25] Y. S. Badyal, M.-L. Saboungi, D. L. Price, S. D. Shastri, and D. R. Haeffner, Journal of Chemical Physics **112** (21), 9206 (2000).

[26] K. Coutinho, R. C. Guedes, B. J. Costa Cabral, S. Canuto, Chemical Physics Letters **369**, 345 (2003).

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

[28] L. Delle Site, A. Alavi, R. M. Lynden-Bell, Molecular Physics **96** (11), 1683 (1999).

[29] K. Watanabe, M. L. Klein, Chemical Physics **131** (2-3), 157 (1989).

[30] P. L. Silvestrelli, M. Parrinello, Journal of Chemical Physics **111** (8), 3572 (1999).

[31] N Marzari, D Vanderbilt, Physical Review B **56** (20), 12847 (1997).

[32] G. Lamoureux, B. Roux, A. D. MacKerell, Journal of Chemical Physics **119** (10), 5185 (2003).

[33] P. Drude, *The Theory of Optics*. (Longmans, Green and Co., New York, 1902).

[34] T. A. Halgren, W. Damm, Current opinion in structural biology **11** (2), 236 (2001).

[35] D. D. F. Loo, E. M. Wright, and T. Zeuthen, Journal of Physiology-London **542** (1), 53 (2002).

[36] J. E. Desnoyers, C. Jolicoeour, *Modern Aspects of Electrochemistry*. (Plenum Publishing Corp, New York, 1969).

[37] J. M. P. Debye, Z. Phys. Chem **25**, 22 (1925).

[38] B. E. Conway, J. E. Desnoyers, A. C. Smith, Phil. Trans. Roy. Soc (London) **A256**, 389 (1964).

[39] G. Tammann, Z. Anorg. Allg. Chem **158**, 1 (1926).

[40] W. F. McDevit, F. A. Long, Journal of the American Chemical Society **74**, 1773 (1952).

[41] P. K. Grover, R. L. Ryall, Chemical Reveiw **105** (1), 1 (2005).

[42] F. Hofmeister, Arch Exp. Pathol Pharmacol **24**, 247 (1888).

[43] J. M. Broering, A. S. Bommarius, J. Phys. Chem. B **109**, 20612 (2006).

[44] N. Muller, Journal of Solution Chemistry **17** (7), 661 (1988).

[45] M. G. Cacace, E. M. Landau, and J. J. Ramsden, Quarterly Reviews of Biophysics **30** (3), 241 (1997).

[46] O. S. Lawal, Food Chemistry **95** (1), 101 (2006).

[47] H. Ohtaki, T. Radnai, Chemical Reveiw **93**, 1157 (1993).

[48] A. K. Soper, Journal of Physics: Condensed Matter **9**, 2717 (1997).

[49] B. M. Rode, C. F. Schwenk, A Tongraar, Journal of Molecular Liquids **110**, 105 (2004).

[50] S. Rajamani, T. Ghosh, S. Garde, Journal of Chemical Physics **120** (9), 4457 (2003).

[51] J. C. Rasaiah, R. M. Lynden-Bell, Philosophical Transactions of the Royal Society of London A **359**, 1545 (2001).

[52] R. M. Lynden-Bell, J. C. Rasaiah, J. P. Noworyta, Pure and Applied Chemistry **73** (11), 1721 (2001).

[53] J. Chandrasekhar, D. C. Spellmeyer, and W. L. Jorgensen, Journal of the American Chemical Society **106** (4), 903 (1984).

[54] G. Palinkas, W. O. Riede, and K. Heinzinger, Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences **32** (10), 1137 (1977).

[55] B. M. Pettitt and P. J. Rossky, Journal of Chemical Physics **84** (10), 5836 (1986).

[56] L. X. Dang, D. E. Smith, Journal of Chemical Physics **99** (9), 6950 (1993).

[57] D. H. Herce, L. Perera, T. A. Darden, C. Sagui, Journal of Chemical Physics **122**, 024513 (2005).

[58] P. B. Petersen, R. J. Saykally, Annual Review Physical Chemical **57**, 333 (2006).

[59] M. Jonsson, M. Skepo, P. Linse, J. Phys. Chem. B **110**, 8792 (2006).

[60] W. Kunz, P. Lo Nostro, B. W. Ninham, Curr Opin Colloid Interface Science **9**, 1 (2004).

[61] S. McLaughlin, Annual Review of Biophysics and Biophysical Chemistry **18**, 113 (1989).

[62] K. D. Collins, Biophysical Journal **72**, 65 (1997).

[63] B. Honig, K. Sharp, and A. S. Yang, Journal of Physical Chemistry **97** (6), 1101 (1993).

[64] P. Jungwirth and D. J. Tobias, Journal of Physical Chemistry B **105** (43), 10468 (2001).

[65] C. W. Bauschlicher, S. R. Langhoff, H. Partridge, J. E. Rice, A. Komornicki, Journal of Chemical Physics **95** (7), 5142 (1991).

[66] L. Perera, M. L. Berkowitz, Journal of Chemical Physics **95** (3), 1954 (1991).

[67] L. Perera, M. L. Berkowitz, Journal of Chemical Physics **96** (11), 8288 (1992).

[68] L. X. Dang, B. C. Garrett, Journal of Chemical Physics **99** (4), 2972 (1993).

[69] L. Perera, M. L. Berkowitz, Journal of Chemical Physics **100** (4), 3085 (1994).

[70] G. Markovich, R. Giniger, M. Levin, O. Cheshnovsky, Journal of Chemical Physics **101** (11), 9344 (1994).

[71] T. Head-Gordon, G. Hura, Chemical Reveiw **120** (8), 2651 (2002).

[72] J. E. Enderby, Chemical Society Reviews, 159 (1995).

[73] J. T. Watson, *Introduction to mass spectrometry*. (Raven Press, New York, 1985).

[74] A. K. Soper, F. Bruni, M. A. Ricci, Journal of Chemical Physics **106** (1), 247 (1996).

CHAPTER II

OVERVIEW OF SIMULATION METHODOLOGIES

2.1 Classical Simulations

Theory

Classical molecular dynamics (MD) simulations apply Newtonian mechanics to predict the motion of atoms in a system, thus leading to prediction for the behavior of the modeled species. MD simulations are based on solving Newton's second law of motion,

$$\frac{d^2\vec{r}_i}{dt^2} = \sum_{j=1,n} \frac{\vec{F}_{ij}(r_{ij})}{M_i} \qquad [2\text{-}1]$$

where M_i is the mass of particle i, $\vec{r}_i$ is the position in real space, the force $\vec{F}_{ij}(r_{ij})$ is calculated for the positions and orientations of the modeled particles relative to each other. By numerically integrating the above equation through a timestep Δt, the velocities of each molecule can be obtained. Integration of the new velocities obtain the new positions, starting the cycle over again with the determination of the forces associated with the new atomic positions. As the mass for a given species is constant, differing models only have different representations for the force. The differing complexities of different models, often at differing computational cost, are an attempt to improve the representation of the particle-particle interactions (i.e. forces) of the models to be consistent with the behavior of real molecules.

Classical forcefields

Most classical models contain similar methods for representing the van der Waals interactions and electrostatic interactions. The attractive interactions arise from London forces, due to transient dipoles neighboring particles. This is usually represented via a theoretically based r^{-6} term[1]. As the particles become closer, the electron orbitals start to overlap. As electrons of the same spin become close to each other, strong exchange forces repulse the two particles. This is the basis of the Pauli exclusion principal[2]. This force is exponential with respect to distance, but is often approximated with a r^{-12} term. Historically, the repulsive r^{-12} term was chosen because of simplification of the calculation of the second virial coefficient, see Figure 2-1. In practice, it is frequently used despite it having no physical basis. Combining these two terms yields the 12-6 Lennard-Jones potential,

$$U_{LJ}(r) = 4\varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right) \qquad [2\text{-}2]$$

$$\vec{F}_{LJ}(r) = -\frac{dU_{LJ}}{dr}\hat{r}$$

where U_{LJ} is the energy and F_{LJ} is the force between particle of type i and a particle of type j, σ_{ij} is a measure of the diameter of the interaction between i and j particles. ε_{ij} is the mixed well depth and r is the distance between particle i and j. The force is the negative differential of the energy with respect to position. The method of creating LJ parameters for the interactions between different components, from pure components σ_i, ε_i is often via combination rules. Combining rules have very little physical basis for real molecules, and the most popular are the Lorentz-Berthelot[3] combining rules.

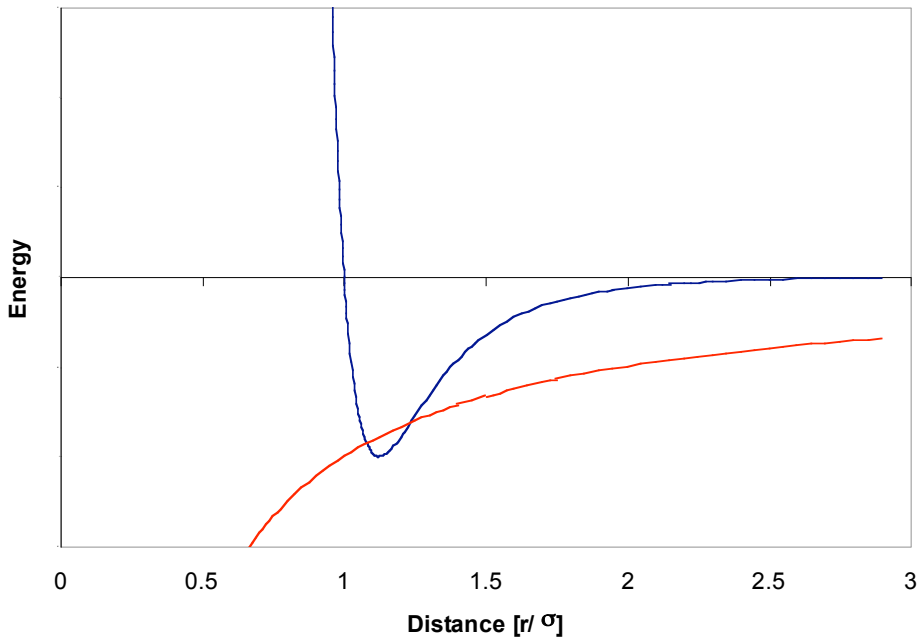


Figure 2-1: *(Blue) Approximation of the short range energy between particles. Positive gradients correspond to attractive interactions, and negative gradients correspond to repulsive interactions. (Red) Electrostatic potential for oppositely charged ions, the gradient is always positive.*

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2},\ \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \qquad [2\text{-}3]$$

To reduce computational time during molecular simulations, only molecules inside a chosen distance from each other (cut off radius) are used for calculating short range interactions for between particles. At large distances, the combined effect of the van der Waals forces on a particle are often assumed to be negligible as these can be approximately cancelled out due to the particle being surrounding by a uniform density of particles outside the cutoff. This is the homogenous solution assumption, and is based on the idea that two equal but opposite forces pushing against an object results in no change of velocity. Even though these long range interactions would have little effect on the equations of motion they have a significant effect on the pressure and energy calculation.

Corrections of the long-range dispersions forces become non-trivial for multiphase simulations if the interface is outside the cut-off. As a result there would be a net force pulling the interfacial molecules towards the denser phase. To account for the energy excluded due to the use of a cut-off radius for a homogeneous solution, we use the expression,

$$U_{lj}^{Corr} = 2\pi\rho \int_{r=R_c}^{r=\infty} 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) r^2 \cdot dr = \frac{2}{3} 4\varepsilon\rho\pi\left(\frac{\sigma^{12}}{3R_c^{\,9}} - \frac{\sigma^{6}}{R_c^{\,3}}\right) \qquad [2\text{-}4]$$

where R_c is the cut off radius and ρ is the density of the bulk system. This corresponds to assuming a uniform structure given by $g(r) = 1$, beyond the cut-off.

The electrostatic potential for point charges are calculated via,

$$U_{ele}(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \qquad [2\text{-}5]$$

where q_i is the charge of particle i and r_{ij} is the distance between particle i and j. ε_0 is the vacuum permittivity of free space. As the electrostatic energy is proportional to r^{-1}, the total electrostatic energy of a system does not decay to zero as the number of particles increase as r^3. Thus other methods are required to account for long range electrostatic interactions. Below is a summary of the long range correction methods used in this thesis.

Reaction field method

In the reaction field method[4], the forces acting on a molecule are separated into two parts. One is a short-range inner sphere, where the electrostatics are calculated via the Coulombic potential in equation 2-5, and the other with an outer dielectric continuum region, with a dielectric constant of ε_s. The effective pair potential for the electrostatics interactions is[4],

$$U(r) = \frac{1}{4\pi\varepsilon_0} q_i q_j \left(\frac{1}{r_{ij}} + \frac{A r_{ij}^2}{2R_C^3} \right)$$
$$A = \frac{2(\varepsilon_s - 1)}{2\varepsilon_s + 1} \qquad [2\text{-}6]$$

where ε_s is the dielectric constant in the continuum, assumed to be equal to that of the fluid. A problem with the reaction field method is that when a molecule crosses from the inner sphere to the outer region, there is a jump in the energy due to direct interactions within the inner sphere and in the reaction field contributions; as a result the energy is not exactly conserved. The benefit of this method is that it has a low computational cost of order N where N is the number of molecules. The computational efficiently, of a simulation technique is often reported, as how the computational costs increase with an

increasing number of molecules in a simulation. Order N $O(N)$, means the computational cost increase directly proportional to the number of molecules. A direct calculation of all Coulombic interaction in the simulation cell would have a cost of $O(N^2)$, due to the sum,

$$U = \sum_{i}^{N} \sum_{j>i}^{N} U_{ij}(r_{ij}) \qquad [2\text{-}7]$$

Problems with the reaction field method are that it tends to over-emphasize the continuum nature of a polar fluid. Another is that the reaction field method requires that the outer dielectric of the continuum region be known. As the dielectric constant of the fluid varies dramatically with the addition of other solutes, especially with the addition of ions, the reaction field method can be inaccurate for solutions, therefore, the reaction field method is used only in the simulation of pure water simulations. In this thesis, it is used for the simulations of pure water described by the Gaussian charge polarizable model.

Ewald summation

The Ewald sum method[5] is a technique for efficiently summing the interactions between an ion and all of its periodic images.

$$U = \sum_{i}^{N} \sum_{j>i}^{N} \frac{q_i q_j}{|r + Lk|} \qquad [2\text{-}8]$$

when k are the lattice vectors, i.e. the number of repeated cells away from the central cell and L is the length of the simulation cell. As this result is conditionally convergent, the Ewald sum method decomposes equation 2-8 into two rapidly convergent sums. The first term is the point charges, with each charge surrounded by a Gaussian distribution of the same magnitude but with the opposite sign. This makes the first term short-ranged and

therefore converges quickly. The second term is a second set of Gaussian distributions, with an opposite sign to the first set. The Fourier transform of the second term convergences, and therefore the summation is done in reciprocal space. A self correction term is required for the self interaction between the point charge and its screening distribution. More details can be found in Allen and Tildesley[3].

The downside of the Ewald sum is that is scales to $O\left(N^{1.5}\right)$[6], resulting in a high computational cost for large systems. Particle-Particle-Particle Mesh (PPPM)[7] and Particle-Mesh Ewald (PME)[8] summation methods are similar to Ewald but use fast Fourier transforms and scales as $O\left(N \ln N\right)$[6] with minimal deviation in results[9]. The use of PPPM and PME are becoming increasingly common in packaged codes.

The fact that Ewald-summation-based methods duplicate the instantaneous dipole fluctuation instead of damping out in the infinite replicated system has been criticized, on the grounds that it tends to overemphasize the periodic nature of the fluid[3,10]. The Ewald sum is popular is package codes, and is used for the water simulations carried out in Car-Parrinello MD[11] and in LAMMPS[12], and for the BDT-MD simulations on a gold surface.

Wolf method

The Wolf method[13] is based on neutralizing the charge in the volume surrounding a charged atom. As a result, charges at long distance see a neutral volume, thus no interaction. It employs a damping factor α, to aid convergence, and a cutoff R_C. The Wolf method is a fast long-range method, which has the same accuracy as the Ewald sum method[14], but is of $O\left(N\right)$ and can be included in the short range interaction calculations if a tabulated force interaction technique is used. In this thesis, the Wolf method is used for

the calculations of the chemical potential for the solutes into water and aqueous systems. A description of the Wolf method, and the derivation of the equations for its application to water and solute forcefields, with and without Gaussian charges, is provided in Appendix A.

2.2 *Ab initio* methods

Background

Ab initio methods differ from classical simulations as they are based on fundamental quantum mechanics, as opposed to empirically defined functions for the atomic interactions. They require solving for the wavefunction, which contains all the information of the system. It represents the probability amplitude, where the square is the electron probability density, i.e. the probability of an electron being found at a given position. Integration of the system wavefunction over all space will result in the number of electrons contained in that space.

$$\Psi(x,y,z)^2 = \rho(x,y,z)$$
$$\int_{\Re^3} \Psi(x,y,z)^2 \cdot dV = n \qquad [2\text{-}9]$$

where Ψ is the wavefunction of a system, $\rho(x,y,z)$ is the probability density of finding an electron at a point and n is the number of electrons in the system. *Ab initio* calculations (time-independent) and simulations (time-dependent) are extremely useful since they can provide reliable predictions of properties and energies that are often too difficult and in some cases impossible to obtain experimentally.

Ab initio calculations allow for the computation of bonding energies and non-bonded molecular interactions, useful in the creation of classically based potentials. The

downside of *ab initio* calculations it that they require significantly more computational resources compared to classical simulations. As a result such simulations are limited to very small systems and very short simulation times if time-dependent calculations are desired. Below is a summary of *ab initio* methods.

Hartree -Fock and Post-Hartree-Fock methods

The Hartree-Fock method[15] (HF) is the central method for *ab initio* quantum chemistry calculations, where the wavefunction for an N-body system Ψ is assumed to approximate by a single Slate determinant, given by

$$\Psi(x_1, x_2, ..., x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) \cdots \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) \cdots \chi_N(x_2) \\ \vdots & \vdots \quad \vdots \\ \chi_1(x_N) & \chi_2(x_N) \cdots \chi_N(x_N) \end{vmatrix} \qquad [2\text{-}10]$$

where $\chi_i(x_j)$ is the molecular wavefunction i for particle j. If any of the wavefunctions are the same, the system wavefunction Ψ is zero, satisfying the Pauli exclusion principle, that no two electrons can be at the same position at the same point of time.

The greatest assumption with the Hartree-Fock method is that electron correlation is ignored. Electron correlation is the effect one electron has on the locations of other nearby electrons. An uncorrelated system is where the probability for finding a given electron is independent to the locations of other electrons. In reality, electrons are more closely correlated due to direct Coulombic repulsion of electrons. For example, around electron i, there would be a local region of space that other electrons would avoid. This space around an electron is called the Coulombic hole. In HF, electrons are treated in an averaged way, resulting in a particular electron interacting with an electron cloud,

excluding the local electron-electron interaction. A good review on electron correlation is presented by Knowles *et al.*[16].

Dispersion forces (or van der Waals forces) come from the electron correlation interactions. Thus, their correct calculation is extremely important for intermolecular interactions, especially in neutral molecules, where Coulombic interactions are not dominant. Post-Hartree-Fock methods were developed to improve on the Hartree-Fock, by adding the effect of electron correlations. The only post-Hartree-Fock method used in this thesis is Moller-Plesset perturbation method[17].

Moller-Plesset perturbation method starts with Hartree-Fock method and adds in excited electron states (electrons moved into high energy orbital). The addition of these excited state $\lambda^n \Psi^n$ are treated as a perturbations from the lowest energy state, where the full wavefunction can be shown as

$$\Psi = \Psi^0 + \lambda\Psi^1 + \lambda^2\Psi^2 + \ldots \qquad [2\text{-}11]$$

For a second order level theory, (MP2), doubly-excited Slater determinants are added, where the doubly excited configuration "interact with" the ground configuration (HF). There is no MP1 energy as there can not be a single-excited state.

$$E \approx E_{HF} + E_{MP2} \qquad [2\text{-}12]$$

With regard to high levels of perturbation theory, Leininger *et al.*[18], reported than the MP*n* energies and properties display rapid or slow convergence, monotonic or oscillatory decay, highly erratic or regular behavior, or early or late divergence, all depending on the chemical system or the choice of one-particle basis set. Helgaker *et al.*[19], stated that molecular properties calculated with MP3 and MP4 were not consistently better than those calculated using MP2. Bearing this in mind, and the additional computational cost

for high levels of theory, all calculations were conducted at the MP2 level. MP2 calculations were carried out for the determination of the van der Waals forces for the ion-solute forcefield parameterization in chapter IV.

Density functional theory

The Hartree-Fock method and perturbation theory are based on N-body wavefunctions (multiple electrons). Density functional theory (DFT) replace the N-body wavefunction $\Psi(r_1, r_2, ..., r_N)$, which is a function of 3N variables (x, y, z for each atomic wavefunction) with an electronic density $\rho(r)$, which is a function of 3 variables (x, y, z for the system). The energy levels of the system are the eigenvalues of the Schrodinger equation

$$H\Psi = E\Psi = \left[\sum_i^N -\frac{h^2}{2m}\nabla_i^2 + \sum_i^N V(\vec{r}_i) + \sum_i^N U(\vec{r}_i, \vec{r}_j)\right]\Psi \qquad [2\text{-}13]$$

The first term, is from the kinetic energy of the system. $V(\vec{r}_i)$ is static external potential operator which is dependent on each system, and $U(\vec{r}_i, \vec{r}_j)$ an operator for the electron-electron interactions. The kinetic energy and $U(\vec{r}_i, \vec{r}_j)$ are system-independent operators. It was proposed by Kohn and Sham[20] that

$$\rho(\vec{r}) = N\int d\vec{r}_1 ... \int d\vec{r}_N \Psi(r_1, r_2, ..., r_N)\Psi^*(r_1, r_2, ..., r_N) \qquad [2\text{-}14]$$

could be rearranged so that $\Psi = \Psi(\rho(\vec{r}))$. Optimizing $\rho(\vec{r})$ to minimize the energy, results in the ground state wavefunction. Therefore all of the properties of the ground state system can be calculated.

Exchange and correlation functionals

As an alternative to perturbation theory, for the electron corrections, DFT calculations makes used of functionals to account for the exchange energy and correlation energy. The exchange energy is due to the additional attraction and repulsion of electrons with differing angular momentum, called the electron spin. One could explain the exchange energy (albeit simplicity) with two electrons with the same angular momentum that are nearing each other. As they become nearer, they become closer to violating the Pauli exclusion principle, that no two electrons can have the same four quantum numbers at the same time, this resulting in repulsion between the two electrons. The space around an electron which repels electrons with similar angular momentum is called a Fermi hole. There is an attraction between two electrons with differing angular momentum, which is important for chemical bonding, as it allows both electrons to be localized in a intermolecular area, shielding the two positively charge nuclei from repulsion. The exchange effects are approximately 10 times greater than that of electron correlation effects. The correction E_{XC} due to exchange and correlation effect is calculated via

$$E_{XC} = \int f_{XC}\left(\rho(\vec{r}), \nabla\rho(\vec{r})\right)\rho(\vec{r})dr \qquad [2\text{-}15]$$

where f_{XC} is the exchange functional. Five different functionals are used in this proposal, LDA[21], BLYP[22,23], B3LYP[24], PBE0[25] and X3LYP[26]. Functionals are made with an exchange part and a correlation part, with one of the most common being the BLYP, made from the Becke's exchange[22] and Lee, Yang and Parr's correlation functional [23]. LDA (local density approximation) is the simplest approximation, as the electron exchange and correlation energy is only a function of the electron density at any point in

space ~$\rho(r)$. The generalized gradient approximation (GGA) uses both the electron density and the first differential of the electron density (gradient) that is $\rho(r)$ and $\nabla\rho(r)$.

In 1993, hybrid functionals became available and are made from mixing different functionals together. It was found that GGA functional over-binds and over-predicts bond lengths and Hartee-Fock (HF) under-binds and under-predicts bond lengths. If HF is mixed with the GGA functionals, it was hoped that it will obtain properties between the respective methods. The coefficients for a hydrid functional are empirically fitted to high level *ab initio* calculations, or highly accurate experimental data. Both B3LYP and X3LYP are fitted to experimental data sets, and PBE0 is determined by using perturbation-theory. The formula for the combination of their different functionals are

B3LYP: $0.2HF_x + 0.8S_x + 0.72B88_x + 0.19VWN_c + 0.81LYP_c$

PBE0: $0.25HF_x + 0.75S_x + 0.75PBE_x + PW91_x + PBE_c$

X3LYP: $0.218HF_x + 0.782S_x + 0.542B88_x + 0.167PW91_x + 0.129VWN_c + 0.871LYP_c$

where, S, HF, PW, PBE, B88, VWN AND LYP correspond to, Slater[27], Hartee-Fock[15], Perdew-Wang91[28], Perdew-Burke-Ernzerhof[29], Becke's exchange[22], Vosko-Wilk-Nusair 80[30] and Lee, Yang, Parr 88[23]

Multiple functionals have been developed and each functional is more accurate for different properties and different systems. The most recent functional, X3LYP, is best for heats of formation, electron affinity and van der Waals attraction. The hybrid B3LYP is better for the calculation of the electrostatics interactions, which are very important for forcefield development, and is close second to X3LPY for many properties. Comparison of the accuracy of the functionals can be found in the work of Xu and Goddard[26].

Basis sets

Basis sets are a collection of functions which define a space in which a molecular system is solved. They contain functions (basis functions) that explain the orbitals of the atom the particular basis sets was designed for. A basis set can be described by,

$$\phi_i^{BS} = \sum_{j=1}^{N} d_j \phi_j^F \qquad [2\text{-}16]$$

where ϕ_i^{BS} is the basis set of atom i, with is the sum of the basis functions ϕ_j^F. The basis functions are added according to a coefficient d_j. The d_j parameters are fitted to Slater atomic orbitals[27] which are not directly used due to computational difficulties.

Basis functions are mathematic expression used to describe atomic orbitals. Basis functions used in this thesis are Gaussian-type atomic functions, which have a form similar to,

$$\phi_j^F = A f(\vec{r}) \exp(-\alpha r^2) \qquad [2\text{-}17]$$

where A is a constant for the basis function, and $f(\vec{r})$ is a product of direction vectors. For example, for the 1s-orbital, in the x direction

$$\phi_j^F(r) = \left(\frac{2\alpha}{\pi}\right)^{\frac{3}{4}} x \exp(-\alpha r^2) \qquad [2\text{-}18]$$

where α is a constant parameter, x is the $f(r)$, r is the distance away from the atomic nucleus. A 'minimal' basis set contains only one basis function to describe each atomic orbital. For example, the carbon atom has 5 orbitals, namely 1s, 2s, $2p_x$, $2p_y$, $2p_z$ orbitals resulting in 5 basis functions for its 6 electrons. For better representation of the orbitals, more functions can be added to represent each orbital. Increasing the number of functions, increases the accuracy of the calculation but it also increases the computational

cost. By nomenclature convention, double zeta stands for 2 functions for each orbital and triple zeta for 3 functions, and so on.

As the valence (outer) orbitals have the greatest effect on the interatomic interactions, the core orbitals are often modeled with less basis functions compared to the valence electrons. This is called a split-valence basis set, and it can have a dramatic reduction of the computational cost of the calculation, with minimal penalty in accuracy. When there is a great number of electrons/orbitals in an atom, e.g. gold (79 electrons), a single function called an effective core potential (ECP), can approximate a number of the core electrons. For this work, the ECP for gold accounted for 60 electrons.

Extra basis functions are added to the basis set for molecular systems where polarization has a sizeable effect. For an s-orbital, 3 p-orbitals are added (p_x, p_y, p_z) and for a p-orbital, the 5 d-orbitals are added (d_{xy} +d_{yz} +d_{xz} +$d_{x^2-y^2}$+d_{z^2}). By nomenclature convention, an asterisk is added to the basis set's name (*) when extra polarization orbitals are adding to the basis set. If polarization is going to be added to hydrogen atoms, two asterisks are added.

For molecular systems where the electrons are likely to be distributed further away from the nucleus, diffuse functions are added to more accurately represent the tail portion of the basis sets, e.g. a negatively charge ion like Br^-, which has extra electron-electron repulsion. By nomenclature convention, the addition of diffuse functions are represented with a plus sign (+), again if diffuse functions are added to hydrogen atoms, two plusses are used.

To choose which basis set to use is a trade-off between computational time verses accuracy. A molecular dynamic simulation scales between $O(N)$ and $O(N^2)$, where N is

the number of molecules, whereas *ab initio* methods scale to the $O(N^3)$ for plane wave calculations, $O(N^4)$ for DFT and $O(N^5)$ for MP2 where N is the number of functions. So for extended basis sets, the computational cost is immense. For example, the 6-31G basis set for carbon contains 9 basis functions, 6-311++G(3df,3pd) basis set for carbon contains 45 basis functions, increasing the computations time up to 5^5 for MP2 calculations. Structure optimization runs, which are a series of single point energy calculations with small changes in the atomic location to calculate the lowest energy configuration, contain a high number of calculations. Often a small basis set (e.g. 6-31G) is used to get a closer approximation to the higher level structure at a reduced computational cost, before using a more detailed basis set for higher accuracy.

Basis set superposition error (BSSE)

When calculating the energy of two or more particles, a single species can use functions from the other nearby particles. This virtually enables the species to increase the size of the basis set (available functions). This artificially increased basis set is called a 'mixed basis set'. As the distance between the two species separate, this effect is reduced. The problem arises when the energy calculated from using a mixed basis set is compared with the energy of the isolated molecules using unmixed basis set. As a result, the total interaction energy between the two species are artificially lowered energy.

In order to account for this via the counterpoise approach by Boys and Bernardi[31], each configuration has to be recalculated twice using ghost orbitals for each of the species. Ghost orbitals are basis functions not accompanied by atoms, in this way there would be no effect of the neighboring nuclei or extra electrons on the species of interest, just the

available orbitals. The error is subtracted *a posteriori* from the full uncorrected energy, as shown below,

(uncorrected) $$E_{tot} = E_{ab} - E_a - E_b \quad [2\text{-}19]$$

(fully corrected) $$E_{tot} = E_{ab} - E_{a'b} - E_{ab'} \quad [2\text{-}20]$$

where E_{tot} is the total interaction energy, E_{ab} is the full calculation with both full a and b basis sets, $E_{a'b}$ is where molecule a, is using ghost orbitals from b. Not correcting for the BSSE, will result in stronger interaction energies, leading to the development of overly attractive force fields. Work from Kim *et al.*[32] stated that even though the full counterpoise method yields a pure interaction energy, it often over-corrects and employing a half BSSE correction yields closer results to experiment. The half BSSE correction has been used for the high-level theoretical studies of the water dimer by Schutz *et al.*[33]

Partial charges

Calculating the partial charges of a system is one of the more complex properties to achieve. The ability to calculate the partial charges on the atom contained in a molecule is required to calculate the magnitude of the different energy contributions to the total energy for a system. There are two main ways for doing this described below.

The first method, derived by Mulliken[34], is that the electron density is fitted to the wavefunction of the system, where

$$\int \rho(r) dV = \sum_{i}^{N} P_{ii} + \sum_{i}^{N} \sum_{j>i}^{N} 2P_{ij}S_{ij} = \sum_{i}^{N} p_i \quad [2\text{-}21]$$

$$p_i = P_{ii} + \sum_{j,j\neq i}^{N} P_{ij} S_{ij} \qquad [2\text{-}22]$$

$$q_i = Z_i - p_i \qquad [2\text{-}23]$$

Here $\rho(r)$, is the total electron density at location r in a system of volume V, P_{ii} is the electron density only surrounding nuclei, i, S_{ij} is the overlap integral, relating to the "sharing" of electrons between two nucleuses i, j. P_{ij} is the electron density surrounding both nucleuses i, j, Z_j is the defined atomic charge at the nuclei, often the atomic number as the positive charge from the protons in the atomic nucleus, p_i is the calculated electron population from Mulliken population analysis, in equation 2-21. The result is q_i, the partial charge of the atom. The problem with this method is that is it not always correct to divide the electron density equally between two different atomic sites. The atom with the greatest electronegativity draws the electron density closer to itself. Additionally with the Mulliken population analysis, an electron is always attributed (as least partially) to their host atom. This can be erroneous for orbitals larger than s-type because an electron can have a long separation between itself and the host nuclei. The Mulliken population analysis attributes electrons to the host atom when in reality; it will attain little effect from the distant electron.

The second way is using the electrostatic potential (ESP). The electrostatic potential is the energy possessed by a unit positive charge at given point r, caused by the charge of the nuclei and the electrons of a molecule,

$$V(r) = \sum_{i=1}^{N} \frac{Z_i e}{r - R_i} + \int - \frac{e\psi_i^* \psi_i}{r - r'} dr' \qquad [2\text{-}24]$$

where Z, is the atomic charge of the nucleus at position R_i. Partial charges are fitted at the nuclei to optimize the reproduction of the electrostatic potential surface. If the charges are going to be used for a computational model, the partial charges do not change as a function of position, i.e. as a flexible molecule rotates. In the restrictive ESP (RESP) method, constraints are added (in a logical manner) to reduce the number of variables.

In a simulation, the hydrogens on a methyl group (as an example) can freely rotate around the C-C bond, thus the charges according to *ab initio* calculations on the individual hydrogens will change. Leaving a fixed charge on an individual hydrogen would be erroneous, as it rotates to a new configuration. Setting the hydrogen atoms (on the same methyl group) to have the same charge would reduce this error, and the partial charge can be optimized for used with this constraint. The effect of intramolecular electrostatic interaction has to be accounted for with the use of polarizability in the classical forcefield.

Work done by Tang *et al.*[35], showed that partial charges calculated via ESP within DFT with a B3LYP functional on polar molecules, gave a better prediction of the isolated dipole moment than partial charges calculated via Mulliken population. Mulliken charges are sensitive to basis set size, and unphysical charges can be calculated when an extended basis is used[36]. One problem with ESP fitting is that if the charges are fitted to the electrostatic potential, calculated on a shell around the molecule, erroneous values will be produced if there is an atomic surface or buried atoms which are "sheltered" by closer atoms from the shell's surface. The *ab initio* electrostatic potential implicitly includes polarization. If partial charges are fitted to the electrostatic potential, there is an

induced dipole implicitly included, which must be extracted in order to calculate the true partial charge.

Car-Parrinello molecular dynamics

In 1985, Car and Parrinello[37] developed a method which coupled density functional theory with molecular dynamics. Because of the large size differences between the nuclei of an atom and surrounding electrons, the nuclei can move classically and the electrons can then respond essentially instantaneously, remaining in the ground energy state, using fictitious dynamics. The separation of the motion of the electrons and nuclei is known as the Born-Oppenheimer approximation[38].

The second computational time saver follows the same lines as the effective core potentials, namely that all the non-valence orbitals are approximated with one function called a pseudopotential. The valence electrons are calculated via planewave basis sets as opposed to a Gaussian based basis set, which is easier to calculate for systems with periodic boundary conditions. A planewave basis set has the form

$$\phi_i^{PW} = \sum_{j=1}^{N} d_j \exp\left[i(k+G)r\right] \qquad [2\text{-}25]$$

$$G = m\,2\pi \qquad [2\text{-}26]$$

where k is the location in Brillouin zone, (location in the central periodic box), G is the reciprocal lattice vector, m is the integer number of cell lengths away from the central cell. Planewave basis sets also have the added benefit of avoiding BSSE, and are independent of atomic positions.

Dynamic equations of motion are solved for the ions with the inter-ionic forces computed from the valence electron density, which is solved for at the first time step

using DFT, i.e. electronic minimization. For the additional steps, the electron density is solved by fictitious dynamics, using a fictitious mass for the electrons. If the timestep is small, the electron density remains on the ground state, and the electronic wavefunction does not have to be solved via computationally expensive electronic minimization for each timestep. The system in theory is adiabatic, but if the timestep is too large, there will be energy transfer between the electronic motion and the atomic motion, leading to erroneous simulations, as the electron density will leave the ground state. At each time step, one calculates the electron density and the resulting forces and the electrostatic interaction between the ions (nuclei). The nuclei move classically, using the forces computed from the electron density.

A downside of Car-Parrinello, shared with classical models, is that nuclei cannot undergo quantum phenomena such as proton tunneling. CPMD[11] has been described as a higher-level semi-empirical approach, due to its simplification of the quantum mechanics[39]. There has been criticism of the accuracy of Car-Parrinello for ionic simulations when compared with structural data[39]. In the case of water, a CPMD simulation consists of O and H ions with eight valence electrons per molecule. The other two electrons (1s orbital for oxygen), are taken into account via a pseudopotential. As polarization and other short range forces are taken into account automatically within the CPMD, this methodology provides useful data to compare to with the polarization included in polarizable classical models. *Ab initio* Car-Parrinello calculations on water and simple ions have been conducted, resulting in the structure of the first two solvation shells and insight into the nature of the water-ion interaction. Properties such as first water residence time and dipole moments have been computed.

References

[1] J. D. S. D. A. McQuarrie, *Statistical mechanics*. (University Science Books, Sausalito, 2000).

[2] D. A. McQuarrie, J. D. Simon, *Physical chemistry, A molecular approach*. (University Science Books, Sausalito, 1997).

[3] M. P. Allen, and D. J. Tildesley, *Computer Simulation of Liquids*. (Clarendon press, Oxford, 1987).

[4] L. Onsage, Journal of the American Chemistry Society **58**, 1486 (1936).

[5] P. P. Ewald, Ann. Phys. **64**, 253 (1921).

[6] D. R. Wheeler, J. Newman, Chemical Physical Letters **366**, 537 (2002).

[7] R. W. Hockney, *Computer simulation using particles*. (McGraw-Hill International Book Company, New York, 1981).

[8] D. Y. T. Darden, L. Pederson, Journal of Chemical Physics **98**, 10089 (1993).

[9] B. A. Luty, M. E. Davis, I. G. Tironi, W. F. Vangusteren, Mol Sim **14**, 11 (1994).

[10] J. M. G. Barthel, H. Krienk, W. Kunz, *Physical Chemistry of Electrolyte Solutions*. (Springer, 1998).

[11] Copyright IBM Corp 1990-2004 CPMD; 3.9.1 ed., (2004).

[12] S. J. Plimpton, J. Comp. Phys. **117**, 1 (1995).

[13] D. Wolf, P. Keblinski. S.R. Phillpot, J. Eggebrecht, Journal of Chemical Physics **110**, 8254 (1999).

[14] Christopher J. Fennell, J. Daniel Gezelter, Journal of Chemical Physics **124**, 234104 (2006).

[15] F. Z. V. Fock, Physik **61**, 126 (1930).

[16] P. Knowles, M. Schutz, H. Werner, in *Modern Methods and Algorithms of Quantum Chemistry*, edited by J. Grotendorst (Julich, 2000), Vol. 3, pp. 97.

[17] C. Moller, M. S. Plesset, Physical Review **46** (7), 618 (1934).

[18] M. L. Leininger, W. D. Allen, H. F. S. III, and C. D. Sherrill, The Journal of Chemical Physics **112** (21), 9213 (2000).

[19] T. Helgaker, P. Jorgensen, J. Olsen, *Molecular Electronic-Structure Theory*. (John Wiley & Sons Ltd, West Sussex, 2000).

[20] W. Kohn, L. J. Sham, Physical Review **140**, A1133 (1965).

[21] S. H. Vosko, L. Wilk, M. Nusair, Canadian Journal of Physics **58** (8), 1200 (1980).

[22] A. D. Becke, Physical Review A **38** (6), 3098 (1988).

[23] C. Lee, W. Yang, R. G. Parr, Physical Review B **37** (2), 785 (1988).

[24] A. D. Becke, Journal of Chemical Physics **98**, 1372 (1993).

[25] C. Adamo and V. Barone, Journal of Chemical Physics **110** (13), 6158 (1999).

[26] X. Xu and W. A. Goddard, Proceedings of the National Academy of Sciences of the United States of America **101** (9), 2673 (2004).

[27] J. C. Slater, Physical Review **36**, 57 (1930).

[28] Y. Wang and J. P. Perdew, Physical Review B **44** (24), 13298 (1991).

[29] J. P. Perdew, K. Burke, M. Ernzerhof, Physical Review Letters **77** (18), 3865 (1996).

[30] S. H. Vosko, L. Wilk, and M. Nusair, Canadian Journal of Physics **58** (8), 1200 (1980).

[31] S. F. Boys, F. Bernardi, Molecular Physics **19** (4), 553 (1970).

[32] K. S. Kim, P. Tarakeshwar, J. Y. Lee, Chemical Review **100**, 4145 (2000).

[33] M. Schutz, S. Brdarski, P. O. Widmark, R. Lindh, F. Karlstrom, Journal of Chemical Physics **107**, 4597 (1997).

[34] R. S. Mulliken, Journal of Chemical Physics **23**, 1833 (1955).

[35] P. Tang, I. Zubryzcki, Y. Xu, J. Comp. Chem **22** (4), 436 (2001).

[36] J. D. Thompson, J. D. Xidos,T. M. Sonbuchner, C. J. Cramer, D.G. Truhlar, Phys Chem Comm **5**, 117 (2002).

[37] R. Car, M. Parrinello, Physical Review Letters **55** (22), 2471 (1985).

[38] M. Born, R. Oppenheimer, Annalen der Physik **84**, 457 (1927).

[39] B. M. Rode, C. F. Schwenk, A Tongraar, Journal of Molecular Liquids **110**, 105 (2004).

CHAPTER III

THE IMPORTANCE OF POLARIZABILITY IN THE MODELING OF SOLUBILITY: QUANTIFYING THE EFFECT OF SOLUTE POLARIZABILITY ON THE SOLUBILITY OF SMALL NON-POLAR SOLUTES IN POPULAR MODELS OF WATER

3.1 Summary

In recent work by Paschek (D. Paschek JCP 2004, 6674) and others (see H. Docherty *et al*. JCP 2006 074510) it has been suggested that, when coupled with a simple Lennard-Jones (LJ) model for various small non-polar solute molecules, the most common models of water (e.g. SPC/E and TIP4P) fail to reproduce quantitatively the solubility of small non-polar solute molecules in water due in part to failing to account for polarization of the solute molecule. Given the importance of such systems as test-case prototype models of the solubility of proteins and biomolecules, in this work we investigate the impact of using a polarizable solute model with the SPC/E, TIP3P, TIP4P, TIP4P/Ew and TIP4P/2005 rigid water models. Specifically we consider Ne, Ar, Kr, Xe and methane as solutes. In all cases we observe that the use of a polarizable solute improves agreement between experiment and simulations, with the best agreement seen for the largest solutes, Kr, CH_4 and Xe and the modern reparameterizations of the TIP4P model, i.e. the TIP4P/Ew and TIP4P/2005 models. The research reported in this chapter has been published[1].

3.2 Introduction

Recently, methane-water interactions have been receiving increased attention as a result of an awareness of the increasing importance of methane hydrates, due both to their traditional importance in the petroleum industry, such as their role in the blockage of pipelines, as well as more recently for their importance as a source of energy[2]. For example, it is thought that the amount of methane stored in the form of hydrates, typically in deep seas, is many times greater than that currently available from traditional sources[3,4].

Another reason for the intense interest in an accurate model for the solubility of methane in water is that this system may be considered a prototype for non-polar solutes in water as well as for longer alkanes and, currently of great interest, proteins[5]. Thus, it is hoped that an improved understanding of the interactions of methane and water, and the causes of hydrophobic hydration, will lead to a greater understanding of protein folding and other interesting biological phenomena. It is because of this importance that the solubility of methane in water has been subject to intense study involving experiments[6], computational molecular simulations[7] (both classical[8] and quantum-mechanical[9]) as well as equations of state[10].

In terms of molecular simulations, the vast majority of investigations have made use of some form of rigid non-polarizable model of water, such as the highly popular SPC/E[11], TIP3P[12] and TIP4P[12-14] variants. Amongst the reasons for choosing these models are their relatively low computational costs, compared to flexible and polarizable water models, as well as the availability of complementary potentials when studying their interactions with other molecules. In addition to this, they have been, or are easily, incorporated into a number of freely available molecular simulation tools and potentials. With regard to simple non-polar solutes such as methane, a common choice is that of a

single Lennard-Jones interaction site[15], which has been successfully used to reproduce and study much of the pure system behavior of these molecules.

However, despite the success of these models in reproducing the thermodynamic properties of pure systems, when combined with rigid non-polarizable water models they fail to reproduce well several important properties, such as the solubility of methane in water. A contributing factor to this failure is that while in a pure non-polar solute phase (in which the local electric field is small) neglecting polarizability is a reasonable assumption, in an aqueous environment at atmospheric conditions, it is not[16]. Regardless of this weakness, due to their ease of use, they remain the model of choice for most researchers.

A major step forward in understanding the causes of the discrepancies in describing the solubility of non-polar molecules in water was recently made by Paschek[17] who compared the solubility of Ne, Ar, Xe and CH_4 in five popular water models (SPC, SPCE, TIP3P, TIP4P, TIP5P) and noted two key considerations. Firstly he notes that the deviation of the models from the experimental solubility over a temperature range of 275-375K corresponds to a similar trend in the deviation of the density of water, calculated with the models, over the same temperature range. Paschek[17] therefore suggests that, in order to obtain qualitative agreement with experimental values for the excess chemical potential in water at infinite dilution, it is necessary to use a water model capable of reproducing accurately the temperature dependence of the density of water. This hypothesis is supported by the work of Krouskop *et al.*[18] who calculated the solubility of Ne, Ar, Xe and CH_4 in TIP4P/Ew water, known to give a better description of the density of water then the models used by Paschek[17], and observed an improvement in the

calculated excess chemical potential. However, the agreement with experiment was still not qualitative, let alone quantitative.

An even stronger confirmation of Paschek's hypothesis is the work of Docherty *et al.*[19] who used the TIP4P/2005 model of Abascal and Vega[13], which was specifically parameterized to accurately reproduce the density of water over a wide range of temperatures, including the region of density maximum. In this work, they show that using the TIP4P/2005 model of water, with a simple Lennard-Jones model of methane (specifically that of Hirschfelder[15]) results in qualitative agreement with experimental solubility data.

With regard to obtaining quantitative agreement, Paschek[17] considered the effect of polarization of the solute on solubility and, focusing on xenon, showed that while polarizability may be neglected in pure systems, for these simple solutes in water polarizability plays an important, water-model-dependent role. In fact, Docherty *et al*.[19] used polarizability as justification for a deviation to the Lorentz-Berthelot combining rules in order to obtain quantitative agreement.

In the rest of this work we consider the effect of solute polarizability on the solubility of small solutes in SPC/E, TIP3P, TIP4P, TIP4P/Ew and TIP4P/2005 water at infinite dilution. Specifically, we examine the effect of using a point polarizability site on the solute as a method of obtaining quantitatively correct results for the chemical potential of methane and noble gases solvated in water.

3.3 Computational method

Water simulations

In this work, we have calculated the excess chemical potential of the solutes in the water models in three stages. In the first stage, for each model and temperature, we start with a random configuration (translational and orientational) of 256 water molecules in a cubic simulation box at the density predicted by the water model of interest for the conditions being considered (i.e. 1 Bar and temperatures of 275, 300, 325, 350 and 375K) obtained from previously published simulation data. An equilibrated system was then obtained by performing 200,000 0.5fs timesteps (100ps) in a canonical ensemble (NVT) molecular dynamics simulation. As a test of this equilibration technique, we have also considered a 350ps equilibration period and note that this leads to values of the chemical potential well within the error of our simulations.

The second stage of our calculations involved the creation of 50,000 sample configurations for each water model and temperature at 1atm. Starting with the equilibrated configurations from the first stage, we performed $5x10^6$ timesteps of 0.8fs (4ns total), again in a cubic canonical ensemble, storing a sample configuration every 100 timesteps for a total of 50,000 configurations. Here we note that in obtaining these configurations for the pure water systems (stages 1 and 2), we have utilized the LAMMPS[20] computational software to perform molecular dynamics simulations. In all cases (i.e. for all models, both water and solutes), the cutoff in the Lennard-Jones interactions was 9Å with standard tail corrections applied. In addition to this, long-ranged Coulombic interactions were accounted for using the PPPM[21] (particle-particle, particle-mesh) technique. It is worth noting that none of the models used in this work were parameterized using the PPPM method, rather they were parameterized using a standard

Ewald summation[22]. However, the increasingly popular PPPM technique has been used in numerous studies involving these models, with the authors noting results in agreement with the standard Ewald summation technique[23,24]. Also, while here we consider only small non-polar solutes, the goal of this line of research, in common with other workers, is the accurate modeling of the solvation of larger industrially and biologically important molecules. The explicit simulation of such systems is inherently more computationally expensive, requiring the use of many times more water molecules than considered in this work, motivating the use of methods such as the PPPM technique. For this, and other reasons, the PPPM methods is incorporated into many of the major molecular dynamics simulations software, which are becoming the tool of choice for many researchers. Thus we consider our choice of method, when dealing with long-range charge interactions, to be consistent with the real-world uses of the models presented here and thus to be a fair and useful comparison.

The third and final stage was the calculation of the chemical potential of the solute via the Widom insertion technique[25] using in-house code (i.e. LAMMPS was not utilized for this stage of the calculations). As noted below, in section 3.4 and Table 3-4, this code has been validated by reproducing literature values for non-polarizable solutes in various water models. In the Widom method, the solute is randomly inserted into a known water configuration. The excess chemical potential is related to the change in the internal energy of the system caused by the addition of a solute molecule by,

$$\mu_{ex} = -kT \ln\left\langle \exp\left(\frac{-U_{Solute}}{kT} \right) \right\rangle \qquad [3\text{-}1]$$

where μ_{ex} is the excess chemical potential, k is the Boltzmann's constant, T is the absolute temperature, and U_{Solute} is the contribution to the system energy due to solute – water interactions

$$U_{Solute} = U_{Solute}^{vdW} + U_{Solute}^{Pol} \quad [3\text{-}2]$$

U_{Solute}^{vdW} is the contribution to the system energy due to solute – water short range interactions, including standard tail corrections, and U_{Solute}^{Pol} is the contribution from the polarization of the solute due to the surrounding electric field. Here we have used 40,000 attempted insertions for each of the 50,000 water configurations, resulting in 2 x 10^8 insertions for every solute, in all the water models and at every temperature considered here. Using this method, we estimate an accuracy in our calculated excess chemical potentials ranging from 0.1 kJ/mol for neon to 0.25 kJ/mol for xenon. This is consistent with the work of others using similar techniques[17,19].

Water potentials

We test five different water models, specifically the SPC/E[11], TIP3P[12], TIP4P[12], TIP4P/2005[13] and TIP4P/Ew[14] water models. These are all popular rigid non-polarizable water models which have been used extensively. Here we note that each of these models consist of a single Lennard-Jones site located at the centre of the oxygen together with a number of point charges to account for the charged nature of the hydrogen and oxygen atoms. The parameters for these models are shown in Table 1. In addition, due to its importance in the calculation of the chemical potential, the difference between the density of water, as predicted by each of these models, and the experimental values at the conditions considered in this work are shown in Table 3-2.

Solute potentials

For each solute molecule we have considered two models. In the first model, the molecules are described by a single Lennard-Jones site. In the second model, in addition to the Lennard-Jones site from model 1, the molecules have a polarizable site located at their centre of mass. The magnitude of the polarizability is taken from experimental values[26,27] and is shown, together with the Lennard-Jones parameters, in Table 3-3.

The contribution to the system energy due to the polarizability of the solute molecule, U_{Solute}^{Pol} is given by,

$$U_{Solute}^{Pol} = -\frac{1}{2}\alpha\left|\left(\vec{E}\right)^2\right| \qquad [3\text{-}3]$$

where $\vec{E}$ is the electric field at the polarization site and α is the polarizability of solute. The electric field at a given point in a system may be calculated by considering an imaginary point charge at the same location and calculating the force on the charge. In this case the field is given by the equation

$$\vec{E} = \frac{\vec{F}}{q_i} \qquad [3\text{-}4]$$

where $\vec{F}$ is the electric force and q_i is the charge on the fictitious point charge. For a system of charges, interacting via the Coulombic potential, the force on a charge q_i due to all the other charges q_j is given by

$$\vec{F}_i = \sum_j -\frac{q_i q_j}{r_{ij}^3}\vec{r}_{ij} \qquad [3\text{-}5]$$

Note that the Coulombic potential is a conditionally convergent sum and, as such, when performing simulations special care has to be taken to ensure these forces are

accurately represented. In this work, we have accomplished this by means of the Wolf method[28], which may be considered as a damped, truncated and shifted potential[29], and as such requires minimal additional computational time. For a review of this method, as well as evidence that its accuracy is similar to that of the standard Ewald sum technique, see the work of Fennell and Gezelter[29] and Avendaño and Gil-Villegas[30]. A summary of the equations and their derivations is given in Appendix A.

In the Wolf method, the force on a charged site may be written as equation 3-6a below, and thus the field at the point of polarizability by equation 3-6b below. Notice that the field in equation 3-6b does not depend on the fictitious charge of the solute's polarizable site.

$$\vec{F} = \sum_j q_{solute}\, q_j \left[\left(\frac{erfc(\beta r)}{r^2} + \frac{2\beta \exp\left(-\beta^2 r^2\right)}{r\sqrt{\pi}} \right) - \left(\frac{erfc(\beta R_C)}{{R_C}^2} + \frac{2\beta \exp\left(-\beta^2 {R_C}^2\right)}{R_C\sqrt{\pi}} \right) \right], \text{ where } r \le R_C \qquad [3\text{-}6a]$$

$$\vec{E} = \sum_j q_j \left[\left(\frac{erfc(\beta r)}{r^2} + \frac{2\beta \exp\left(-\beta^2 r^2\right)}{r\sqrt{\pi}} \right) - \left(\frac{erfc(\beta R_C)}{{R_C}^2} + \frac{2\beta \exp\left(-\beta^2 {R_C}^2\right)}{R_C\sqrt{\pi}} \right) \right], \text{ where } r \le R_C \qquad [3\text{-}6b]$$

where R_C is the truncation cut-off, and β is a damping parameter. Following the work of Demontis *et al.*[31] and Avendaño and Gil-Villegas[30], we have chosen R_C to be half of the simulation box length, and a corresponding β value of $2/R_C$.

3.4 Results

Values for the chemical potential of Ne, Ar, Kr, Xe and CH_4 are shown in Figures 3-1 to 3-5. In each case, (a) is for a non-polarizable solute and, (b) a polarizable solute. Firstly, as in the work of Paschek[17] and Docherty *et al.*[19], for the models used in this paper, we observe that the ability of the water models to capture qualitatively the effect of temperature on solubility is related to the ability of the model to reproduce the density of water over the same temperature range (comparison of Table 3-2 to Figures 3-1 to 3-5). However, what separates this work from that of Paschek[17] is that, while he considered a polarizable model of xenon noting the magnitude of the change in the solubility for each water model, we answer the question, "is accounting for density and polarizability sufficient to successfully predict the solubility of simple non-polar solutes in common rigid non-polar water models?". To this end we note that, in all cases, inclusion of the polarizability reduces the chemical potential, as expected. Given that most of the water models tend to over-predict the chemical potential of theses solutes at ambient conditions, the inclusion of polarizability tends to lead to improved agreement with experiment.

For the smaller and less polarizable molecules (Ne, Ar), the effect of a polarizable solute is relatively small (*Figures 3-1 and 3-2*) in comparison to the effect seen for the larger solutes (Kr, CH_4, and Xe), which are quite dramatic (*Figure 3-3 to 3-5*). Starting with the effect of polarizability on qualitative agreement with experimental values, in Tables 3-4 and 3-5, we present the ranges of the deviation in solubility for each solute and water model in a format similar to that of Krouskop *et al.*[18]. In Table 3-4, the values correspond to the case of a non-polarizable solute model and are in agreement with those published by Krouskop *et al.*[18].

We note the improvement of the recent reparameterizations of the TIP4P model compared to the SPC/E, TIP3P and TIP4P models. We also note, with caution due to statistical uncertainties in simulations, that in all cases the TIP4P/2005 model provides the best qualitative agreement. Here we wish to highlight that this is in perfect agreement with the hypothesis and work of Paschek[17], as well as the work of Krouskop *et al.*[18]. and Docherty *et al.*[19], and confirms the relationship between density prediction and solubility since the TIP4P/2005 model provides a better, albeit only slightly, description of the density of water over the temperature range considered here. In other words, it is logical that an improvement in the density leads to an improvement in solubility.

In Table 3-5, the values correspond to the case of a polarizable solute model. Interestingly, rather that improving qualitative agreement, as one might have hoped, the inclusion of polarizability is detrimental. Thus we conclude that, if a particular model is incapable of capturing the temperature dependence of the chemical potential, the use of an explicitly polarizable solute molecule will not improve the qualitative agreement. This is particularly true of the TIP3P, TIP4P and SPC/E models, which show the greatest deviation from the experimental trend.

Moving our focus to the consideration of quantitative agreement with experiment, we observe that inclusion of polarizability improves all the water models considered here. Of particular interest are the TIP4P/Ew and TIP4P/2005 models which show similar improvements and, once again, out perform the SPC/E, TIP3P and TIP4P models. In fact, for the larger solutes considered here (i.e. Kr, CH_4 and Xe) these models approach quantitative agreement with experiment when polarizability is included, although at the cost of reduced qualitative agreement.

It is worth noting that while Docherty *et al*[19]. achieved a similar level of accuracy for methane, it required a computationally expensive optimization of cross-interaction parameters to match experimental data. Here we approach quantitative agreement in a purely predictive manner. It is interesting to note that despite TIP3P's extensive use in the study of biomolecular and protein systems, in which hydrophobic hydration in considered to play a very important role, the TIP3P model in unable to correctly predict the solubility of methane. However, it maybe of interest to many that when polarizability is included agreement with experiment is seen at 300K, albeit not at the other temperatures considered here.

Table 3-1: Rigid water model structure, electrostatic charge values and Lennard-Jones potential parameters for the different water models

	SPC/E	TIP3P	TIP4P	TIP4P/Ew	TIP4P/2005
OH (Å)	1.000	0.9572	0.9572	0.9572	0.9572
HOH (deg)	109.47	104.52	104.52	104.52	104.52
OM (Å)	0.000	0.000	0.1500	0.1250	0.1546
σ (Å)	3.1656	3.1506	3.1537	3.1644	3.1589
ε (K)	78.181	76.546	77.941	81.910	93.240
q_h (e)	0.4238	0.4170	0.5200	0.52422	0.5564
q_o (e)	-0.8476	-0.8340	-1.0400	-1.04844	-1.1128

Table 3-2: Deviation from experimental density in g/cm^3 at atmospheric pressure for the SPC/E, TIP3P, TIP4P, TIP4P/Ew, TIP4P/2005.

	275K	300K	325K	350K	375K
SPC/E	0.0092	0.0017	-0.0038	-0.0097	-0.0144
TIP3P	0.0049	-0.0124	-0.0267	-0.0399	-0.0531
TIP4P	0.0055	-0.0030	-0.0106	-0.0194	-0.0286
TIP4P/Ew	-0.0003	-0.0018	-0.004	-0.0067	-0.0106
TIP4P/2005	-0.0002	0.0000	-0.006	-0.0027	-0.0036

Table 3-3: Solute Lennard-Jones potential parameters and polarizability (α), σ and ε from reference [17], and α from references [26,27]

	σ [Å]	ε [K]	α [Å^3]
Neon	3.035	18.6	0.40
Argon	3.415	125	1.64
Krypton	3.675	169	2.48
Xenon	3.975	214.7	4.11
Methane	3.730	147.5	2.56

Table 3-4: Ranges of the deviation from experiment for the predicted solubility of non-polarizable solute molecules (kJ/mol)

	SPC/E	TIP3P	TIP4P	TIP4P/EW	TIP4P/2005
Neon	1.17	2.46	1.59	0.74	0.52
Argon	1.30	2.80	1.81	0.79	0.51
Krypton	1.40	3.02	1.99	0.89	0.54
Xenon	1.00	2.89	1.75	0.64	0.50
Methane	1.26	2.92	1.87	0.74	0.39

Table 3-5: Ranges of the deviation from experiment for the predicted solubility of polarizable solute molecules (kJ/mol)

	SPC/E	TIP3P	TIP4P	TIP4P/EW	TIP4P/2005
Neon	1.29	2.55	1.70	0.87	0.62
Argon	1.60	3.05	2.11	1.12	0.78
Krypton	1.79	3.33	2.36	1.30	0.87
Xenon	1.54	3.31	2.25	1.20	0.55
Methane	1.65	3.24	2.25	1.15	0.73

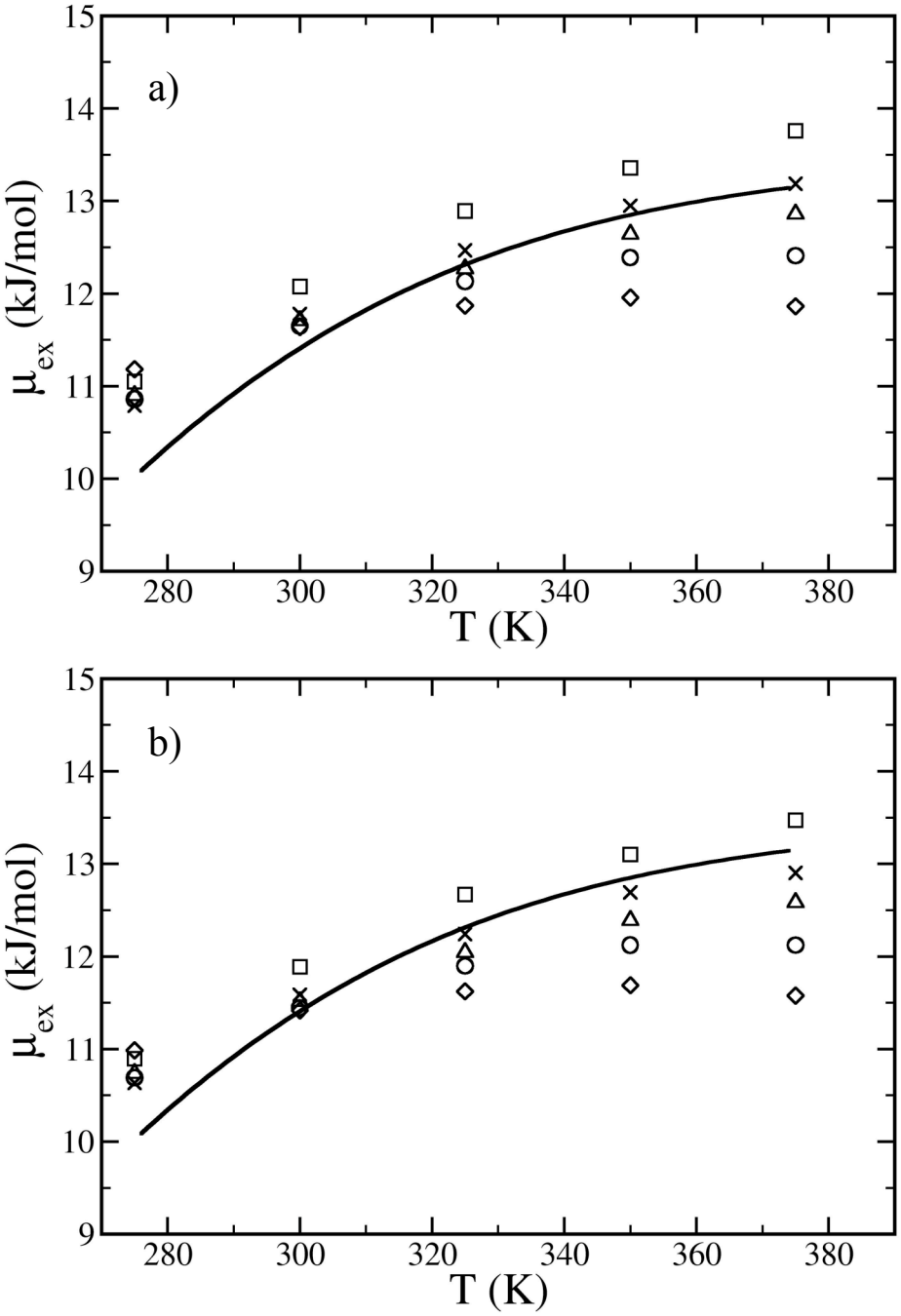


Figure 3-1: *Temperature dependence of the excess chemical potential of neon in water at infinite dilution for (a) A non-polarizable neon model and (b) A polarizable neon model. In both cases the solid line corresponds to experimental values*[17,32] *and the symbols to simulated values. Specifically, diamonds TIP3P, circles TIP4P, triangles SPC/E, crosses TIP4P/Ew and squares TIP4P/2005.*

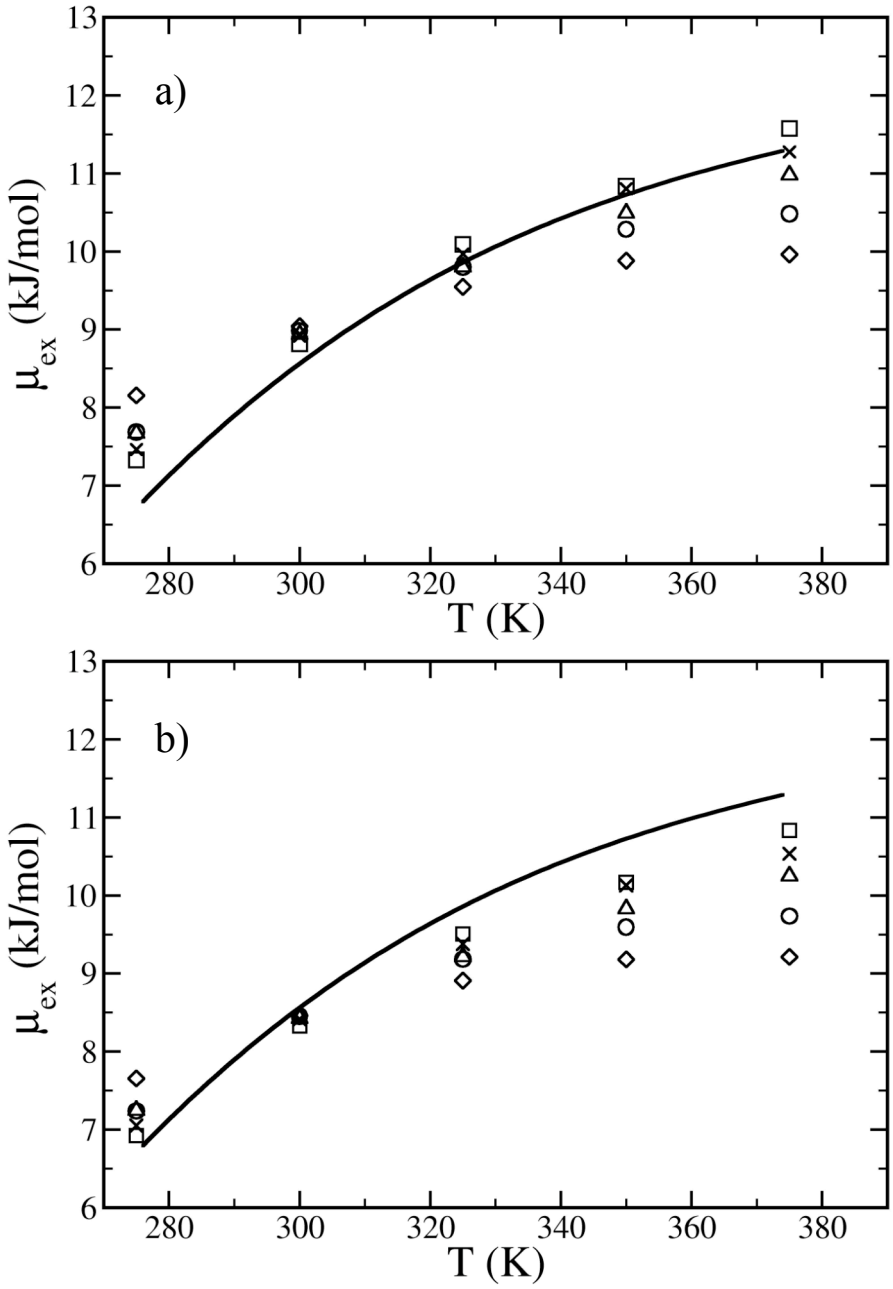


Figure 3-2: *Temperature dependence of the excess chemical potential of argon in water at infinite dilution for (a) A non-polarizable argon model and (b) A polarizable argon model. For explanation of symbols, see Figure 3-1.*

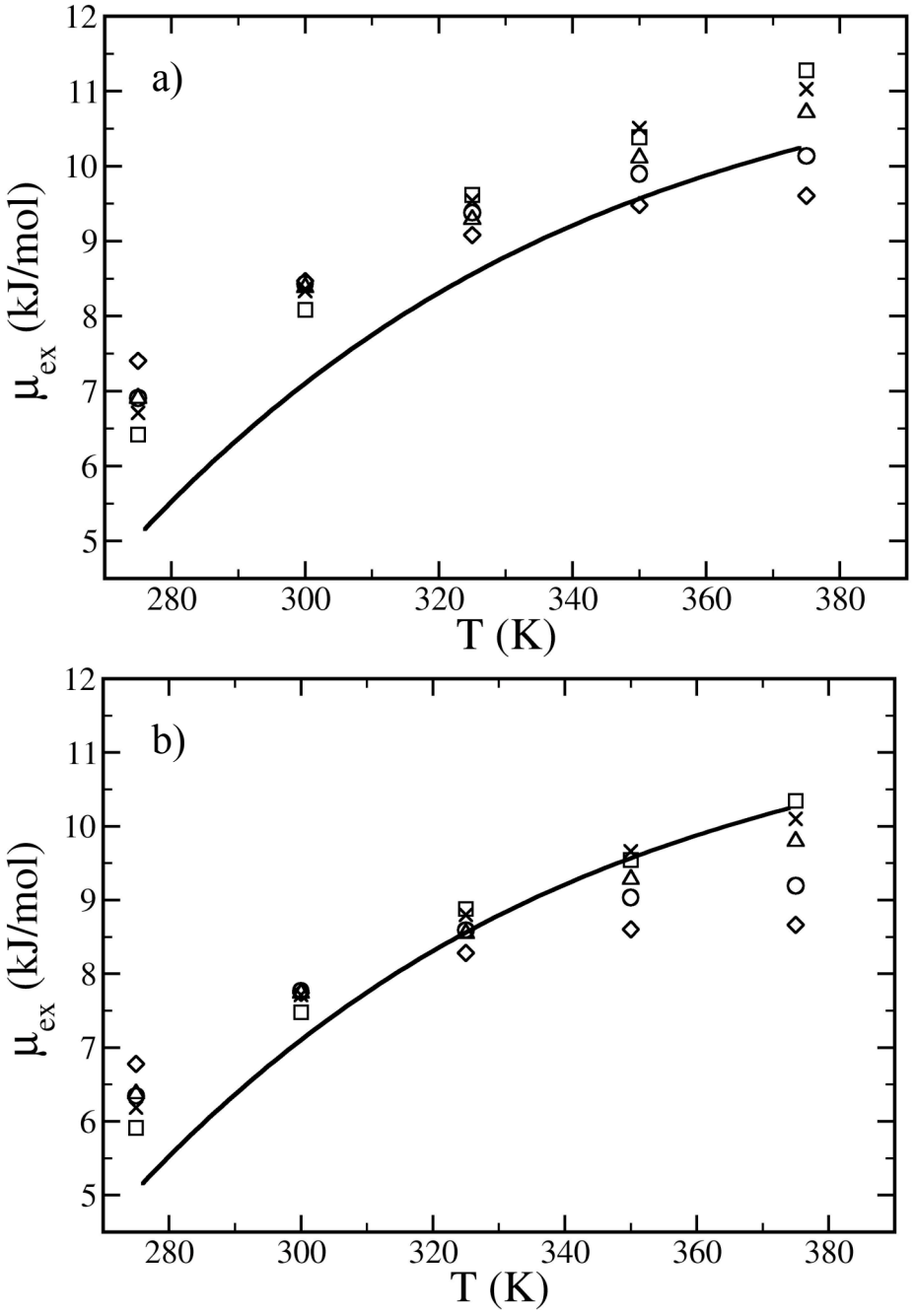


Figure 3-3: *Temperature dependence of the excess chemical potential of krypton in water at infinite dilution for (a) A non-polarizable krypton model and (b) A polarizable krypton model. For explanation of symbols, see Figure 3-1.*

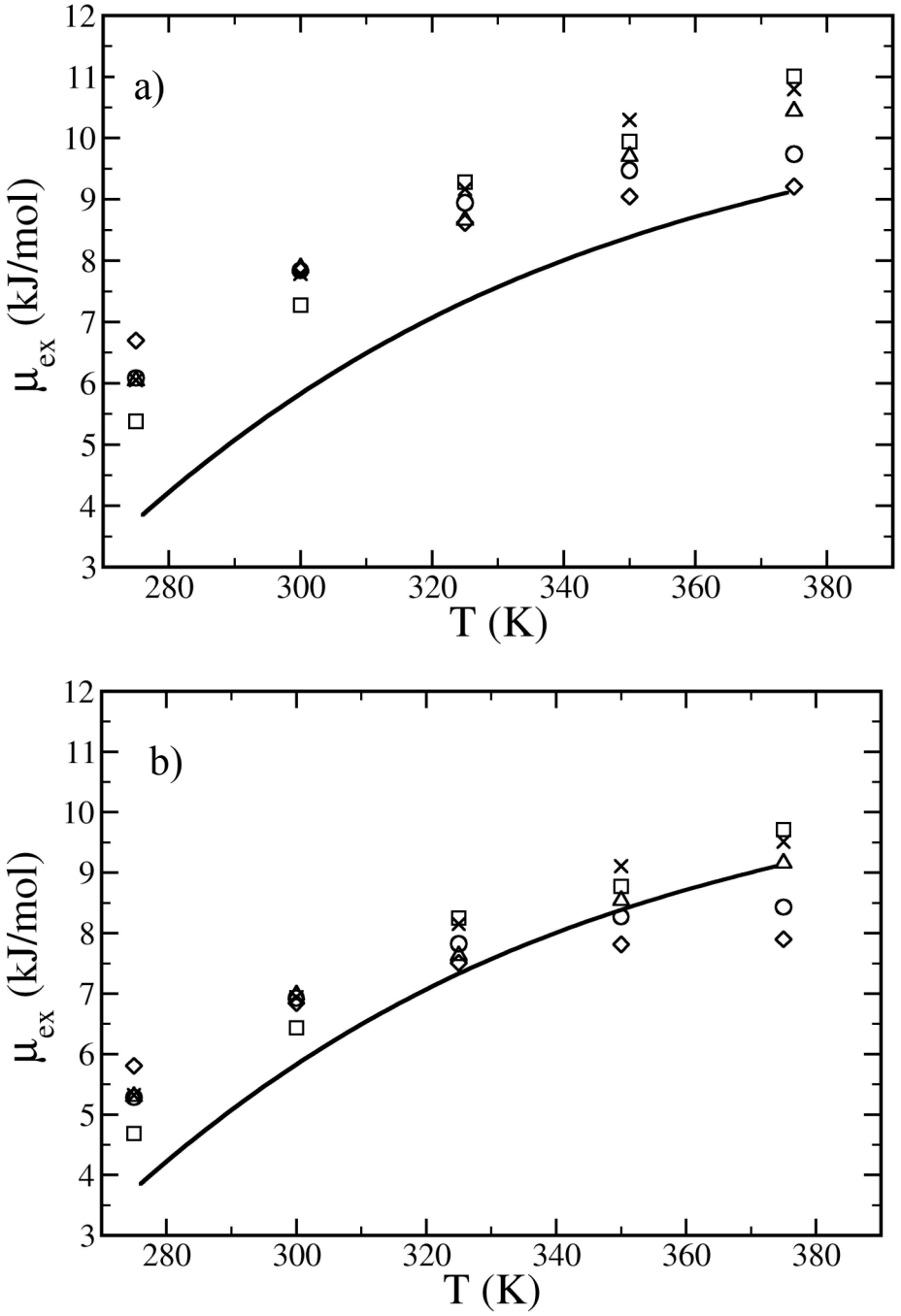


Figure 3-4: *Temperature dependence of the excess chemical potential of xenon in water at infinite dilution for (a) A non-polarizable xenon model and (b) A polarizable Xenon model. For explanation of symbols, see Figure 3-1.*

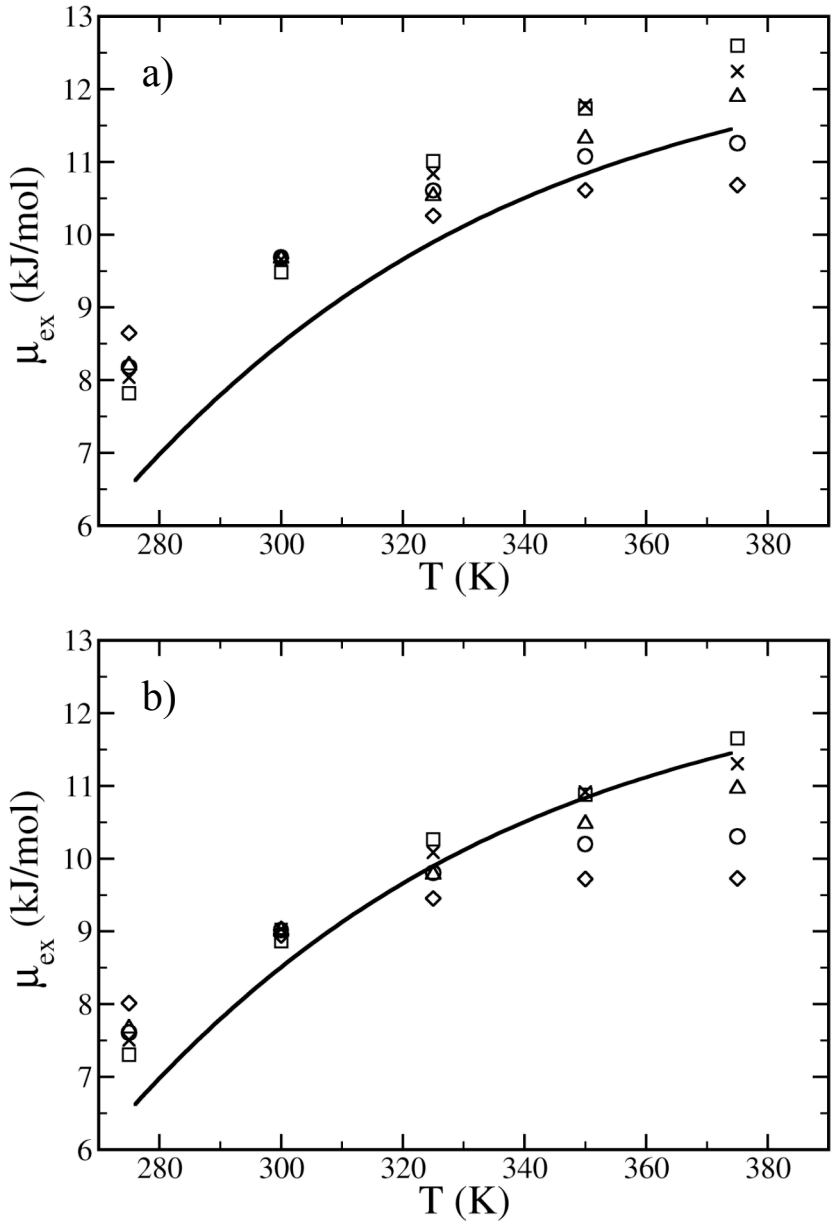


Figure 3-5: *Temperature dependence of the excess chemical potential of methane in water at infinite dilution for (a) A non-polarizable methane model and (b) A polarizable Methane model. For explanation of symbols, see Figure 3-1.*

3.5 Conclusions

Common rigid non-polarizable models of water coupled with single non-polarizable Lennard-Jones site models are unable to quantitatively reproduce the temperature dependence of the solubility of simple non-polar solutes and, with the exception of TIP4P/2005, these models are also unable to capture the relationship qualitatively.

At ambient conditions, these models tend to under-predict the solubility of these solutes. Accounting for the polarizability of the solute molecule increases the predicted solubility of these simple non-polar solutes, improving the agreement with experiment. However, its effect on the temperature dependence of solubility is detrimental and, as such, the SPC/E, TIP3P and TIP4P, models still do not achieve qualitative or quantative agreement. With regard to the TIP4P/Ew and TIP4P/2005 models, inclusion of polarizability leads to equally improved quantitative agreement. However TIP4P/2005, which provides a better description of the density of water, provides a better qualitative description of the solubility for both the non-polarizable and polarizable solute models (Tables 3-4 and 3-5) in agreement with the hypothesis of Paschek[17] and the work of Krouskop *et al.*[18] and Docherty *et al.*[19]

Thus we conclude that with the use of an explicitly polarizable solute, rigid non-polar models of water, such as TIP4P/2005, approach quantitative and qualitative prediction of the solubility of simple non-polar solutes. This means there is no need to "optimize" the mixing rules to experimental data and, one would hope, such a polarizable solute model is better able to describe any density dependence and effect of concentration, i.e, addition of other co-solutes, in particular, charged species.

Finally, this work highlights the importance of accounting for polarizability explicitly, especially in the development of transferable potentials, raising the question of how a polarizable water potential would perform. Attempting to answer this question will be the focus of future work.

References

[1] P. J. Dyer, H. Docherty, and P. T. Cummings, Journal of chemical physics **129**, 024508 (2008).

[2] E. D. Sloan, Nature **426**, 353 (2003).

[3] I. Lerche, E. Bagirov, Energy Exploration and Exploitation **22**, 3 (2004).

[4] T. Appenzeller, Science **252**, 1990 (1991).

[5] S. Harrison, S. Swaminathan, D. Beveridge, R. Ditchfield, Int J Quantum Chem **14**, 319 (1978).

[6] T. D. O'Sullivan, N. O. Smith, J. Phys. Chem **74**, 1460 (1970).

[7] D. Chandler, Nature **437**, 640 (2005).

[8] J. Errington, G. Boulougouris, I. Economou, A. Panagiotopoulos, D. Theodorou, J. Phys. Chem. B **102**, 8865 (1998).

[9] J. Li, R. Car, C. Tang, N. Wingreen, PNAS **104**, 2626 (2007).

[10] S. Dec, K. Bowler, L. Stadterman, C. Koh, E. Sloan, J. Amer. Chem. Soc **128**, 414 (2006).

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

[12] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L, Klein, Journal of Chemical Physics **79** (2), 926 (1983).

[13] J. L. F. Abascal, C. Vega, Journal of Chemical Physics **123** (234505) (2005).

[14] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, Journal of Chemical Physics **120** (20), 9665 (2004).

[15] J. O. Hirschfelder, C. F. Curtiss, R. B. Bird, *Molecular Theory of Gases and Liquids*. (John Wiley, New York, 1954).

[16] B. Guillot, Y. Guissani, J. Chem. Phys **99** (10), 8075 (1993).

[17] D. Paschek, J. Chem. Phys **120** (14), 6674 (2004).

[18] P. E. Krouskop, J. D. Madura, D. Paschek, A. Krukau, J. Chem. Phys **124**, 016102 (2006).

[19] H. Docherty, A. Galindo, C. Vega, E. Sanz, J. Chem. Phys **125**, 074510 (2006).

[20] S. J. Plimpton, J. Comp. Phys. **117**, 1 (1995).

[21] R. W. Hockney, *Computer simulation using particles*. (McGraw-Hill International Book Company, New York, 1981).

[22] P. P. Ewald, Ann. Phys. **64**, 253 (1921).

[23] B. A. Luty, M. E. Davis, I. G. Tironi, W. F. Vangusteren, Mol Sim **14**, 11 (1994).

[24] A. E. Ismail, G. S. Grest, M. J. Stevens, J. Chem. Phys **125**, 014702 (2006).

[25] B. Widom, J. Chem. Phys **39**, 2808 (1963).

[26] A. T. Wong, G. B. Bacskay, N. S. Hush, Mol Phys **74** (5), 1037 (1991).

[27] D. R. Lide, *CRC Handbook of Chemistry and Physics*, 88 ed. (CRC Press, 2007).

[28] D. Wolf, P. Keblinski. S.R. Phillpot, J. Eggebrecht, Journal of Chemical Physics **110**, 8254 (1999).

[29] C. J. Fennell, J. D. Gezelter, J. Chem. Phys **124**, 234104 (2006).

[30] C. Avendano, A. Gil-Villegas, Mol Phys **104** (9), 1475 (2006).

[31] P. Demontis, S. Spanu, G. B. Suffritti, Journal of Chemical Physics **114** (18), 7980 (2001).

[32] R. F. Prini, R. Crovetto, J. Phys. Chem Ref. Data **18**, 1231 (1989).

CHAPTER IV

THE IMPORTANCE OF POLARIZABILITY IN THE MODELING OF SOLUBILITY: QUANTIFYING THE EFFECT OF CHARGED CO-SOLUTES ON THE SOLUBILITY OF SMALL NON-POLAR SOLUTES

4.1 Summary

In Chapter III, we demonstrated that by accounting for the polarizability of small non-polar solutes, in an explicit manner it is possible to approach quantitative agreement with experimental values of the excess chemical potential of the molecules in pure water. Here we have continued this line of research by considering the effects of charged co-solutes on the explicit polarizability model for a variety of small non-polar solutes, development in the previous chapter. In this chapter, we calculate the excess chemical potential, which is related to the solubility on the non-polar solute in the solution, and how the solubility of the hydrophobic solutes vary with changing ion concentration, i.e. salting-out/salting-in effects. We have used the Setchenow parameter, as a measure of the salting-out effect on the solutions. Given the importance of the small solutes as test-case prototype models for the solubility and salting-out effects on proteins and biomolecules, in this work we investigate the impact of using a polarizable solute model with the TIP3P, TIP4P/Ew and TIP4P/2005 rigid non-polarizable water models along with non-polarizable point charge ion models. Specifically we consider Ne, Ar, Kr, Xe and methane as solutes, and Na and Cl as ions.

We have found that the addition of explicit polarizability to the solutes model decreases the excess chemical potential, but it failed to obtain quantative agreement with

experimental values. It was also found that the explicit polarization had minimal effect on the Setchenow parameter and in turn, minimal effect on the salting-out effect. Instead, we have shown that the over-prediction of the excess chemical potential was due to on over repulsive ion-solute interaction potential.

4.2 Introduction

Following on from the work in Chapter III, where we have demonstrated that explicitly accounting for the polarizability of small non-polar solutes, it is possible to approach quantative agreement with experimental values of the excess chemical potential for the solute in pure water for a range of temperatures. In this work we consider the related question of whether or not the addition of explicit polarizability on the solute model is sufficient to account for the qualitative and quantitative disagreement between experiment and simulation values for the modeling of the excess chemical potential for the solute in solutions containing charged co-solvents.

Understanding the solubility of hydrophobic compounds in aqueous solutions is important to traditional engineering situations such as the formation of hydrates in oil pipelines[1,2] and, of the denaturing of proteins[3-5]. The latter is currently the focus of intense research. The effect of the addition of charged species into aqueous solution containing hydrophobic solutions is well known, but mechanism is still under debate. The decrease in the solubility of hydrophobic compounds due to the addition of salt is known as salting-out. Alternatively, if the solubility of the hydrophobic compounds increases it is known as salting-in. Applications of salting-out effects include the separation of

proteins, based on the principle that hydrophobic proteins are less soluble at higher salt concentration. Thus the ability to reproduce the correct behavior of the electrolytic solution is of importance to correctly understand the salting-out mechanism in biological systems.

The salting effect (i.e. salting-out and salting-in) follows what is known as a Hofmeister series[6]. The Hofmeister series was developed from the differing effect of various ions on the solubility and the stability of proteins[7,8]. The basis of the Hofmeister effect has been long thought to be due in part to the effect of the ions on the surround water structure[9,10], i.e. the chaotropes and kosmotropes nature of the ions, but recently, this explanation has been challenged. Zhang and Cremer[11], concluded that changes in the bulk water structure and density will not explain specific ion effects seen in the Hofmeister series. Their reasoning is based on there being a negligible change in the water structure around an ion after the second solvation shell, as revealed in the work of Omta *et al.*[12]. As a result, they postulate that the Hofmeister effects are also partly caused by the direct interactions between the solute and the ions. We are attempting to clarity these theories, by also looking the effect of different regions of the aqueous solution, and its individual contribution to the excess chemical potential. For more details on the mechanisms on the salting-in and salting out, see chapter I.

In terms of modeling the salting-out of methane and other similar small solutes, Docherty *et al.*[13] using the water model TIP4P/2005 with a NaCl model together with a methane-water potential developed to accurately reproduce experimental excess chemical potential in bulk water which implicitly including the effect polarization into the CH_4-

H_2O potential showed that the salting-out effect can be rationalized in terms of a single parameter, the packing fraction, highlighting the role of volume exclusion. They proceed to show that even a model capable of reproducing accurately the solubility of methane in pure water is unable to capture the effect of salt concentration on the methane solubility, i.e. salting-out.

Given the importance of polarization of the solute molecule for the solubility of methane in pure water, they logically suggest that the observed over-prediction of the salting-out effect maybe due to neglecting the effect of the charged co-solvents contribution to the electric field. In this work we test this hypothesis using point polarizable models for a variety of solutes (Ne, Ar, Kr, Xe and CH_4), and show that while polarizability plays a role in the salting-out effect, the dominant force is once again volume exclusion, more specifically, the direct solute-ion interaction. We show that accounting for explicit polarizability in the solute models, is insufficient to account for the discrepancy between simulation and experiment values for the excess chemical potential and the salting-out effects. We will also demonstrate the importance of an accurate representation of the ion-water system, i.e. solute free, for accurately reproducing the experimental excess chemical potential.

4.3 Theory

In studying the effect of the addition of salts on the solubility of hydrophobic molecules, the Setchenow coefficient is common indictor for the degree of salting-out. Setchenow parameter may be written as,

$$\ln\left(\frac{x_{Sol}^{0}}{x_{Sol}}\right) = k_{Salt} M_{Salt} \qquad [4\text{-}1]$$

where, x_{Sol}^{0}, x_{Sol} are the solubility molar faction of the solute in the pure water and aqueous salt solution respectively, k_{Salt} is the Setchenow coefficient and M_{Salt} is the molarity of the aqueous solution. It is worth noting that, while in this work we use molarity and molality, in the literature a variety of concentration scales appear to have been used, often without specification, making comparison difficult.

In terms of calculation the solubility in terms of the mole fraction, the solute's mole fraction is not the most intuitive, or easy to calculate directly from molecular simulations, so we would like to rephrase the Setchenow equation in terms of more easily accessible variables. To start, the mole fraction x_{Sol} can be related to the Henry's coefficient,

$$x_{Sol} = \frac{y_{Sol} P}{H_{Sol}} \qquad [4\text{-}2]$$

where, H_{Sol} is the Henry's constant, and $y_{Sol} P$ is the partial pressure of the solute in the gas phase. Assuming that the solvent does not form a significant fraction of the vapor phase, which is a reasonable assumption for water and aqueous solutions at ambient condition, $y_{Sol} P = P$, and will be the same for all ion concentrations and, thus, equation 4-1 can be written in terms of the Henry's coefficients.

$$\ln\left(\frac{x_{Sol}^{0}}{x_{Sol}}\right) = \ln\left(\frac{H_{Sol}}{H_{Sol}^{0}}\right) = k_{Salt} M_{Salt} \qquad [4\text{-}3]$$

In turn, the Henry's coefficients can be determined from the excess chemical potential of the solute at infinite dilution

$$H_{Sol} = k_B T\rho \exp\left(\frac{\mu_{Sol}^{ex}(\infty)}{k_B T}\right) \quad [4\text{-}4]$$

where k_B is the Boltzmann constant, T is the temperature and ρ is the density of the solution. As in previous work, we calculate the excess chemical potential using the Widom insertion method[14] which can be expressed as

$$\mu_{Sol}^{ex}(\infty) = -k_B T \ln\left\langle \exp\left(\frac{-U_{Sol}}{k_B T\rho}\right)\right\rangle \quad [4\text{-}5]$$

Thus,

$$H_{Sol} = \frac{k_B T\rho}{\left\langle \exp\left(\frac{-U_{Sol}}{k_B T}\right)\right\rangle} \quad [4\text{-}6]$$

and the Setchenow coefficient may now be written in terms of a change in internal energy U_{Sol}, caused by the addition of a solute particle

$$\ln\left(\frac{x_{Sol}^0}{x_{Sol}}\right) = \ln\left(\frac{H_{Sol}}{H_{Sol}^0}\right) = \ln\left(\frac{\rho^0\left\langle \exp\left(\frac{-U_{Sol}}{k_B T}\right)\right\rangle}{\rho\left\langle \exp\left(\frac{-U_{Sol}^0}{k_B T}\right)\right\rangle}\right) = k_{Salt}M_{Salt} \quad [4\text{-}7]$$

The contribution of the solute to the internal energy of the system maybe considered as the sum of a van der Waals and polarizability component.

$$U_{Sol} = U_{Sol}^{vdW} + U_{Sol}^{Pol} \quad [4\text{-}8]$$

where U_{Sol}^{vdW} is the energy due to the van der Waals interactions and U_{Sol}^{Pol} is the contribution to the system energy from the polarizability of the solute molecule and is given by

$$U_{Sol}^{Pol} = -\frac{1}{2}\alpha\left|\left(\bar{E}\right)^2\right| \quad [4\text{-}9]$$

where $\vec{E}$ is the electric field at the polarization site and α is the polarizability of the solute. For full details of the calculation of the electric field see Chapter III.

4.4 Computational Method

Methodology

Our calculation of the excess chemical potential of each solute in solutions of NaCl of varying concentration consisted of three stages. In all stages, for all models and systems, interactions were truncated at 9Å and standard tail corrections applied. In the first two steps, i.e. steps 1 and 2, simulations were performed using the LAMMPS[15] computational software with long-ranged Coulombic interactions calculated via the particle-particle, particle-mesh (PPPM)[16] technique. Finally, in step 3, we used our in-house code that utilizes the Wolf method[17] to account for the long-range nature of electrostatic interactions, i.e. the calculation of the electric field. Following the work of Demontis *et al*.[18], and Avendano and Gil-Villages[19], in the Wolf method we choose a damping factor (β) of $2/R_C$, where R_C is the Coulombic truncation distance (9Å). For an explanation of our use of the Wolf method, as all as a justification of the choice of the PPPM method in our LAMMPS simulations, we refer the reader to chapter III, and appendix A.

Step one: Aqueous solution simulations

In the first step of our calculations, the density, as predicted by each model at a variety of concentrations, was determined from isothermal-isobaric molecular dynamics simulations (NPT), at a pressure of 1 bar, and a temperature of 298K. These simulations

consisted of 278 water molecules together with 1, 2, 5, 10, 20, 25 or 30 NaCl molecules, corresponding to concentrations of 0.200, 0.399, 0.998, 1.997, 2.995, 3.993, 4.992, 5.990 mol/kg (mol of solute/kg solvent) molality respectively. Starting with a randomly orientated and transitionally distributed configuration at a density close to experimental, each system was first subjected to a simulation period of 300ps with a time step of 0.5fs in order to obtain an equilibrated configuration. A 1ns production run was then performed in which the mean density was calculated. The densities for the solution are shown in Figure 4-1.

Step two: Generation of sample configurations

Starting with the final configuration from step 1, each system was adjusted in volume to the average density for each system (a minor adjustment) and subjected to an equilibration period of 5ps of constant volume, constant temperature molecular dynamics simulation (NVT) in order to relax the system following the volume change. Sample configurations were then generated by performing 7ns of NVT molecular dynamics simulations, recording a configuration every 70fs for a total of 100,000 configurations for each water model and ion concentration.

Step Three: Calculation of the excess chemical potentials

The third and final stage was the calculation of the chemical potential of the solute via the Widom insertion technique[14] using in-house code. In this method, a phantom solute molecule is inserted 100,000 times into each configuration, resulting in 10^{10} insertions for every water model and ion concentration.

4.4 Molecular models

In this section we describe in more detail the molecular models used for each of the species considered in this work, namely water, NaCl and solute molecules (Ne, Ar, Kr, Xe and CH_4). We begin by noting that all cross-species interactions are approximated using the Lorentz-Berthelot combining rules

$$\sigma_{ij} = \left(\frac{\sigma_{ii} + \sigma_{jj}}{2} \right), \quad \varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \qquad [4\text{-}10]$$

Water potential

In this work we consider three popular models of water. The TIP4P/Ew[20] and TIP4P/2005[21] models are chosen due to their superior performance in predicting the solubility of small non-polar solutes in water (Chapter III). In addition to these, the TIP3P[22] model is evaluated due to its extensive use in the study of biological systems, in which polarizability and hydrophobic hydration are believed to play an important role. All of theses are rigid non-polarizable water models consisting of a single Lennard-Jones site located at the center of the oxygen, together with a number of point charges to account for the charged nature of the hydrogen and oxygen atoms. The parameters for these models are shown in Table 4-1.

Co-solvent potential

Unpolarizable Na and Cl ions were modeled using the Smith and Dang[23] potential models for ions, i.e. they are considered as dissociated ions, described by charged single-site Lennard-Jones spheres. This model was chosen as it has previous been shown to reproduce well a high level of accuracy for the density of NaCl solutions[13]. The parameters for the ions are presented in Table 4-2.

Solute potential

Following our previous work, we consider two models for the solute molecules. In the first, the molecules are described by a single Lennard-Jones site whilst, in the second, in addition to the Lennard-Jones site for the first model, the molecules have a polarizable site located at their center of mass. The solute polarizability is taken from experimental values[24,25], and is shown together with the Lennard-Jones parameters for each model in Table 4-2. For more detail, refer to Chapter III.

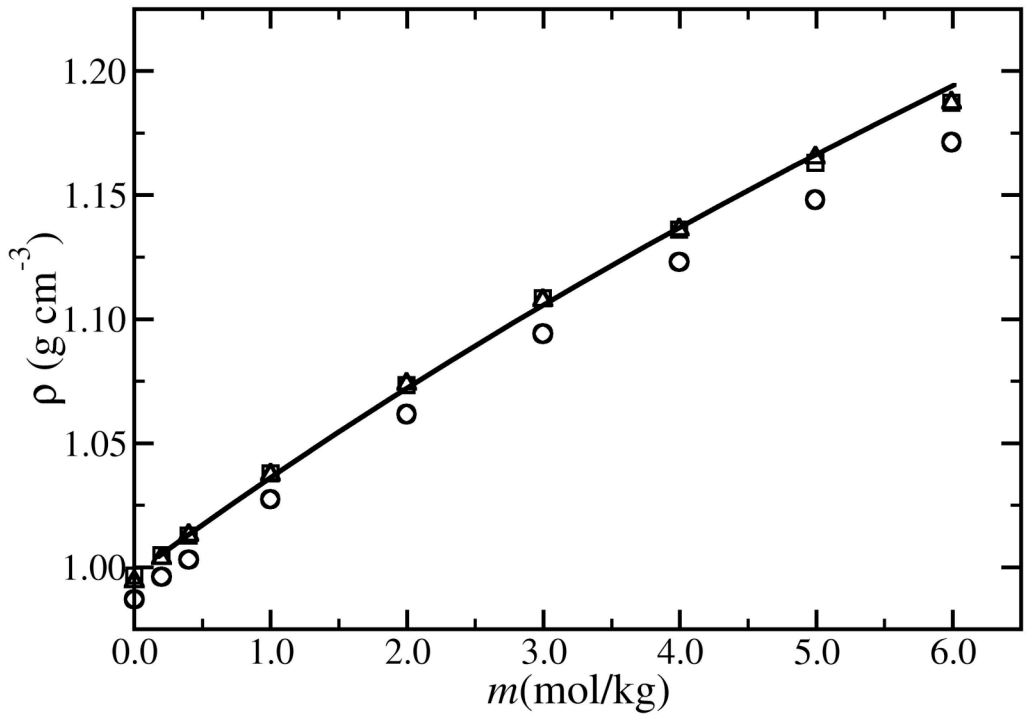


Figure 4-1: *Density of the aqueous systems as a function of NaCl ion concentration at 298K and 1 bar for TIP4P/2005, TIP4P/Ew and TIP3P water models. The solid line represents experiment densities*[26]*. Squares are TIP4P-2005, triangles are TIP4P/Ew, and circles are TIP3P.*

Table 4-1. Lennard-Jones interaction potential parameters for the aqueous solution

	TIP3P	TIP4P/2005	TIP4P/Ew	Na	Cl
σ	3.1506	3.1589	3.1644	2.35	4.40
ε	76.546	93.240	81.910	65.42	50.32
q	q_h 0.4170	q_h 0.5564	q_h 0.52422	q_i 1.0	q_i -1.0
	q_o -0.8340	q_o -1.1128	q_o -1.04844		

Table 4-2. Solute potential Lennard-Jones parameters and polarizability (α)

	σ [Å]	ε [K]	α [$Å^3$]
Neon	3.035	18.6	0.40
Argon	3.415	125	1.64
Krypton	3.675	169	2.48
Xenon	3.975	214.7	4.044
Methane	3.730	147.5	2.56

4.6 Results

Density

First, given the importance of the solution density for the solubility of hydrophobic molecules as shown in previous work[27], as well as work by Paschek[28] and Docherty *et al.*[13], we test the ability of each water model to reproduce the density-concentration relationship for NaCl solutions of varying concentration. In Figure 4-1, we show the density as a function of ion concentration, predicted by the water models. We note that TIP4P/2005 and TIP4P/Ew models provide essentially identical values and an excellent description of the density of NaCl solutions. TIP3P, however, significantly under-predicts the densities of the solutions considered in this work. Note that this will have a fairly dramatic effect on the solute solubility, which will be demonstrated later.

Chemical potential

As in the previous chapter, we begin by showing the effect of polarizability on the excess chemical potential for each solute species considered. In Figures 4-2 to 6, we have shown the excess chemical potential for the different solutes using both the polarizable and non-polarizable models. As expected, the inclusion of explicit polarizability reduces the excess chemical potential, i.e. increases solubility, although it does not provide enough of a decrease to achieve quantitative agreement with experiment. Interestingly, while the inclusion of explicit polarization improves quantative agreement of the temperature dependences of solubility at a cost of reduced qualitative agreement (chapter III) here we note an improvement of both qualitative and quantative agreement.

Salting-out effects

In Figures 4-7 to 11, the Setchenow coefficient is the gradient of the line as function of molarity, in accordance to equation 4-1. The Setchenow coefficient is overestimated in each case. The effect of explicitly accounting for polarizability is shown in Figure 4-9a and b. As expected, the addition of explicitly point polarizability to the solute leads to a decrease in the chemical potential. However, the decrease is not sufficient to obtain quantative agreement. This of interest to a number of workers, Paschek and Docherty *et al.*[13], who have suggested that the lack of accounting for explicit polarizable is the cause of the over-prediction of the Setchenow coefficient.

As a result, the Setchenow coefficient calculated with and without polarizability is essentially the same, with only a small decrease in the gradient as the ion concentration increases. It is clear that the addition of polarizability is not the main contributing factor involved in the ability of the solute, water and ion system to reproduce the correct salting out behavior. Given that this hypothesis is incorrect; the question is now, what causes the over-prediction in the Setchenow coefficient?

Our attention now comes back to the excluded volume of the system. As the water-ion potential attains the correct density for the TIP4P/2005 and TIP4P/Ew models, and the water-solute potential approaches quantative agreement, our focus is on the ion-solute potential. We hypothesis, that the over-prediction of the Setchenow, and in turn, the over prediction of the excess chemical potential is due to the overly repulsive potential between the ion and solutes. The reasons for this hypothesis are given below.

TIP3P

In Figure 4-1 we show that the TIP3P water model significantly under-predicts density of the solvent. The TIP3P water model is the best at reproducing the excess chemical potential and the correct salting-out behavior. However, this agreement is not due to a better description of the physics of the systems, rather, the agreement between TIP3P with experimental values is due to cancellation of errors. As shown in the work of Paschek[28], Docherty *et al.*[29] and in Chapter III, the lower density of the solvent, relating to a lower packing density, and the energy between the solutes and the aqueous solution are further down the short-range repulsive wall. This in turn leads to a decrease in the excess chemical potential. Given that all of these models typically over-predict the excess chemical potential, any reduction leads to improved agreement with experiment. In Figure 4-12, we show the excess chemical potential for the TIP3P model obtained from simulations performed at the densities predicated by the TIP4P/Ew model of water. When the densities predicted by TIP4P/Ew are used; the predicted chemical potential is similar to the other models presented here. This is typical of what was seen for all species and is in support of the idea that salting-out is excluded volume driven.

Brake down the contribution to the excess chemical potential

In order to confirm our hypothesis, that the overly repulsive ion-solute potential is the cause of the over-prediction of the salting-out effects, we divided the solvent into three regions. The first region is the first hydration layer, we defined as the volume between the ion locations and the first minima in the radical distribution function, the second region is the second hydration shell, which is between the first and second ion-oxygen minima and the final region is the bulk water, which is any point outside the

second region. The radical distribution function for Na-O and Cl-O are shown in Figure 4-13 and 4-14 respectively. As a single insertion point can occur in both a 1st and 2nd water layers due to multiple ions, we have defined that in this situation, it will be counted as being in the 1st water layer. For this work, only the TIP4P/2005 water structure was studied. Our objectives for studying the different solvent regions are to determine the nature of the fluid away from the ions, the excess chemical potential in each region and to look at the long range effect of the ion-solute potential.

The probability of the solutes being inserted to the different region is shown in Figure 4-15. As the ion concentration increases, the probability of the inserting into the bulk region rapidity decreases, as the water molecules are being reallocated into the hydration regions. The 2nd region increases quickly due it greater volume that the 1st hydration region. After a molality around 2.0 mol/kg, the volume of the 2nd region decreases as the water molecules in the 2nd region are being reallocated into the 1st hydration region. After a molality around 3.0 mol/kg, the volume of the bulk region is very small and does not contribute significantly to the system.

The density of each region of the system can be determined by counting the number of water molecules in each region and using a Monte Carlo sampling of space in the cell to determine the average volume of each region. The effect of ion concentration on the localized density in units of g/cc of each region is shown in Figure 4-16. We see that the local density of the first hydration layer increases as the ion concentration increases, due to the increase in the number of ions in each other hydration layer. What is of interest is that the local density in the bulk region decreases from 0.997g/cc for pure water, to 0.941g/cc for a molality of 1.997 mol/kg. At the same time, the density of the second hydration shell also has decreased from 1.00 g/cc for 0.200 mol/kg, to 0.871g/cc

at the 5.99 mol/kg. This shows for the both the 2nd hydration and the bulk regions that there is not an increase in the structure of the fluid away from the 1st hydration layer.

We begin assessing the separate regions by considering the excess chemical potential in the bulk region. As an example, we have chosen the TIP4P/2005 water model with the non-polarizable krypton solute model. The excess chemical potential dependence on the concentration for the separate regions is shown in Figure 4-17, which we note that the excess chemical potential for the bulk region is approximately constant at about 8 kJ/mol. The excess chemical potential of both the 1st and 2nd hydration region increases linearity as the ion concentration increases.

The effect of the density on the excess chemical potential from both pure water, and bulk water in the aqueous solution is shown in Figure 4-18. The pure water solvent has a significant reduction in the excess chemical potential due to a decrease in the density opposed to the bulk water solute where the excess chemical potential is constant. As the only difference in the two systems is the ion-solute interactions, so we can conclude that the ion-solute potential has a long range effect on the chemical potential, and that the ion-solute interaction increase the excess chemical potential.

The regional density of the second hydration layer also decreases as the ion concentration is increased. Again, we would expect that the excess chemical potential will decrease but, the excess chemical potential increases. This again we expect that this is due to the increasing amount of ions interacting with the solute, even if it is inserted outside the first hydration layer. That is to say, at low concentration, the 2nd hydration shell surrounds only 1 ion, whereas at high concentrations, an increasing number of insertions will be into the second hydration layers containing two or more ions. A similar

effect is seen for the first hydration energy, where the excess chemical potential also increases linearly as the ion concentration increases.

In regards to the increase in the overall excess chemical potential, this is caused the high proportion of the sampling occurring at the 1st hydration layers, which has a higher value compared to the 2nd hydration layer and the bulk. As the excess chemical potential in the bulk is constant, the cause of the increase in the excess chemical potential can not be due to an increase in structuring or order in the bulk region.

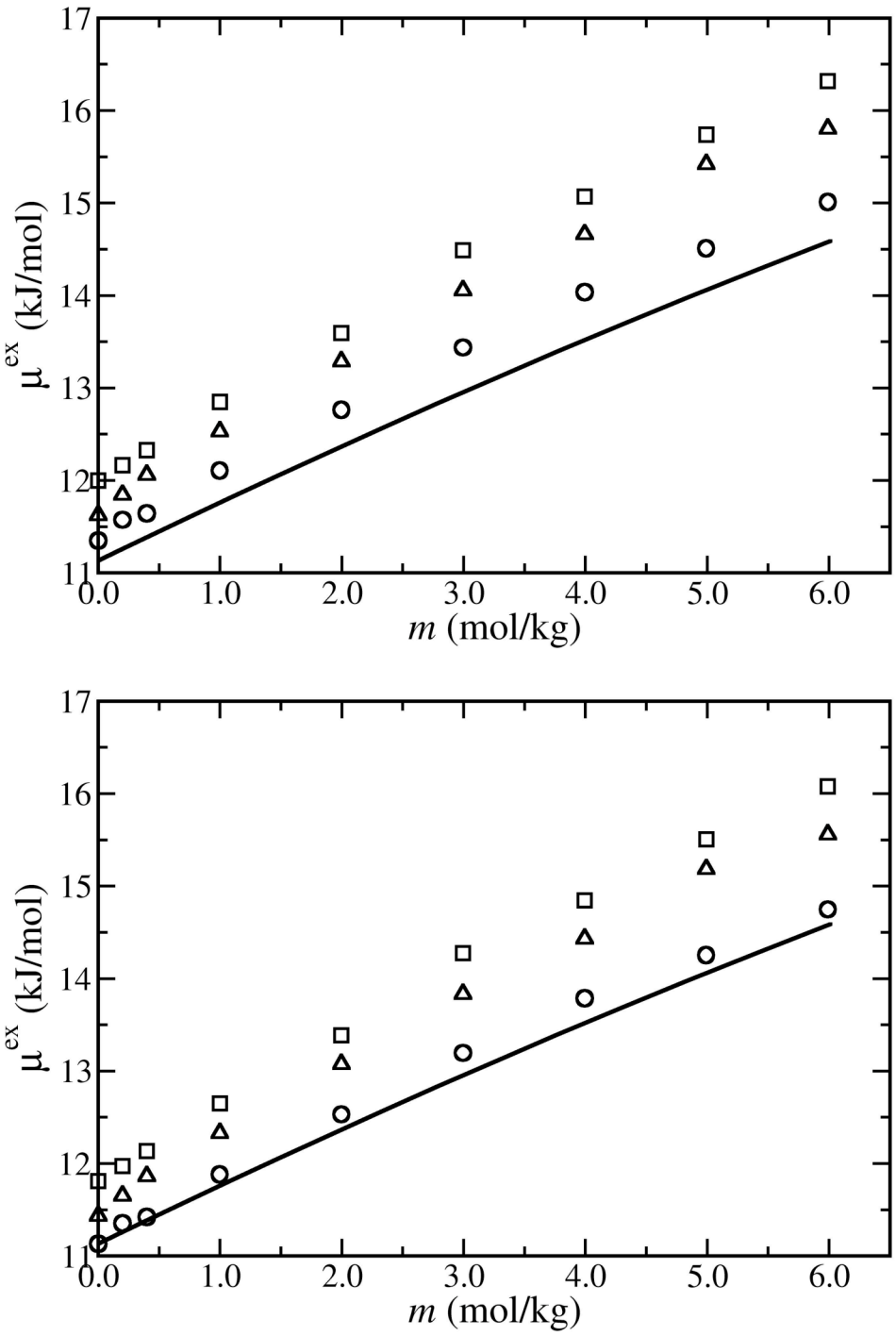


Figure 4-2: *NaCl*, c*oncentration dependence on the excess chemical potential of neon in aqueous solution at infinite dilution for (a) A non-polarizable neon model and (b) A polarizable neon model. Squares are for TIP4P/2005, triangles are for TIP4P/Ew, circles for TIP3P*

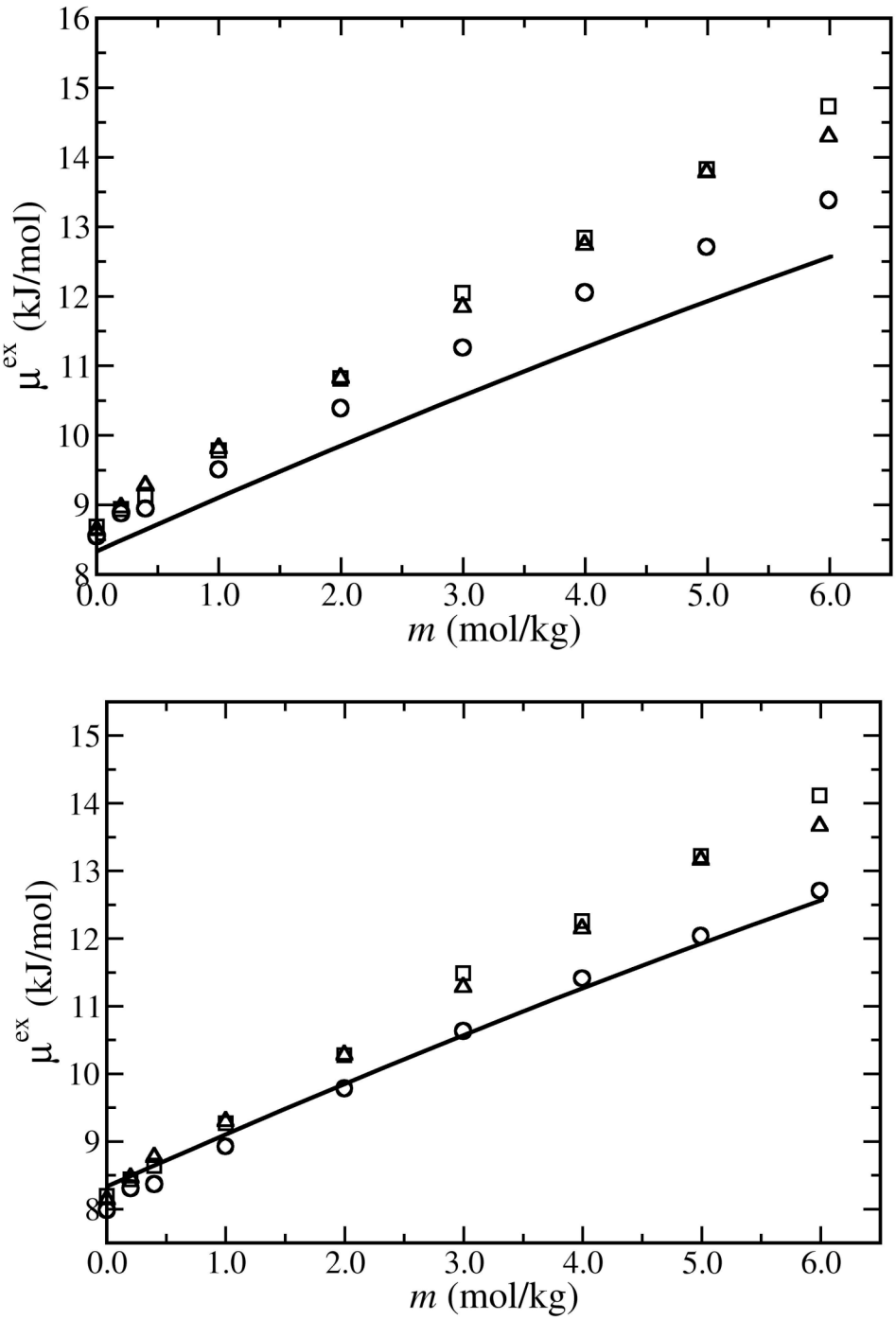


Figure 4-3: *NaCl*, c*oncentration dependence on the excess chemical potential of argon in aqueous solution at infinite dilution for (a) A non-polarizable argon model and (b) A polarizable argon model. For explanation of symbols, see Figure 4-2*

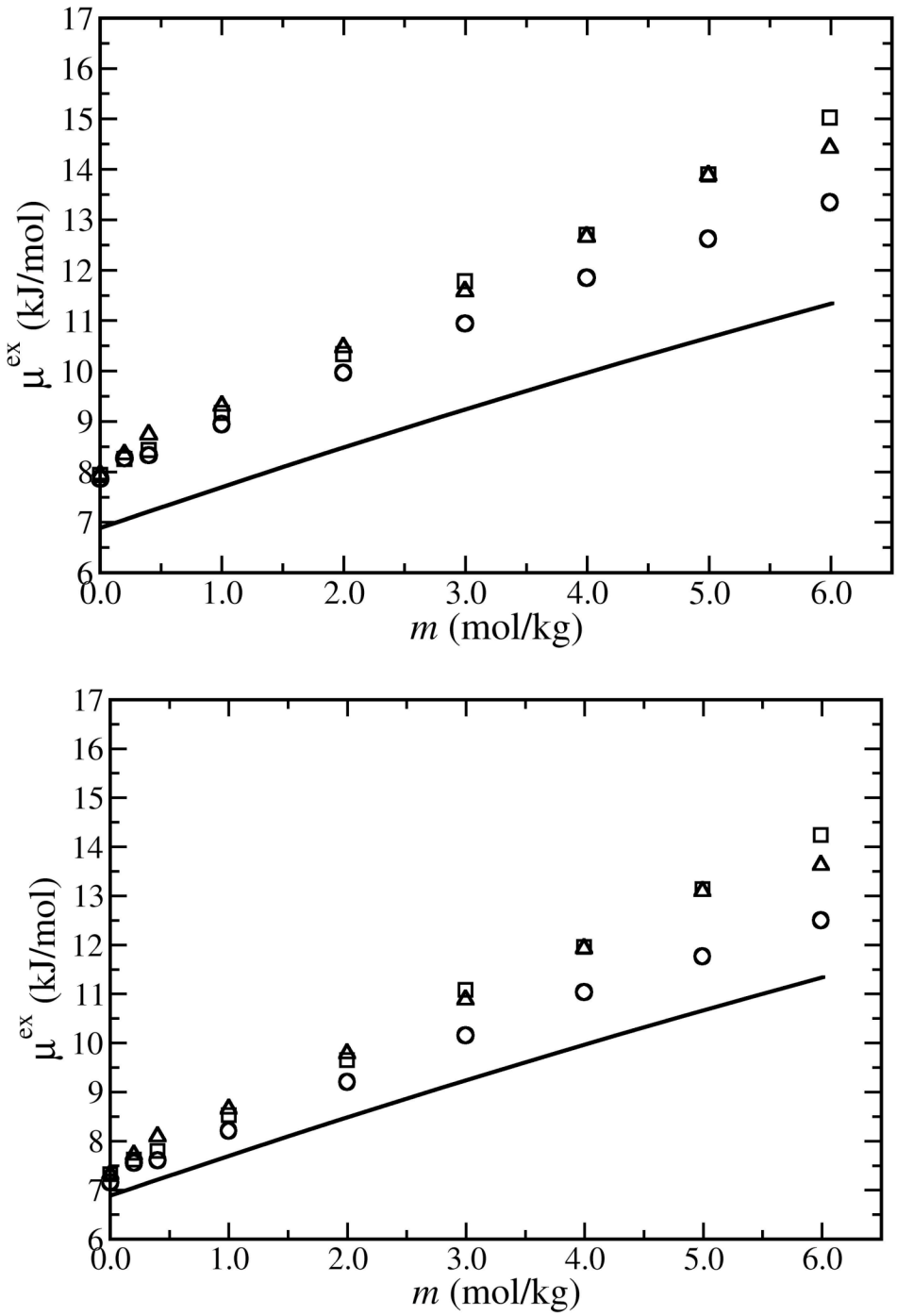


Figure 4-4: *NaCl*, c*oncentration dependence on the excess chemical potential of krypton in aqueous solution at infinite dilution for (a) A non-polarizable krypton model and (b) A polarizable krypton model. For explanation of symbols, see Figure 4-2*

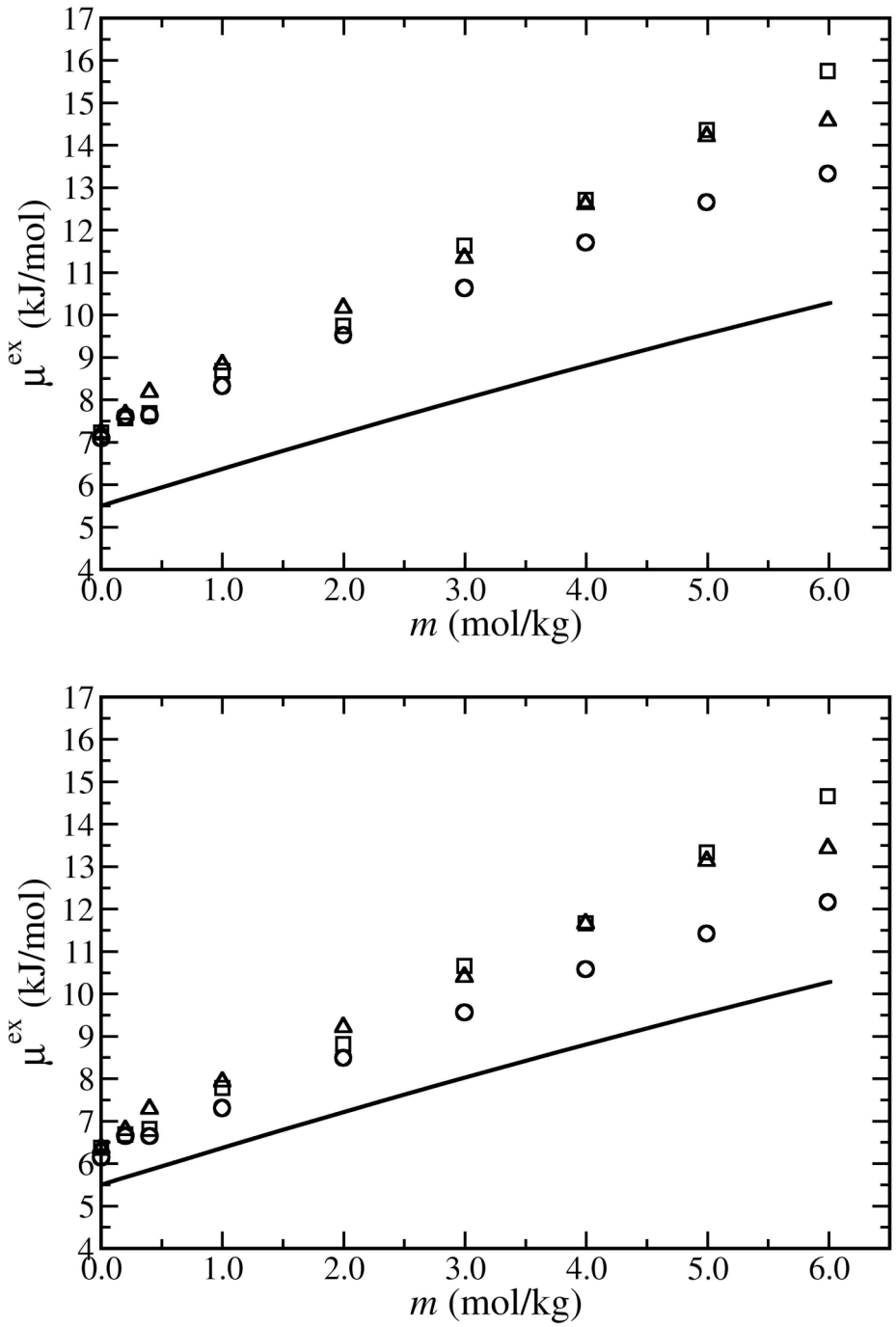


Figure 4-5: *NaCl,* c*oncentration dependence on the excess chemical potential of xenon in aqueous solution at infinite dilution for (a) A non-polarizable xenon model and (b) A polarizable xenon model. For explanation of symbols, see Figure 4-2*

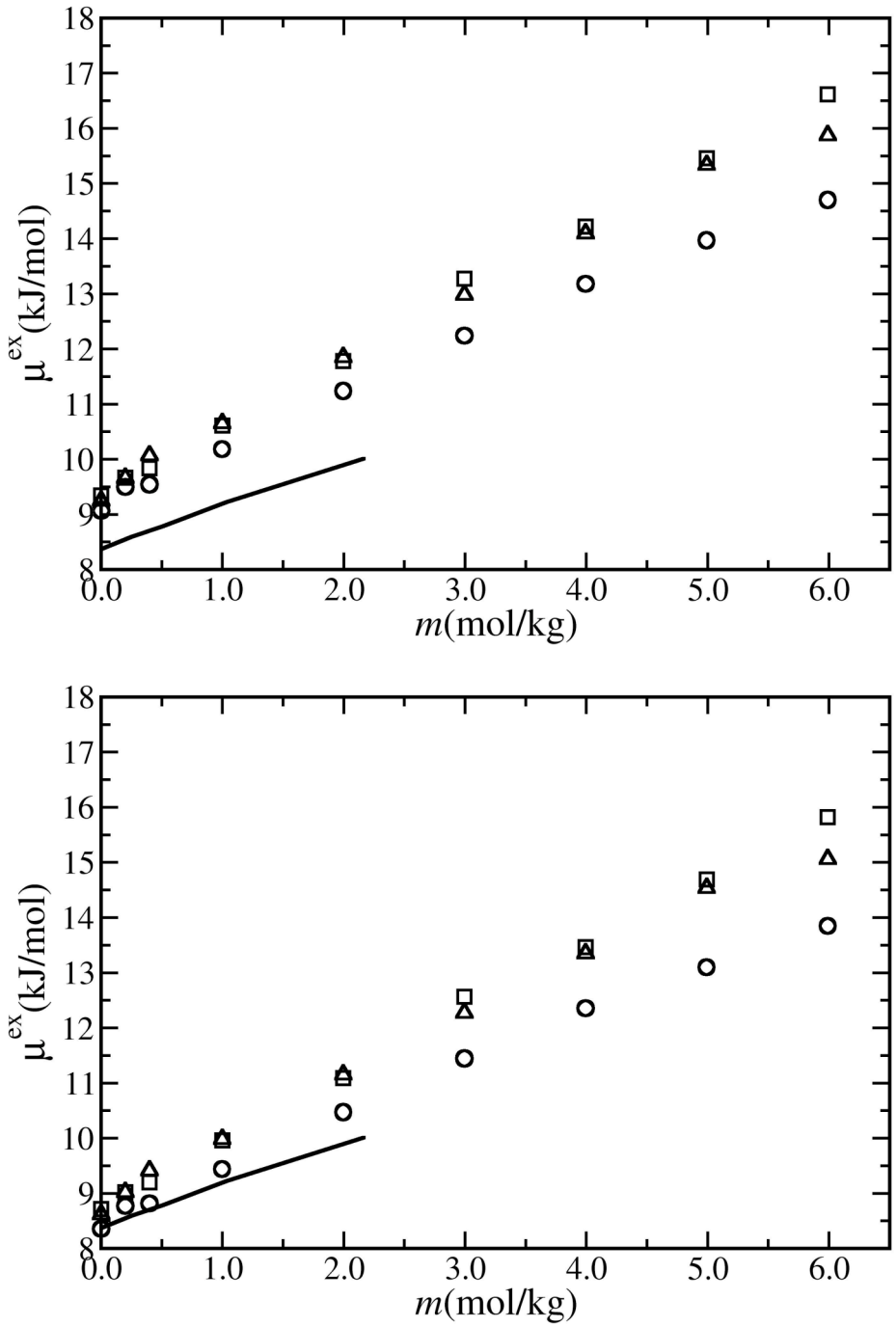


Figure 4-6: *NaCl*, c*oncentration dependence on the excess chemical potential of methane in aqueous solution at infinite dilution for (a) A non-polarizable methane model and (b) A polarizable methane model. For explanation of symbols, see Figure 4-2*

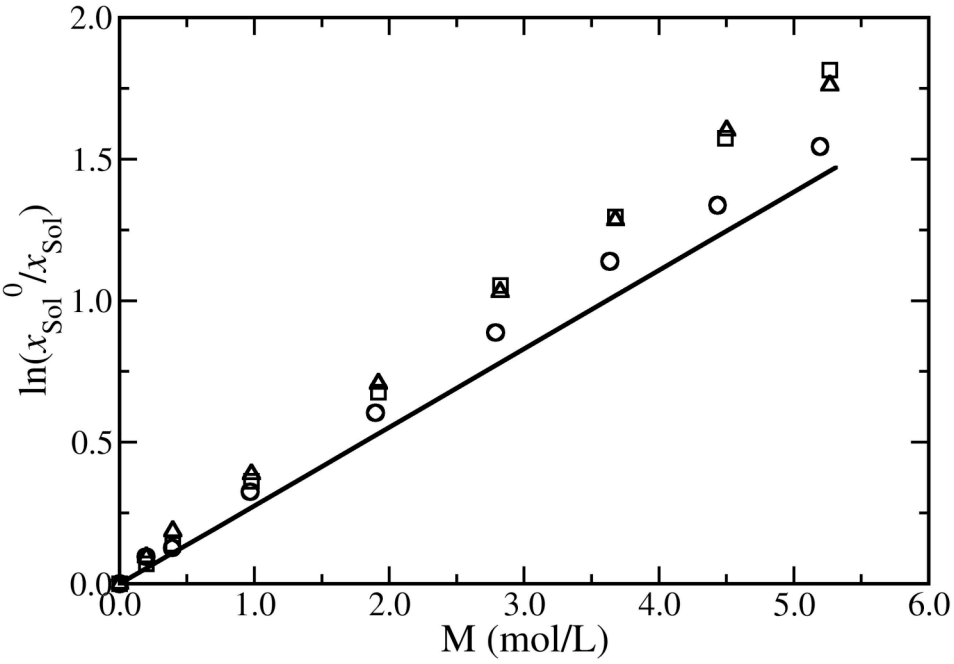


Figure 4-7: *NaCl*, c*oncentration dependence on the Setchenow parameter of neon in aqueous solution at infinite dilution for different water models. For explanation of symbols, see Figure 4-2*

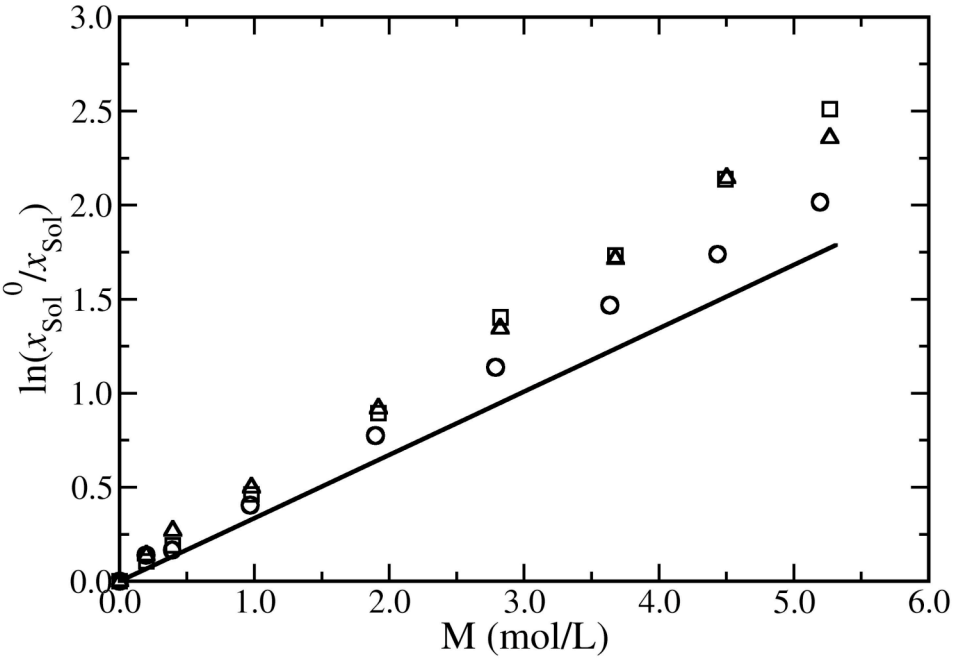


Figure 4-8: *NaCl*, c*oncentration dependence on the Setchenow parameter of argon in aqueous solution at infinite dilution for different water models. For explanation of symbols, see Figure 4-2*

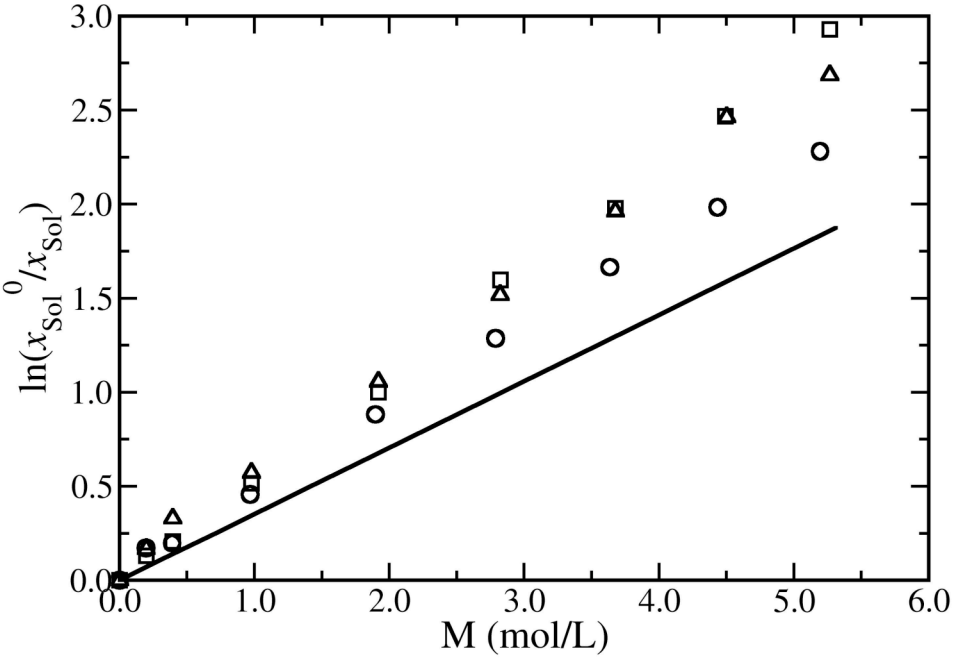


Figure 4-9a: *NaCl*, c*oncentration dependence on the Setchenow parameter of krypton in aqueous solution at infinite dilution for different water models, without polarizability. For explanation of symbols, see Figure 4-2*

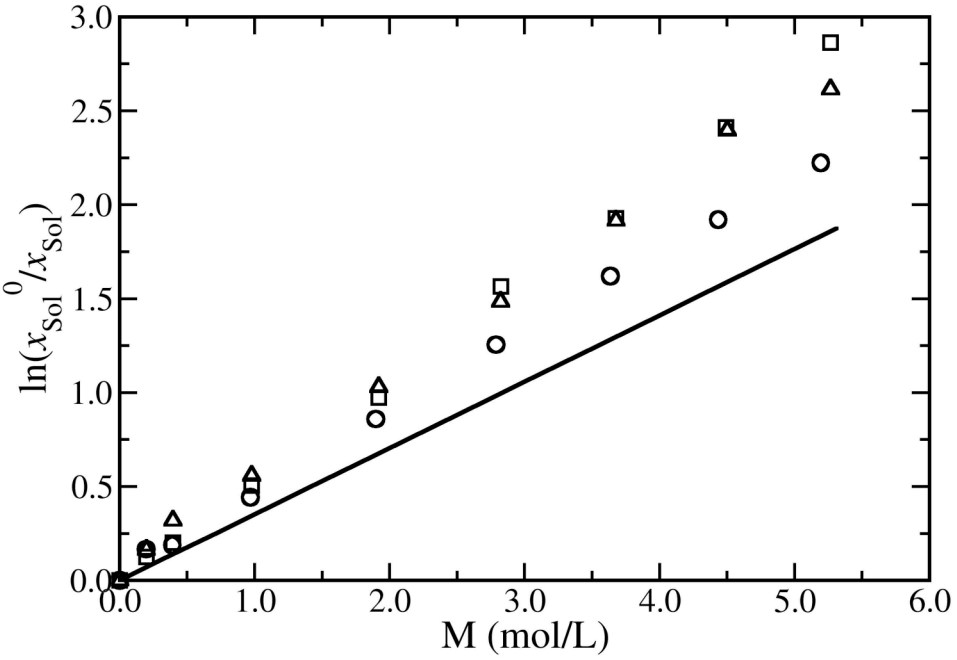


Figure 4-9b: *NaCl*, c*oncentration dependence on the Setchenow parameter of polarizable krypton in aqueous solution at infinite dilution for different water models. with polarizability. For explanation of symbols, see Figure 4-2*

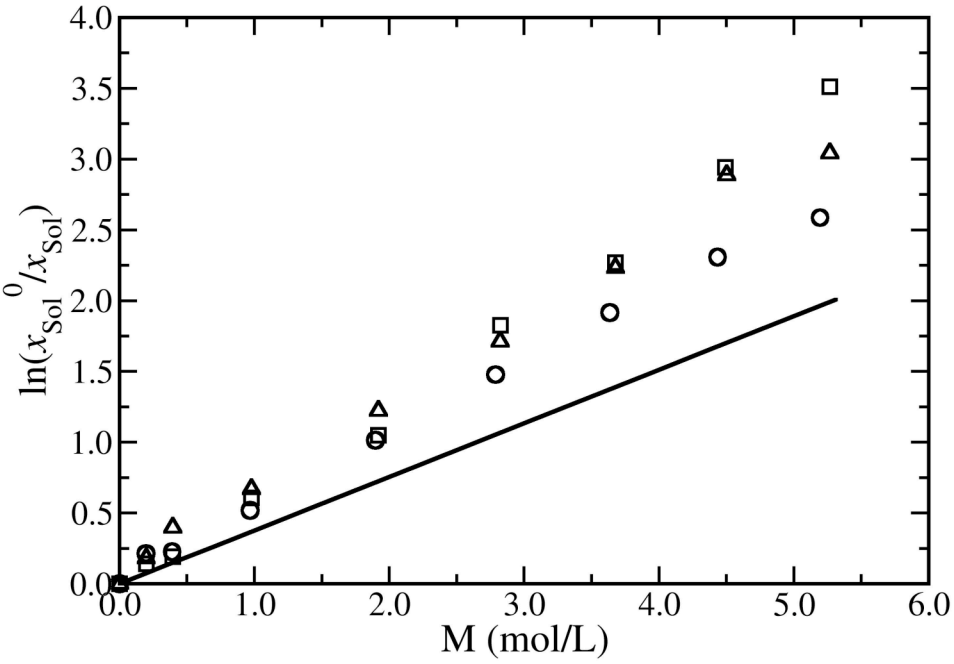


Figure 4-10: *NaCl*, c*oncentration dependence on the Setchenow parameter of xenon in aqueous solution at infinite dilution for different water models. For explanation of symbols, see Figure 4-2*

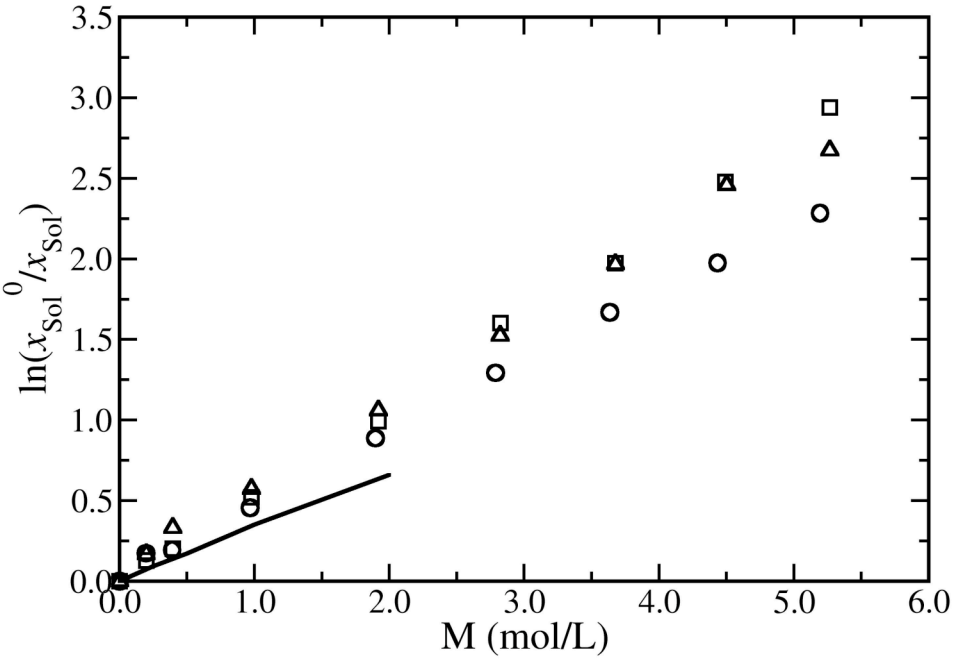


Figure 4-11: *NaCl*, c*oncentration dependence of the Setchenow parameter of methane in aqueous solution at infinite dilution for different water models. For explanation of symbols, see Figure 4-2*

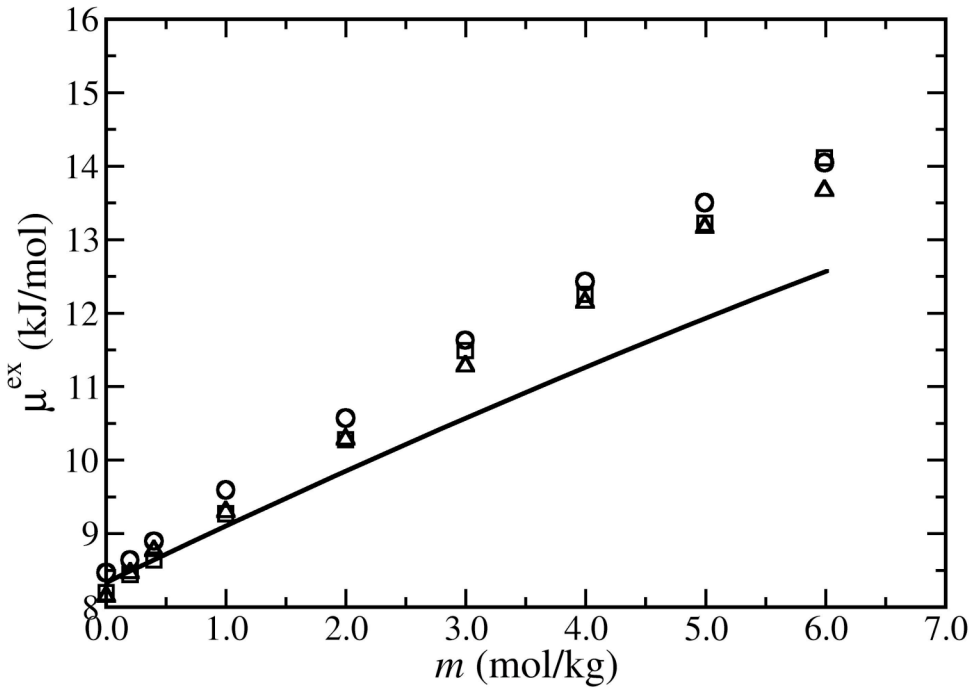


Figure 4-12: *NaCl*, c*oncentration dependence of the excess chemical potential of argon in aqueous solution at infinite dilution. With TIP3P at the same density of TIP4P/Ew, Squares are for TIP4P/2005, triangles are for TIP4P/Ew, circles for TIP3P*

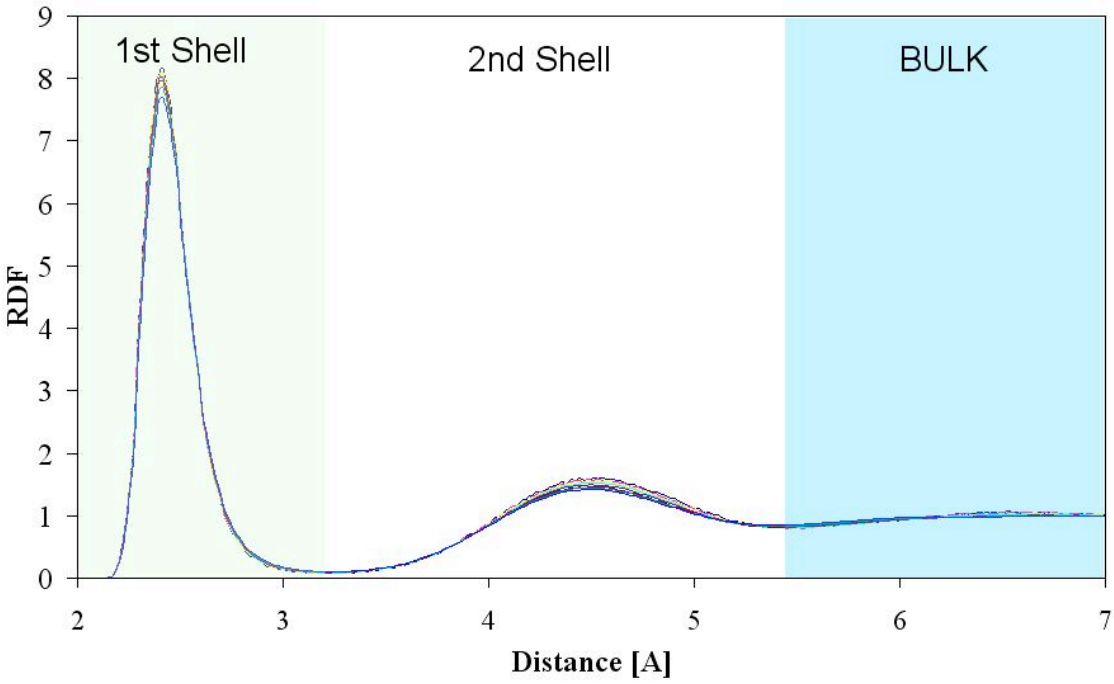


Figure 4-13: *Sodium-oxygen radial distribution function of a range of NaCl concentrations*.

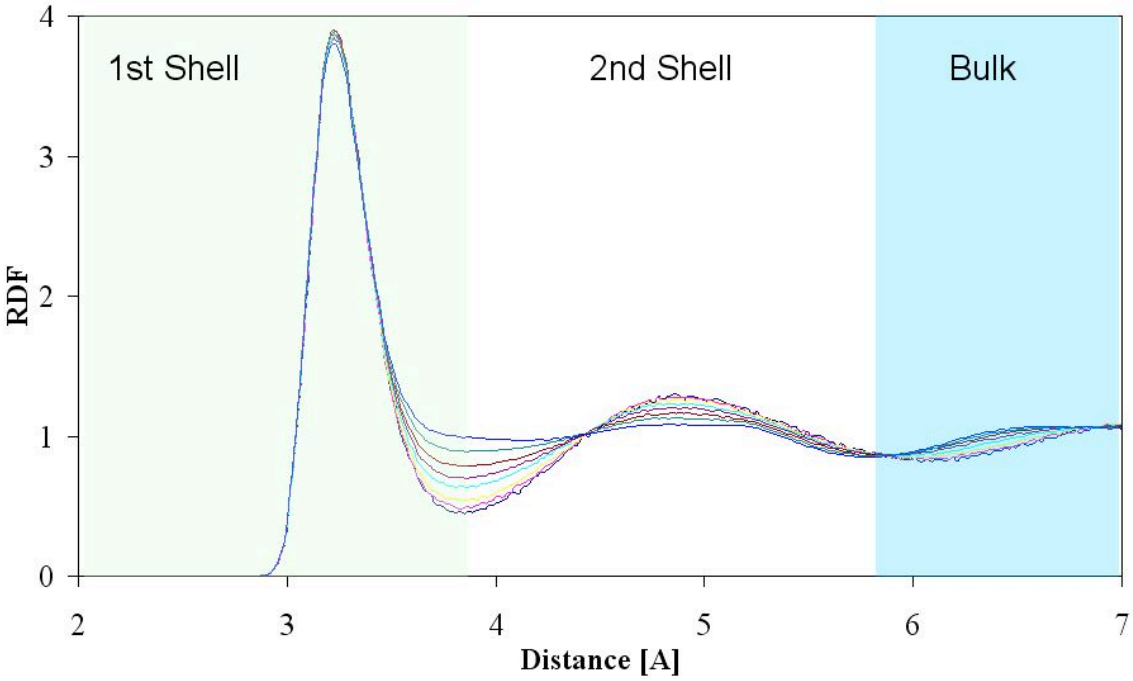


Figure 4-14: *Chloride-oxygen radial distribution function of a range of NaCl concentrations.*

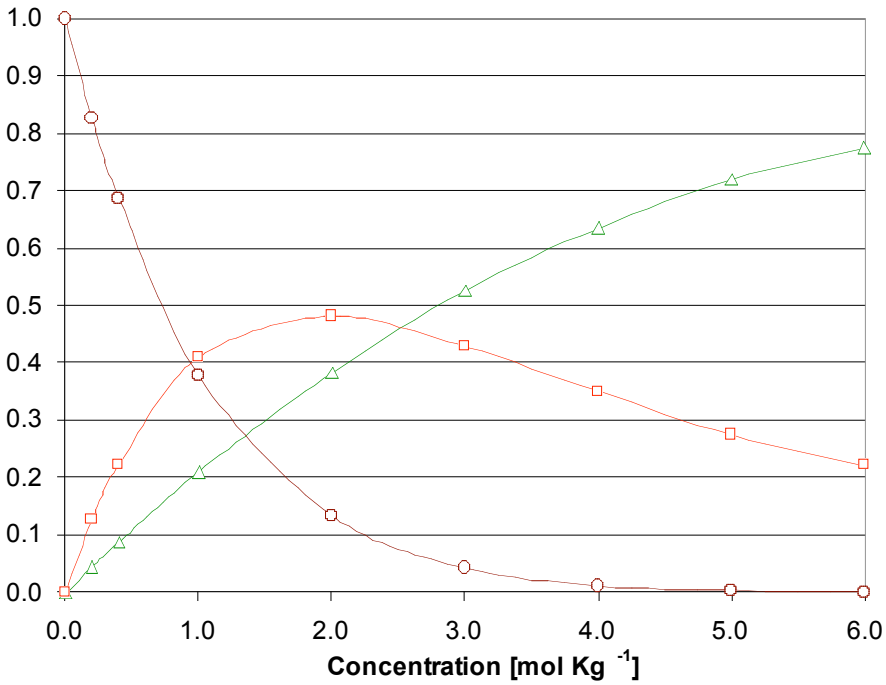


Figure 4-15: *Probability of the solute being inserted into a particular region, 1st hydration shell (triangles), 2nd hydration shell (squares), and the bulk (circles).*

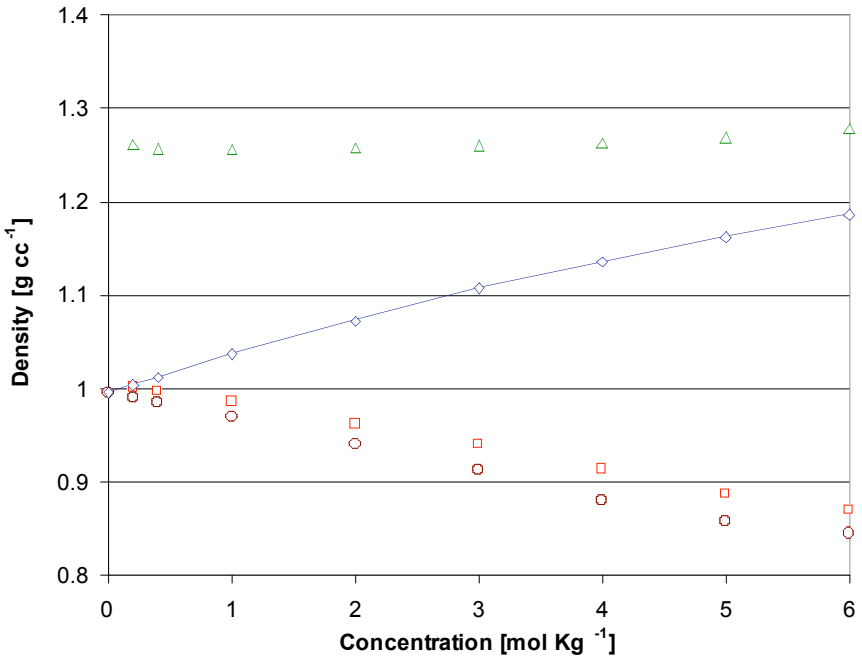


Figure 4-16: *Regional densities for aqueous solution with varying concentration.1st hydration shell contains ions (triangles), 2nd hydration shell (squares), and the bulk (circles). The total density on the solution (diamonds).*

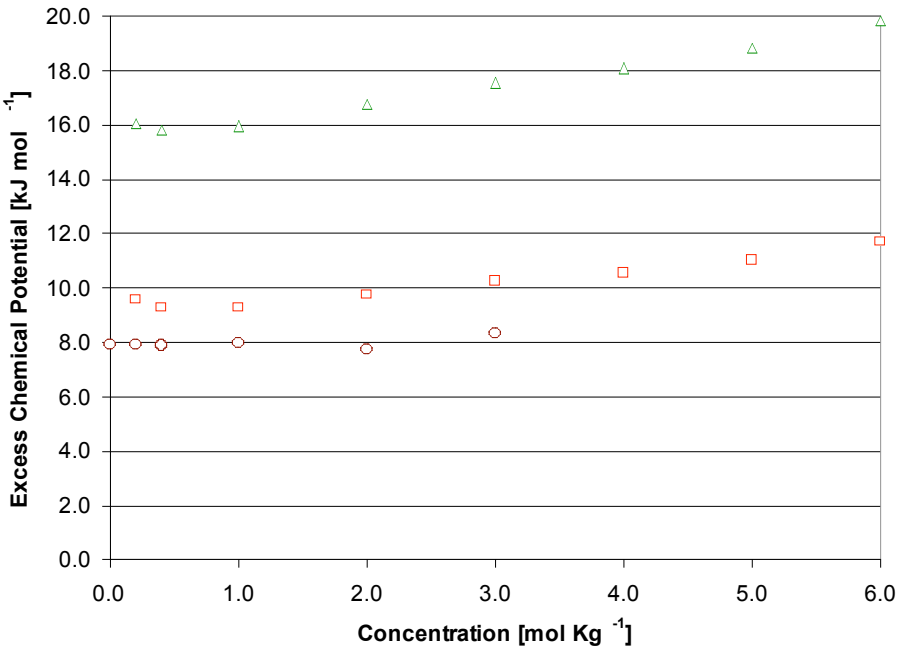


Figure 4-17: *Excess chemical potential from the insertion of krypton into the different region individually. 1st hydration shell (triangles), 2nd hydration shell (squares), and the bulk (circles).*

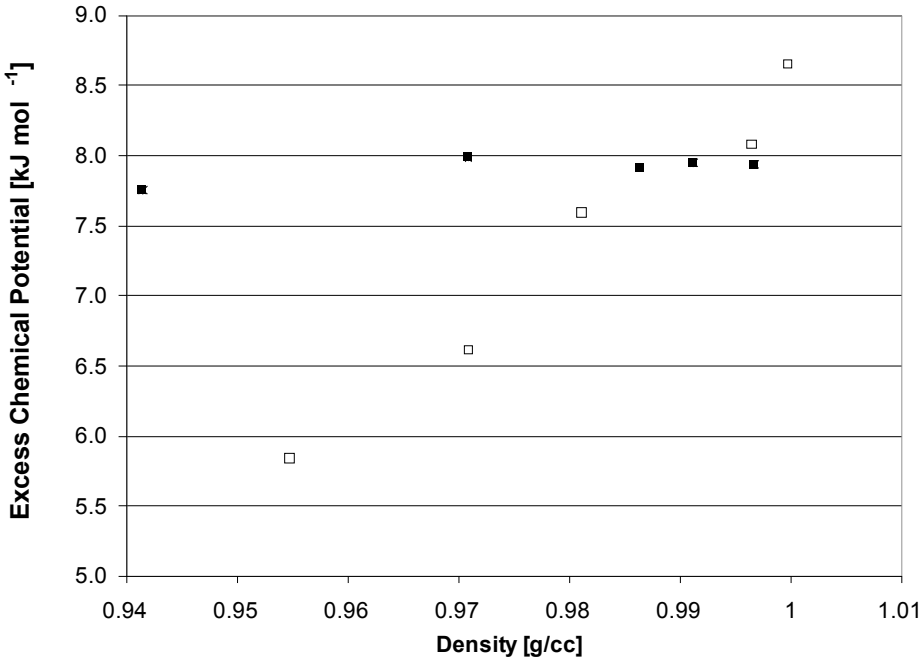


Figure 4-18: *Effect of density on the excess chemical potential. The black squares are the excess chemical potential from the bulk region on the ion-water simulations. The hollow squares are from the excess chemical potential calculated from pure water simulations are artificial densities.*

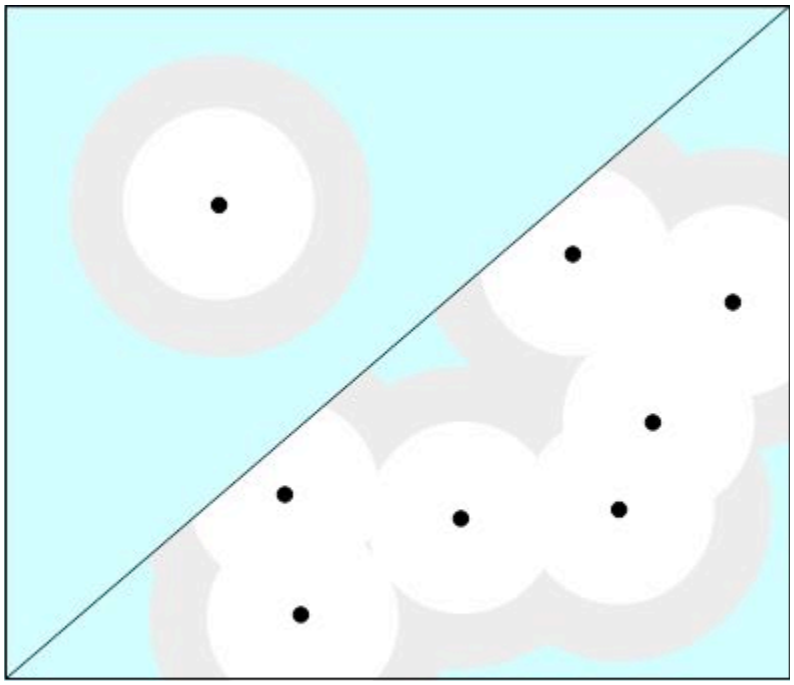


Figure 4-19: *Diagram of the different region in the aqueous solution, where the bulk is blue, 1st hydration layer is white, and the 2nd hydration layer is gray. The ions are represented as dots. Top, at low density, the there is one ion in the centre of the two hydration layer. Bottom, at the higher concentrations, there are multiple ion around each shell, contribution to high energy in the hydration shells.*

4.7 Force field from *ab initio* calculations

Having shown that the over-prediction for the salting-out effect is due to the ion-solute potential, it seam logical that the next step in attempting to correctly reproduce the salting-out effect is to development of improved solute-ion potential.

Perhaps the most obvious and straight forward method of accomplishing this would be to fit the solute-ion potential to experiment. However, apart from being computationally costly, is requires experimental data which, even for a common molecule such as CH_4 in NaCl solution, is very limited. As an alternative to this, we have chosen to fit the ion-solute potentials to *ab initio* calculations as this requires no experimental data and, as such, is a fully predictive method.

In an attempt to more accurately model the overly repulsive potential, we choose to model, the new short-range interactions are calculated via the Buckingham exp-6 potential,

$$U(r) = \frac{\varepsilon_{ij}}{1-6/\gamma_{ij}}\left(\frac{6}{\gamma_{ij}}\exp\left(\gamma_{ij}\left(1-\frac{r_{ij}}{\sigma_{ij}}\right)\right)-\left(\frac{\sigma_{ij}}{rij}\right)^6\right) \qquad [4\text{-}13]$$

which contains an exponential repulsive wall which is less steep than the r^{-12} of the Lennard-Jones. σ_{ij} is the location of the minimum energy, the ε_{ij} is the minimum energy and γ_{ij} controls the gradient of the repulsive wall. The Buckingham exp-6 potential obtains a better approximation of the potential wall, which is important for Monte Carlo simulations as the wall potential is frequently sampled. The Buckingham exp-6 parameters $\sigma_{ij}, \varepsilon_{ij}, \gamma_{ij}$ are optimized to reproduce the *ab initio* MP2 energy for ion-solute dimer, using a gradient descent optimization for the energy.

Force field development

For the refinement of the interactions between the solute and ions, the van der Waal interactions were fitted to *ab initio* calculations. The fitting method is based on calculations of the dimer energies for a range of configurations, i.e. with the ion-solute separated at different distances. In this work, the MP2 energy is calculated by using single point energy calculations at 40 different distances. The point energies are calculated with the distances chosen for each molecule starting with an ion-solute separation approximately 0.6Å closer than the overall energy minima and then for an addition 3.3Å at an interval of 0.1Å. The *ab initio* energy is made up from four components, as shown by,

$$U_{MP2} = U_{electro} + U_{Pol} + U_{vdW} + U_{BSSE} \qquad [4\text{-}14]$$

where $U_{electro}$ is the electrostatic energy, (zero, because the solute does not have a electrostatic charge), U_{Pol} is the polarization energy. U_{BSSE} is the basis set superposition error, and U_{vdW} is the van der Waals energy. Figure 4-20 shows an example of the parameterization, for the van der Waals energy interaction of Na^+-Ne.

The polarizability in the parameterization stage is vital as there is no shielding from the electric field by other molecules, as would be the case in a condensed phase solution. The ion-solute dimer has a very large electric field generated from the ion, and without the removal of the polarization effect, the van der Waal interaction will be highly overestimated because the van der Waal energy is calculated from the residue from the MP2 energy less all the other energy contributions. The size of the polarization energy from been seen in Figure 4-20, which is for a solute is a small polarization. The effect is greater of larger molecules with a higher polarizability.

For our study, all the *ab initio* calculations were performed using the computational package NWCHEM[30]. The 6-311++G(3df,3pd) basis set was used for the calculation of the Hartree-Fock energy, then added electron correlation effects to the MP2. basis set superposition error (BSSE) was corrected for via the half counterpoise method[31,32]. For a summary of the *ab initio* methods, refer to Chapter II.

The values of the new ion-solute interaction potentials are shown in Table 4-3. As before, the electrostatic charges on the ions and solutes were still set to their respectively isolated value ($^{+}$1e for Na^{+} and $^{-}$1e for Cl^{-}) and the polarizability was set to the relevant experimental value[24,25,28]. The parameters for the new ion-solute potentials are shown in Table 4-4. All of the figures showing the parameterization are shown in Appendix B.

Results from new potentials

The new potentials were tested by comparing values for the excess chemical potential and the Setchenow coefficients to those calculated using the original parameters, as well as experiment. Figure 4-21 to 23, shows the Setchenow coefficient for both the new and original potentials using both the TIP4P/2005 and TIP4P/Ew water models. We note that for the argon and krypton solutes, shown in Figure 4-22 and 23 respectively, both show a significantly improvement over the Lennard-Jones parameters taken from the literature, shown in Table 4-2. On the other hand, for neon, shown in Figure 4-21 had poor agreement. Is was due to the fact that the Ne-Cl potential produced from the *ab initio* calculations obtains a more repulsive potential for distances greater than 3.1Å. As a proof of concept, the new potential works exceeding well.

Table 4-3: Solute potential exp-6 parameters and polarizability (α), from fitting to MP2 point energies

		σ [Å]	ε [K]	λ	α [Å^3]
Neon	Na^+	2.833	142.666	12.765	0.40
	Cl^-	5.228	1.516	16.750	
Argon	Na^+	3.414	161.691	12.783	1.64
	Cl^-	4.326	105.278	12.620	
Krypton	Na^+	3.597	300.777	11.741	2.48
	Cl^-	4.597	59.927	13.904	

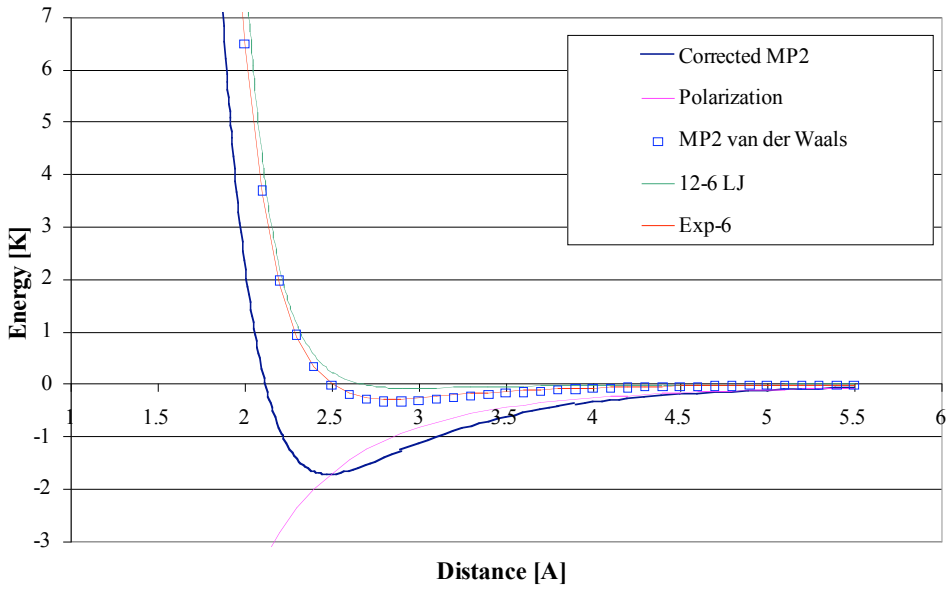


Figure 4-20: *Neon-sodium, dimer energy from the ab initio MP2 point energy calculations. Polarization energy, estimated van der Waals interaction and new exp-6 potential fitted to the van der Waals component of the ab inito MP2 calculation. The Dang-Hirshfielder potential generated from pure components and mixing rules are shown as a comparison.*

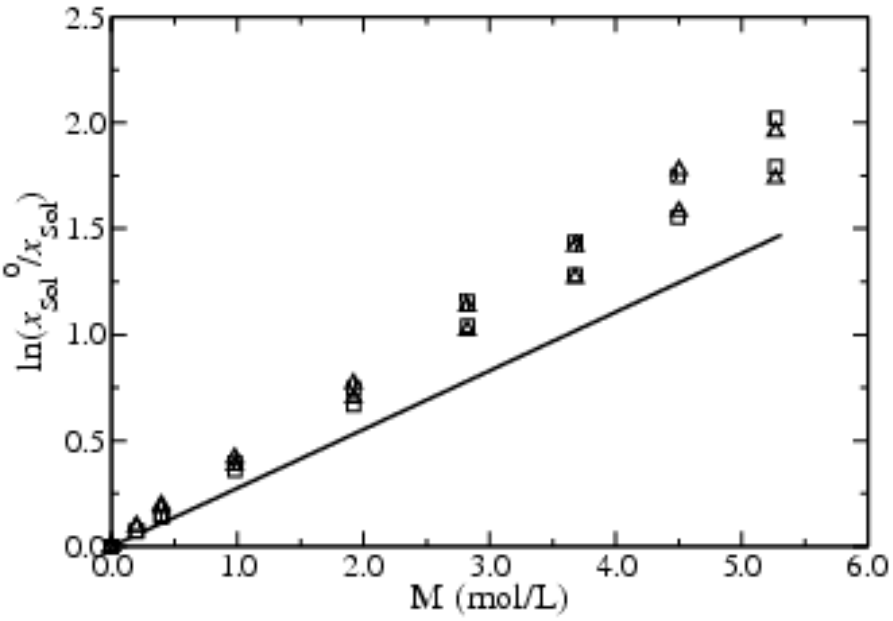


Figure 4-21: *NaCl concentration dependence of the Setchenow parameter of neon in aqueous solution at infinite dilution comparing the effect of the different ion-solute forcefields. Squares are for TIP4P/2005 and triangles for TIP4P/Ew. The bottom two rows are the old potentials and the top two rows are the new potentials*

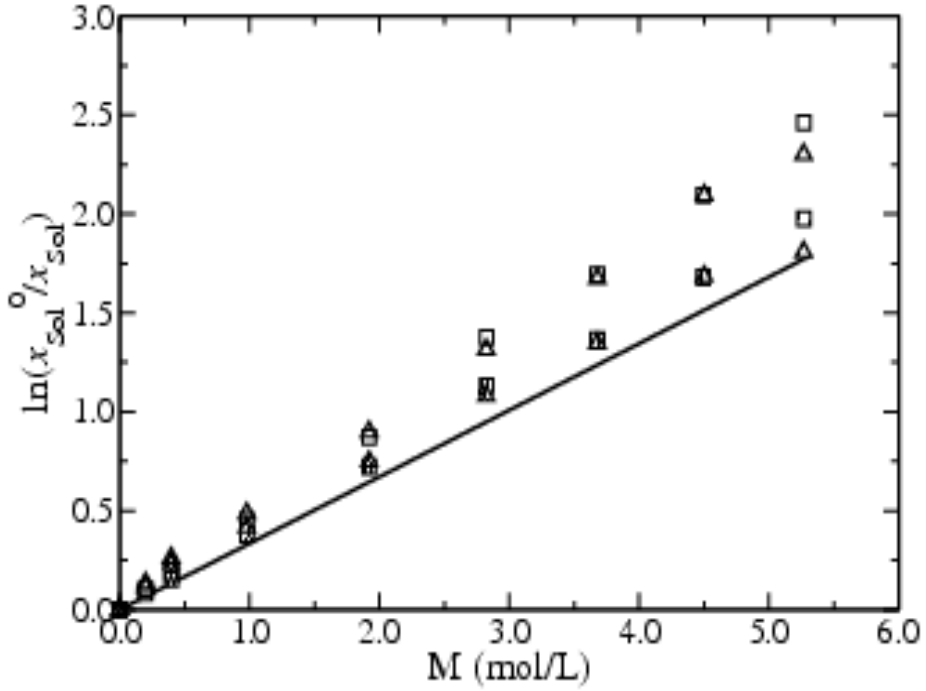


Figure 4-22: *NaCl concentration dependence of the Setchenow parameter of argon in aqueous solution at infinite dilution comparing the effect of the different ion-solute forcefields. Squares are for TIP4P/2005 and triangles for TIP4P/Ew. The top two rows are the old potentials and the bottom to rows are the new potentials*

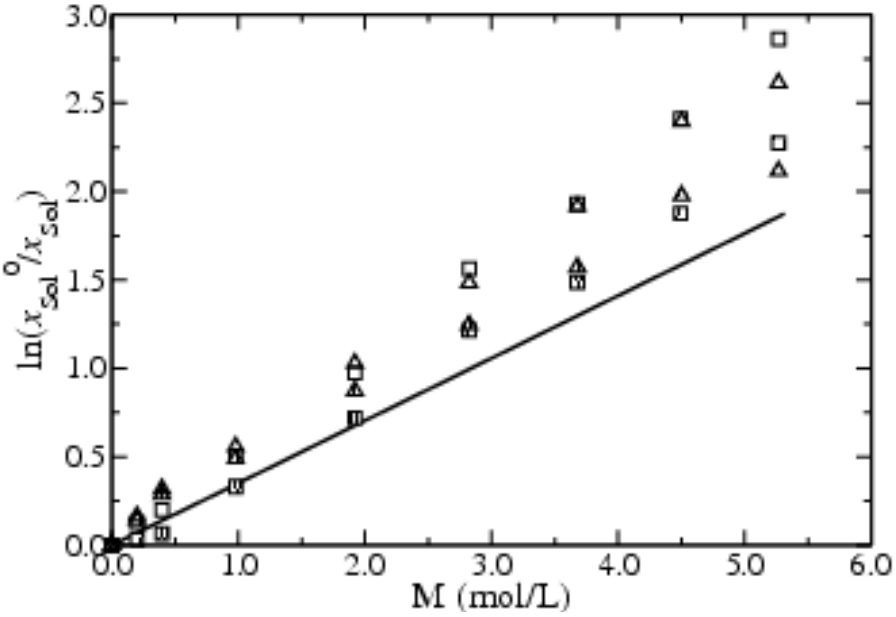


Figure 4-23: *NaCl concentration dependence of the Setchenow parameter of krypton in aqueous solution at infinite dilution comparing the effect of the different ion-solute forcefield, Squares are for TIP4P/2005 and triangles for TIP4P/Ew. The top two rows are the old potentials and the bottom to rows are the new potentials*

4.8 Conclusions

In this chapter, we began by showing that, when coupled with common rigid molecules, single site non-polarizable models of small non-polar solutes fail to capture either qualitatively or quantatively the effect on solubility of the charged co-solvents. Given that in chapter III we showed that such models fail to reproduce the temperature dependence in pure water, this is unsurprising. In fact, as shown in literature, these models over-predict the salting-out effects. Following on from our working in chapter III, we investigate the common hypothesis that this over-prediction in due to not accounting for the polarization on the solute explicitly. However, our investigation shows that polarization improves both the qualitative and quantitative agreement. In regards to the salting-out effect, it is a subtle improvement and not sufficient to provide the desired agreement with experiment.

In evaluation these results, we note that TIP3P appears to give better description of these solutes in aqueous NaCl solutions, we show that the cause of this agreement is the models under-prediction of the solution density. Furthermore, we demonstrate that if the solution density is fixed, all models used in this chapter provide similar values for the excess chemical potential, highlighting once again the importance of accurately reproducing the solvent density.

Turning our focus to the cause of the disagreement between experiments and simulations, we have considered the effect of salt on the structure of the fluid. Dividing the system into three regions, 1^{st} hydration layer, 2^{nd} hydration layer and bulk with respects to the ion-oxygen distance, we found that the bulk density decreases with increasing ion concentration. However, this does not result in a reduction in the excess chemical potential as would be expected from pure water results. We also see a similar

situation with the 2nd hydration layer, as the density also goes down but the excess chemical potential goes up. Given that the excess chemical potential in the bulk remains constant, underlying cause of salting-out can not be enhanced structure of the bulk fluid.

The 1st hydration layer shows an increase in the excess chemical potential as the ion concentration increases, so we conclude that the salting-out occurred because we replacing solute-water interactions with solute-ion interactions, which are less attractive. Thus we propose that the course of the over-prediction of salting-out by the model studied here is due to inappropriate ion-solute potentials.

As proof of this hypothesis we have developed now ion-solute potentials based upon *ab initio* calculations. As such, these potentials are purely predictive and do not required experimental data which is often scares even for common molecules such as CH_4. We compare the Setchenow coefficient using these new models with values obtained earlier in this work. We note they provide better prediction of krypton and argon. While we hoped for better agreement experiment, this models are pure predictive.

References

[1] E. D. Sloan, Nature **426** (6964), 353 (2003).

[2] I. Chatti, A. Delahaye, L. Fournaison, J. Petitet, Energy Conversion and Management **46**, 1333 (2005).

[3] C. Scharnagl, M. Reif, Josef Friedrich, Biochimica et Biophysica Acta **1749** (2), 187 (2005).

[4] J. M. Broering, A. S. Bommarius, J. Phys. Chem. B **109**, 20612 (2006).

[5] K. D. Collins, Biophysical Journal **72** (1), 65 (1997).

[6] F. Hofmeister, Arch Exp. Pathol Pharmacol **24**, 247 (1888).

[7] A. C. Dumetz, A. M Snellinger-O'Brien, E. W. Kaler, A. M. Lenhoff, Protein Science **16** (9), 1867 (2007).

[8] P. H. von Hippel, T. Scheich, Accounts of Chemical Research **2** (9), 257 (1969).

[9] A. K. S. R. Lenerman, Nature **378** (6555), 364 (1995).

[10] P. M. Wiggins, Cell Mol. Biol **47**, 735 (2001).

[11] Y. Zhang, P. S. Cremer, Current Opinion in Chemical Biology **10**, 658 (2006).

[12] A. W. Omta, M. F. Kropman, S. Woutersen, H. J. Bakker, Science **301**, 347 (2003).

[13] H. Docherty, A. Galindo, E. Sanz , and C. Vega, Journal of Physical Chemistry B **111**, 8993 (2007).

[14] B. Widom, J. Chem. Phys **39**, 2808 (1963).

[15] S. J. Plimpton, J. Comp. Phys. **117**, 1 (1995).

[16] R. W. Hockney, *Computer simulation using particles*. (McGraw-Hill International Book Company, New York, 1981).

[17] D. Wolf, P. Keblinski. S.R. Phillpot, J. Eggebrecht, Journal of Chemical Physics **110**, 8254 (1999).

[18] P. Demontis, S. Spanu, G. B. Suffritti, J. Chem. Phys **114**, 7980 (2001).

[19] C. Avendano, A. Gil-Villegas, Mol Phys **104** (9), 1475 (2006).

[20] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Jura, T. Head-Gordon, J. Chem. Phys **120**, 9665 (2004).

[21] J. Abascal, C. Vega, J. Chem. Phys **123**, 234505 (2005).

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

[23] D. E. Smith, L. X. Dang, J. Chem. Phys **100** (5), 3757 (1994).

[24] A. T. Wong, G. B. Bacskay, N. S. Hush, Mol Phys **74** (5), 1037 (1991).

[25] D. R. Lide, *CRC Handbook of Chemistry and Physics*, 88 ed. (CRC Press, 2007).

[26] Robert H. Perry, and Don W. Green, *Perry's Chemical Engineers' Handbook*, 7th ed. (McGraw-Hill International Editions, 1997).

[27] P. J. Dyer, H. Docherty, and P. T. Cummings, Journal of chemical physics **129**, 024508 (2008).

[28] D. Paschek, J. Chem. Phys **120** (14), 6674 (2004).

[29] H. Docherty, A. Galindo, C. Vega, E. Sanz, J. Chem. Phys **125**, 074510 (2006).

[30] T. P. Straatsma, A Computational Chemistry Parkage for Parallel Computers **Version 4.7** (Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA) (2005).

[31] S. F. Boys, and F. Bernardi, Molecular Physics **19**, 553 (1970).

[32] K. S. Kim, P. Tarakeshwar, J. Y. Lee, Chemical Review **100**, 4145 (2000).

CHAPTER V

HYDROGEN BONDING AND INDUCED DIPOLE MOMENTS IN WATER: PREDICTIONS FROM THE GAUSSIAN CHARGE POLARIZABLE MODEL AND CAR-PARRINELLO MOLECULAR DYNAMICS

5.1 Summary

We compare a new classical water model, which features Gaussian charges and polarizability (GCPM) with *ab initio* Car-Parrinello simulations (CPMD). We compare the total dipole moment, the total dipole moment distribution and degree of hydrogen bonding at ambient to supercritical conditions. We also compare the total dipole moment calculated from both the electron density ('Bader' approach) and from the center of localized Wannier functions (WFC). Compared to CPMD, we found that GCPM over-predicts the dipole moment derived by the 'Bader' approach and under-predicts that obtained from the WFCs, but exhibits similar trends and distribution of values. We also found that GCPM predicted similar degrees of hydrogen bonding compared to CPMD and has a similar structure. The research reported in this chapter has been published[1].

5.2 Introduction

Water is an essential and ubiquitous component of our natural environment crucial to many chemical, biological and physical processes. The importance of water has led to the development of many models for the interaction between water molecules (i.e. forcefields)[2]. Many of these models, most of which are rigid and non-polarizable, reproduce a subset of the important physical properties of water.

Recently, the Gaussian charge polarizable model for water (GCPM) was introduced[3]. This model has been shown to predict accurate structural, dielectric, thermodynamic and transport properties over the entire fluid range (including isolated dimers) as well as accurately reproducing the vapor-liquid equilibria. Additionally, GCPM has recently been shown to reproduce both the second and third virial coefficients[4]. The isolated GCPM water molecule has been fixed to the experimental gas phase geometry[5]. The center of negative charge (M) is located 0.27Å, along the θ_{HOH} bisector and the charges are set for the permanent dipole to equal the experimental value of 1.855D[6]. A point polarizable dipole is located at the center of mass of the molecule, and the polarizability is given by the experimental value ($\alpha = 1.44$ Å^3)[7]

The accuracy of the GCPM suggest that it may be representing the properties of the water molecule faithfully, both in the low density (isolated and dimer) states and in the condensed liquid state. To test the accuracy of the GCPM further, we present in this chapter comparisons of the GCPM predictions for the induce dipoles moment, quadruple moments and hydrogen bonding with results obtained from Car-Parrinello molecular dynamics[8] (CPMD).

CPMD simulations represent one of the most fundamental approaches to modeling condensed phases. Simulations have been conducted at both standard[9-11] and elevated conditions.[12,13] Dynamic equations of motion are solved for the ions with the inter-ionic forces computed from the valence electron density, which is solved for at each time step using density functional theory. In the case of water, a CPMD simulation consists of O and H ions and 8 valence electrons per molecule. *Ab initio* simulations do not restrict the atoms into rigid molecules, allowing the atoms to move freely. The ground state electron density calculated for the system is computed at each time step. As polarization and other

short range forces are taken into account automatically within the CPMD, this provides useful data to compare to the polarization included in the GCPM model.

Previous studies[9,14] have shown that the bulk water structure can be described by CPMD. Despite this success, CPMD remains too expensive computationally for the study of large and complex systems, since even for a relatively small (32-64 molecules) simulation, CPMD requires approximately 4-5 orders of magnitude greater computation time than an equivalent classical MD simulation.

Recent x-ray diffraction measurements[15] have resulted in an estimate of the bulk water dipole of 2.9 ± 0.6D at ambient conditions. However, *ab initio* results calculated using MP2 yield total dipole moments of 2.65D[16] and 2.7D[17,18].

In order to calculate the effective dipole moments of water molecules from electron density data, the boundary of the water molecule has to be defined. Laasonen *et al.*[9] used a spherical volume centered at the center of charge of a water molecule with a radius half the average O-O intermolecular distance, determined from the radial distribution function. The electron density contained within this volume was taken as being associated with the water molecule and the dipole computed accordingly, with the result that the average dipole moment $\bar{\mu}$ was found to be 2.66D at 300K and 1g/cc. Bader[19] defines the extent of the molecule as the surface where the electron density flux is zero, which we shall refer to as the Bader approach. Delle Site *et al.*[20] used this definition and calculated a total dipole moment of 2.47D in a simulation containing 32 water molecules at 300K and 1g/cc. We have adopted a similar philosophy in the results described below.

However, defining a molecule via the partitioning of the bulk electron density has been criticized as yielding incorrect results in regards to the total dipole whenever there is significant molecular overlap[21]. A solution to the lacking a formal definition of the molecular 'space' or volume has been proposed that uses maximally localized Wannier functions.[22,23] Wannier functions are a set of orthogonal functions used to represent the real-space picture of localized orbitals. As the Wannier functions are strongly non-unique, maximally-localized Wannier functions (MLWF)[24,25] constrain then system to find the Wannier functions which minimize their distribution about their center. The center of a MLWF can be thought of classically as the location of an electron pair. This method was used by Silvestrelli and Parrinello,[10] yielding a total dipole moment of 3.0D for bulk water at 318K and 1.0g/cc. For comparison, in addition to the Bader approach, we also calculate the dipole and quadrupole moments using Wannier functions centers (WFC).

The effect of temperature on the induced water dipole has not been extensively studied. Gubskaya and Kusalik[11], reporting the total molecular dipole moment between 263 and 373K. Only limited density dependence of the effective dipole moment has been reported. In this paper, we will present comparisons of the effective dipole moments by CPMD as a function of density and temperature outside the two-phase region, and compare these with the predictions of the GCPM

Another very important aspect of water structure is hydrogen bonding. For example, the high melting and boiling point of water are due to the high degree of hydrogen-bonding at ambient pressure. Likewise, the density maximum at 4^0C is the result of water freezing into a tetrahedral hydrogen-bonded network. In order to calculate the degree of hydrogen bonding in a simulation, Mezei and Beveridge[26] developed a geometric criterion for determining a hydrogen bond. Although other definitions (based on

energetic or combination energetic/geometric criteria) exist, the geometric definition of Mezei is preferred by us since it is easily implemented within CPMD, where the pair interaction energy is problematic to isolate. Using the Mezei and Beveridge geometrical criterion, we compare the distribution of the number of the hydrogen bonds obtained from GCPM and CPMD simulations as a function of temperature and density.

For a comparison to the GCPM and CPMD values for the dipole, we are using the simple rigid non-polarizable SPC/E[27] water model with point charges on each atom. Given its simplicity compared to GCPM and CPMD, SPC/E water model gives accurate predictions for the structure, phase equilibria and dielectric constant, particularly at ambient conditions. In this study, SPC/E is used as prototypical of the popular non-polarizable, rigid model and is compared to GCPM and CPMD results. Being non-polarizable and inflexible, SPC/E has a constant dipole moment of 2.35D.

In this paper, we will first present our method for simulating the water systems, followed by the calculations of the dipole moment, quadruple moment and the extent of hydrogen bonding. We conclude with discussion of the results.

5.3 Simulation Details

GCPM computational details

The GCPM water simulations were conducted between the temperatures of 300 and 1000 K and density of 0.1 and 1.0g/cc with 100K and 0.1g/cc intervals respectively. However, simulations were not performed in the two-phase vapor-liquid equilibrium region. A constant number of molecules, constant volume and constant temperature (NVT) ensemble of 256 water molecules was used with an equilibration time of 100ps followed

by a 100ps production run where data on dipole moment and hydrogen bonding were collected for statistical analysis. For these simulations, a 1.0fs timestep was used.

In regards to the calculation of the dipole moment, at each time step the induced dipole was iterated over all the molecules until convergence, creating a self-consistent field (SCF). The dipole moment was calculated from

$$\vec{\mu}_i = \vec{\mu}_i^{\,0} + \alpha\left(\vec{E}_i^{\,perm} + \vec{E}_i^{\,induced}\right) \qquad [5\text{-}1]$$

where $\vec{\mu}_i$ is the total dipole moment, $\vec{\mu}_i^{\,0}$, is the permanent dipole moment for electrostatic charges, α is the polarizability for water (1.44Å)[7], $\vec{E}_i^{\,perm}$ is the electric field caused from electrostatic charges, and $\vec{E}_i^{\,induced}$ is the electric field caused from the induced dipole.

CPMD computational details

The simulation was carried out with version 3.9.1 of the computational code CPMD[28]. In each simulation, 32 water molecules were simulated in a periodic cubic cell with a temperature between 300 and 1000K and density of 1.0g/cc. Initial samples were prepared by a 100ps initialization with an NVT ensemble using the GCPM at the same conditions. Due to its agreement with experiment[14], the Troullier-Martins type pseudopotentials[29] and the gradient-corrected density functional BLYP [30,31] were used.

The Kohn-Sham orbitals were expanded in a planewave basis set up to an energy cutoff of 70Ry, which has been shown to be accurate by the work of Silverstrelli and Parrinello[10]. A fictious mass for the electrons of 600 a.u. was used in the thermosetting equation of motion. Each simulation began with a 1ps relaxation run, involving 0.5ps simulation with velocity rescaling followed by 0.5ps with the temperature controlled with the Nose-Hoover thermostat. A timestep of 0.1fs was used for all the simulations. At

each temperature and density, there was a 5.0ps production runs where the electron density of the system was outputted every 0.5ps and the WFCs were calculated every 1fs. For the simulations at 1.0g/cc, the temperature was set at 300, 500, 700 and 1000K and for the simulations at 0.6g/cc the temperature was at 700 and 1000K.

In comparing the GCPM predictions with CPMD, we must keep in mind one limitation of the CPMD results. The GCPM model has been constructed as a classical model for use in classical simulation and so fits experimental data directly. This means that effects due to quantum dynamics in water that result in higher diffusivity, reduced hydrogen bonding, lower dielectric constant and a more relaxed structure[32-36] have already been incorporated into the classical GCPM intermolecular potential implicitly. By contrast, in CPMD simulations the dynamics is classical, but the interactions, derived from density functional theory applied to each configuration, have not been empirically adjusted to incorporate the effects of quantum dynamics. Hence, if the quantum-corrected properties of the CPMD are expected to agree well with experiment, one should expect that CPMD will yield lower diffusivities, higher dielectric constants, higher degrees of hydrogen bonding and a more rigid structure than real water. This has been confirmed with respect to lower diffusivities[37] and enhanced structure[37] and the degree of hydrogen bonding[38].

Isolated CPMD water molecule

A CPMD simulation of a single water molecule was carried out at 300K in a 12Å periodic cubic cell. Statistics were taken over 10ps, with an electron density snapshot every 0.5ps and WFC calculated every 1fs. The dipole was calculated using both the Bader approach and Wannier functions. In the Bader approach, the dipole was calculated

by centering the box on the water molecule, using the box walls to define the limits of the molecule.

Dipole moment calculations from electron density

The electron density generated from the CPMD simulations was mapped onto a three-dimensional grid with a 0.09Å spacing in each direction. In the Bader method (*Figure 5-1 a, b*), a boundary point of a water molecule was defined as the minimum in the electron density between two water molecules. A numerical spline interpolation was used to find minima in the electron density, corresponding to points of zero electron flux. The electron density is only calculated from the valence electrons, i.e. the 1s electrons are included with the oxygen nucleus. Integration over the molecule's space returns the number of valence electrons associated with that molecule. In principle, there are 8 valence electrons, six for the oxygen and 1 for each of the hydrogens. The number of electrons found within the boundary defined by the zero-flux condition is 8 ± 1% which is in excellent agreement with the expected value. The water dipole moment was then calculated from

$$\vec{\mu}_r = \int_{v_i} \vec{r}\, q(\vec{r}) d\vec{r} \qquad [5\text{-}2]$$

where $q(r)$ is the electron density and v_i is the volume around molecule i whose boundary is defined by the zero-electron-flux condition.

Wannier functions

In the WFC method (*Figure 5-1c*), the 8 valence electrons from each water molecule are divided into 4 double-occupied localized Wannier functions, each with a

charge of 2 electrons, concentrated at the WFCs. The positive charge is located at the atom nuclei, with a charge equal to its number of valance electrons, i.e. $6e^+$ on the oxygen and $1e^+$ on each of the hydrogens.

Quadrupole moment

The quadrupole moment of a water molecule is the next term in the multipole expansion of the electrostatic energy following the dipole moment[39]. Like the total dipole moment, the quadrupole moment of water is also unknown for condensed water, as there has not been much investigation into the value of the quadrupole moment. However, there has been limited investigation into effect of the quadrupole moment on the properties of water obtained using a variety of common models for water. Abascal *et al.*[40] have shown that quadrupole interactions play an important role in common water models, by comparing the magnitude of the quadruple moments of several rigid water models to the predicted melting temperature of ice. Tu and Laaksonen,[16] calculated the quadruple moments of water in liquid water by simulating 256 TIP3P molecules at 298K and a density of 0.997g/cc. The electrostatic potential (ESP) of the configuration is solved as a point energy calculation using *ab initio* MP2 level of theory. Finally the point charges are fitted to the electrostatic potential and determined the dipole and quadrupole moments.

The quadrupole moment for a molecule with discrete charges, for example using WFC, maybe calculated via,[39]

$$Q_{ij} = \frac{1}{2}\sum_{n=1}^{N} q_n \left(3x_i x_j - r^2 \delta_{ij}\right) \qquad [5\text{-}3]$$

where, Q_{ij} is the traceless quadruple moment in the *ij* direction, q_i is the electrostatic charge at location i, x_i is the distance vector i, i.e. x, y, z. r is the distance from the charge site to the reference point. The value of the quadrupole moment is dependant on the location the quadrupole is calculated relative too, i.e. where r equals zero. To be able to compare with Tu and Laaksonen[16] the quadrupole moment was calculated using the at the center of mass of the water molecule as the reference point. δ_{ij} is the Dirac function, where

$$\begin{aligned} \delta_{ij} &= 1, \quad when \quad i = j \\ \delta_{ij} &= 0, \quad when \quad i \neq j \end{aligned} \qquad [5\text{-}4]$$

In calculating the quadrupole moment from integration of the electron density values obtained from the CPMD simulations, we used the same approach as for determination of the total dipole moment, i.e.

$$Q_{ij} = \frac{1}{2}\int_{v_i} q(\vec{r})\left(3\vec{r}\vec{r} - r^2\delta_{ij}\right)d\vec{r} \qquad [5\text{-}5]$$

Hydrogen bonding

The criteria of a hydrogen bond defined by Mezei and Beveridge,[26] has the distance between the oxygen atoms $R_{OO} \leq 3.3$Å , angles between the proton donor or the lone pair of electrons and the oxygens as θ_{HOO}, $\theta_{LPOO} \leq 45^0$ and the torsion between the proton and the lone pair as $\theta \leq 180^0$. The location of the lone pair was taken as the inflection of the OH vectors and rotated 90^0. The distance of the lone pair to the nuclei is not involved in the definition. The number of hydrogen bonds was computed every 200fs over the 100ps production run for the GCPM and every 1.0fs, over a 5ps production run for CPMD.

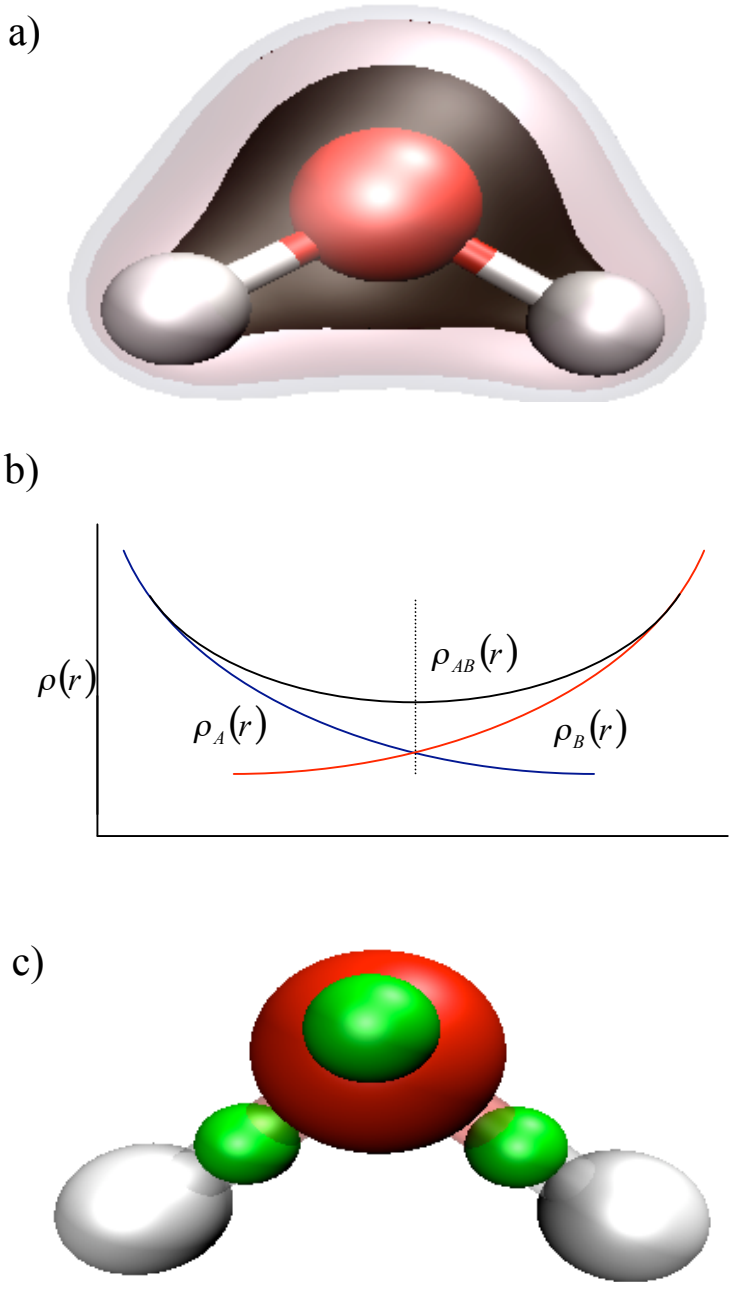


Figure 5-1: *Diagrams of the two different methods of calculating the total dipole moment from CPMD simulations. a), Bader method the electron density around a water molecule in the XZ plane. The dark region in the center has the highest electron density, which decreases as the distance away from the water atomic sites increases. b) Division of the electron density between molecules A and B, where the dotted line is the division of the electron density from the Bader method.* $\rho_i(r)$ *is the electron from molecules i functional. c) The centers of the Wannier functions (green) form a tetrahedral around the water molecule, with 1 center approximately on each of the OH bonds and the other 2 around the locations of electrostatic minima.*

5.4 Results

Dipole moment

In Figure 5-2, the total dipole moments of CPMD and GCPM water are compared at 300K, 500K, 700K and 1000K at a density of 1.0g/cc. It can be seen that GCPM water results exhibit similar trends to the CPMD water results, with the GCPM results consistently 0.1 to 0.2 Debye higher in magnitude than the CPMD-Bader results and 0.2 to 0.3 Debye lower than the CPMD-WFC results. At a density of 0.6g/cc, the GCPM again predicts total dipoles that are 0.1 Debye above the CPMD-Bader results. For lower temperatures at a density of 0.6g/cc, water enters a two-phase region and so we do not consider these states. The CPMD-Bader result at 300K and 1.0 g/cc is consistent with that calculated by Delle Site *et al.*[20] using essentially the same method for determining the effective dipole moment, validating our procedure.

Since the total dipole moment for the water molecule fluctuates, the ability to reproduce the total dipole moment distribution is likely to be an important ingredient in developing an accurate model for water. In Figure 5-3, we have plotted the dipole distribution for CPMD using both Bader and WFC methods together, with that obtained from GCPM at two state conditions, 300K and 700K, both at 1g/cc.

The standard deviations at 300K and 1.0 g/cc were 0.22D, 0.30D and 0.19D for CPMD- Bader, CPMD-WFC and GCPM respectively. At 700K and 1.0g/cc the standard deviations were 0.27D, 0.33D and 0.24D for CPMD-Bader, CPMD-WFC and GCPM respectively. For a density of 0.6g/cc and 700K, the standard deviation were 0.23D, 0.27D and 0.21D for CPMD- Bader, CPMD-WFC and GCPM respectively.

The width of the distribution of the dipole was always found to be larger for the CPMD simulations, using either method, than for the GCPM simulations. The narrower

distribution of GCPM could be due to the rigid nature of the model, i.e. neither the O-H bond length nor the HOH angle is able flex. Flexing of the molecule enhances the occurrence of dipoles at the tails of the distribution. This is consistent with the finding of Allesch[41], who found a narrower total dipole moment distribution is obtained from a rigid molecule compared to the total dipole moment distributions from a freely moving molecule using Car-Parrinello simulations.

In general, we find that the mean dipole moment calculated using the GCPM is bounded above by the WFC method and below by the Bader approach. The one exception to this general observation is shown in Figure 5-4 where Boero *et al.*[13] also simulated CPMD water in the supercritical regime at 653K for densities of 0.32 and 0.73g/cc. These results suggests that at low density (0.32 g/cc) and high temperature (673 K) the GCPM prediction for the dipole moment may be higher than that obtained from the WFC method applied to CPMD by Boero *et al.*[13]. However we note that the CPMD results are somewhat noisy at this state condition, so that this conclusion is at best tentative. For the lower density of 0.32g/cc, both CPMD and GCPM have a maximum probability around 2.06D, corresponding to the dipole moment of a water dimer (2.1D)[18]. The multiple peaks of the CPMD curve at low density are likely due to insufficient simulation time, results in poor statistics. For the higher density of 0.73g/cc, both GCPM and Boero *et al.* results have a similar maximum dipole value, but with GCPM having a smaller distribution.

Since for GCPM it is relatively easy to determine the effective dipole moment over a broad range of temperature and densities, in Figure 5-5 we present these results. It is clear from these results that the dipole moment calculated from the GCPM has a near-linear relationship to the inverse temperature. A similar trend was observed, over a smaller temperature difference by Gubskaya and Kusalik[11].

Isolated water dipole

The isolated water dipole was calculated from CPMD-Bader to be 1.84D, with a standard deviation of 0.06D, compared to experiment of 1.855D[6], which corresponds to the value chosen for GCPM. CPMD-WFC yielded a dipole of 1.86D, with a standard deviation of 0.08D. Two conclusions can be inferred from this: first, that CPMD (using either method) yields a good prediction of the isolated water dipole; and second, CPMD and GCPM converge to the same value at lower densities (GCPM due to design). Consistent with this is the fact that the differences between CPMD and GCPM at the lower density of 0.6g/cc are less than the differences at 1.0g/cc (*Figure 5-3b, c*).

Quadrupole moment

As the Q_{zz} quadrupole moment is quite small, and the quadrupole is traceless, it has been suggested that a convenient measure of the strength of the quadrupole is Q_T, the average magnitude of the quadrupole moment, defined by,

$$Q_T = \frac{(Q_{xx} - Q_{yy})}{2} \qquad [5\text{-}6]$$

by Abascal and Vega[42]. This reduces the quadrupole moment to a single value enabling easier comparison between results. The values of the quadrupole moment of water calculated via CPMD, seen in Table 5-1, are in good agreement with published results which have been attained using similar methods[10,20]. Calculation of the Q_T using WFCs obtained a high value for the traceless quadrupole moment compared to the integration of the electron density over the molecule volume, consistent with the fact that the dipole moment calculated via WFCs was also higher than using the Bader method.

Also the results from Tu and Laakenson[16] are closest to the result obtained via the Bader method, which is again consistent with the fact that the dipole moment is also the closest to the value obtained via the Bader method. As the method used by Tu and Laakenson is significantly different to the Bader method, this starts to build a case for the true value of the quadrupole moment.

The quadrupole moment GCPM does not change, so is fixed as it isolated value. The quadrupole value for water is close to the experimental isolated values, but as it can not change, is significantly lower than the condensed phase values prediction from either method for CPMD of the results from Tu and Laakenson.

Structure

The radial distribution functions for GCPM and CPMD are shown in Figure 5-6 for comparison. Due to the small number of water molecules, the CPMD results do not extend beyond the second water shell. The agreement of the CPMD results with GCPM results is very good; the latter in turn are in very good agreement with experiments[3]. Note that CPMD predicts slightly enhanced structures (indicated by higher peaks) for water than GCPM, consistent with our discussion in Section 5.3 – that is, we expect CPMD to over-predict structure slightly compared to experiment.

As can be seen from Figure 5-7, the degree of the hydrogen bonding in GCPM water is similar to that calculated via CPMD. CPMD predicts a slightly more structured fluid as the temperature decreases, shown by a higher degree of hydrogen bonding. This is consistent with our remarks in Section 5.3 concerning CPMD, in that we expect it to have a higher degree of hydrogen bonding compared to experiment.

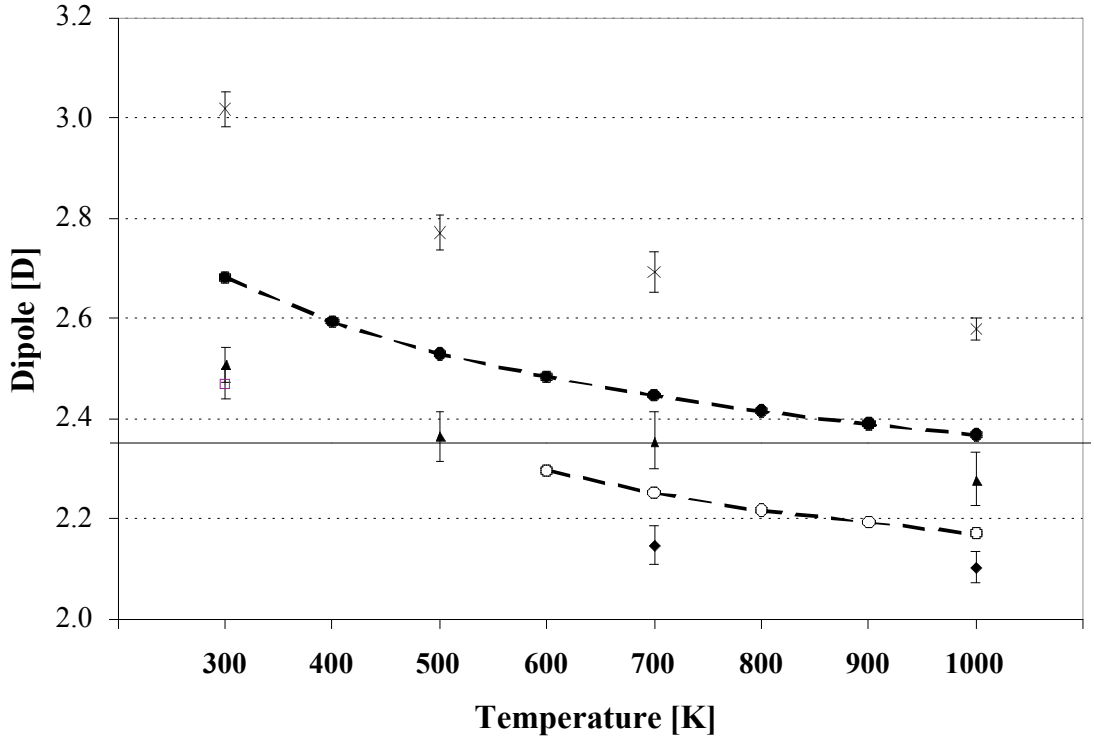
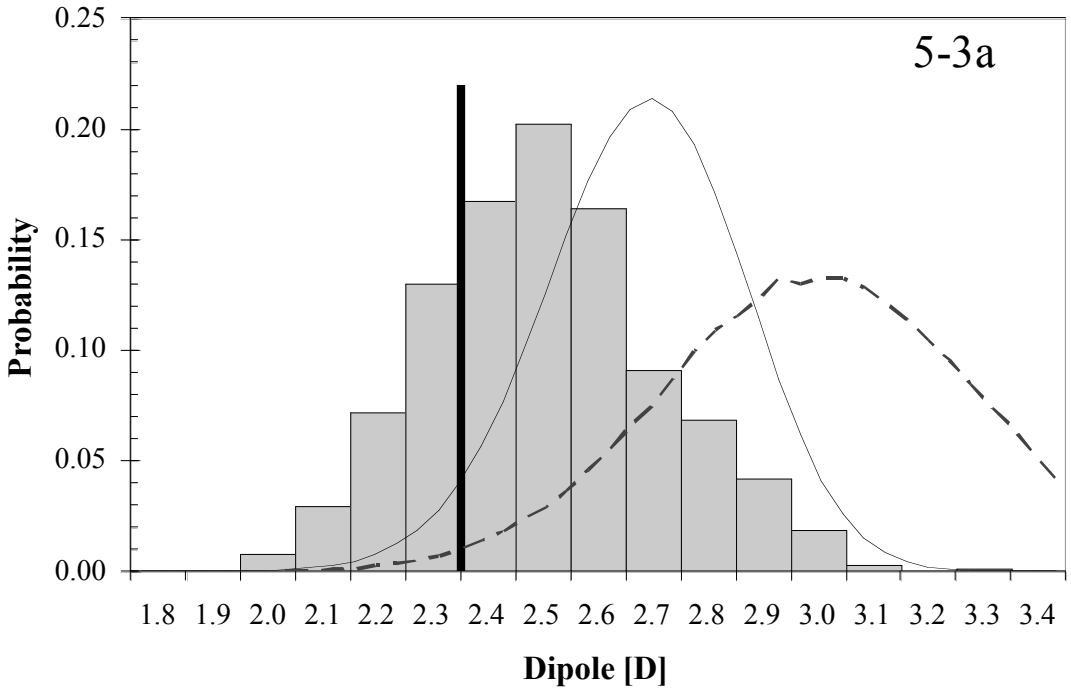
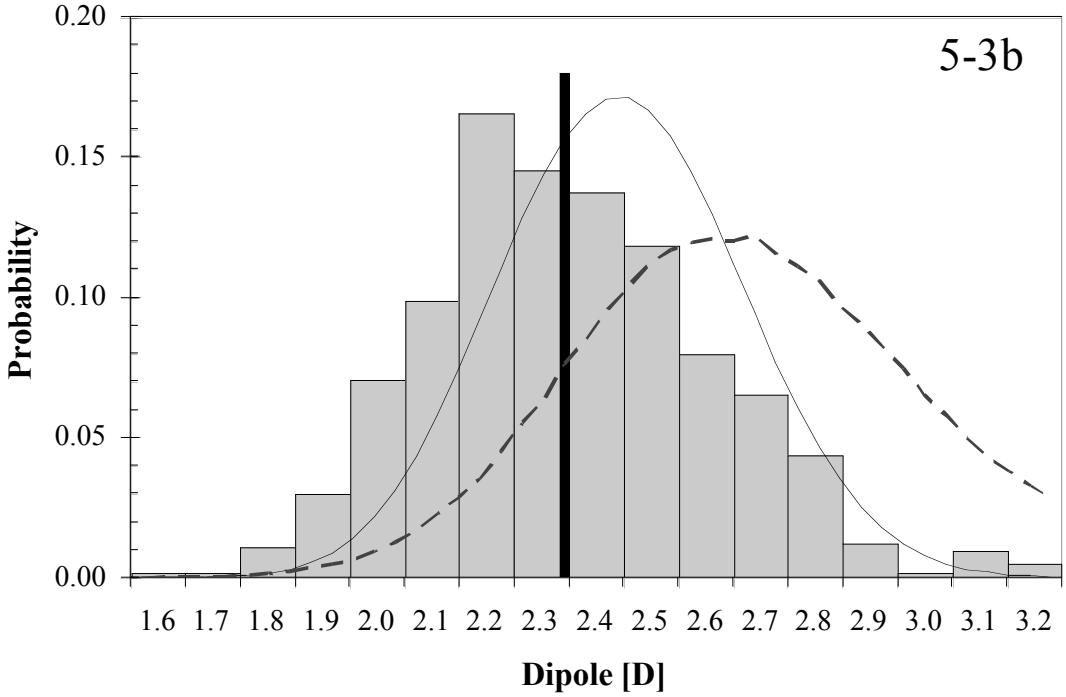


Figure 5-2: *Comparing the dipole moment of GCPM and CPMD water simulation at 1.0 g/cc and 0.6g/cc. The GCPM results of 1 g/cc and 0.6 g/cc are shown as filled and hollow circles respectively with a dashed line. The CPMD-Bader results at 1.0 g/cc and 0.6 g/cc are shown as filled triangles and diamonds respectively. The CPMD-WFC results at 1 g/cc are shown as crosses. Note that the result by Delle Site et al.[20] at ambient conditions is shown as an open square. The (constant) SPC/E value is shown as a solid horizontal line.*


5-3a
Probability
0.25
0.20
0.15
0.10
0.05
0.00
1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9 3.0 3.1 3.2 3.3 3.4
Dipole [D]


5-3b
Probability
0.20
0.15
0.10
0.05
0.00
1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9 3.0 3.1 3.2
Dipole [D]

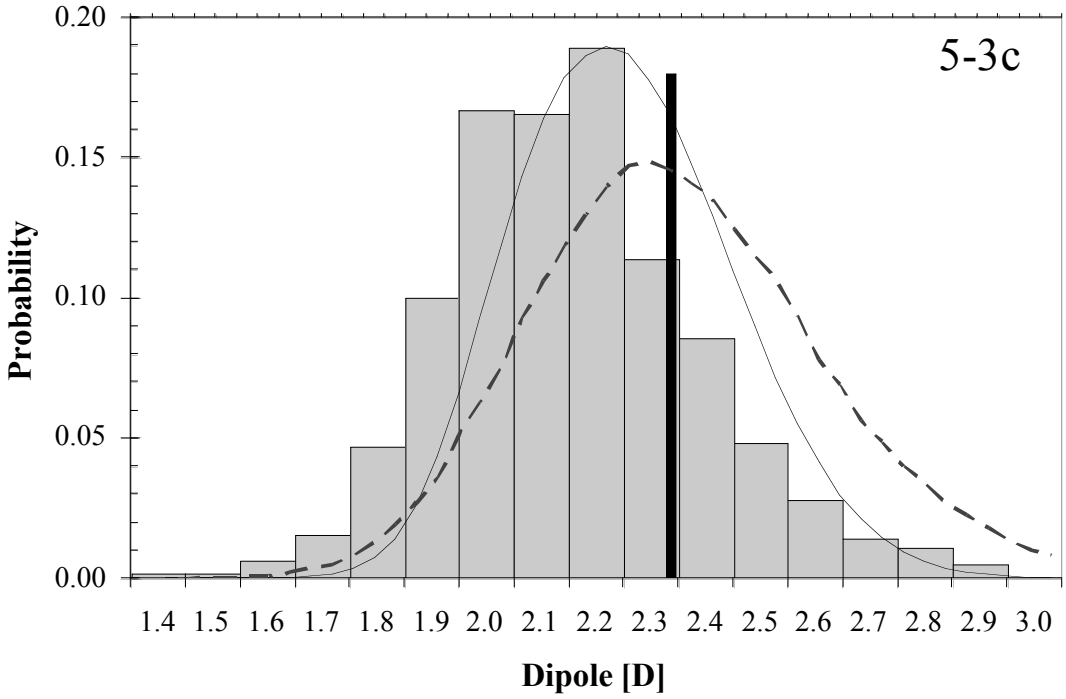


Figure 5-3: *Distributions of the dipole for water at 1 g/cc and temperatures (a) 300K, (b) 700K. (c) 0.6 g/cc at 700K. The CPMD-Bader results are shown as bars while the GCPM as smooth curves. The CPMD-WFC results are shown as a dashed line. For reference, the fixed dipole moment of the SPC/E model is shown as a single vertical line.*

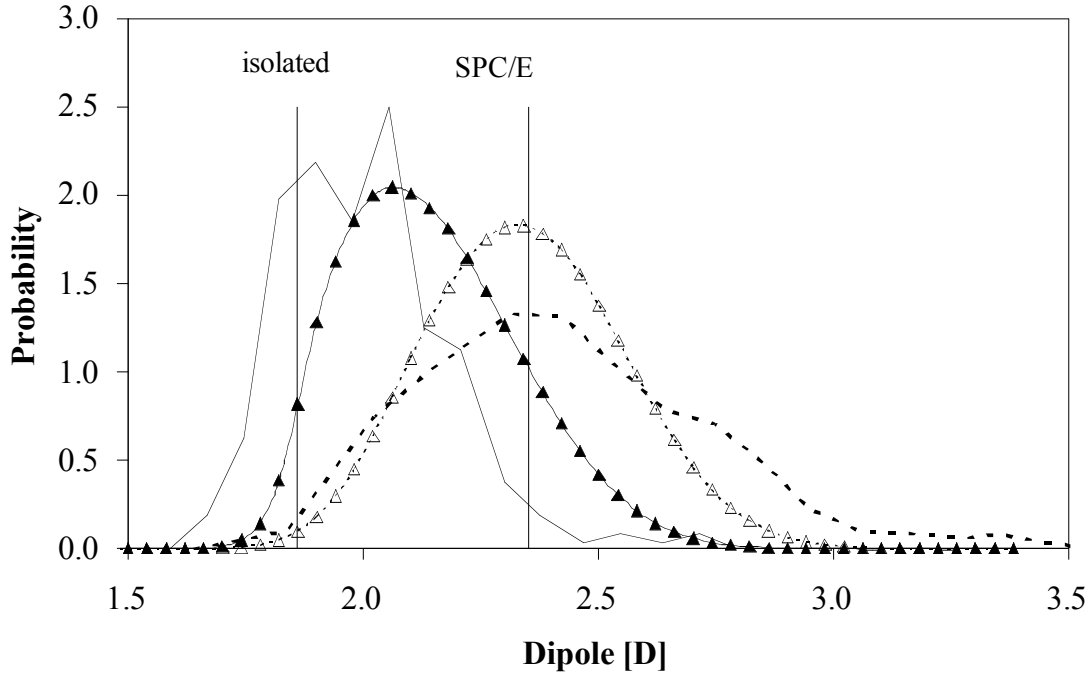


Figure 5-4: *Comparing dipole distribution at 0.32g/cc and 0.73g/cc and 653K for GCPM (triangles, solid 0.32 and hollow 0.73g/cc) and the data from Boero et al.*[13]*, (line, solid 0.32 and dash 0.73g/cc). The two vertical lines indicate the isolated water molecule value (1.85D) and the SPC/E value (2.35D)*

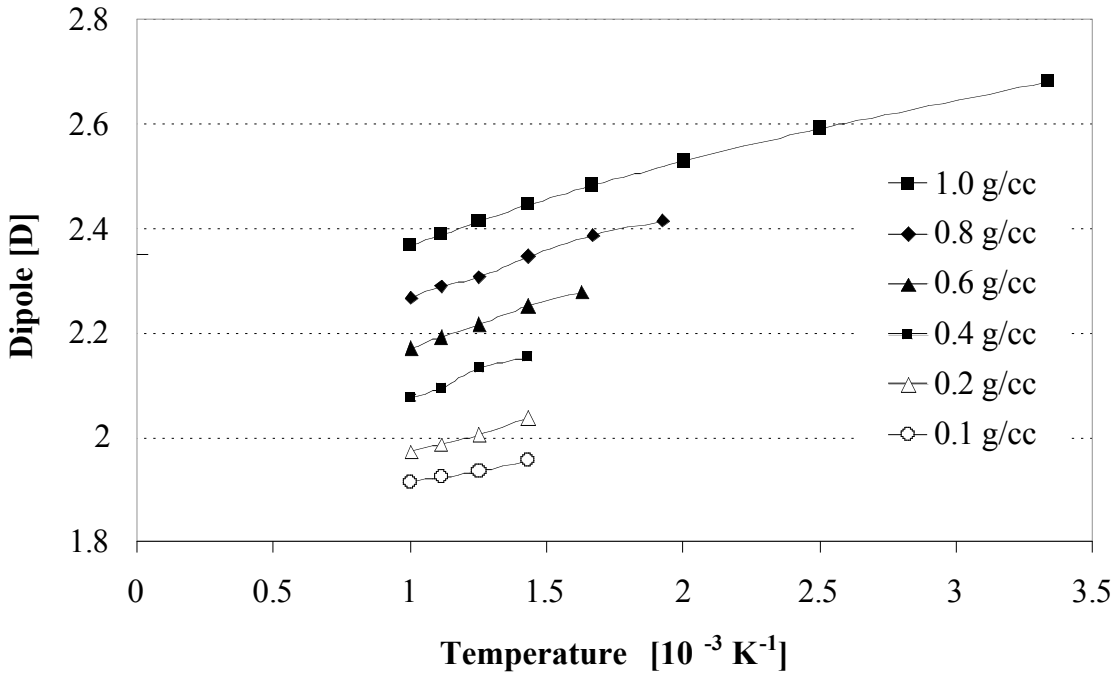
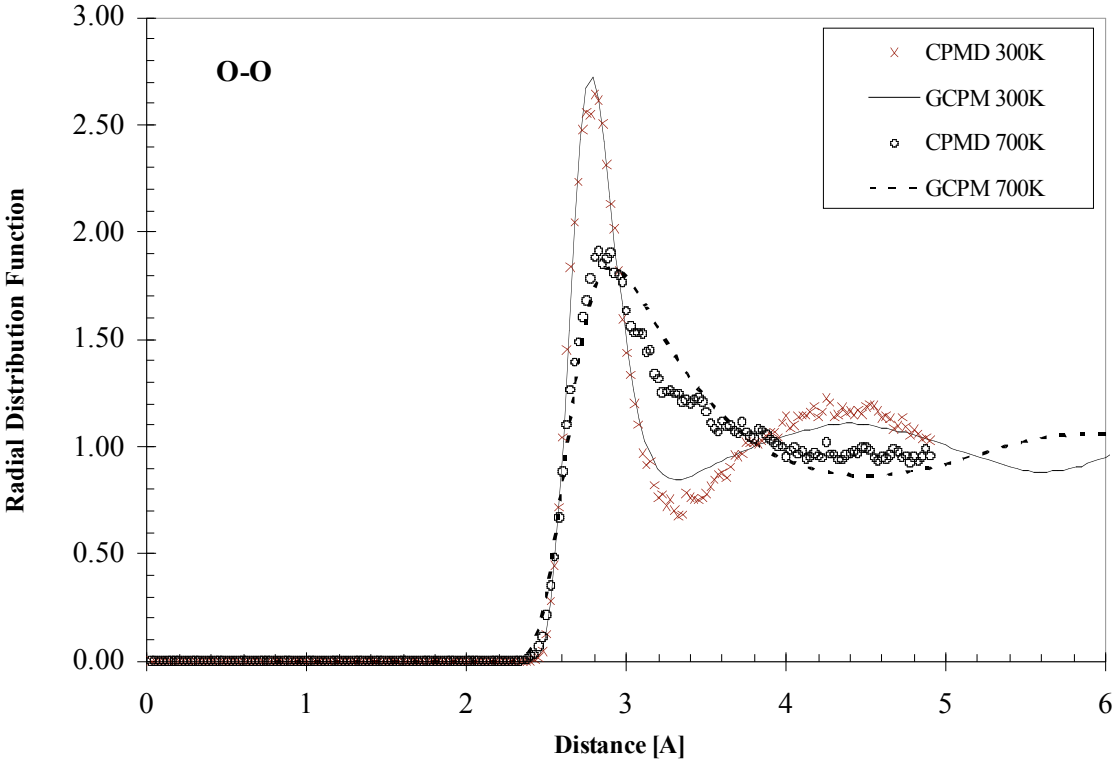
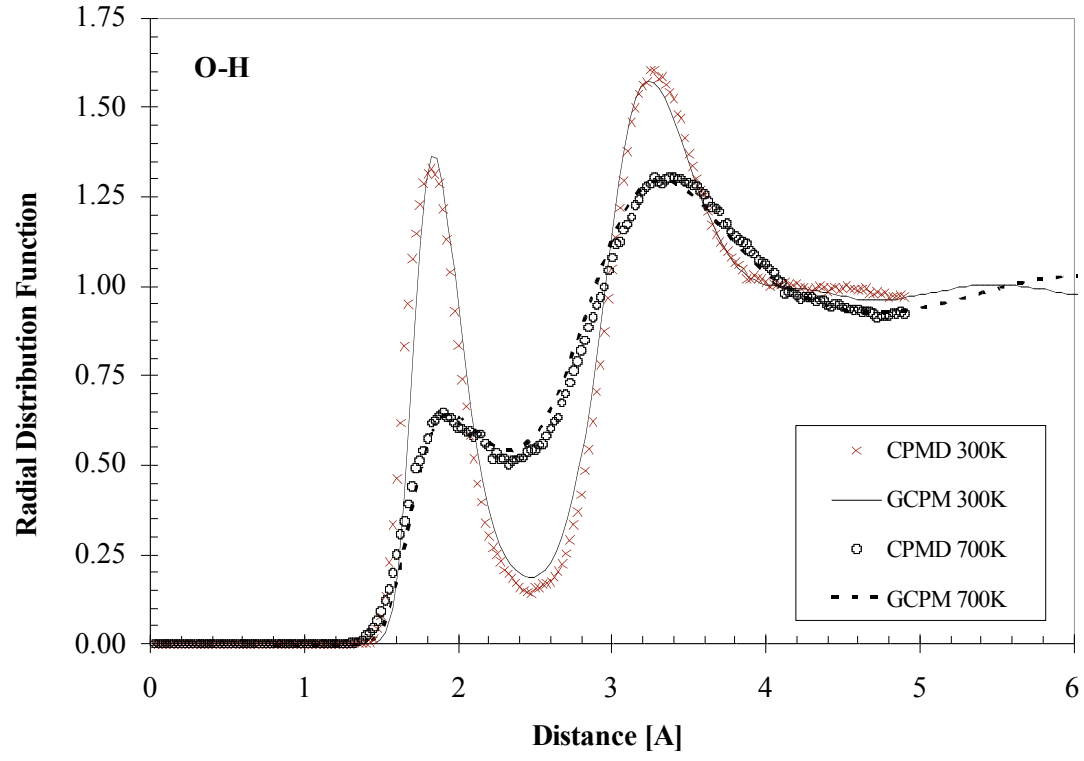


Figure 5-5: *GCPM water dipole moment as a function of inverse temperature for several densities, ranging from highest at the top to lowest at the bottom. SPC/E is shown as a solid line at 2.35D*


O-O
CPMD 300K
GCPM 300K
CPMD 700K
GCPM 700K
Radial Distribution Function
Distance [A]
3.00
2.50
2.00
1.50
1.00
0.50
0.00
0
1
2
3
4
5
6


O-H
CPMD 300K
GCPM 300K
CPMD 700K
GCPM 700K
Radial Distribution Function
Distance [A]
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0
1
2
3
4
5
6

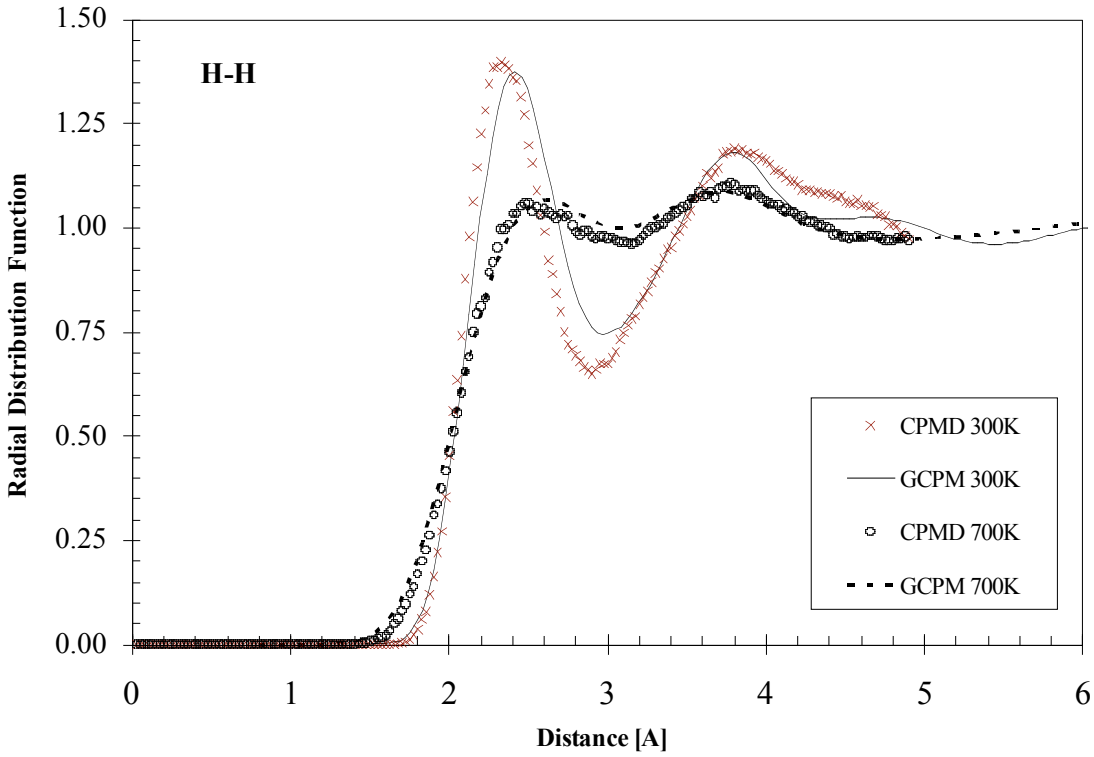


Figure 5-6: *Radial distribution function of GCPM and CPMD at 300 and 700K at 1.0g/cc. a) OO, b) OH, c) HH. The GCPM results are shown as solid line and dotted line for 300 and 700K respectively. The CPMD results are shown as 'x' and 'o' for 300 and 700K respectively.*

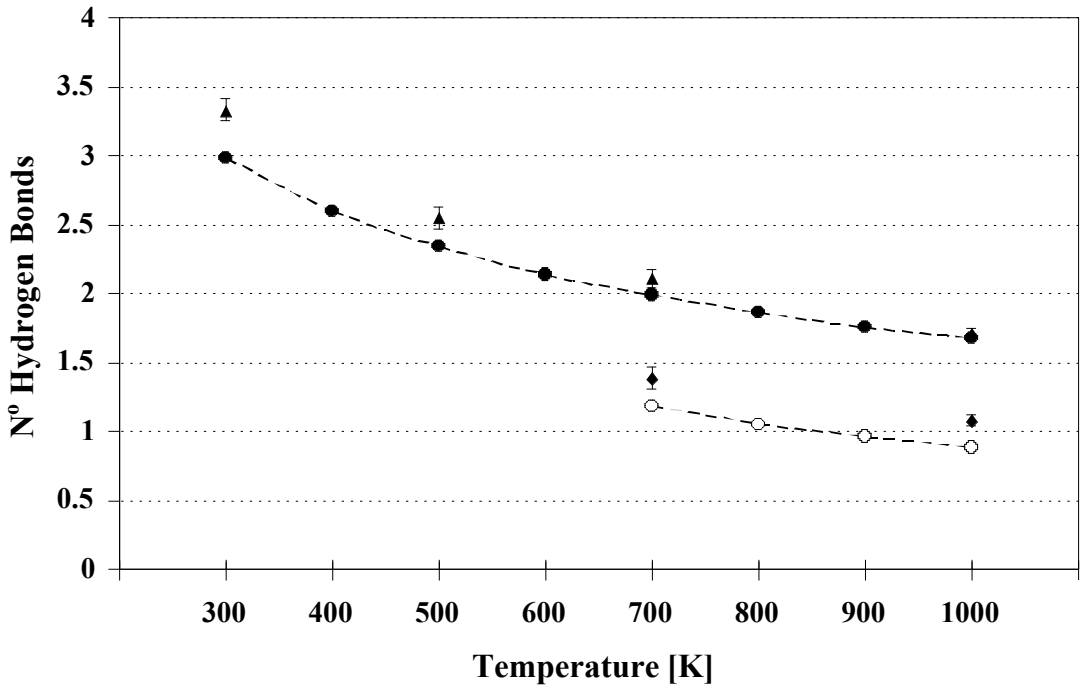


Figure 5-7: *Hydrogen bonding as a function of temperature. GCPM results of 1.0 and 0.6g/cc are shown as filled and hollow circles respectively. The CPMD results at 1.0 and 0.6g/cc are shown as filled triangles and diamonds respectively.*

Table 5-1: Quadrupole moment of water [DÅ]

	Q_{zz}	Q_{xx}	Q_{yy}	Q_T
Isolated Molecule[43]	-0.13	2.63	-2.5	2.57
TIP4P/2005[42]	-0.1256	2.3597	-2.2341	2.92
GCPM (isolated)[3]	-0.366	2.692	-2.355	2.49
CPMD (Bader)	-0.20 (sd 0.33)	2.86 (sd 0.37)	-2.66 (sd 0.18)	2.76
CPMD[20]	-0.10	2.77	-2.67	2.72
CPMD (WFC)	-0.07 (sd 0.45)	3.43 (sd 0.20)	-3.36 (sd 0.47)	3.35
CPMD[10]	-0.22	3.38	-3.16	3.27
MD/MP2[16]*	0.19±0.06	2.89±0.07	-2.70±0.17	2.80

* Results adjusted into the traceless quadrupole $q\vec{r}^{\,2} \rightarrow 0.5q\left(3\vec{r}^{\,2} - r^2\right)$

5.4 Conclusion

In this chapter, we have compared predictions for the dipole moment, quadrupole moment and hydrogen bonding from GCPM water to CPMD as a range of temperatures and densities. In general, the agreement between the two models is very good. Specifically, we have shown that GCPM yields predictions of the total dipole moment and the total dipole moment distribution, for a range of temperatures and densities that are in good agreement with the CPMD results. The GCPM values and are in fact bracketed by the values obtained from the two different methods from CPMD. The distribution of the induced dipole moment as similar between all three methods of calculation, but the GCPM was consistently lower. We expect that this is due to the rigid nature of the GCPM model, which prevents sampling at the extremes of the dipole moment distribution. We conclude that given the variation in the dipole moment, the value predicted by GCPM is an accurate as the total dipole moment obtained from CPMD simulations.

As expected, the total water dipole moment calculated using both GCPM and CPMD decreases with increasing temperature in the fluid region. We have shown that there is a near linear relationship between the dipole moment and inverse of the temperature.

The GCPM predicted a similar, but consistently slightly lower, degree of hydrogen bonding compared to GCPM over the whole range of state conditions studied. Likewise, the water structure predicted by both models is in good agreement, with the CPMD slightly more structured than GCPM.

Both CPMD and GCPM exhibit the experimental value for the dipole moment of the isolated water molecule (in the case of GCPM, this is by construction). The value of the induced dipole moment using CPMD was highly dependant on the methods of

calculation. Integration over the molecules electron density, (Bader method) on the molecular gave consistently lower results compared to the values calculated via maximally localized Wannier functions in the condensed phases. For the isolated case, both methods converged to the experimental values of the isolated molecule.

The quadrupole moment calculated by CPMD using the two different computational methods gave consistent values from other published values using similar methods. Like for the calculation of the dipole moment, the quadrupole moment calculated by the WFC was larger than the quadrupole moment from the Bader method. The values obtained used the Bader method are similar to the results from Tu and Laaksonen[16], even though it is obtained from significantly different calculation method.

References

[1] P. J. Dyer, and P. T. Cummings, J. Chem. Phys **125**, 144519 (2006).

[2] B. Guillot, Journal of Molecular Liquids **101**, 219 (2002).

[3] P. Paricaud, M. Predota, A. A. Chialvo, P. T. Cummings, Journal of Chemical Physics **122** (24), 4511 (2005).

[4] K. M. Benjamin, A. J. Schultz, and D. A. Kofke, Journal of Physical Chemistry C **111** (43), 16021 (2007).

[5] S. A. Clough, Y. Beers, G. P. Klein, Journal of Chemical Physics **59** (5), 2254 (1973).

[6] W. S. Benedict, N. Gailar, and E. K. Plyler, Journal of Chemical Physics **24**, 1139 (1956).

[7] F. H. Stillinger, *The Liquid State of Matter: Fluids Simple and Complex*. (North-Holland, 1982).

[8] R. Car, M. Parrinello, Physical Review Letters **55** (22), 2471 (1985).

[9] K. Laasonen, M. Sprik, M. Parrinello, Journal of Chemical Physics **99** (11), 9080 (1993).

[10] P. L. Silvestrelli, M. Parrinello, Journal of Chemical Physics **111** (8), 3572 (1999).

[11] A. V. Gubskaya, P. G. Kusalik, Journal of Chemical Physics **117** (11), 5290 (2002).

[12] E. Schwegler, G. Galli, F. Gyfi, R Q. Hood, Physical Review Letters **87** (26), 5501 (2001).

[13] M. Boero, K. Terakura, T. Ikeshoji, C. C. Liew, M. Parrinello, Journal of Chemical Physics **115** (5), 2219 (2001).

[14] M. Sprik, J. Hutter, M. Parrinello, Journal of Chemical Physics **105** (3), 1142 (1996).

[15] Y. S. Badyal, M.-L. Saboungi, D. L. Price, S. D. Shastri, and D. R. Haeffner, Journal of Chemical Physics **112** (21), 9206 (2000).

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

[17] A. V. Gubskaya, P. G. Kusalik, Molecular Physics **99** (13), 1107 (2001).

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

[19] R. F. W. Bader, *Atoms in molecules*. (Oxford University Press, New York, 1994).

[20] L. Delle Site, A. Alavi, R. M. Lynden-Bell, Molecular Physics **96** (11), 1683 (1999).

[21] I. Souze, R. M. Martin, N. Marzari, X. Zhao, D. Vanderbilt, Physical Review B **62** (23), 15505 (2000).

[22] G. H. Wannier, Physical Review **52**, 191 (1937).

[23] W. Kohn, Physical Review **115**, 809 (1959).

[24] N Marzari, D Vanderbilt, Physical Review B **56** (20), 12847 (1997).

[25] P. L. Silverstrelli, N. Marzari, D. Vanderbilt, M. Parrinello, Solid State Communication **107**, 7 (1998).

[26] M. Mezei, D. L. Beveridge, Journal of Chemical Physics **74** (1), 622 (1981).

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

[28] Copyright IBM Corp 1990-2004, CPMD (2004).

[29] N. Troullier, J. L. Martins, Physical Review B **43** (3), 1993 (1991).

[30] A. D. Becke, Physical Review A **38** (6), 3098 (1988).

[31] C. Lee, W. Yang, R. G. Parr, Physical Review B **37** (2), 785 (1988).

[32] G. S. Del Buono, P. J. Rossky, J. Schnitker, Journal of Chemical Physics **95** (5), 3728 (1991).

[33] J. Lobaugh, G. A. Voth, Journal of Chemical Physics **106** (6), 2400 (1997).

[34] H. A. Stern, B. J. Berne, Journal of Chemical Physics **115** (16), 7622 (2001).

[35] L. H. De La Pena, P. G. Kusalik, Molecular Physics **102** (9-10), 927 (2004).

[36] T. F. Miller III, D. E. Manolopoulos, Journal of Chemical Physics **123**, 154504 (2005).

[37] I. W. Kuo, C. J. Mundy, M J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack, M. Parrinello, Journal of Physical Chemistry B **108**, 12990 (2004).

[38] S. Izvekov, G. A. Voth, The Journal of Chemical Physics **123**, 044505 (2005).

[39] C. G. Gray, K. E. Gubbins, *Theory of Molecular Fluids Fundamentals*. (Clarendon Press, Oxford, 1984).

[40] J. L. F. Abascal, C. Vega, Physical Chemistry Chemical Physics **9**, 2775 (2007).

[41] M. Allesch, Journal of Chemical Physics **120** (11), 5192 (2004).

[42] J. L. F. Abascal, C. Vega, Physical Chemistry Chemical Physics **9**, 2775 (2007).

[43] J. Verhoeven, A. Dymanus, Journal of Chemical Physics **52** (6), 3222 (1970).

CHAPTER VI

EFFECT OF GAUSSIAN CHARGE AND POLARIZATION ON ION SOLVATION: BEHAVIOR OF A CHARGED SOLUTE AT INFINITE DILUTION AND IN CLUSTERS

6.1 Summary

We developed a polarizable Gaussian-charge forcefield for monovalent ions to be used in conjunction with the Gaussian charged polarized model (GCPM) for water. We found that the induced dipole moment and location of the ion in an ion-water cluster depended on whether the ion was a kosmotrope or a chaotropic. We found that within the first hydration shell, the induced water dipole moment is influenced by the ion, and that the magnitude of the induced water dipole moment is related to the size and charge of the ion. We see that the water-ion orientation still exhibits the same classical effects seen is other classical water models, which is still stronger, that the orientation seen from *ab initio* simulations.

6.2 Introduction

This chapter examines the effect of using a Gaussian distribution for modeling the electrostatic interactions of simple ions in aqueous solution. Previous polarizable classical water-ion models have been developed by Dang *et al.*[1], using a polarizable water model (RPOL)[2,3], Lamoureux *et al.*[4], using the polarizable water model (SWM4-DP)[5], both parameterize the ion-water models to reproduce the binding energies of gas-phase monohydrates and for Lamourux *et al.* also using the hydration free energies in the bulk liquid. Carrillo-Tripp *et al.*[6] developed a polarizable model for the hydration of Na^+ and

K^+ using *ab initio* calculations, using the polarizable and flexible water model (MCDHO)[7].

As the ion particles are parameterized to specific water model, there is a lack of transferability of ion-water potentials between water models. This is especially true given that the polarization can be accounted for using vastly different method. For the either Drude oscillators[8] used by Lamourux *et al* and Carrillo-Tripp *el al.* or polarization from the electric field, as used by Dang *et al.* Due to these transferability issues, it was necessary for us parameterize the ion-water interactions to used with the GCPM water model.

The Gaussian charge polarizable model for water (GCPM)[9] has been shown to yield accurate structural, vapor-liquid equilibria, saturated vapor pressure and transport properties over the entire fluid range. To accompany this accurate water model, a new model for ions is required, featuring polarizability and Gaussian charges to describe the electrostatic interactions.

The first objective is to look at the effect of using Gaussian distribution to model the ion electrostatic charges. We will do this by looking at differing widths of the Gaussian distribution, with differing size particles and differing magnitude and sign on the charge. The second objective of this paper is to parameterize an ion-water model compatible with the GCPM representing the electrostatic charge using a Gaussian function. Here we provide Gaussian parameters for the simple monovalent cations Li^+, Na^+, K^+ and the anions F^-, Cl^- and Br^-. The Gaussian distribution parameters were fitted to ion-water cluster energies[10-12], *ab initio* structure data[13-18], and ionic radii[19]. The third objective is in two parts. First to study the behavior of the ions in water clusters, and second, to examine the behavior of the ion in the bulk water

Forcefield

The model we propose has a positive Gaussian distribution for the cations, and a negative Gaussian distribution for the negative ions. The charge distribution $\rho_i(r)$ about a charge i is modeled by

$$\rho_i(r) = \frac{q_i}{\left(2\pi\Omega_i^2\right)^{\frac{3}{2}}} \exp\left(\frac{-r^2}{2\Omega_i^2}\right) \qquad [6\text{-}1]$$

where, Ω_i is the width of the Gaussian distribution about *i*, r is the distance from the center of the charge (nuclei of the ion) and q_i is the magnitude of the charge A point charge has a Ω_i value of 0. The electrostatic interaction energy between two charge distributions *i* and *j* is represented by

$$U(r) = \frac{q_i q_j}{2\left(4\pi\varepsilon_0\right)r} erf\left(\frac{r}{\left(\sqrt{2\left(\Omega_i^2 + \Omega_j^2\right)}\right)}\right) \qquad [6\text{-}2]$$

As these ion models are using polarizable, we will follow the same method as in the GCPM, which is via polarization due to the electric field. As in the GCPM, the induced dipole moment is calculated from,

$$\vec{\mu}_i = \vec{\mu}_i^0 + \alpha\left(\vec{E}_i^0 + \vec{E}_i^p\right) \qquad [6\text{-}3]$$

where $\vec{\mu}_i$ is the induced dipole moment on the ion, α is the ion's polarizability, $\vec{E}_i^0$ is the permanent electric field, from the electrostatic charges on the water molecules, and $\vec{E}_i^p$ is the induced electric field, from the dipole on neighboring molecules. $\vec{\mu}_i^0$ is the permanent dipole moment is calculated from the electrostatic charges and geometry of the molecule. For the simple ions, the permanent dipole moment has a value of 0. The induced dipole moment is iterated over all the molecules (water plus ions) at every time

step until convergence, creating a self-consistent field. The polarizability parameters for the ions were taken from the experimental data of Pyper *et al.*[20]

From equation 6-2, one can see that the electrostatic ion-water interaction is reduced as the width of the Gaussian distribution is increased. The magnitudes of the electrostatic charge on the ions are kept at their isolated values of +1e for the cation and -1e for the anion.

As these are still classical models, charge transfer are not accounted for, even though a lower magnitude value for the electrostatic charges for ions in solutions would be more realistic. The work of Tongarra and Rode[21] using QM/MM simulations on hydrated anions obtained values for the electrostatic charges to be -0.94e ± 0.01e and -0.89e ± 0.01e for F^- and Cl^- respectively. Implicity accounting for the reduced charge has the effect of reducing the electrostatic interactions. We hope to implicitly partially account for the effect of electron transfer through the use of Gaussian distributions, which reduces the electrostatic interactions on neighboring atoms.

6.3 Effect of using a Gaussian charges for the electrostatic interactions

For the first objective, we assess the effect of the width of the Gaussian distribution for the electrostatic charges on the solvation free energy of the ion in water. The effect of the width of the Gaussian distribution (Ω) for charged particle on the Helmholtz free energy of solvation was calculated for a range of atomic charges. The electrostatic charge on the ions were varied between -2e and 2e, and the width of the Gaussian distribution varied between 0 (point charge) and 1.5Å. The free energy is calculated using thermodynamic integration using the method of Beutler *et al.*[22], by integrating along a

reversible path from state 1 (bulk water with no ion-water interaction) to state 2 (infinite dilute ion-water solution, in a constant volume, and constant temperature ensemble, the Helmholtz free energy given by,

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU(\lambda)}{d\lambda} \right\rangle \cdot d\lambda - A_{Lat} \qquad [6\text{-}4]$$

where λ is the Kirkwood's coupling parameter. This parameter is defined so that $U(\lambda = 0)$ yields pure water simulation and $U(\lambda = 1)$ corresponds to a simulation with a fully realized ion in the water. A_{lat} is the energy from the ions being in a lattice due to the system having repeated images through the periodic boundary conditions. This method was used by Lynden-Bell and Rasaiah[23] for looking at hydrophobic to hydrophilic behavior of ions in solutions.

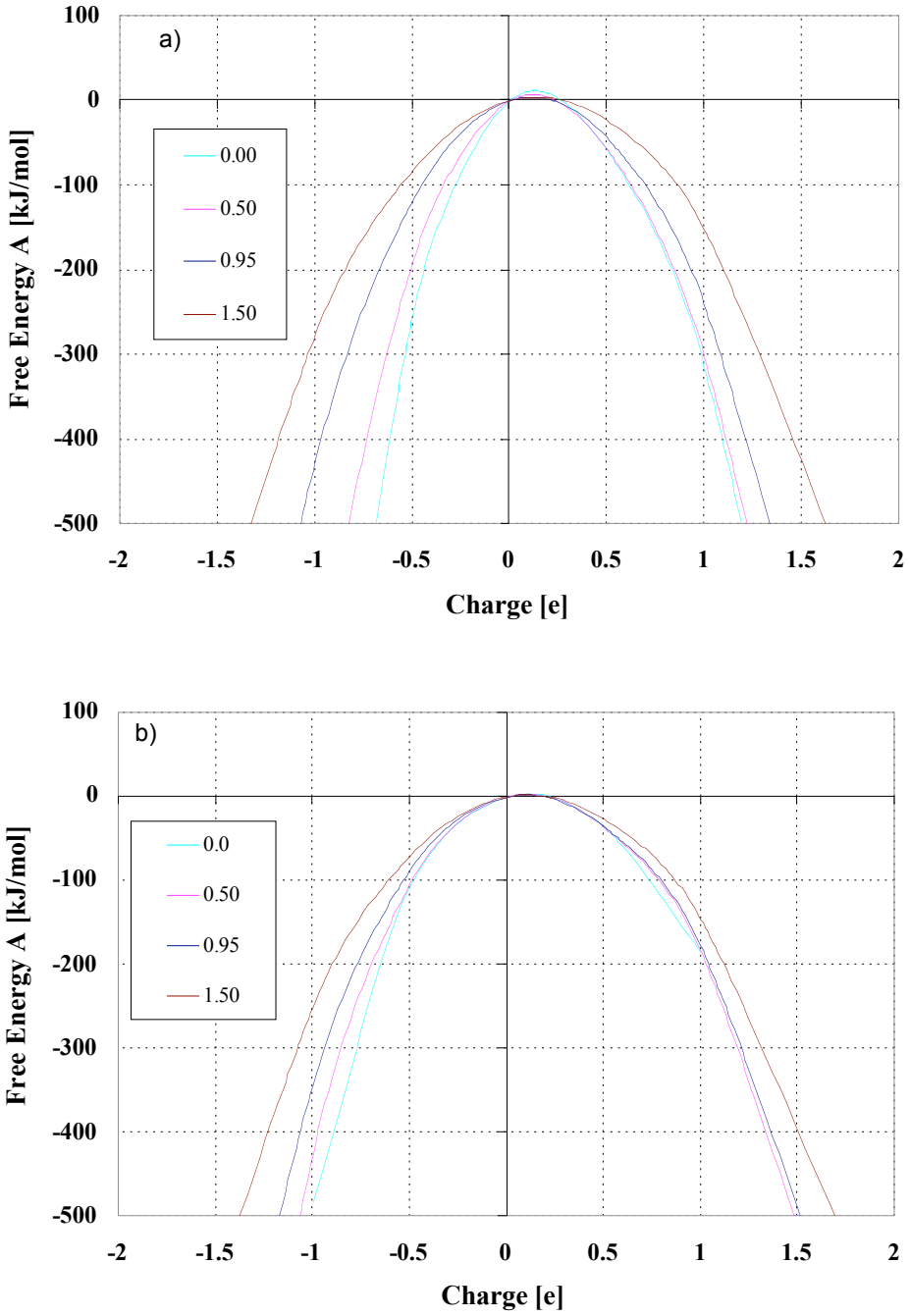


Figure 6-1: *The effect of different widths for the charge distribution on the Helmholtz free energy of solvation as a function of ionic charges for particle sized approximately to (a) sodium and (b) chlorine atoms.*

Simulation details

The Helmholtz free energy calculations were carried out in a simulation box contained 107 water molecules and 1 ion in an NVT ensemble in a simple periodic cubic cell. The density was 1.0 gcc^{-1} and the temperature was set to 300K using a Gaussian isokinetic thermostat[24]. A timestep of 0.5fs was used. A 100ps equilibration period was conducted, starting from a lattice configuration, during which no statistics were recorded. Starting from the end of the equilibration period a 200ps production run was performed where statistics were recorded. The long range corrections for electrostatics interactions were calculated via the reaction field method. The Helmholtz free energy was determined by growing in the ion, using thermodynamic integration[22], using 10 simulations corresponding to with a coupling parameter λ values of 0.1 to 1.0 in intervals of 0.1. An identical simulation containing 255 water molecules and 1 ion was used to assess the effect of system size, since 107 water molecules is a relativity small system.

Results

In Figure 6-1, the Helmholtz free energy is shown for particles of different electrostatic charge and differing widths for the Gaussian distribution. Greater the magnitude of the free energy, (more negative the free energy) relates to an increase in the ability of the ion to be dissolved in the solvent. The effect of changing the width of the Gaussian distribution (Ω) on the solvation energy of ions differs depending on the sign of the charge of the ion. We have found that changing the width of the Gaussian distribution has a more dramatic effect on smaller, negative ions than the larger, positive ions. In fact, as there was so little effect on the positive ions, for ion-water parameterization purposes, the width of the cation's Gaussian distribution might as well be set to a point charge. The

asymmetric nature of the solvation energy due to the magnitude and sign of the charge is due to the asymmetric charge distribution in a water molecule[23]. We also expect that this asymmetric nature of water is also responsible for the differing effect of the width of the Gaussian distribution.

6.4 Parameterization with ion-water clusters

The second objective is to parameterize ion-water forcefield. The ion-water forcefield was parameterized through fitting to experimental gas-phase enthalpies for small ion-water clusters from Dzidic and Kabarle[10] for the alkali ions, Hiraoka *et al.*[12] and Arshadi *et al.*[11] for the halides ion, and to ion-water structural data[13-18].

Water model

The details of the GCPM water model have been presented by Paricaud *et al.*[9] In summary, the negative charge (M) is set 0.27Å on the θ_{HOH} bisector and the permanent dipole is set to the experimental value of 1.855D. The exp-6 potential were used by the pair interactions, centered on the oxygen atom or ion. The width of the Gaussian charge distribution, Ω_M = 0.610Å and Ω_H = 0.455Å, and the polarizability is equal to the experimental value of 1.44Å^3.

Simulation details for ion-water clusters

To parameterize the potentials, we performed simulations of the water cluster energies. The short-range interaction parameters and the width of the Gaussian distribution were determined from small ion-water clusters. The enthalpy of the clusters were compared to experimental values of Dzidic and Kabarle[10] for the alkali ions,

Hiraoka *et al.*[12] and Arshadi *et al.*[11] In order to parameterize the potentials, the water cluster energies calculations were performed on large 30x30x30Å^3 cubic cells, containing one to six water molecules and the ion. For each water-ion cluster, using the same code as in section 6-3, an equilibration period was performed for 250ps, followed by a 250ps production runs at 300K during which statistics were recorded. In these cluster simulations, there was no long range electrostatics or short range energy corrections to be accounted for due the cutoff being larger that the size of the water cluster. The cluster sizes varied from 1 to 6 water molecules. The model parameters were adjusted until simulation energies were in agreement with experiment. The fit to the experimental data is shown in figure 6-2. The parameters are shown in Table 6-1.

Having obtained the potentials, the third objective is to investigate the behavior on the ion with water clusters. Like in the parameterization stage, we preformed ion-water clusters simulations, but this time containing between one and 25 water molecules in a 30x30x30Å^3 cubic cell. An equilibration period was performed for 250ps, followed by a 250ps production run, at 300K.

Simulation details for ion in bulk water solution

The second part of the third objective is to investigate the behavior of the ion in infinite dilution in bulk water. Simulations of ions in bulk water were performed in a simulation box contained 255 water molecules and 1 ion in an NVT ensemble in a simple periodic cubic cell. The density was 1.0 gcc^{-1} and the temperature was set to 300K. A timestep of 0.5fs was used, and the 100 ps equilibration from a lattice was prior a 1 ns production run. This was repeated 10 times from different starting configurations to obtain better statistics.

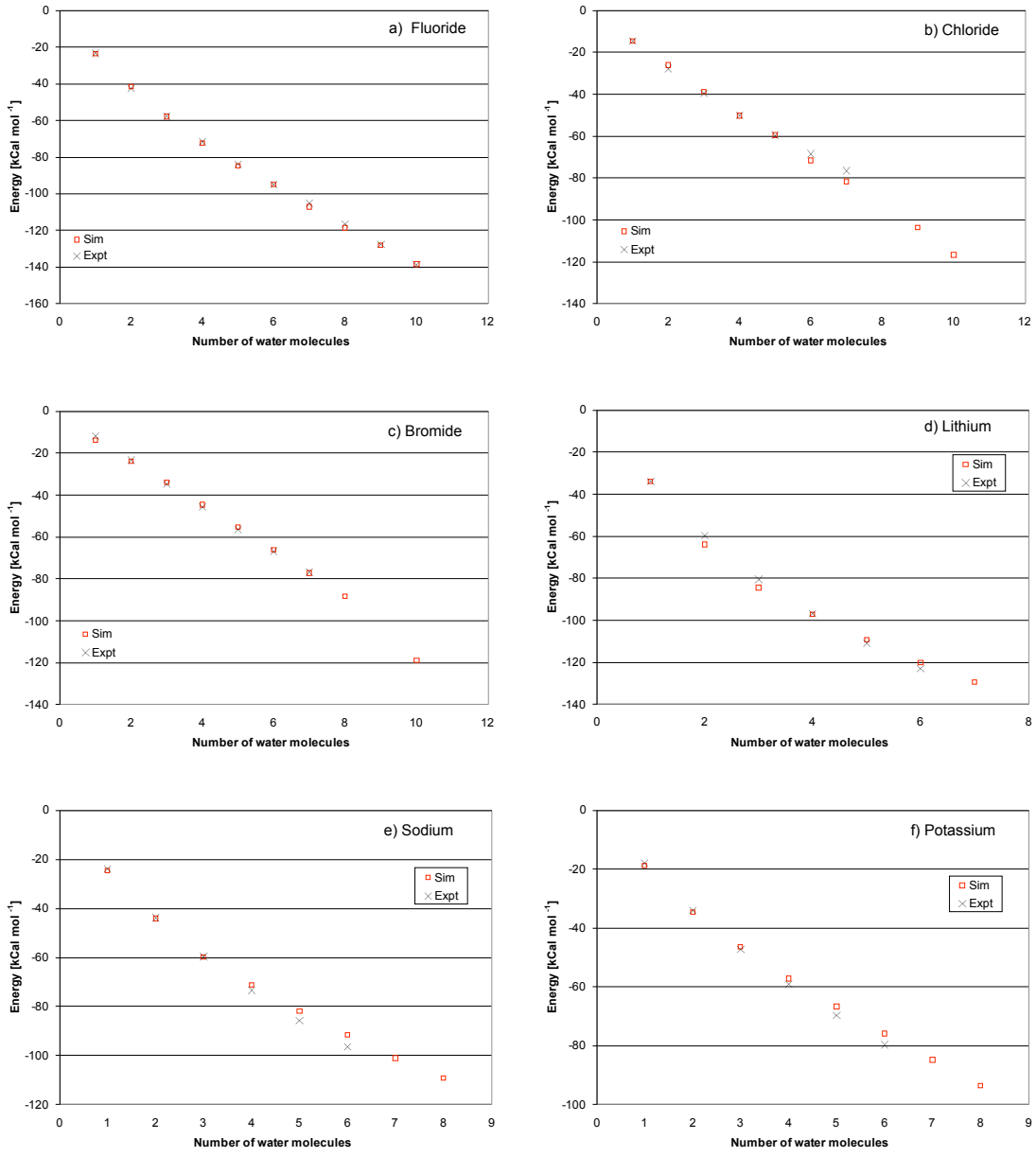


Figure 6-2: *Comparison of the enthalpy of the water-ion clusters. A) F^-, b) Cl^- c) Br^-, d) Li^+, e) Na^+, f) K^+. Experimental data for taken from Hiraoka et al.[12] and Arshadi et al.[11] for the halides ion, and Dzidic et al.[10] for the alkali ions.*

Table 6-1: Ion-water interaction parameters for Buckingham-six potential and the Gaussian charge distribution. The polarizability constants are taken from Pyper *et al.*[20]

i, j	Sigma [Å]	Wall depth ε [K]	λ	Ω_i [Å]	α_i [Å^3] [20]
O-O	3.69	110	12.75	0.61	1.444*
Li-O	2.8985	104.88	12.75	0.00	0.0285
Na-O	3.6004	46.90	12.75	0.00	0.1485
K-O	3.9204	46.90	12.75	0.00	0.7912
F-O	3.1808	66.33	12.75	0.8724	1.3100
Cl-O	3.7196	104.88	12.75	1.1036	3.7565
Br-O	3.8880	128.45	12.75	1.0909	5.0709

6.5 Results

Induced dipoles moment in water-ion clusters

Herce *et al.*[25], reported that the inclusion of explicit polarizability in both the water and ion models is vital when determining the location of the ion inside the water cluster. As we have polarizable potentials for both the water and ion species, we expect that we should be able to reproduce the same phenomena. The location of the ion in water clusters is dependant on the charge density of the ion and polarizability, where the smaller ions with a high charge density and lower polarizability, the kosmotropes e.g. F^-, Li^+, Na^+ and K^+ should undergo bulk solvation, i.e. the ion is surrounding by water molecules. Alternatively, the larger ions with a lower charge density and higher polarizability, the chaotropes e.g. Cl^- and Br^-, should undergo surface solvation, i.e. the ion stays at the edge of the ion-water cluster.

In figure 6-3, we have defined the z-axis as the vector between the location of the center of mass of the cluster and the location of the ion. Slicing the cluster in the xy plane, and plotting the location of the ion and the water molecules shows the position of the ion in the cluster. The chaotropes are shown in Figure 6-3a, where the ion is located at the surface of the cluster, where both Cl^- and Br^- obtain a similar profile for their location in the clusters. For the kosmotropes shown in Figure 6-3b, all the Li^+, Na^+, K^+ and F^- ions were located on the inside the water cluster. These results are all consistent with the results from Herce *et al.*[25]

Given that polarizability is the reason given by Herce *et al* that the chaotropes exhibit the surface solvation, we investigated the mean dipole moment for both the water cluster and the ion. Figure 6-4 shows the change in the mean dipole moment of the water cluster and for the induced ion as the water cluster increases in size. The kosmotropes, in

Figures 6-4, a-d, all shown a similar tread, which is that the induced dipole moment of the ion is relatively high for the ion-water dimer and it decreases as the number of water molecules increase. This is due to the water molecules surrounding the ions, pulling the electric field from the water molecules in multiple directions, which will have a canceling effect on the electric field at the ion. Therefore, as the number of water molecules increase, the induced dipole moment on the ion decreases until it reaches the bulk solution dipole moment.

For the chaotropic ions shown in Figure 6-4 e-f, there is a significant change in the trend of the induced dipole moment of the ion. In these cases, the ion in the ion-water dimer has a relativity low induced dipole moment, where the ion dipole moment increases as the number of water molecules in the cluster increase. This is due to the water molecules not completely filling the first hydration layer before starting filling the second hydration layer. This leads to the ion remaining at the edge of the water cluster. The electric field generated from the water clusters increases as the number of water molecules increase, increasing the induced dipole on the ion.

The induced dipole moments of the water molecules show the same trend as the size of the cluster increases, regardless of the ion's nature. The induced dipole moment for the water is the highest for the ion-water dimer cluster, then decreases as the induced as the water cluster increases in size.

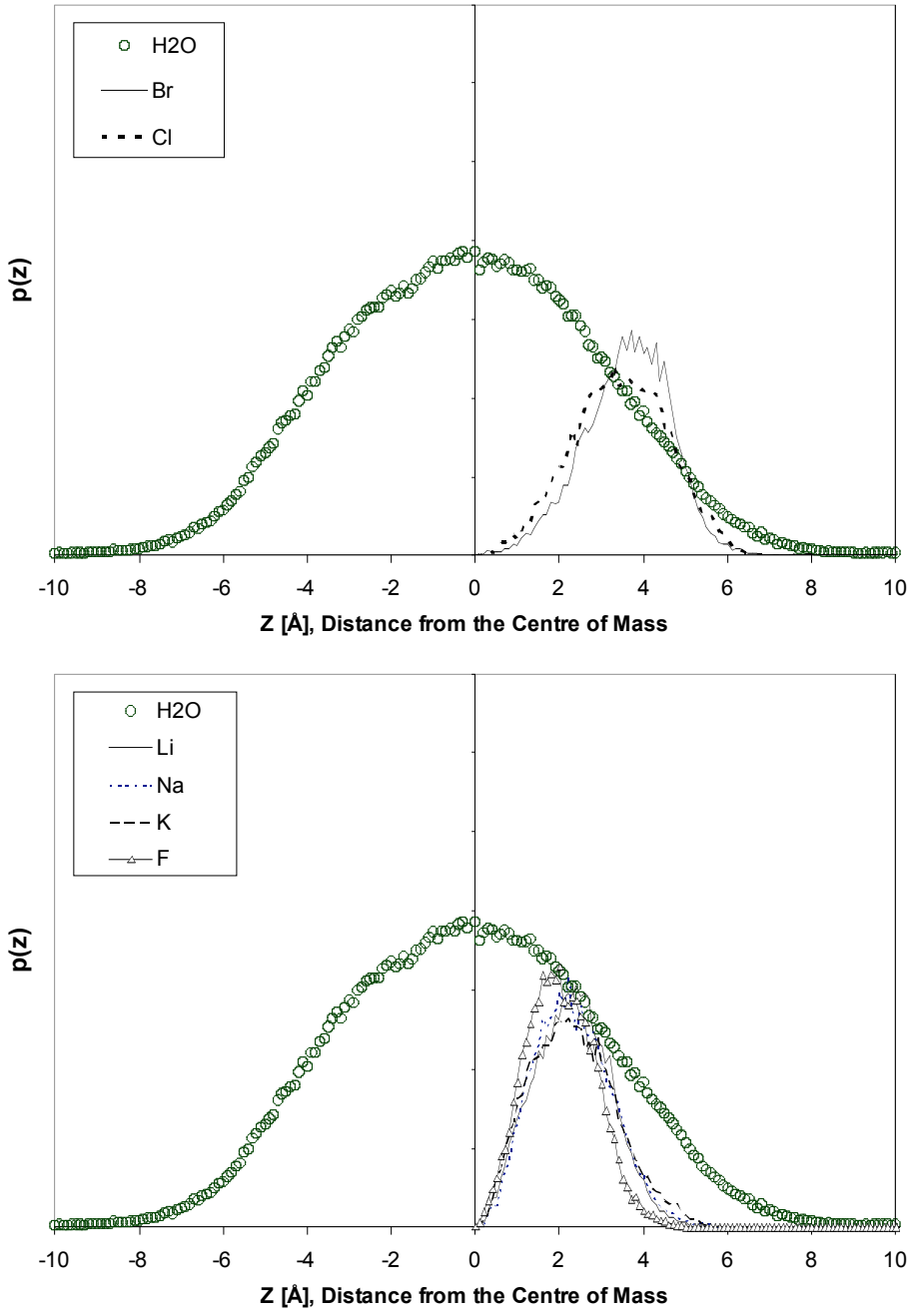
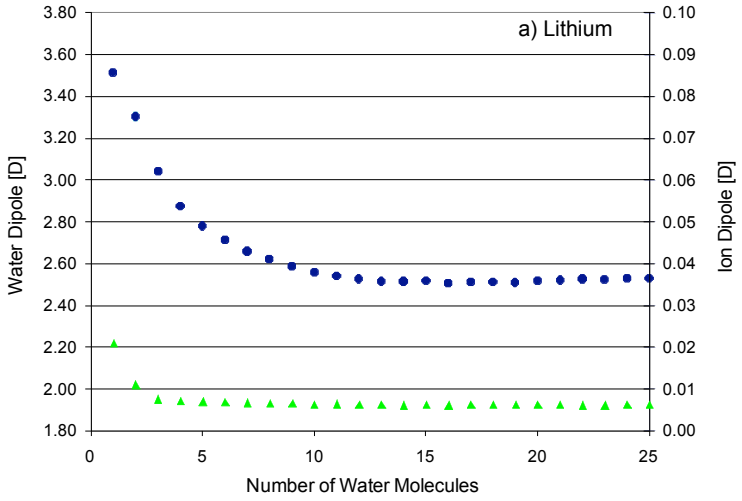
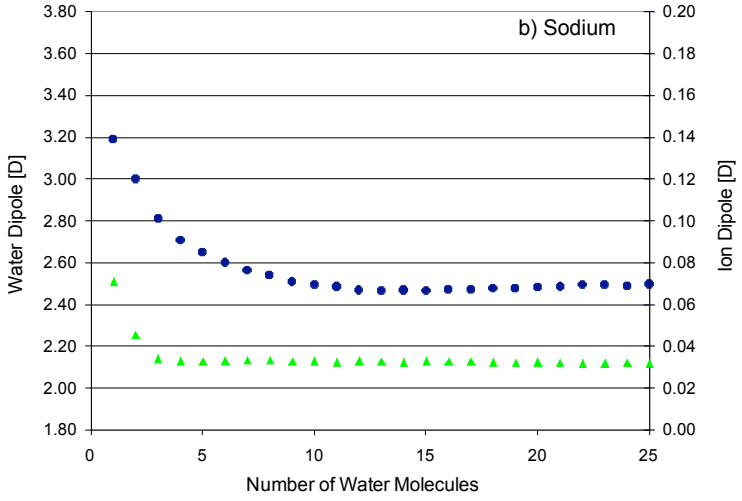
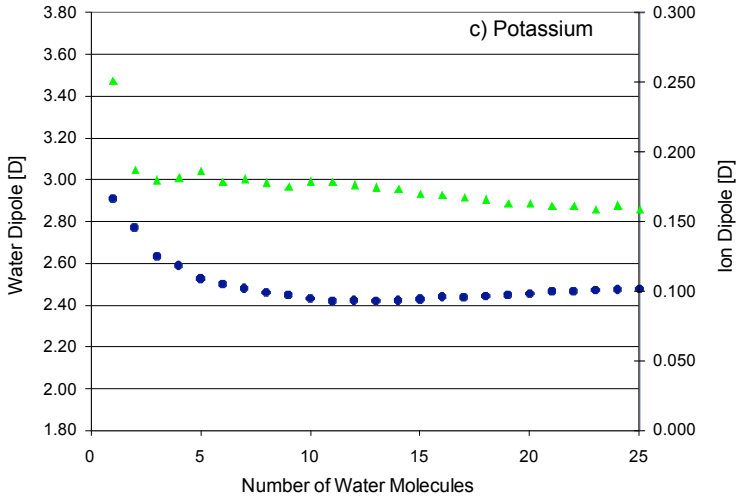


Figure 6-3a and 6-3b: *Location of the ion with respect to the center of mass on the water ion clusters. Cl^- and Br^- (Top) and Li^+, Na^+, K^+ and F^- (Bottom). The Z axis is defined as the vector between the center of mass and the ion. The Z axis is cut in the slices in the XY plane. These clusters contain 32 water molecules.*


a) Lithium
Water Dipole [D]
Ion Dipole [D]
Number of Water Molecules
3.80
3.60
3.40
3.20
3.00
2.80
2.60
2.40
2.20
2.00
1.80
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
25


b) Sodium
Water Dipole [D]
Ion Dipole [D]
Number of Water Molecules
3.80
3.60
3.40
3.20
3.00
2.80
2.60
2.40
2.20
2.00
1.80
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
25


c) Potassium
Water Dipole [D]
Ion Dipole [D]
Number of Water Molecules
3.80
3.60
3.40
3.20
3.00
2.80
2.60
2.40
2.20
2.00
1.80
0.300
0.250
0.200
0.150
0.100
0.050
0.000
0
5
10
15
20
25

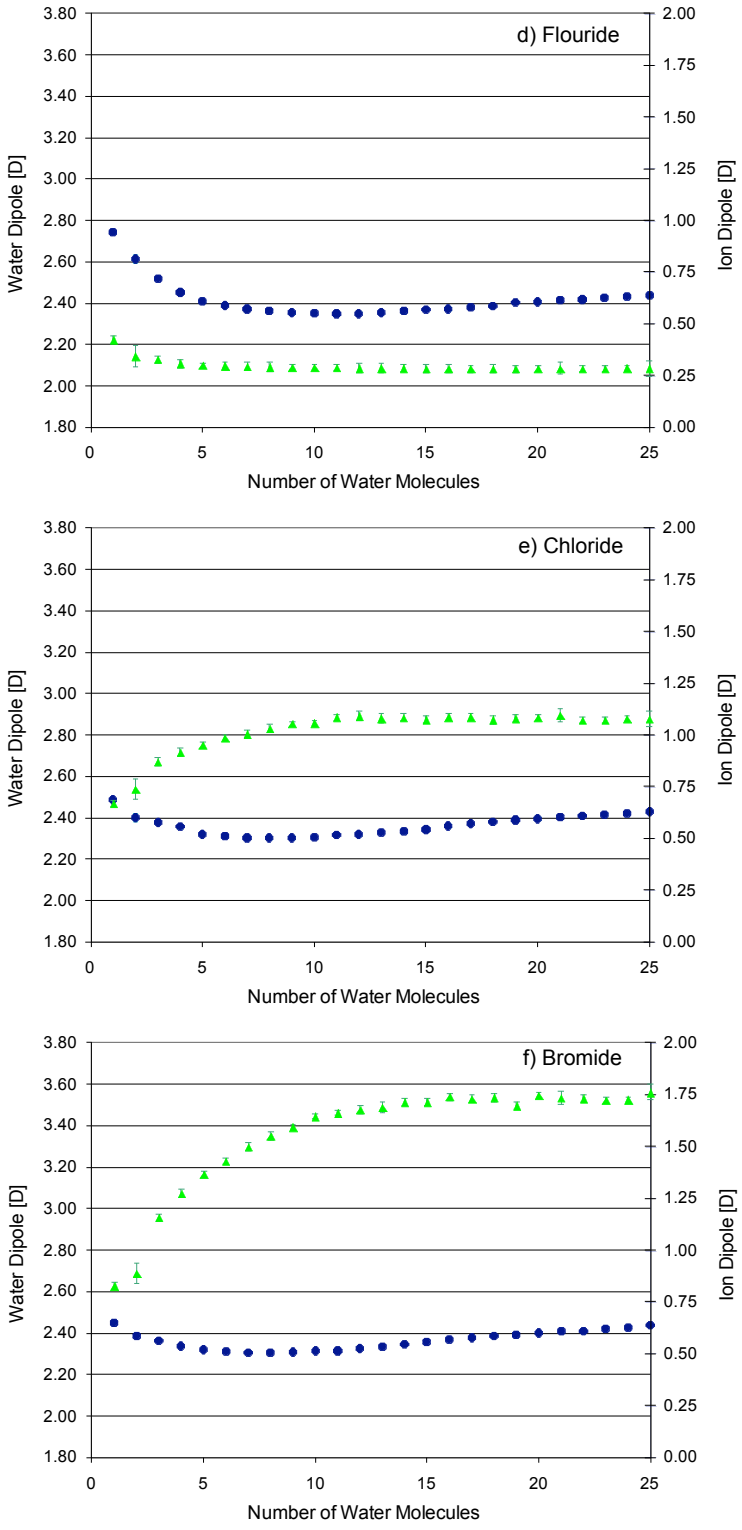


Figure 6-4: *Effect of the water cluster size on the water and ion dipole moments. Left axis, the induced dipole moment for water is circles (•) and right axis, and the induced dipole moment for ion is triangles (▲).*

Table 6-2: Structure Properties for the GCPM and ion models from the radial distribution function, comparing to experiment.

Ions	G(r) 1st max	G(r) 1st min	G(r) 2nd max	G(r) 2nd min	CN
Lithium	IO: 7.83; 2.05Å; (1.90-2.28Å)* IH: 2.93; 2.65Å; (2.50-2.73Å)*	IO: 0.13; 2.80Å IH: 0.50; 3.40Å	IO: 1.49; 4.25Å IH: 1.27; 4.85Å	IO: 0.87; 5.15Å IH: 0.92; 5.95Å	4.24; (4-6)*
Sodium	IO: 4.23; 2.40Å; (2.40-2.50Å)* IH: 2.33; 2.95Å	IO: 0.42; 3.25Å IH: 0.69; 3.85Å	IO: 1.30; 4.45Å IH: 1.18; 5.15Å	IO: 0.90; 5.40Å IH: 0.96; 6.20Å	5.47; (4-8)*
Potassium	IO: 2.92; 2.65Å; (2.70-2.95Å)* IH: 2.02; 3.15Å	IO: 0.64; 3.55Å IH: 0.78; 4.20Å	IO: 1.18; 4.65Å IH: 1.12; 5.40Å	IO: 0.92; 5.60Å IH: 0.99; 6.50Å	6.51; (4-8)*
Fluoride	IO: 4.65; 2.40Å; (2.62-2.92Å)* IH: 5.34; 1.50Å	IO: 0.41; 3.20Å IH: 0.53; 2.30Å	IO: 1.30; 4.55Å IH: 1.53; 2.80Å	IO: 0.89; 5.50Å IH: 0.87; 3.30Å	5.57; (4-6)*
Chloride	IO: 3.07; 2.90Å; (3.10-3.45Å)* IH: 2.64; 2.00Å	IO: 0.78; 3.75Å IH: 0.65; 2.65Å	IO: 1.10; 4.90Å IH: 1.34; 3.30Å	IO: 0.94; 5.85Å IH: 0.99; 3.85Å	7.43; (4-8)*
Bromide	IO: 2.95; 3.05Å; (3.12-3.85Å)* IH: 2.48; 2.15Å	IO: 0.83; 3.80Å IH: 0.56; 2.80Å	IO: 1.09; 4.95Å IH: 1.37; 3.45Å	IO: 0.94; 6.10Å IH: 0.96; 5.45Å	7.46; (4.2-8)*

CN, is the coordination number, the average number of water molecules in the first solvation shell.
* Experimental results, shown in brackets, taken from ion-water review of Ohtaki and Radnai[26]

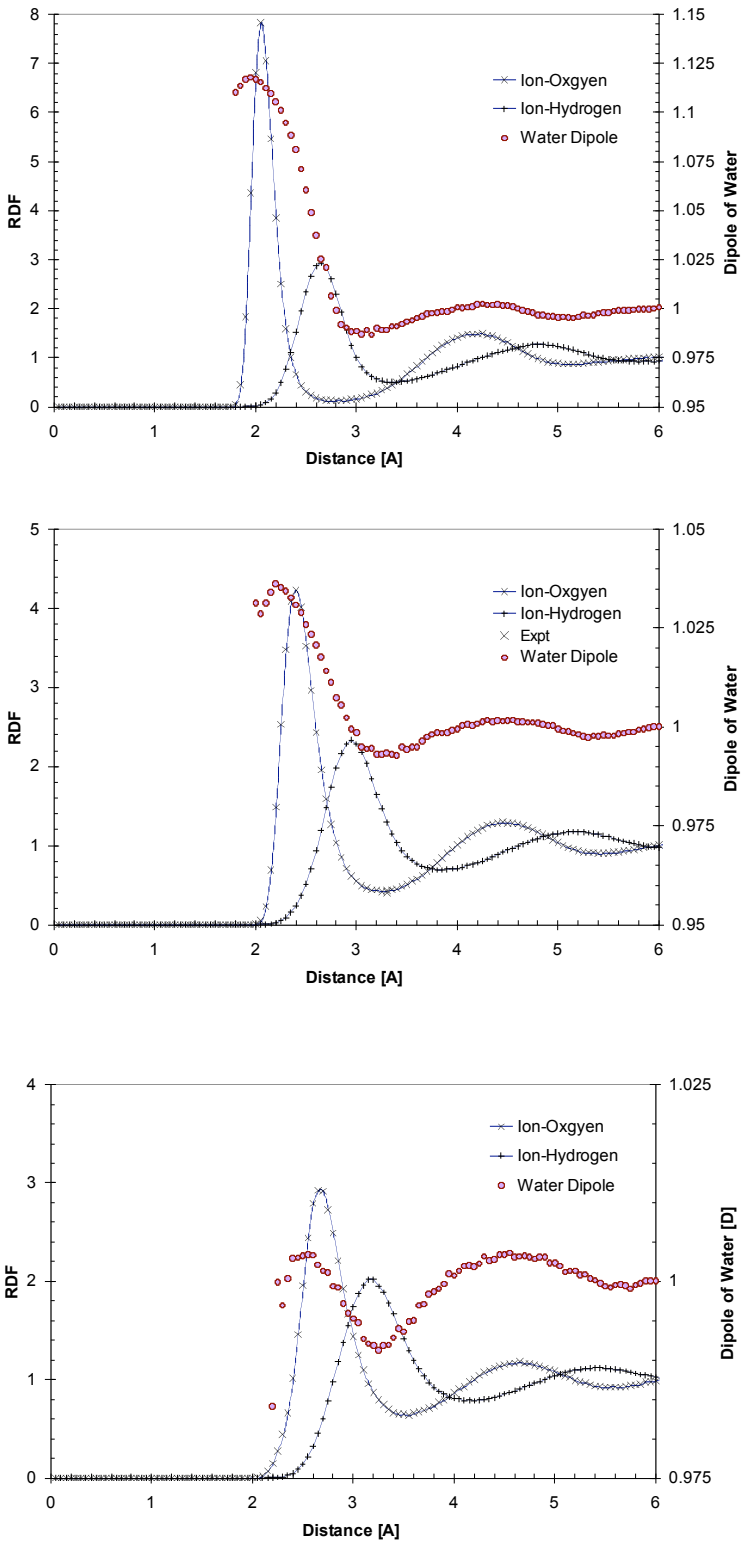


Figure 6-5: *Radial distribution functions of Li^+-water, Na^+-water and K^+-water, and the effect of the induced water dipole moment as a function of distance from the ion.*

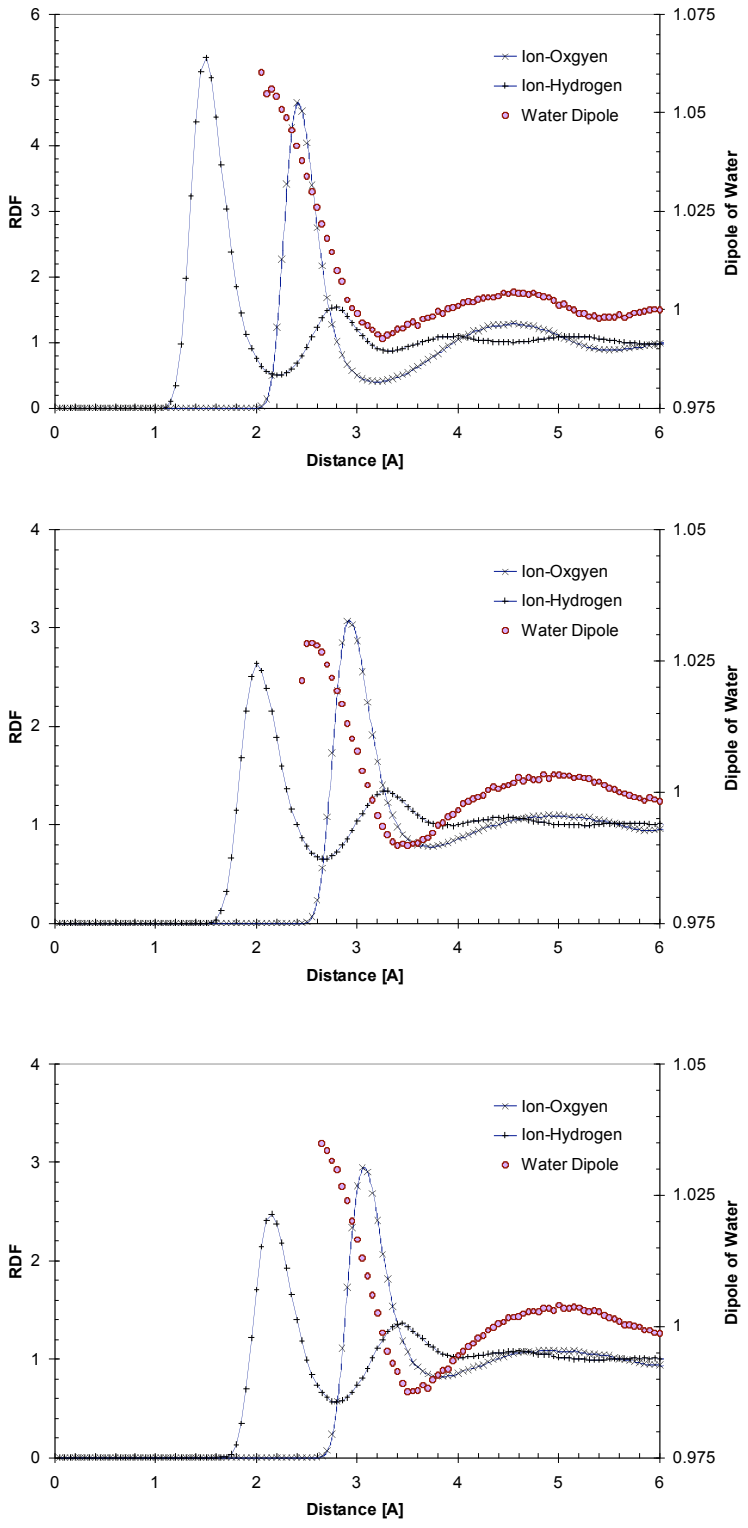


Figure 6-6: *Radial distribution functions of F^--water, Cl^--water and Br^--water, and the effect of the water dipole as a function of distance from the Ion. The dipole moment is in terms of relative to the bulk dipole moment.*

Induced dipole moment for water due to ionic species

The last objective of this chapter is to look at the behavior of ion at infinite dilution. Following on from the induced dipole of the ion and water molecules in the clusters, we are going to look at the effect on the induce dipole moment on the ion and water molecules in a bulk solution. Previous *ab initio* simulations for single charge cations have been found to have either a small or negligible effect on water dipole moment in the first solvation shell and no influence on the bulk water molecules. The only stated influence on the first hydration layer is for cations ions with a very high charge density, i.e. Mg^{++} and Ca^{++}, which has a noticeable effect of around 0.2-0.3D[13-18,27-29]. Given the uncertainty for the induced dipole moment from *ab initio* calculations in Chapter III for bulk solutions, let alone as a function of the solution structure, classical simulations due to their increased computational speed, can get a more detailed picture of how the induced dipole moment of water molecules is affected by the ion. In Figure 6-5 and 6, the ion-water structure is shown as the radial distribution function, and effect of the ion on the induced dipole of the water molecules as a function of separation from the ion are shown. Figure 6-5 shows the cation ions, where lithium, the smallest cation with the highest charge density, showed a maximum of a 12 percent increase in the induced water dipole moment for water in the first hydration shell. The ion influence on the surround water decreased as the charge density decreases. For sodium, the simulations predict a 4 percent change, and for potassium ion a negligible effect is prediction on the surrounding water molecules. The effect is short-ranged, as water molecules between the first and second hydration shell showing a decrease in the induced dipole moment. Carrillo-Tripp *et al.*[6] reported a decrease in the first hydration shell in the induced dipole moment in the first hydration shell around a K^+ ion relative to the bulk. That same

conclusion could be applied to the results here, due to the large decrease in the induced dipole moment in the tail end of the first solvation shell, but in the our opinion, this in insignificant due to the variation the induced dipole moment relative to the bulk value.

Likewise, with the halide ions, shown in Figure 6-6, the smallest ion with the highest charge density has the largest effect on the induced water dipole moment. The fluorine ion had a maximum 6.5 per cent increase. Chloride and bromide increased the water's dipole 3.5 percent increase.

It was interesting to note that for all the ions, water's induced dipole moment oscillated, concurrently with the troughs and peaks in the water structure, where a high induced dipole moment present in the higher area of water concentration and a lower induced dipole moment in the spaces in between. We expect that this is due to the water molecule breaking the fluid structure as the water molecules diffuses from one water layer to another. During this transitional phase, the water structure is temporary broken, the decrease in the ordering decreases the electric field and therefore the induced dipole moment.

Induced dipole moment for ions in bulk solution

The next step is to look at the ability of the ions in water solutions to reproduce the induce dipole moment of ion compared to values calculated via *ab initio* calculations. Like water, the induced dipole moments for the ions are also relatively unknown. The calculated induced dipole moment and their standard deviation of dipole moment for the ions are shown in Table 6-3. As a reference point, the standard deviation for the induced dipole moment in GCPM water is 0.19D[30]. The values of the induced dipole moment of the anions are very similar to values calculated from a*b initio* simulations. The induced

dipole moments of the cations calculated by the classical model are very small, so it is understandable that induced dipole moments are not reported using *ab initio* simulations.

Water – ion tilting angle

Continuing with the structure on the ion-water solution, the next step is to look at the orientations of the water molecules surrounding the ion. We have defined the tilting angle (θ) is defined as the angle between the HOH bisector vector and ion-oxygen vector, as shown in Figure 6-7. Classical simulations tend to show that ions have a strong effect on the water orientation, while *ab initio* calculations show a softer attraction, with the water molecules showing a strong tendency to hydrogen bond with themselves not the ion[16,31]. White[16] showed in *ab initio* simulations for Na^+ a plateau in the tilting angle between 180 and 120 degrees. It is hoped that the used of Gaussian distributions for the charges will damping the neighboring ion-water electrostatic interactions, bring the ion-water orientations closer to the *ab initio* predictions.

Figure 6-9a, shows the probability of the water molecules orienting around the cation's first hydration shell. The water molecules in the first hydration layer are oriented around the cation in such a way that the oxygen atom is directly facing the ion, i.e. $(Cos\theta = -1)$. As expected, the probability on the water molecule's oxygen atoms pointing towards the ion decreases as the charge density of the ion decreases. These results are still inline with the normal classical ion-water models observations. Figure 6-9b shows the probability of the water molecules orienting around the second hydration shell. For a reference, the probability of a water molecules orienting around another water molecules is shown. It is clear to see that the effect of the ions on the second

hydration shell orientation is the same as that for the water-water. These results are in agreement to the observations of Carrillo-Tripp *et al.*[6]. They show that the direct structure effects of the ion are limited to the first hydration shell, as the second shell is purely an artifact of the water-water interactions, not due to the ion interaction.

In Figure 6-10a, shows the probability of the water molecules orienting around the anion's first hydration shell. The water molecules in the first hydration layer are oriented around the anion in such a way that the hydrogen atom is point towards the ion i.e. $(\theta \sim 50)$. Therefore, for both the cations and anions, the softer tilting angle in the first hydration shell for Na^+, K^+, F^- and Cl^- from the *ab initio* simulations of Tongarra *et al.*[21,31] are not reproduced. We expect that the main cause for the cations is due to the fact that the GCPM water model has the lowest region for the electrostatic potential at the oxygen side of the hydrogen-oxygen bisector, i.e. still in the zx plane. For real water, (and *ab initio* water) the electrostatic potential minima's are around the position of the "lone electron pairs". The inclusion of the electrostatic minima should enhance the tilt, and leading to closer reproduction of the *ab initio* results.

Autocorrelation functions of first water shells

As the orientation on the water molecules around the ion are similar to the previous classical simulations, the next question is whether the dynamically properties of the first hydration shell are changed. The residence time of the first hydration shell is the length of time the original water molecules are present in the solvation shell before diffusing away. The stronger the charge density of the ion, stronger the attraction between the water-ion molecules, thus longer the water molecules should be present in the first

hydration layer. The residence time, τ, is obtained by fitting the correlation function to the exponential decay function,

$$R(t) = \frac{1}{N}\sum_{i=1}^{N}\left(\theta_i(r,t)\theta(r,0)\right) \cong \exp\left(\frac{-t}{\tau}\right) \qquad [6\text{-}5]$$

where $\theta_i(r,t)$ is a Heaviside unit function, which equals 1 if the original water molecules are present in the first hydration shell, and 0 elsewhere. N is the average number of water molecules in the first hydration shell, which is computed using the 1st minima from the ion-water radial distribution functions (*Figures 6-5, 6-6*). If a water molecule leaves the hydration shell, it is excluded from the original water molecules. In the case of a water molecule moving out of the defined hydration shell temporarily and return without molecules entering the bulk water phase, it should not be excluded from the original hydration shell. We accounted for this primitively with a 2ps grace period where the water molecules can exit the hydration layer and return without it being excluded from the original hydration layer molecules. This approximation technique has been used by Impey *et al.*[32].

Table 6-4 and 6-5 shows the residence time for the cation and anions respectively, and other published results. As expected, the lithium ion has the longest residence time (τ = 17.10 ps), as it has the strongest attraction to the surrounding water molecules. Sodium and potassium ions have decreasing residence times (τ = 6.30 ps, τ = 4.79 ps respectively) as the charge density of the ions decreases. The same trend is shown for the anions, with the fluoride ion obtaining the longest residence time (τ = 7.06 ps). The chloride and bromide ions have decreasing residence times (τ = 4.79 ps, τ = 2.13 ps respectively). One thing that is consistent, is that the residence times calculated by the new ion-water potentials, are lower than the residence times calculated via the other

classical simulation, yet still inside the residence times measure by experiment[33]. This indictors a weaker attraction to the first hydration shell.

Free energy of solvation

Results for the free energy of solvation are shown in Table 6-6. The Helmholtz free energy calculated for the cations is in poor agreement compared with experiment. For the cations, the magnitude of the free energy of solvation is underestimated, corresponding to a less hydrophilic attraction. On the other hand, anion solvation free energies are in reasonable good agreement with experimental values, although they did have a slight error on the side of greater hydrophilicity.

Table 6-3: Mean induced ion dipole in bulk water from molecular dynamics

	This Work	Ab initio
Lithium	0.007D	-
Sodium	0.030D	-
Potassium	0.14D, (sd 0.04)	-
Fluoride	0.30D, (sd 0.05)	0.39D [13]
Chloride	0.90D, (sd 0.26)	0.80D [34]
Bromide	1.49D, (sd 0.41)	0.95D, (sd 0.80) [18]

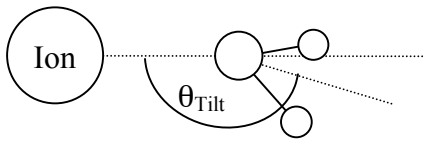


Figure 6-7: *Tilting angle, between the oxygen-ion vector and the oxygen-hydrogen bisector*

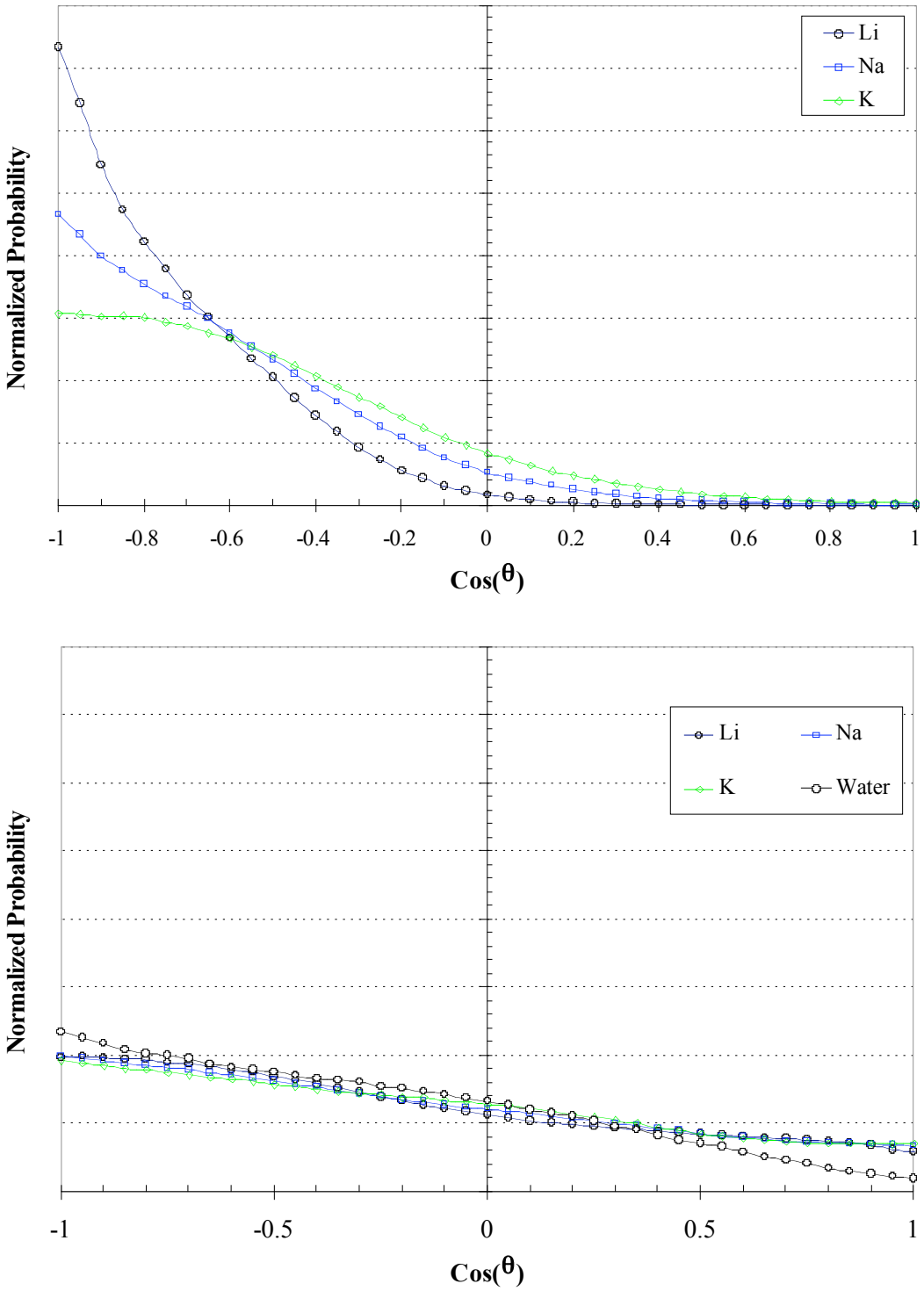


Figure 6-8a and 6-8b: *(Top) Tilting distribution of the first solvation shell water molecules surrounding lithium, sodium and potassium, and (Bottom) the tilting angle for the distribution of the second solvation shell of water*

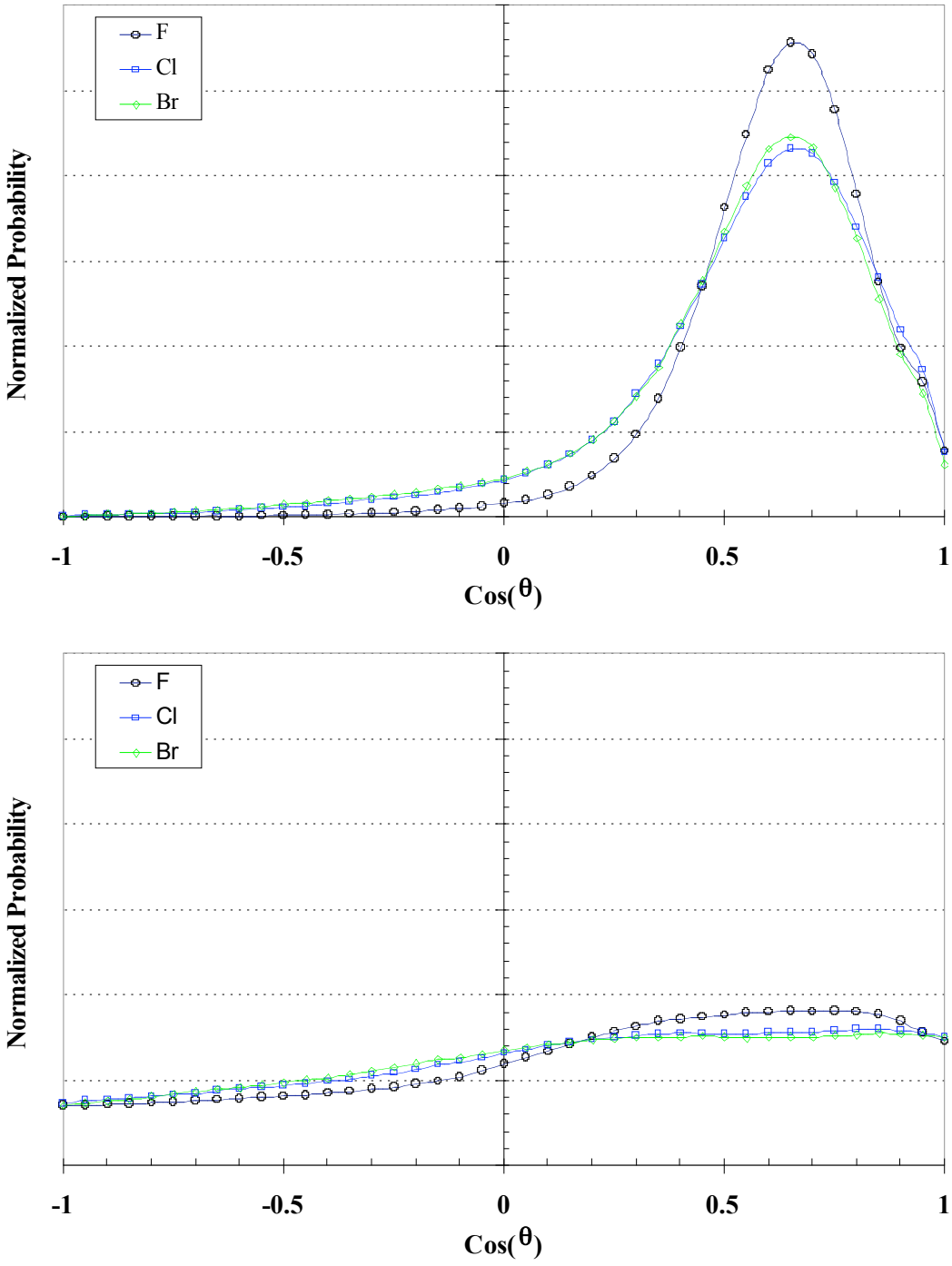


Figure 6-9a and 6-9b: *(Top) Tilting distribution of the first solvation shell water molecules surrounding fluoride, chloride and bromide, and (Bottom) the tilting distribution of the second solvation shell water molecules surrounding fluoride, chloride and bromide*

Table 6-4: Mean residence time for the first hydration shell for cation [ps]

Ion	This Work	Experiment[33]	Impey[32]	Obst[35]	Others
Lithium	17.10	8-40	33.3 (278K)	41.4	25.5[36]
Sodium	6.30	5-30	9.9 (282K)	14.7	29[37]
Potassium	4.79	3-10	4.8 (274K)	8.3	

Table 6-5: Mean residence time for the first hydration shell for anion [ps]

Ion	This Work	Experiment[33]	Impey *et al* [32]	Ab initio	Smith[37]
Fluoride	7.06	6-60	20.3 (278K)	16.0[13]	
Chloride	4.79	3-10	4.5 (287K)	12[17]	12
Bromide	2.13	2-7		19 ± 5[18]	

This Work, (300K/255GCPM), Impey *et al.*[32] (~Varied Temperature/64MCY), Obst and Bradaczek[35] (300K/525TIP3P), Egorov *et al*[36] (298K/SPC/E), Smith[37] (300K/214RPOL)

Table 6-6: Free energy of solvation for ions at infinite dilution in GCPM bulk water
All values are in units of kJ mol^{-1}

Ions	This Work	Expt [38]	Friedman [39]	Conway [40]
Lithium	-318.1	-475		
Sodium	-259.7	-365	-371	-372
Potassium	-225.0	-295	-298	-298
Fluoride	-468.4	-465	-394	-441
Chloride	-348.3	-340	-277	-324
Bromide	-337.5	-315	-263	-310

6.6 Conclusions

The effect of the Gaussian distribution on the electrostatic charges for the Helmholtz free energy showed an asymmetric nature with regards to the magnitude and sign of the charges. The width of the Gaussian distribution has a limited effect for the cations compared to the anions. Due to this, the cations were parameterized using a point charges.

In conclusion, we have developed classical ion-water potentials using Gaussian distributions for representing the anions' electrostatic charges and using polarizability for both the water and ion. We used the new potentials to investigate the ion-water behavior in water clusters. We have shown that the smaller ions, with a higher charge density undergo bulk solvation, i.e. brought into the center of the water cluster. As a result, the induced dipole moment of the ion decreasing as the size of the cluster increases as the electric field water molecules is cancelled out by water molecules surrounding the ion.

The larger anions, with a lower charge density and higher polarizability undergo surface solvation, i.e. the ion is kept at the surface of the cluster. As a result, the induced dipole moment of the ion increases as the size of the cluster increases. This is due the increasing electric field cause by the water cluster, which is positioned on the edge of the ion. This is in agreement with *ab initio* observations[25].

In regards to the ions solvated in bulk water, we have investigated the structural and dynamical properties of the solution. We have shown that there is a small effect of the ions on the induced water dipole moment in the 1st water solvation shell, but it only goes one solvation shell deep. We have also shown good agreement for the induced dipole moment of the ions compared to *ab initio* simulations.

The uses of GCPM as the water model still shows the same trends as other classical water models with regards to the ion-water orientation. We expect that the development of the classical water model will required some form of acknowledgement of the electrostatic potential minima associated with lone pair electrons in order to reproduce the ion-water orientation exhibited in *ab initio* simulations.

The use of Gaussian charges in the water-ion interactions has shown that it can reduce the attachment of the ion to the surrounds water molecule. This is shown by the lower residence times for the ions first hydration layer compared to other classical simulations.

References

[1] L. X. Dang, D. E. Smith, Journal of Chemical Physics **99** (9), 6950 (1993).

[2] L. X. Dang, Chemical Physics Letters **200** (1), 21 (1992).

[3] L. X. Dang, Journal of Chemical Physics **97** (4), 2659 (1992).

[4] G. Lamoureux, B. Roux, Journal of Physical Chemistry B **110**, 3308 (2006).

[5] G. Lamoureux, E Harder, L V. Vorobyov, B. Roux, A. D. MacKerell, Chemical Physics Letters **418**, 245 (2006).

[6] M. Carrillo-Tripp, H. Saint-Martin, I. Ortega-Blake, Journal of Chemical Physics **118** (15), 7062 (2003).

[7] H. Saint-Martin, J. Hernandez-Cobos, M. I. Bernal-Uruchurtu, I. Ortega-Blake, and H. J. C. Berendsen, Journal of Chemical Physics **113** (24), 10899 (2000).

[8] P. Drude, *The Theory of Optics*. (Longmans, Green and Co., New York, 1902).

[9] P. Paricaud, M. Predota, A. A. Chialvo, P. T. Cummings, Journal of Chemical Physics **122** (24), 4511 (2005).

[10] I. Dzidic, P. Kabarle, Journal of Chemical Physics **74** (7), 1466 (1970).

[11] M. Arshadi, R. Yandagni, P. Kebale, Journal of Chemical Physics **74** (7), 1475 (1970).

[12] K. Hiraoka, S. Mizuse and S, Yamabe, Journal of Physical Chemistry **92** (13), 3943 (1988).

[13] J. M. Heuft, E. J. Meijer, Journal of Chemical Physics **122**, 094501 (2005).

[14] L. M. Ramaniah, M. Bernasconi, M. Parrinello, Journal of Chemical Physics **111** (4), 1587 (1999).

[15] A. P. Lyubartsev, K. Laassonen, A. Laaksonen, Journal of Chemical Physics **114** (7), 3120 (2001).

[16] J. A. White, E. Schwegler, G. Galli, F. Gygi, Journal of Chemical Physics **113** (11), 4668 (2000).

[17] J. M. Heuft, E. J. Meijer, Journal of Chemical Physics **119** (22), 11788 (2003).

[18] S. Raugei, M. L. Klein, Journal of Chemical Physics **116** (1), 196 (2002).

[19] J. E. Huheey, E. A. Keiter, R. L. Keiter, *Inorganic Chemistry: Principles of Structure and Reactivity*, 4th ed. (HarperCollins College Publishers, 1993).

[20] N. C. Pyper, C.G. Pike, P. P. Edwards, Molecular Physics **76** (2), 353 (1992).

[21] A. Tongraar, B. M. Rode, Physical Chemistry Chemical Physics **5**, 357 (2003).

[22] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, W. F. van Gunsteren, Chemical Physical Letters **222**, 529 (1994).

[23] R. M. Lynden-Bell, J. C. Rasaiah, Journal of Chemical Physics **107** (6), 1981 (1997).

[24] D. J. Evans, G. P. Morriss, Physical Letters A **98**, 433 (1983).

[25] D. H. Herce, L. Perera, T. A. Darden, C. Sagui, Journal of Chemical Physics **122**, 024513 (2005).

[26] H. Ohtaki, T. Radnai, Chemical Reveiw **93**, 1157 (1993).

[27] I. Bako, J. Hutter, G. Palinkas, Journal of Chemical Physics **117** (21), 9838 (2002).

[28] D. Marx, M. Sprik, M. Parrinello, Chemical Physics Letters **273**, 360 (1997).

[29] F. C. Lightstone, E. Schwegler, R. Q. Hood, F. Gygi, G. Galli, Chemical Physics Letters **343**, 549 (2001).

[30] P. J. Dyer, P. T. Cummings, Journal of Chemical Physics **125** (14), 144519 (2006).

[31] A. Tongraar, K. R. Liedl, B. M. Rode, Journal of Physical Chemistry A **102**, 10340 (1998).

[32] R. W. Impey, P. A. Madden, I. R. McDonald, Journal of Physical Chemistry **87**, 5071 (1983).

[33] G. Jancso, P. Bopp, K. Heinzinger, Hungarian Academy of Sciences, Report No, KFKI-1997-101 (1977).

[34] D. J. Tobias, P. Jungwirth, M. Parrinello, Journal of Chemical Physics **144** (16), 7036 (2001).

[35] S. Obst, H. Bradaczek, Journal of Physical Chemistry **100**, 15677 (1996).

[36] A.V. Egorov, A. V. Komolkin, V. I. Chizhik, P. V. Yushmanov, A. P. Lyubartsev, A. Laaksonen, Journal of Physical Chemistry **107**, 3234 (2003).

[37] D. E. Smith, Journal of Chemical Physics **100** (5), 3757 (1994).

[38] Y. Marcus, Journal of the Chemical Society Faraday Transactions **87** (18), 2995 (1991).

[39] H. L. Friedman, , Krishnan, C. V., *In Water: A Comprehensive Treatise*. (Plenum, New York, 1973).

[40] B. E. Conway, J. Solution Chemistry **7**, 721 (1978).

CHAPTER VII

CALIBRATION OF CHEMICAL BONDING BETWEEN BENZENEDITHIOLATE AND GOLD: THE EFFECTS OF GEOMETRY AND SIZE OF GOLD CLUSTERS

7.1 Summary

The effects of the geometry and size of gold clusters on the chemical bonding between benzenedithiolate (BDT) molecule and gold clusters have been evaluated for several different BDT-nAu complexes. The original potentials for the BDT-Au interactions have been developed based on BDT-1, 2 Au complexes (Y. S. Leng *et al.* Journal of Chemical Physics, 122, (24), 244721 2005). Here we look at the effect of the geometry and size of the gold clusters on the chemical bonding and how it effects the parameterization of the bonding potentials. Density functional theory (DFT) calculations are employed to parameterize the bond-stretching behavior between BDT and gold atoms. It was found that to some extent, the bonding curves depends on the gold cluster geometry. However this variation in the bonding curves does not change molecular packing structure on Au (111) surface and only has minor effect on local bonding geometry. The Mulliken charge distribution at the bonding interface (i.e. among bonded atoms) also does not show any impact on the global packing structure. The research reported in this chapter has been published[1].

7.2 Introduction

The bonding and packing structure of organic self-assembled monolayers (SAMs) on metal surfaces or clusters are fundamental issues in many nanotechnology

applications, such as catalysis, structural materials, electronic materials, and the development of molecular electronics devices[2]. Since the original suggestion[3] that 'donor-bridge-acceptor' (metal-molecule-metal) structure might be useful for constructing molecular electronic circuits and devices, many fundamental studies[4-10] have been accomplished to find appropriate molecular wires. Self-assembled monolayers (SAMs) composed of many single-molecule devices are being conceived as candidates for such kind of molecular wires[4]. Electron transport through single-molecule junctions is directly related to the local electronic structure of nanoscale region that involves the molecule and a number of metal atoms in proximity contacts. However, the difficulty in getting a thorough understanding to the problem lies in the unknown nature of the molecular bonding and the geometry of the SAMs packing structure. Large-scale *ab initio* quantum mechanical calculations is quite computationally expensive, leaving classical molecular simulations is an alternative way to investigate this problem, provided that a good force field for the intermolecular potential and particularly the organic/metallic chemical bonding potentials are well developed.

For thiolate molecules adsorbed on Au (111) surface or clusters, recent *ab initio* density functional theory (DFT) calculations showed that sulfur head groups prefer to bind at bridge or bridge-like binding sites with a strong chemical bonding[11-13], instead of the face centered cubic (fcc) hollow sites[14]. DFT calculations for the binding between thiolate molecules and gold clusters also find that sulfur forms strong chemical bonds with only one or two gold atoms,[15-18] corresponding to the 'on-top' or 'on-bridge' bonding on an extended gold surface. The strong Au-S covalent bonding is indicated by the large concentration of electron density between S and bonded Au atoms[12,13,18] and by

the distinctly directional Au-S-C bond[15]. Given the evidence of the very local chemical bonding between S and Au atoms, Leng *et al.*[19], recently developed a series of chemical bonding potentials for BDT-1 ,2Au complexes based on DFT calculations. Molecular dynamics (MD) and Monte Carlo (MC) simulations[19,20] has demonstrated that the BDT SAMs have a well-ordered herringbone structure. When simulating the self-assembly of BDT molecules on Au (111) surface, the total interaction between BDT and Au (111) can be separated into two parts: the bonded interaction involving sulfur and one or two gold atoms and the non-bonded interaction between thiolate molecule and other gold atoms. The latter includes van der Waals interactions represented by the universal force field (UFF)[21] and the electrostatic interactions. The partial charges are determined by Mulliken[22] population analysis. A fundamental question needed to be answered concerning the validity of the chemical bonding potentials, since the electronic structure of individual gold clusters may strongly depend on the geometry and size of clusters. When thiolate molecules come into contact with different gold clusters, we need to understand how the geometry and size of gold clusters influence the bonding curves. Leng *et al.*[19] in a earlier study found that the dominant contribution to the local bonding comes from the bond stretching. For this reason, this current study concentrates on the bond stretching behavior between BDT and a few different gold clusters. We then perform MD simulations using different BDT-Au chemical bonding potentials for the BDT binding on to the Au (111) surface. The general conclusion is that whereas the local bonding depends more or less on the geometry of Au clusters, this effect has much less impact on the BDT SAMs packing structure.

7.3. Computational details

Density functional theory

All the *ab initio* calculations were performed by DFT using the NWChem package[23] The Gaussian valence triple zeta basis set 6-311G[24] is used for S, C and H atoms in BDT molecule and the effective core potential (CRENBL-ECP) and associated basis set[25] is used for the Au atoms. This is the same basis set used by Leng *et al.*[19] and has been shown already to have satisfactory convergence. To avoid spin multiplicity complications, this work is limited to closed shell system with singlet spin state. Figure 7-1 shows the molecular models of neutral BDT-nAu complexes with n is the number on gold atoms in the clusters, i.e n = 3 and 7. Specifically, for BDT-3Au complex, three different bonding geometries are considered (*Figure 7-1 a-c*), which covers 'on-top' (BDT-3Au12) and 'on-bridge' (BDT-3Au21) cases. The linear monoatomic chain of Au cluster (BDT-3Au111) with one gold atom bonded with BDT (*Figure 7-1c*) is an extreme case[26], which may correspond to the molecule-metal lead configuration in the measurements of conductance through single BDT molecules by scanning tunneling microscope technique[27]. In the case of BDT-7Au complex (*Figure 7-1d*), we only consider the 'on-top' bonding configuration. The 7-Au cluster was initially optimized from a planar structure in the bulk by the classical tight-binding second-moment approximation (TB-SMA) potential[28] and was further optimized by DFT using the NWChem package. The geometry of this cluster is quite similar to the three-dimensional one found in recent extensive DFT calculations of gold clusters, but not the global minimum[29]. For the correction due to the electron exchange and correlation energyies, three DFT functionals are used, firstly the local density approximation (LDA)[30], secondly

the PBE0[31], which is a hybrid functional, which does not contain any adjustable parameters and thirdly, the X3LYP[32], an extended hybrid, generalized gradient functional combined with Lee-Yang-Parr[33] correlation functional. The X3LYP combines both the B88[34] and PW91[35] functionals for the exchange energy. For more details on functionals, refer to chapter II.

Molecular dynamics

The molecular dynamics simulation details are based on the work of Leng *et al*[36]. The forcefield is based on the universal forcefield (UFF)[21]. The total energy is based on sum of bonded and non-bonded interactions,

$$U_{tot} = U_{bond} + U_{angle} + U_{tors} + U_{inv} + U_{vdW} + U_{qq} + (U_{constr}) \qquad [7\text{-}1]$$

where U_{bond} is the linear bonded interactions, U_{angle} is the angle bending, U_{tors} is the torsion, U_{inv} is the inversion energy, U_{vdW} is the contribution from the van der Waals forces, modeled by a 12-6 Lennard Jones potential and U_{qq} is the electrostatic interactions. U_{constr} is the constraint potential for BDT molecule sliding along Au-Au direction in the case of on-bridge bonding. Previous work by Leng *et al*[19], BDT-2 Au (like BDT-3Au21, without the tail gold) complexes shows that the major bonding term is the bond stretching which largely controls the local bonding geometry.

The bond stretching is modeled by a harmonic potential,

$$U_{bond} = \frac{1}{2} k_{ij} \left(r - r_{ij} \right)^2 \qquad [7\text{-}2]$$

where k_{ij} is the bond stretching force constant, and r_{ij} is the equilibrium bond length. For the thiol bond, the newly parameterized bond stretching potential was used. The bond angle bending for atoms i and k bonded to atom j with angle θ, is modeled by

$$U_{angle} = \frac{k_{ijk}}{2\sin^2\theta_0}\left(\cos\theta - \cos\theta_0\right)^2 \qquad [7\text{-}3]$$

where k_{ijk} is the angle bending force constant, and θ_o is the equilibrium bond angle.

The torsion is the angle between the bonds between the atoms i, j and atoms k, l perpendicular to the bond j, k. The energy is calculated via

$$U_{tors} = V_\phi\left(1 - \cos n\phi_o \cos n\phi\right) \qquad [7\text{-}4]$$

where V_ϕ is the rotation barrier, n is the periodicity of the potential and ϕ_o is the equilibrium angle. The values used for the bond, angle and torsion potential are shown in Table 7-1. Inversion energy is for an atom i which is bonded to three other atoms, i, j, k. ω_{ijkl} is the angle between the bond between atoms i and l, and the plane defined by atoms i, j, k. The inversion energy is calculated by

$$U_\omega = k_{ijkl}\left(1 - \cos\omega_{ijkl}\right) \qquad [7\text{-}5]$$

The long-range columbic interactions where treated using the 3-D Ewald summation technique in 2-D, given by Yeh and Berkowitz[37].

Table 7-1. UFF Bonded Potential Parameters for BDT and gold

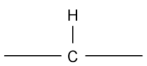
Bond stretching	r_{ij} (Å)	k_{ij} (kcal/mol Å^2)
C-C	1.379	925.83
C-H	1.085	708.61
C-S	1.800	588.45
S-H	1.429	438.3
Angle bending	θ_0 (^{0})	k_{ijk} (kcal/mol rad^2)
C-C-C	120	222.72
C-C-H	120	114.23
C-C-S	120	201.01
C-S-H	92.1	102.16
Torsion	ϕ_0 (^{0})	V_ϕ (kcal/mol)
X-C-C-X	180	13.474
X-C-S-X	90	3.9528
Inversion	ω_0(^{0})	k_{IJKL}
	0	6

Table 7-2 UFF Parameters for the Lennard-Jones 12-6

Atom type	D_{ii} (kCal/mol)	r_{ii} (Å)
C	0.105	3.851
H	0.044	2.886
S	0.274	4.035
Au	0.039	3.293

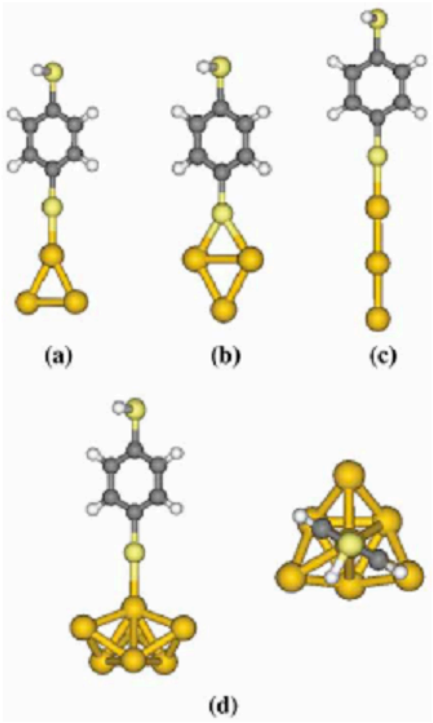


Figure 7-1 Molecular geometries for different BDT-*n*Au (*n* = 3 and 7) complexes. (a) 3Au12, (b) 3Au21, (c) 3Au111, (d) 7Au.

7.4 Results

Our current study only focuses on the bond stretching behavior between BDT and gold clusters. Further calculations for BDT-3Au complex show that the angle bending curves are very similar to those for BDT-2Au complexes. The deviations from equilibrium values of bond angle and torsion increase significantly due to intermolecular interactions of BDT SAMs on Au (111) surface[19].

The bond stretching curves

For BDT-3Au complexes, the Au-S distance and Au clusters were first optimized prior to DFT bond stretching energy calculations. During the optimization, the BDT molecule is kept rigid. For the 'on-top' bonding case (BDT-3Au12), the Au-S bond lengths from different DFT functionals are: 2.259 (LDA), 2.302 (PBE0), and 2.311Å (X3LYP), respectively. These values are comparable to other calculated results reported in the literature[15,17], as well as the results in our previous calculations for BDT-1Au complex[19]. The bottom and side Au-Au distances are significantly less than the bulk value (2.877 Å). These values vary from 2.576 ~ 2.674 Å, depending on the DFT functional used. This is also true for the Au-Au distances in BDT-3Au21 and BDT-3Au111 complexes, except the 'on-bridge' case (BDT-3Au21) where the upper Au-Au distance slightly increases to 2.759 (LDA) ~ 2.866 Å (X3LYP) due to the 2-fold coordination bonding between BDT and gold. The single Au-S bond length in this 'on-bridge' bonding case increases substantially to 2.430 (LDA) ~ 2.499 Å (X3LYP), which is consistent with the results for small thiolate-gold clusters (2.420 ~ 2.583 Å) using

Gaussian 98 program[16], but a little larger than those given by CPMD software package (2.352 ~ 2.392 Å)[17].

Figure 7-2 shows the on-top bond stretching curves for BDT-3Au12 complex from the three DFT functionals. Compared with the results for BDT-1Au complex[19], all the three DFT functionals give more consistent results with a little higher bonding energies. In contrast, the BDT-3Au111 stretching curves show a little more divergence (not shown here) with relatively low bonding energies (*Table 7-3*). This shows that BDT-3Au12 is a more favorable bonding configuration.

Table 7-3: BDT-*n* Au on-top bond stretching potentials. The units of parameters E_0, α and r_0 are in *kcal/mol*, $Å^{-1}$ and Å, respectively.

Complex	Parameter	LDA	PBE0	X3LYP	Average	Variation %
BDT-1Au	E_0	20.11	34.54	25.58	26.74	54.0
	α	2.157	2.242	2.505	2.301	15.1
	r_0[a]	2.309	2.253	2.257	2.273	2.47
BDT-3Au12	E_0	47.95	51.29	48.94	49.39	6.76
	α	1.823	1.667	1.671	1.720	9.07
	r_0	2.257	2.302	2.311	2.290	2.36
BDT-3Au111	E_0	22.74	14.34	12.49	16.52	62.0
	α	1.999	2.159	2.134	2.097	7.63
	r_0	2.347	2.402	2.434	2.394	3.63
BDT-7Au	E_0	21.53	22.58	18.20	20.77	16.03
	α	2.057	2.237	2.266	2.187	9.56
	r_0	2.358	2.341	2.369	2.356	1.19

Table 7-4: BDT-*n* Au on-bridge bond stretching potentials. The units of parameters E_0, α and r_0 are in *kcal/mol*, $Å^{-1}$ and Å, respectively. The parameters in parentheses are for the single bond Au-S Morse potentials.

Complex	Parameter	LDA	PBE0	X3LYP	Average	Variation %
BDT-2Au	E_0	20.99 (9.298)	74.18 (29.35)	75.25 (27.44)	56.8 (22.03)	95.5 (91.0)
	α	1.949 (2.664)	1.136 (1.611)	1.105 (1.626)	1.397	60.4
	r_0[a]	1.905 (2.387)	1.982 (2.449)	2.000 (2.462)	1.962	4.84
BDT-3Au21	E_0	64.10 (28.27)	65.63 (30.01)	63.25 (27.90)	64.33 (28.73)	3.70 (7.34)
	α	1.302 (1.787)	1.162 (1.618)	1.152 (1.594)	1.205	12.4
	r_0[a]	1.986 (2.399)	2.046 (2.456)	2.046 (2.472)	2.026	2.96

For BDT-7Au complex, the on-top bond-stretching curve is based on the optimized 7Au cluster. Figure 7-3 shows that the three DFT functionals also yield more consistent bonding curves compared with those for BDT-1Au complex. However, the bonding energy is comparably low. This indicates that the strength of BDT-Au chemical bonding depends more or less on the geometry and size of gold clusters. We fit the bond-stretching curves with a shifted Morse potential

$$E_{str} = E_0 e^{-\alpha(r-r_0)}\left(e^{-\alpha(r-r_0)} - 2\right) + E_0 \qquad [7\text{-}2]$$

where r is the distance between the sulfur atom and the gold atom, E_0 is the bonding energy, i.e. the energy well-depth, α is a parameter in Morse potential which control the shape and r_0 is the equilibrium bond length, i.e. the location of the minimum energy. These parameters for new BDT-nAu complexes, together with the results for BDT-1Au in previous work[19], are listed in Table 7-3. Comparing with the isotropic Morse potential for thiolate molecules on Au(111) surface[38], the current bonding energies between BDT molecule and Au clusters are still 1 ~ 4 times larger than the well depth of isotropic Morse potential (8.763 kcal/mol). The average values and total variations (defined as the difference between the largest and the smallest values divided by the average value) of the fitted force field parameters from the three DFT functionals are also show in Table 7-3. The largest variation is in the well-depth E_0, however, for BDT-3Au12 and BDT-7Au complexes, the three DFT functionals give more consistent results. Furthermore, the two hybrid functionals obtain similar binding energy, but the PBE0 consistently results in tighter bonding, shown by a deeper well depth, and smaller equilibrium bonding length.

Figure 7-4 shows the on-bridge bond stretching curves for BDT-3Au21 complex. Similar to the BDT-2Au complex[19], the 'bond' is denoted by X-S where X is the mid-

point between the upper 2 Au atoms (*Figure 7-1b*). Here we see that the three DFT functionals also give more consistent results compared with the BDT-2Au results. The parameters for the Morse potential in this case are shown in Table 7-4. By assuming that the bonding energy of individual BDT-Au bond is equal to the half of the total DFT energy, we refit the single Au-S bond energies and represent them in parentheses in the same table.

The Au-S bond lengths are already close to bonding length between thiolate bonded to the bridge or bridge-like sites on bulk gold surfaces. Published results from planewave *ab initio* calculations of Thiol binding distance are 2.55[11], 2.49~2.56[12,39] (depending on the coverage, where the bond length increases with higher percentage of surface coverage), and 2.5 Å[40,41]. These numbers are also consistent with other results for small thiolate-gold clusters (2.42 ~ 2.58 Å)[16]. From Tables 7-3 and 7-4 the average bonding energy data, a crude number of the single Au-S bond strength in BDT-*n*Au complex can be estimated as in the range of 10 ~ 50 kcal/mol, depending on the contact geometry between BDT molecules and Au clusters.

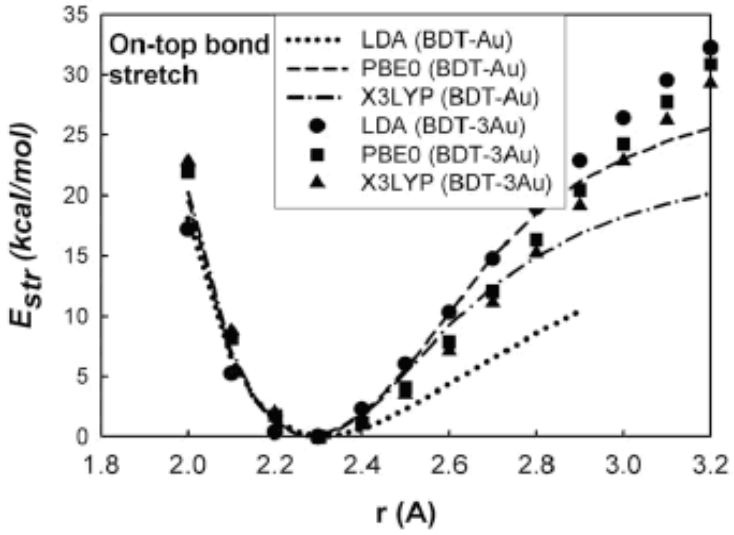


Figure 7-2: *Comparisons of bond stretching curves for BDT-1Au and BDT-3Au on-top bonding from the LDA, PBE0 and X3LYP DFT functionals.*

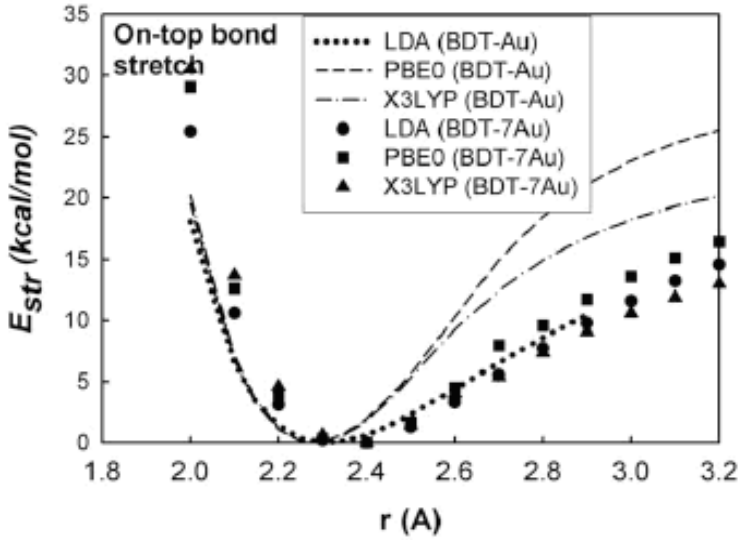


Figure 7-3: *Comparisons of bond stretching curves for BDT-1Au and BDT-7Au on-top bonding from the LDA, PBE0 and X3LYP DFT functionals.*

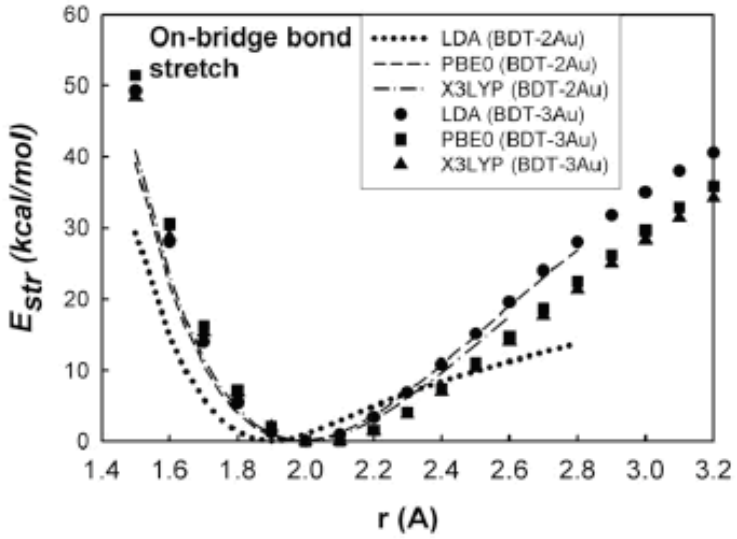


Figure 7-4: *Comparisons of bond stretching curves for BDT-2Au and BDT-21Au on-bridge bonding from the LDA, PBE0 and X3LYP DFT functionals.*

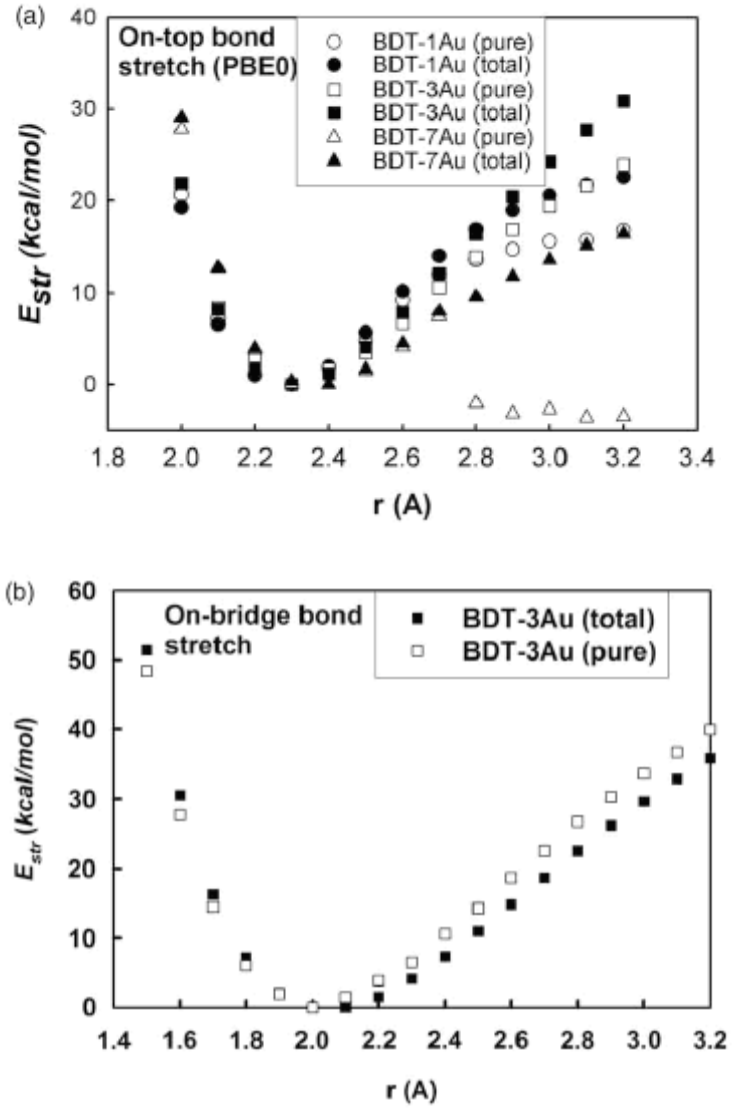


Figure 7-5: *The pure chemical bonding potentials derived from the total DFT energies subtracting the electrostatic interactions between nonbonded atoms. (a) Top, is for the BDT on top of the Au atoms, and (b) Below, is for the BDT between two Au atoms. These potentials for BDT-nAu (n = 1, 3 and 7) essentially show the same bonding behavior near the energy minima. Beyond the Au-S distance of 2.7~2.8Å the electrostatic interaction contributes significantly to the total DFT energy. This is particularly true for BDT-7Au complex.*

The pure chemical bonding

The bond stretching curves obtained are based on the total DFT energy calculation that in principle includes all the bonded and non-bonded interactions between BDT molecules and gold clusters. The net chemical bonding energy described in equation 7-1 is believed to be the main component of this total DFT energy, which in fact only involves the bonded atoms. To give a quantitative measurement of this 'pure' chemical bonding interaction, we should subtract the non-bonded component from the total DFT energy, i.e.

$$U_{DFT} = U_{Binding} + U_{Non-Binding} \qquad [7\text{-}3]$$

where $U_{Binding}$ is the energy associated with the chemical bond between the gold, ad sulfur and neighboring atoms. Following the UFF forcefield method[21] in which the interaction between atoms separated by more than two neighboring atoms are non-bonded and should be excluded from the total DFT energy (the 1-2 bond and 1-3 angle interaction exclusion). The assumption that 1-4 interaction is taken as a non-bonded term comes from our early investigation in which we note that torsion barrier in Au-S-C-C bond as being very low[19]. The $U_{Non-Binding}$ energy is due to the van der Waals interactions and electrostatic energy between the other atoms in the BDT molecules not involved in the chemical bonding with the Au clusters. Currently, DFT calculations does not include the van der Waals interaction properly[13], therefore the only non-bonded component needed to be subtracted from the total DFT energy is the electrostatic interaction. Figure 7-5 shows the total DFT (the same as those in figure 7-2 and 7-3) and pure bond stretching curves for the on-top bonding case. Here we only present the results given by PBE0 functional. The other two functionals yield very similar phenomenon. Very

interestingly, the total and pure bond-stretching curves overlap remarkably well until the Au-S bond length reaches ~2.7 Å, beyond which the two curves begin to split. This suggests that the net chemical bonding dominates the local bonding property. Specifically, there is a significant jump down to the negative value in the pure bond-stretching curve of BDT-7Au, indicating that at this point the BDT molecule and 7-Au cluster is nonbonded. The overall information we obtained from this figure is that for all the three BDT-*n*Au (n = 1, 3, 7) complexes the Au-S bond lengths are quite close to each other and the pure chemical bonding property near equilibrium bond length can be described by the total DFT energy variations. However, at larger distance the main contribution to the total DFT bonding energy is from the electrostatic interactions between the BDT molecules and Au clusters. We keep in mind that the splitting point is only an qualitatively estimate since the Mulliken charge calculations depend on the basis set used and usually the atomic charge is not uniquely defined[16,42]. This bond rupture point at ~2.7 Å seems consistent with the Car-Parrinello MD simulation result of the break between thiolate molecule and gold dimmer during the mechanically pulling of the molecule away from a stepped gold surface[43].

Bond stretching potentials on packing structure

To evaluate the effect of bond stretching potential on the local bonding geometry and BDT SAMs packing structure, we investigate two extreme cases where the well depths of Morse bond stretching potential take the maximum and minimum as shown in Table 7-1. These correspond to X3LYP BDT-3Au111 (E_0 = 12.49 kcal/mol) and PBE0 BDT-3Au12 (E_0 = 51.29 kcal/mol) cases. The equilibrium Au-S bond lengths are 2.434

and 2.302Å, respectively. The on-bridge bonding parameters are taken from the BDT-3Au21 configuration. Atomic partial charges for the bonded atoms are simply taken from the ∠Au-S-C = 180° configuration where the bonded Au and S charges are quite different from those in the most probable configurations[19]. It turns out that this difference in charge reassignment has almost no effect on the BDT packing structure. It was found that BDT molecules still keep the well-ordered herringbone structure as shown in Figure 7-6. The probability distributions of the on-top bond length for the two extreme cases are shown in figure 7-7. At room temperature (298K) the distribution curves for both 'weak' (X3LYP BDT-3Au111) and 'strong' (PBE0 BDT-3Au12) bonding potential are quite close to the corresponding Boltzmann distributions.

The packing structure of the BDT on the Au (111) surface, as predicted by MD simulations is shown in Figure 7-8. The other difference we note from the MD simulation from the variation in the bonding potentials is the distance distributions between the S ad-atoms and the first-layer gold atoms. The result from the strong bonding (PBE0 BDT-3Au12) still shows the first small peak at 2.3Å as observed by Leng *et al.*[19] that represents a small amount of BDT-Au binding at the on-top site. With the use of a weaker bonding potential, (X3LYP BDT-3Au111) the first small peak is totally disappeared. The most significant adsorption peak corresponds to the Au-S distance at ~2.5 Å, consistent with other theoretical investigations[40]. This indicates that when all the components of intermolecular interactions are considered, the Au-S bond length approaches the universal value of 2.5 Å. Obviously, most of these Au-S bond lengths come from the on-bridge or bridge-like bonding sites, but may also come from the on-top bonding sites if the actual on-top bonding energy is not so strong. The third small peak at

2.7-2.8 Å corresponds to the non-bonded Au-S distance between the bridge S atom and the third nearest Au atom around the hollow site. As we mentioned before, the total DFT bonding energies beyond this distance (*Figure 7-5*) are largely from the electrostatic intermolecular interactions.

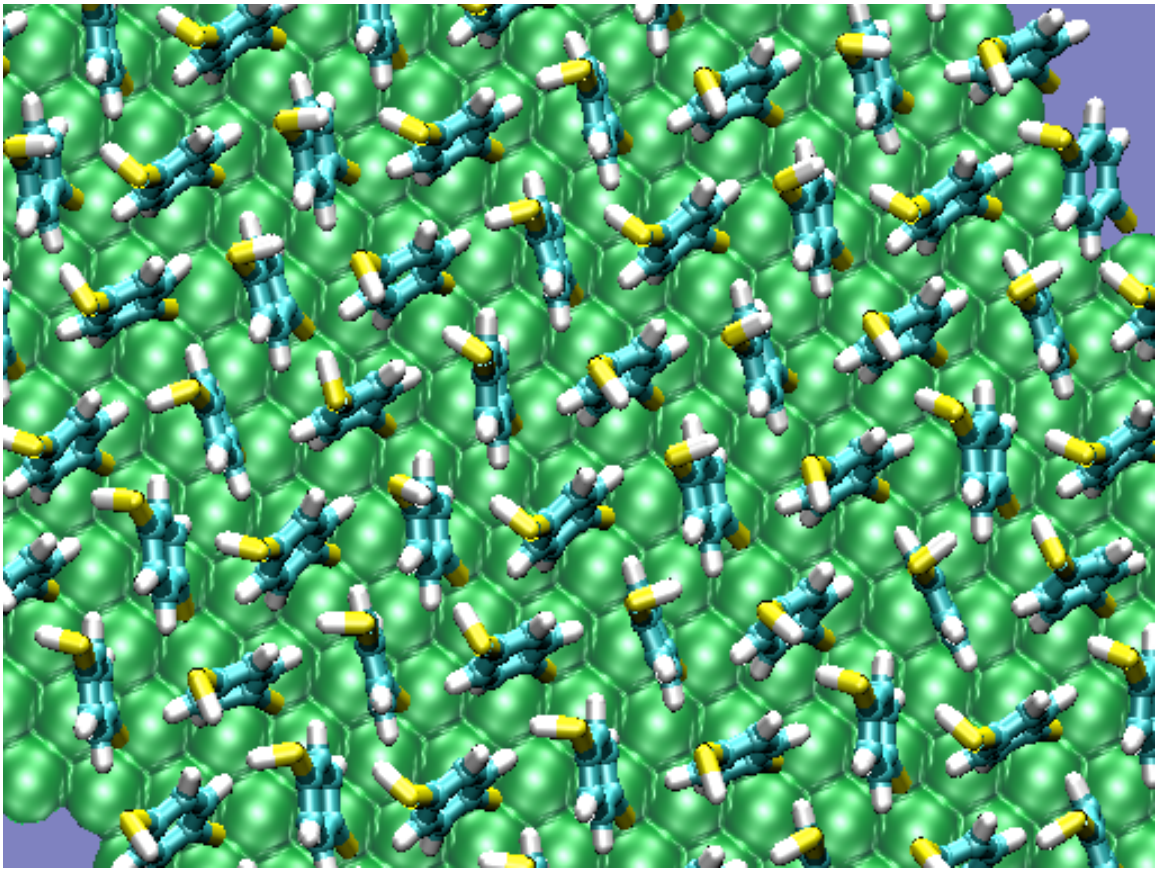


Figure 7-6: *Herringbone structure of the BDT molecules on the gold 111 surface predicted from molecular dynamic simulations.*

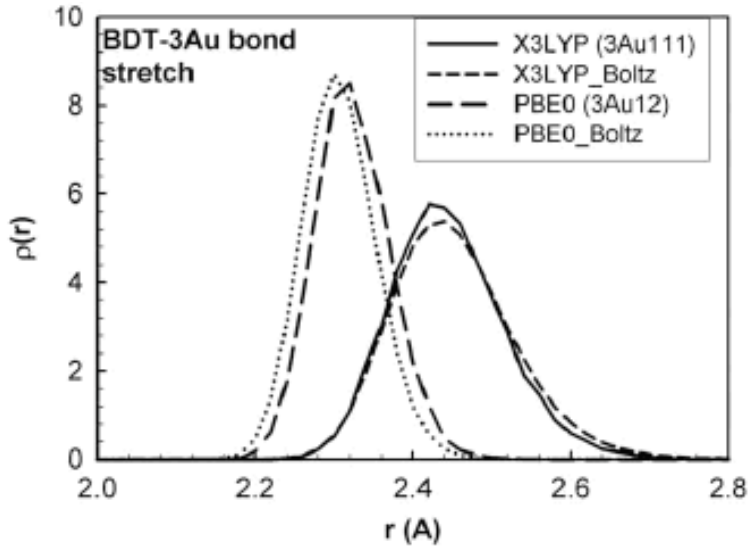


Figure 7-7: *The normalized probability distributions of Au-S bond stretching for the strong (PBE0-3Au12) and the weak (X3LYP-3Au111) bonding cases from MD simulations. The Boltzmann distributions are obtained based on individual bond-stretching potentials.*

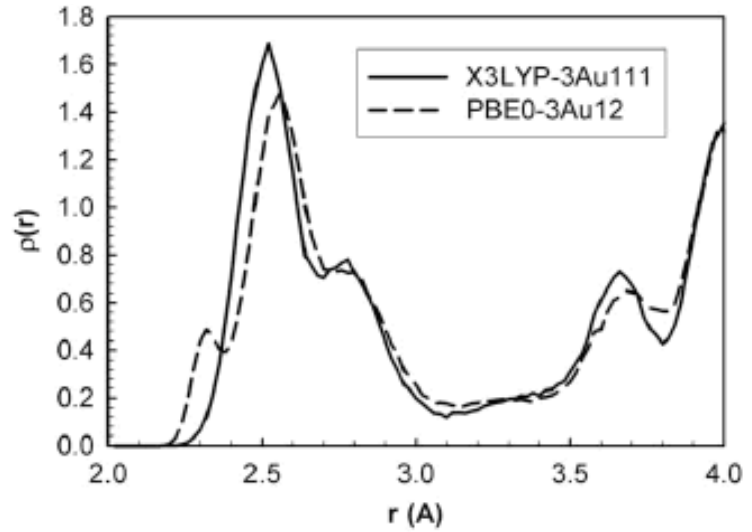


Figure 7-8: *The normalized probability distributions of the Au-S distance between the S ad-atoms and the first-layer gold atoms for the strong (PBE0-3Au12) and the weak (X3LYP-3Au111) bonding cases from MD simulations.*

7.5 Conclusions

We have investigated several BDT-*n*Au complexes to evaluate the Au-S chemical bonding properties. The bond stretching curve is fitted by the three-parameter Morse potential. We find that the bonding energy depends on the gold cluster size and geometries. For BDT-3Au complex, the bonding energy of Au-S single bond in the 'top-binding' case can vary from 12.49 kcal/mol for a Au mono-atomic chain configuration to 51.29 kcal/mol for a triangle Au cluster. However, for BDT-7Au complex, the Au-S single bond energy is approximately 20 kcal/mol and comparable to those in BDT-1Au complex. Molecular dynamics simulations show that this variation in the bond energy has much less effect on the BDT SAMs packing structure and even has limited influence on the local bonding geometry. The pure chemical bonding between the bonded atoms, i.e., the 1 or 2 bonded Au atoms and the bonded S and C atoms in BDT molecule has also been derived. Within the Au-S bonding distance, the net bonding potential overlaps with the total bonding curve. This suggests that the pure chemical bonding is dominant near local minimum and also validates our original suggestion that the BDT-Au interaction can be represented by the local chemical bonding which only involves 1-2 Au atoms and the far field non-bonded interaction described by the UFF force parameters. We have also found that the charge distribution among bonded atoms is not critical to the BDT SAMs structure. The current studies take a further step towards the development of a consistent force field for the large-scale molecular simulations of self-assembly of thiolate molecules on Au (111) surface.

References

[1] Y. S. Leng, P.J. Dyer, P.S. Krstic, R. J. Harrison, P. T. Cummings, Molecular Physics **105** (2-3), 293 (2007).

[2] A. Nitzan and M. A. Ratner, Science **300** (5624), 1384 (2003).

[3] A. Aviram and M. A. Ratner, Chemical Physics Letters **29** (2), 277 (1974).

[4] M. A. Reed, C. Zhou, C. J. Muller, T. P. Burgin, and J. M. Tour, Science (Washington, D. C.) **278** (5336), 252 (1997).

[5] C. Joachim, J. K. Gimzewski, and A. Aviram, Nature (London) **408** (6812), 541 (2000).

[6] J. Chen, M. A. Reed, A. M. Rawlett, and J. M. Tour, Science (Washington, D. C.) **286** (5444), 1550 (1999).

[7] B. Q. Xu and N. J. J. Tao, Science **301** (5637), 1221 (2003).

[8] J. M. Seminario, A. G. Zacarias, and J. M. Tour, Journal of Physical Chemistry A **103** (39), 7883 (1999).

[9] M. Di Ventra, S. T. Pantelides, and N. D. Lang, Physical Review Letters **84** (5), 979 (2000).

[10] Y. Q. Xue, S. Datta, and M. A. Ratner, Journal of Chemical Physics **115** (9), 4292 (2001).

[11] J. Gottschalck and B. Hammer, Journal of Chemical Physics **116** (2), 784 (2002).

[12] Y. Yourdshahyan and A. M. Rappe, Journal of Chemical Physics **117** (2), 825 (2002).

[13] J. Nara, S. Higai, Y. Morikawa, and T. Ohno, Journal of Chemical Physics **120** (14), 6705 (2004).

[14] H. Gronbeck, A. Curioni, and W. Andreoni, Journal of the American Chemical Society **122** (16), 3839 (2000).

[15] D. Kruger, H. Fuchs, R. Rousseau, D. Marx, and M. Parrinello, Journal of Chemical Physics **115** (10), 4776 (2001).

[16] H. Basch and M. A. Ratner, Journal of Chemical Physics **119** (22), 11926 (2003).

[17] M. Konopka, R. Rousseau, I. Stich, and D. Marx, Journal of the American Chemical Society **126** (38), 12103 (2004).

[18] S. Letardi and F. Cleri, Journal of Chemical Physics **120** (21), 10062 (2004).

[19] Y. S. Leng, P. S. Krstic, J. C. Wells, P. T. Cummings, and D. J. Dean, Journal of Chemical Physics **122** (24), 244721 (2005).

[20] X. C. Zhao, Y. S. Leng, and P. T. Cummings, Langmuir **22** (9), 4116 (2006).

[21] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, Journal of the American Chemical Society **114** (25), 10024 (1992).

[22] R. S. Mulliken, Journal of Chemical Physics **23**, 1833 (1955).

[23] R. A. Kendall, E. Apra, D. E. Bernholdt, E. J. Bylaska, M. Dupuis, G. I. Fann, R. J. Harrison, J. L. Ju, J. A. Nichols, J. Nieplocha, T. P. Straatsma, T. L. Windus, and A. T. Wong, Computer Physics Communications **128** (1-2), 260 (2000).

[24] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, Journal of Chemical Physics **72** (1), 650 (1980).

[25] R. B. Ross, J. M. Powers, T. Atashroo, W. C. Ermler, L. A. Lajohn, and P. A. Christiansen, Journal of Chemical Physics **93** (9), 6654 (1990).

[26] S. H. Ke, H. U. Baranger, and W. T. Yang, Journal of Chemical Physics **123** (11) (2005).

[27] X. Y. Xiao, B. Q. Xu, and N. J. Tao, Nano Letters **4** (2), 267 (2004).

[28] F. Cleri and V. Rosato, Physical Review B **48** (1), 22 (1993).

[29] L. Xiao, B. Tollberg, X. K. Hu, and L. C. Wang, Journal of Chemical Physics **124** (11) (2006).

[30] S. H. Vosko, L. Wilk, and M. Nusair, Canadian Journal of Physics **58** (8), 1200 (1980).

[31] C. Adamo and V. Barone, Journal of Chemical Physics **110** (13), 6158 (1999).

[32] X. Xu and W. A. Goddard, Proceedings of the National Academy of Sciences of the United States of America **101** (9), 2673 (2004).

[33] C. Lee, W. Yang, R. G. Parr, Physical Review B **37** (2), 785 (1988).

[34] A. D. Becke, Physical Review A **38** (6), 3098 (1988).

[35] J. P. Perdew, K. Burke, M. Ernzerhof, Physical Review Letters **77** (18), 3865 (1996).

[36] B. J. K. Y. Leng, P. T. Cummings, Journal of Physical Chemistry **107**, 11940 (2003).

[37] I. C. Yeh, M. L. Berkowitz, Journal of Chemical Physics **111**, 3155 (1999).

[38] R. Mahaffy, R. Bhatia, and B. J. Garrison, Journal of Physical Chemistry B **101** (5), 771 (1997).

[39] M. C. Vargas, P. Giannozzi, A. Selloni, and G. Scoles, Journal of Physical Chemistry B **105** (39), 9509 (2001).

[40] D. Fischer, A. Curioni, and W. Andreoni, Langmuir **19** (9), 3567 (2003).

[41] V. De Renzi, R. Di Felice, D. Marchetto, R. Biagi, U. del Pennino, and A. Selloni, Journal of Physical Chemistry B **108** (1), 16 (2004).

[42] Y. Karzazi, X. Crispin, O. Kwon, J. L. Bredas, and J. Cornil, Chemical Physics Letters **387** (4-6), 502 (2004).

[43] D. Kruger, H. Fuchs, R. Rousseau, D. Marx, and M. Parrinello, Physical Review Letters **89** (18) (2002).

CHAPTER VIII

CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

In this work we have investigated the effect of including explicit polarization for the solute, water and ion forcefields. We have also investigated intermolecular and intramolecular forcefield development from *ab initio* calculations. Starting with chapter III, the importance of the inclusion of polarization into molecular forcefields was shown where using an explicitly polarization in the model for small hydrophobic solutes enabled us to reproduce the experimental excess chemical potential of the solutes in pure water for a range of temperatures. It also showed that polarization did not affect the gradient of the excess chemical potential or, in other words, it did not significantly affect the qualitative agreement with experimental values.

In addition, this work shows that for pure systems where the effect of polarizability plays an insignificant role in their bulk fluid properties due to weak electric field strength, it is still necessary to consider the effect of polarization when these molecules are solvated in water, which has a strong electric field. This leads us to suggest that in the development of transferable molecular potentials, inclusion with explicit polarization is essential.

The investigation of the role of polarizability is continued in chapter IV, where we show that the effects of polarizability on the excess chemical potential and the salting-out effect with the investigation of the effect of the explicit polarization on the solute forcefield in an aqueous solutions containing charged particles. Accounting for

polarization explicitly in small hydrophobic solute models reduced the excess chemical potential of the solute in water, which improves the agreement with experiment. However, the addition of explicit polarizability to the solute model had very little effect on the salting-out effects, i.e. the decrease in the solubility of the solute from the addition of salts into the aqueous solution remained over-estimated. We have shown that the ion-solute forcefields are responsible for the inability of these potentials to accurately reproduce the correct salting-out effects, meaning that exclusion of polarizability is not the main cause as has been suggested by other workers.

We note that while the modern reparameterizations of the TIP4P model (i.e. TIP4P/2005 and TIP4P/Ew) provide significantly better predictions than TIP3P of the excess chemical potential in pure water shown in chapter III, for the case of a charged co-solvent TIP3P obtains a better trend. This is due to TIP3P underestimating the solution density, where TIP4P/2005 and TIP4P/Ew are in good agreement with experiment. The importance of obtaining the correct density of solvent solution was thus reinforced and, additionally, it was noted that simulations at incorrect densities can give rise to a correct numerical results but via the wrong mechanism due to a cancellations of errors. This could give rise to incorrect validation of results in many simulations if these system properties are left unchecked, including incorrect interpretations of the causes of salting-out.

New solute-ion interaction parameters have been fitting to *ab initio* calculations, for the purpose of obtaining better agreement with experimental results. The new interaction parameters obtained from these predictions work just as well, if not better, than those derived from the mixing rules of the pure species for the reproduction of the excess chemical potential.

We have shown that explicit polarizability is important for the correct reproduction of the solubility of solutes in solution, the next step in the use of polarizable solutions. Before investigations into polarizable solutions can be conducted, polarizable water models need to be developed and shown to be able to accurately reproduce the physical system. From comparing molecular dynamics simulations of GCPM water to *ab initio* CPMD simulations of the water, in Chapter V, it was shown that polarization and, in particular, its inclusion in the Gaussian distributed polarizable model for water, was able to reproduce the total dipole moment and the distribution for the total dipole moment within the accuracy of the *ab initio* Car-Parrinello simulations. We also saw that the method used to calculate the induced dipole moment from the Car-Parrinello simulation has a large effect on the value of the mean total dipole moment.

Chapter VI continued with the development of polarizable models, looking into the ion-water forcefields. With the empirically parameterization of the ion-water interactions from ion-water cluster energy, for used with the GCPM water model, it was shown that with the inclusion of explicit polarizability, we achieved good agreement with *ab initio* simulations for the behavior of ion in water clusters, and ion in bulk solution. We have also been shown that we can get the correct kosmotropic and chaotropes behavior with regard to bulk or surface solvation. We found that within the first hydration shell, the induced water dipole moment is influenced by the ion, and that the magnitude of the induced water dipole moment is related to the size and charge of the ion. However, the use of GCPM as the water model still shows the same trends as other classical models of water with regards to the ion-water orientation.

Following on from the used of *ab initio* calculation for the development of intermolecular forcefields, we have used *ab initio* DFT calculation for the development of

intramolecular bonding potentials. The bonding potentials used to describe the surface chemistry between organic molecules and metallic surfaces is dependant on the shape and size of the clusters used to approximate the surface. The variance in the bonding potentials has minimal effect on the structuring of a self-assembled monolayer of BDT on the gold surface was insensitivity to the variation in the possible bonding potentials.

8.2 Recommendations for future work

As is the nature of research, in attempting to understand one problem several others are brought to light. Unfortunately, there is never sufficient time to investigate all the things we might like and as such many of the interesting questions raised during this work remain unanswered. Here we outline several of the most interesting ideas we believe would make productive area for future work.

In summary, in this work we showed that the explicit inclusion of polarizability in molecular models may have a dramatic effect on the predicted properties of small non-polar solutes in aqueous solution. It follows from this that explicit polarizability will also play an important role in the solvation of more complex biological molecules. Given the available computational resources and their continued advanced in this field, the development of such models poses an appealing challenge.

Parameterization of molecular forcefields is not a trivial activity, especially as experimental data is limited for unique and complex molecules. New methods of producing transferable forcefields are required and should be developed in the future to prevent having to parameterize new interaction potentials for every possible combination of molecules. The development of interaction parameters directly from *ab initio*

calculations has great promise and has the ability to avoid mixing rules, which have already been shown to be problematic.

The transferability of forcefields makes it possible for the forcefield to be used in a range of different situations. In the work of Chapter III, we have shown that explicit polarizability is a key element in the development of transferable potentials.

The next step following Chapter III and IV is to investigate the effect of explicit polarization on the hydrophobic solute in polarizable aqueous solutions, and should be the focus of future work. Following on from this work, a polarizable water model with polarizable ions and solute should continue to be developed. Included is the use the better methods of handling the long range electrostatic interactions. This will allow the ability to simulation systems away from pure water and infinite dilution, i.e. simulating concentrated solution, and multi-component mixtures.

From the inability of the GCPM-ion forcefields to correct reproduce the ion-water orientation in the first hydration layer in Chapter VI, the development of a classical water model with some form of acknowledgement of the electrostatic potential minima would be required to reproduce the *ab initio* results for the ion-water orientation. We believed this is important for reproduction of the correct behavior of water in biological situations, especially due to the importance of the water's orientation for the folding of proteins. The development of a water model which is fitted to the quadrupole moment, with a two negative charge points, separated in the y direction, has started to be developed. This work should be continued in order to develop a water model with the correct ion-water structure as well the correct water-water structure.

The development of bonding potentials for organic molecules bonded to metallic surfaces from *ab initio* calculations of organic molecules with metallic clusters suffers

from convergence of the binding energies as the metallic cluster varies in size and geometry. Even though, for the situation described in Chapter VII, where the variations in the pure binding energy and the resulting forcefields had minimal effect on the packing structure of the organic molecules on the gold surface. Work has been done on the structural and electronic properties of gold clusters, and how the structure and average atomic energies vary with clusters size, but an understanding of how this affects bonding potentials has not been investigated. Convergences of the binding energies or at least an understanding of why there are not convergent for small clusters, would be of great interest and should be studied further.

APPENDIX A

MOLECULAR DYNAMICS ENERGY, FORCE EQUATIONS USING THE WOLF METHOD FOR LONG RANGE INTERACTIONS OVER GAUSSIAN CHARGE AND DIPOLES

A1 The Wolf Method

Even though the electrostatics energy for a pair interaction decay to the r^{-1}, when the number different interaction increase r^3. As a result, the total amount of electrostatic interaction increase to the r^2, mean a simply truncation is erroneous. In the Wolf method[1], each electrostatic charges have an another equal but opposite neutralizing charge, or an image charge of opposite sign projected onto a truncated sphere (surface of the volume surrounding the charge i). Simply put, the Wolf method is a fancy type of truncated and shifting potential, so the $U_{wolf} = 0$ at $r = Rc$.

$$U_{shifted} = U(r) - U(R_c) \qquad \text{[A1-1]}$$

The problem is that for molecular dynamic simulations, the forces are also required for the equation of motion. Using the Wolf method directly, there is a discontinuity at $r = Rc$, making the energy conversation a problem. To solve this problem, Fennel[2] also shifted the force, so that the force also equals zero at $r = Rc$. A force and energy shifted potential is shown below

$$U_{Force\,Shifted} = U(r) - U(R_c) - \left(\frac{dU(r)}{dr}\right)_{r=R_c}(r - R_c) \qquad \text{[A1-2]}$$

where, $U(r)$ is the energy potential for charge-charge, charge-dipole and dipole-dipole interactions. Note the potential is not always only a function of distance, as the direction of the dipole moment are also important for charge-dipole interactions but for clarity the dipole direction has been omitted.

A2 Derivation of the Wolf Method for a system point charges

The electrostatic potential for two point charges can be separated into two terms,

$$U_{QQ} = \frac{1}{2}\sum_{i=1}^{N}\sum_{i\neq j=1}^{N}\frac{q_i q_j}{r_{ij}} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j\, erfc(\alpha r_{ij})}{r_{ij}} + \frac{q_i q_j\, erf(\alpha r_{ij})}{r_{ij}} \qquad [A2\text{-}1]$$

as

$$1 = erfc(x) + erf(x) \qquad [A2\text{-}2]$$

The first term $erfc(x)$ can be made to be the dominant contribution with the right value for α, the Wolf damping parameter. The second term $erf(x)$ is normally small, but can be large if a large value of α is chosen. In order to exclude the second term, a correction is used; the self interaction term $(i = j)$ is subtracted from the first term and added to the second term. The second becomes small, and is considered a systemic error in using a damp potential. As a self interaction has zero distance, we take the limit of when $r_{ij} \to 0$. The self correction term is therefore,

$$\lim_{r_{ij}\to 0}\left[\frac{1}{2}\sum_{i=1}^{N}\frac{q_i q_i}{r_{ij}}erf(\alpha r_{ij})\right] = \frac{\alpha}{\sqrt{\pi}}\sum_{i=1}^{N}q_i^2 \qquad [A2\text{-}3]$$

The electrostatic interaction energy would therefore become.

$$U_{QQ} = \frac{1}{2}\sum_{i=1}^{N}\sum_{i\neq j=1}^{N}\frac{q_i q_j\, erfc(\alpha r_{ij})}{r_{ij}} - \frac{\alpha}{\sqrt{\pi}}\sum_{i=1}^{N} q_i^2 + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j erf(\alpha r_{ij})}{r_{ij}} \qquad \text{[A2-4]}$$

Third term is identical to the reciprocal-space energy in the Ewald sum. This will be ignored, and used as a systematic error, from replacing the Coulombic potential by a damped potential. As the Wolf method is a truncated, Coulombic potential, we much remove the energy of the neutralization sphere. The total Coulombic energy using the Wolf method is shown below,

$$U_{qq}^{Total} = U_{qq} - U_{qq}^{neutral} \qquad \text{[A2-5]}$$

where the $U_{qq}^{neutral}$ represents the charges neutralizing the charged volume. The neutralization term is calculated by taking the is limit of U_{qq} as r_{ij} tends to the cutoff R_c.

$$U_{qq}^{neutral} = \lim_{r_{ij}\to R_c}\left(\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{i\neq j=1\\ r_{ij}<R_c}}^{N}\frac{q_i q_j\, erfc(\alpha r_{ij})}{r_{ij}} + \frac{1}{2}\sum_{i=1}^{N}\frac{q_i q_i erfc(\alpha r_{ij})}{r_{ij}}\right) \qquad \text{[A2-6]}$$

$$U_{qq}^{neutral} = \left(\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{i\neq j=1\\ r_{ij}<R_c}}^{N}\frac{q_i q_j\, erfc(\alpha R_c)}{R_c} + \frac{erfc(\alpha R_c)}{2R_c}\sum_{i=1}^{N} q_i^2\right) \qquad \text{[A2-7]}$$

So the total electrostatic interaction energy using A2-5, A2-4 and A2-7 becomes,

$$U_{qq}^{Total} = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{i\neq j=1\\ r_{ij}<R_c}}^{N} q_i q_j\left(\frac{erfc(\alpha r_{ij})}{r_{ij}} - \frac{erfc(\alpha R_c)}{R_c}\right) + \left(\frac{-erfc(\alpha R_c)}{2R_c} - \frac{\alpha}{\sqrt{\pi}}\right)\sum_{i=1}^{N} q_i^2 \qquad \text{[A2-7]}$$

A self interaction term can be extracted to ease molecular dynamic calculations as this term is constant, independent of the system molecular configuration and density.

$$U_{self}(r) = -\left(\frac{erfc(\alpha R_c)}{2R_c} + \frac{\alpha}{\sqrt{\pi}}\right)\sum_{i=1}^{N} q_i^2 \qquad \text{[A2-8]}$$

A3 Derivation of the Wolf Method for Gaussian charges, using shift force potential

If the electrostatic charges are represented by a Gaussian distribution instead of a point charge, for example the GCPM. A Gaussian charge is represented by,

$$\rho_{q_i}(r) = \frac{q_i}{\left(2\pi\Omega_i^2\right)^{\frac{3}{2}}} \exp\left(\frac{-|r - r_i|^2}{2\Omega_i^2}\right) \qquad \text{[A3-1]}$$

where $\rho_{q_i}(r)$ is the charge density, Ω_i is the width of the Gaussian distribution (i.e. standard deviation of a normal curve), r_i is the center of the charge site, and q_i is the magnitude of the charge. The electrostatic interaction energy $U_{Q_iQ_j}$ between two charged Gaussian distributions is, (equation [A3-2])

$$U_{Q_iQ_j} = \iint \frac{q_i}{\left(2\pi\Omega_i^2\right)^{\frac{3}{2}}} \exp\left(\frac{-|r_i' - r_i|^2}{2\Omega_i^2}\right) \frac{q_j}{\left(2\pi\Omega_j^2\right)^{\frac{3}{2}}} \exp\left(\frac{-|r_j' - r_j|^2}{2\Omega_j^2}\right) \frac{1}{|(r_i' - r_i) - (r_j' - r_j)|} dr_i' dr_j'$$

meaning that at every point of distribution i, at a distance r_i' from the center of charge r_i, interacts with every point of distribution j, at a distance r_j' from the center of charge r_j. This equation can be simplified down a useful form,

$$U_{Q_iQ_j} = \frac{q_i q_j}{r_{ij}} erf\left(\frac{r_{ij}}{\sqrt{2\left(\Omega_i^2 + \Omega_j^2\right)}}\right) \qquad \text{[A3-3]}$$

In order to use the Wolf method of a systems contained Gaussian charges, all the electrostatic potentials have to be derived, using a similar method used in section A2. In equation A1-2, the correct force potential is required for the calculation the potential. To get around this, as the derivative of the potential at the cutoff is a constant, it can be solved for at the end. The starting columbic potential for a system of Gaussian charges, using a shift force potential is

$$U_{QQ} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j \left(\frac{erf(\beta r_{ij})}{r_{ij}} + (Const) r_{ji} \right) \qquad \text{[A3-4a]}$$

$$\beta = 1/\sqrt{2(\Omega_i^2 + \Omega_j^2)} \qquad \text{[A3-4b]}$$

We define the β, the combined width of the distributions, for clarity in the following equations. The constant term will be solved later. In order to minimize the self correction for the *erf* term (as it equation A2-6), the self-interaction term is calculated by taking the limit of $r_{ij} \to 0$. Unlike the point charge case, as there is an error function due the Gaussian charges, so this term goes to zero.

$$\lim_{r_{ij}\to 0}\left(\frac{erf(\alpha r_{ij}) erf(\beta r_{ij})}{r_{ij}} \right) = 0 \qquad \text{[A3-5]}$$

The force shifted term can just be included fully into the first energy term. The resulting columbic energy (analog to equation A2-4) can be shown below.

$$U_{QQ} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j \left(\frac{erf(\beta r_{ij}) erfc(\alpha r_{ij})}{r_{ij}} - r_{ij}(Const) \right) \qquad \text{[A3-6]}$$

where the *Const* is defined as,

$$Const = \frac{d}{dr}\left(\frac{erf(\beta r_{ij}) erfc(\alpha r_{ij})}{r_{ij}} \right) \qquad \text{[A3-7a]}$$

$$Const = \frac{erfc(\alpha R_c)}{R_c}\frac{2\beta}{\sqrt{\pi}}\exp\left(-\beta^2 R_c^2\right) - erf(\beta R_c)\left(\frac{erfc(\alpha R_c)}{R_c^2} + \frac{2\alpha}{R_c\sqrt{\pi}}\exp\left(-\alpha^2 R_c^2\right)\right) \quad \text{[A3-7b]}$$

The total interaction energy is calculated by equation A2-4, where the $U_{qq}^{neutral}$ is limit of U_{qq} as r_{ij} tends to R_c, at the cutoff. Where the potential is shifted, it is the equivalent to the neutralization of the i, j interactions, but not i,i self interactions which will be calculated independently to the shifting.

$$U_{qq}^{neutral} = \lim_{r_{ij}\to R_c}\left(\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{i\neq j=1 \\ r_{ij}<R_c}}^{N} q_i q_j\left(\frac{erf(\beta r_{ij})erfc(\alpha r_{ij})}{r_{ij}} - r_{ij}(Const)\right) + U_{qq}^{Self}\right) \quad \text{[A3-7]}$$

$$U_{qq}^{Self}(R_c) = \left(\frac{erf(\beta R_c)erfc(\alpha R_c)}{2R_c}\right)\sum_{i=1}^{N} q_i^2 \quad \text{[A3-8]}$$

So the total electrostatic interaction energy becomes, for Gaussian charges using the Wolf method becomes,

$$U_{QQ}^{Total} = \frac{1}{2}\sum_{i=1}^{N}\sum_{i\neq j=1}^{N} q_i q_j\left(\frac{erf(\beta r_{ij})erfc(\alpha r_{ij})}{r_{ij}} - \frac{erf(\beta R_c)erfc(\alpha R_c)}{R_c} - (r_{ij} - R_c)(Const)\right) - \frac{1}{2}\left(\frac{erf(\beta R_c)erfc(\alpha R_c)}{R_c}\right)\sum_{i=1}^{N} q_i^2 \quad \text{[A3-9]}$$

A4 Gaussian charges and dipoles

The formula for the charge density of a Gaussian charge, or a s-Gaussian[3] is

$$\rho_S(r) = q_i\left(\frac{\beta_i}{\sqrt{\pi}}\right)^3 \exp\left(-r^2\beta^2\right) \qquad \text{[A3-1]}$$

where $\beta_i = \frac{1}{\sqrt{2}\Omega_i}$, Ω_i is the width of the Gaussian distribution of the charge, and q_i is the magnitude of the charge. Integration of equation 3A, over all space, will obtain the magnitude q_i. Figure A4-1 shows what a typical s-Gaussian looks like.

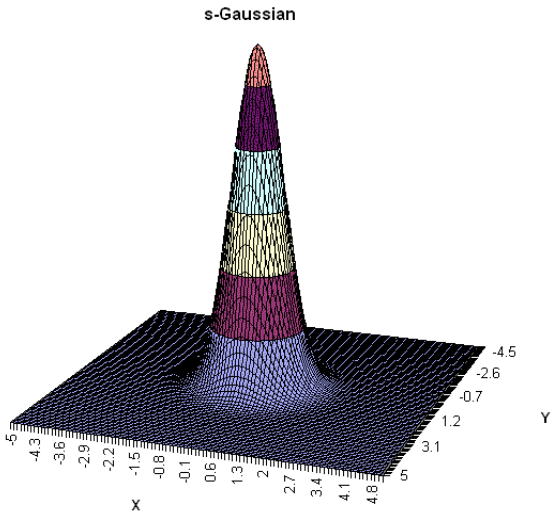


Figure A4-1, Charge density for a spherical Gaussian of charge 1.22e, and Ω_i of 1.4 Å

A5 Gaussian Dipole

The formula for the charge density of a Gaussian dipole, or a p-Gaussian[3] is

$$\rho_p(r) = \bar{\mu}_i \nabla_R \left(\frac{\beta_i}{\sqrt{\pi}} \right)^3 \exp\left(-r^2 \beta_i^{\,2} \right) \qquad \text{[A5-1]}$$

Where the resulting charge densities can seen below in Figure A5-1, where the distance between the two peaks is $2\Omega_i$. Integration over the Gaussian Dipole would result in a zero charge as a dipole is neutral.

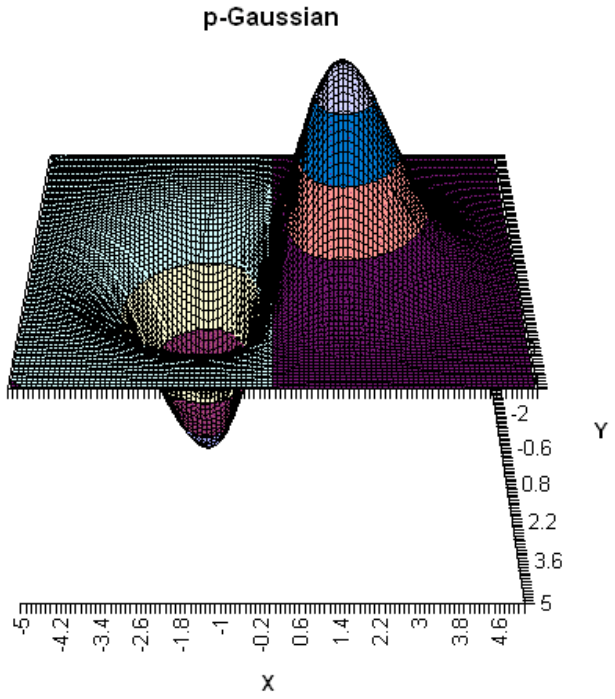


Figure A5-1, Charge density for a p-Gaussian of dipole 5D and Ω_i of 1.4 Å, where the dipole is only in the x direction

A6 Charge-Charge interactions (s-s type)

The uncorrected charge-charge interaction for long range interactions can be calculated from

$$U_{QQ}(r) = \frac{q_i q_j erf(\beta r) erfc(\alpha r)}{r} \qquad \text{[A6-1]}$$

where $\beta = \frac{1}{\sqrt{2(\Omega_i^2 + \Omega_j^2)}}$, and α is the Wolf damping parameter, set to $2/R_c$, in accordance to the work of Demontis *et al.*[4] and Avendaño and Gil-Villegas[5]

So the Wolf charge-charge interaction less self correction from equation [A3-9] is below in equation [A6-2]

$$U_{QQ}^{Total} = \frac{1}{2}\sum_{i=1}^{N}\sum_{i\neq j=1}^{N} q_i q_j \left(\frac{erf(\beta r_{ij}) erfc(\alpha r_{ij})}{r_{ij}} - \frac{erf(\beta R_c) erfc(\alpha R_c)}{R_c} - (r_{ij} - R_c)(Const) \right) \qquad \text{[A6-2a]}$$

$$Const = \frac{erfc(\alpha R_c)}{R_c}\frac{2\beta}{\sqrt{\pi}}\exp(-\beta^2 R_c^2) - erf(\beta R_c)\left(\frac{erfc(\alpha R_c)}{R_c^2} + \frac{2\alpha}{R_c\sqrt{\pi}}\exp(-\alpha^2 R_c^2)\right) \qquad \text{[A6-2b]}$$

The electrostatic force due to a potential using the Wolf long-range correction is

$$F_{wolf}(r) = -\frac{dU_{wolf}(r)}{dr} \qquad \text{[A6-3]}$$

Equation [A6-4]

$$F_{QQ}^{Wolf}(r) = -q_i q_j \left[\frac{erfc(\alpha r)}{r}\left(\frac{2\beta}{\sqrt{\pi}}\exp(-\beta^2 r^2)\right) - erf(\beta r)\left(\frac{erfc(\alpha r)}{r^2} + \frac{2\alpha}{r\sqrt{\pi}}\exp(-\alpha^2 r^2)\right) - \left(\frac{erfc(\alpha R_c)}{R_c}\frac{2\beta}{\sqrt{\pi}}\exp(-\beta^2 R_c^2) - erf(\beta R_c)\left(\frac{erfc(\alpha R_c)}{R_c^2} + \frac{2\alpha}{R_c\sqrt{\pi}}\exp(-\alpha^2 R_c^2)\right)\right)\right]$$

A7 Charge-Dipole interactions (s-p type)

The uncorrected Charge-Dipole interaction potential for long range interactions can be calculated from,

$$U_{QD} = q_i\left(\bar{\mu}_j \bullet \hat{r}_{ij}\right)\frac{\beta}{r}\left(\frac{erf\left(\beta r\right)}{\beta r} - \frac{2}{\sqrt{\pi}}\exp\left(-\beta^2 r^2\right)\right) \qquad \text{[A7-1]}$$

Simply adding the Wolf damping factor $erfc\left(\alpha r\right)$, as the self interactions are again zero

$$U_{QD} = q_i\left(\bar{\mu}_j \bullet \hat{r}_{ij}\right)\frac{\beta}{r}\left(\frac{erf\left(\beta r\right)}{\beta r} - \frac{2}{\sqrt{\pi}}\exp\left(-\beta^2 r^2\right)\right)erfc\left(\alpha r\right) \qquad \text{[A7-2]}$$

Using the Wolf regime will obtain equation A7-3 for the potential energy and equation A7-4 for the force. There is no self neutralization term as a dipole is overall neutral (see section A5).

$$U_{QD}^{Wolf} = q_i\left(\bar{\mu}_j \bullet \hat{r}_{ij}\right)\begin{bmatrix}\left(\frac{erf\left(\beta r\right)}{r^2} - \frac{2}{r\sqrt{\pi}}\beta\exp\left(-\beta^2 r^2\right)\right)erfc\left(\alpha r\right) - \left(\frac{erf\left(\beta R_c\right)}{R_c^{\,2}} - \frac{2}{R_c\sqrt{\pi}}\beta\exp\left(-\beta^2 R_c^{\,2}\right)\right)erfc\left(\alpha R_c\right) \\ -\left(Const\right)\left(r - R_c\right)\end{bmatrix}$$

$$F_{QD}^{Wolf} = -q_i\left(\bar{\mu}_j \bullet \hat{r}_{ij}\right)\begin{bmatrix}\left(\frac{-2erf\left(\beta r\right)}{r^3} + \frac{4\exp\left(-\beta^2 r^2\right)}{\sqrt{\pi}}\left(\frac{\beta}{r^2} + \beta^3\right)\right)erfc\left(\alpha r\right) + \\ \left(\frac{erf\left(\beta r\right)}{r^2} - \frac{2\beta}{r\sqrt{\pi}}\exp\left(-\beta^2 r^2\right)\right)\left(\frac{-2\alpha}{\sqrt{\pi}}\exp\left(-\alpha^2 r^2\right)\right) \\ -\left(Const\right)\end{bmatrix}$$

$$Const = \begin{pmatrix}\left(\frac{-2erf\left(\beta Rc\right)}{R_c^{\,3}} + \frac{4}{\sqrt{\pi}}\exp\left(-\beta^2 R_c^{\,2}\right)\left(\frac{\beta}{R_c^2} + \beta^3\right)\right)erfc\left(\alpha R_c\right) + \\ \left(\frac{erf\left(\beta\, R_c\right)}{R_c^{\,2}} - \frac{2\beta}{R_c\sqrt{\pi}}\exp\left(-\beta^2 R_c^{\,2}\right)\right)\left(\frac{-2\alpha}{\sqrt{\pi}}\exp\left(-\alpha^2 R_c^{\,2}\right)\right)\end{pmatrix}$$

A8 Dipole-Dipole Interactions (p-p type)

The uncorrected dipole-dipole interaction potential for long range interactions can be calculated from,

$$U_{DD} = \bar{\mu}_i \bar{\mu}_j \frac{\beta}{r^2}\left[\frac{erf(\beta r)}{\beta r} - \frac{2}{\sqrt{\pi}}\exp\left(-\beta^2 r^2\right)\right] - \left(\bar{\mu}_i \bullet \hat{r}_{ij}\right)\left(\bar{\mu}_j \bullet \hat{r}_{ij}\right)\left[3\frac{erf(\beta r)}{r^3} - \frac{2\beta}{\sqrt{\pi}}\left(\frac{3}{r^2} + 2\beta^2\right)\exp\left(-\beta^2 r^2\right)\right] \quad [8A\text{-}1]$$

Excluding the Wolf correction was excludes as dipole-dipole potential interactions decay to r^{-6}.^{6}. Therefore the force,

$$F_{DD} = -\frac{dU_{DD}}{dr} = -\bar{\mu}_i \bar{\mu}_j\left[-3\frac{erf(\beta r)}{\beta r^4} + \frac{2\beta}{\sqrt{\pi}}\left(\frac{3}{r^3} + \frac{2\beta^2}{r}\right)\exp\left(-\beta^2 r^2\right)\right] + \left(\bar{\mu}_i \bullet \hat{r}_{ij}\right)\left(\bar{\mu}_j \bullet \hat{r}_{ij}\right)\left[-9\frac{erf(\beta r)}{r^4} + \frac{2\beta}{\sqrt{\pi}}\left(\frac{9}{r^3} + \frac{6\beta^2}{r} + 4\beta^4 r\right)\exp\left(-\beta^2 r^2\right)\right] \quad [8A\text{-}2]$$

A9 Electrostatic Field

The electrostatic field for a point charge '*i*' acting at point 'j' is

$$\vec{E} = \hat{r}\frac{\dfrac{-dU(q_i,q_j)}{dr}}{q_j} = \frac{\vec{F}(q_i,q_j)}{q_j} = \frac{q_i}{r^2} \qquad [9A\text{-}1]$$

where $\vec{E}$ is the electrostatic field generated from the point charge i at a point j. Likewise, for a electrostatic field from a Gaussian charge i action at a spherical Gaussian j is

$$\vec{E} = \frac{\dfrac{-dU(q_i,q_j)}{dr}}{\int_V q_j dv = q_j} = \frac{q_i erf(\beta_{ij} r)}{r^2} \qquad [9A\text{-}2]$$

Where the β_{ij} term contain both the Gaussian distribution information of the charge generating the field i, and the Gaussian being effected by the field j. At the limiting case, where Gaussian j has an infinite width, i.e. $\beta_j << 1$, there is no net electric field on Gaussian j, as one would expect.

As the U_{QQ} using the Wolf method for long range electrostatic correction, is shown by equation 9A-3

$$\vec{E}_{QQ} = \frac{\vec{F}_{QQ}}{q_j}$$

$$= -q_i \hat{r} \left[\begin{array}{l} \frac{erfc(\alpha r)}{r}\left(\frac{2\beta}{\sqrt{\pi}}\exp\left(-\beta^2 r^2\right)\right) - erf(\beta r)\left(\frac{erfc(\alpha r)}{r^2} + \frac{2\alpha}{r\sqrt{\pi}}\exp\left(-\alpha^2 r^2\right)\right) - \\ \left(\frac{erfc(\alpha R_c)}{R_c}\frac{2\beta}{\sqrt{\pi}}\exp\left(-\beta^2 R_c^2\right) - erf(\beta R_c)\left(\frac{erfc(\alpha R_c)}{R_c^2} + \frac{2\alpha}{R_c\sqrt{\pi}}\exp\left(-\alpha^2 R_c^2\right)\right)\right) \end{array} \right]$$

This is used for the electric field between a s-Gaussian charge to the site of another s-Gaussian. For the electric field between a s-Gaussian on to a p-Gaussian, i.e. Gaussian charge to an dipole, equation 9A-5 is shown below.

$$U_{QD}^{Wolf} = E_{Qd}^{Wolf} \bullet \vec{\mu}_i,\ U_{QD}^{Wolf} = \mu_i \bullet [f(r)],\ E_{QD}^{Wolf} = f(r)$$

$$\vec{E}_{QD}^{Wolf} = q_i\ \hat{r}_{ij} \left[\begin{array}{l} \left(\frac{erf(\beta r)}{r^2} - \frac{2}{r\sqrt{\pi}}\beta\exp\left(-\beta^2 r^2\right)\right) erfc(\alpha r) - \left(\frac{erf(\beta R_c)}{R_c{}^2} - \frac{2}{R_c\sqrt{\pi}}\beta\exp\left(-\beta^2 R_c{}^2\right)\right) erfc(\alpha R_c) \\ -(Const)(r - R_c) \end{array} \right]$$

This is basis on the fact that we know the direction of the field form a single charge (radial), thus created a 1-diminional problem, and removes the dot product.

For the electrostatic field generated from a p-Gaussian on to another p-Gaussian, i.e. dipole-dipole interactions. T_{ij} is a tensors of the dipole positions and Gaussian distributions. The electric field can be calculated from[7]

$$\vec{E}_i^{DD} = \vec{T}_{ij}\vec{\mu}_j \qquad [9A\text{-}6]$$

$$\bar{T}_{ij} = \frac{1}{r_{ij}^3}\left(3f\,\hat{r}_{ij}\,\hat{r}_{ij} - gI\right) \qquad [9A\text{-}7]$$

where,

$$f = erf\left(\beta r_{ij}\right) - \frac{2}{\sqrt{\pi}}\left[\beta r + \frac{2}{3}\beta^3 r^3\right]\exp\left(-\beta^2 r^2\right)$$
$$g = erf\left(\beta r_{ij}\right) - \frac{2}{\sqrt{\pi}}\beta r \exp\left(-\beta^2 r^2\right) \qquad [9A\text{-}8]$$

It can be shown that with equation 8A-1, and 9A-6,

$$U_{ij}^{DD} = -\frac{1}{2}\bar{\mu}_i \bar{T}_{ij} \bar{\mu}_j = -\frac{1}{2}\bar{\mu}_i \bar{E}_i^{induced} \qquad [9A\text{-}9]$$

resulting in an quicker calculation of the induced dipole-induced dipole interaction energy.

References

1 D. Wolf, P. Keblinski. S.R. Phillpot, J. Eggebrecht, Journal of Chemical Physics **110**, 8254 (1999).

2 Christopher J. Fennell, J. Daniel Gezelter, Journal of Chemical Physics **124**, 234104 (2006).

3 D. Elking, T. Darden, and R. J. Woods, J. Comp. Chem **28** (7), 1261 (2007).

4 P. Demontis, S. Spanu, G. B. Suffritti, Journal of Chemical Physics **114** (18), 7980 (2001).

5 C. Avendano, A. Gil-Villegas, Mol Phys **104** (9), 1475 (2006).

6 J. Prausnitz, *Molecular Thermodynamics of Fluid-Phase Equilibria*, 3rd ed. (Prentice Hall, Upper Saddle River, 1999).

7 P. Paricaud, M. Predota, A. A. Chialvo, P. T. Cummings, Journal of Chemical Physics **122** (24), 4511 (2005).

APPENDIX B

PARAMETERIZATION OF THE SOLUTE-ION INTERACTION PARAMETERS FROM MP2 CALCULATIONS

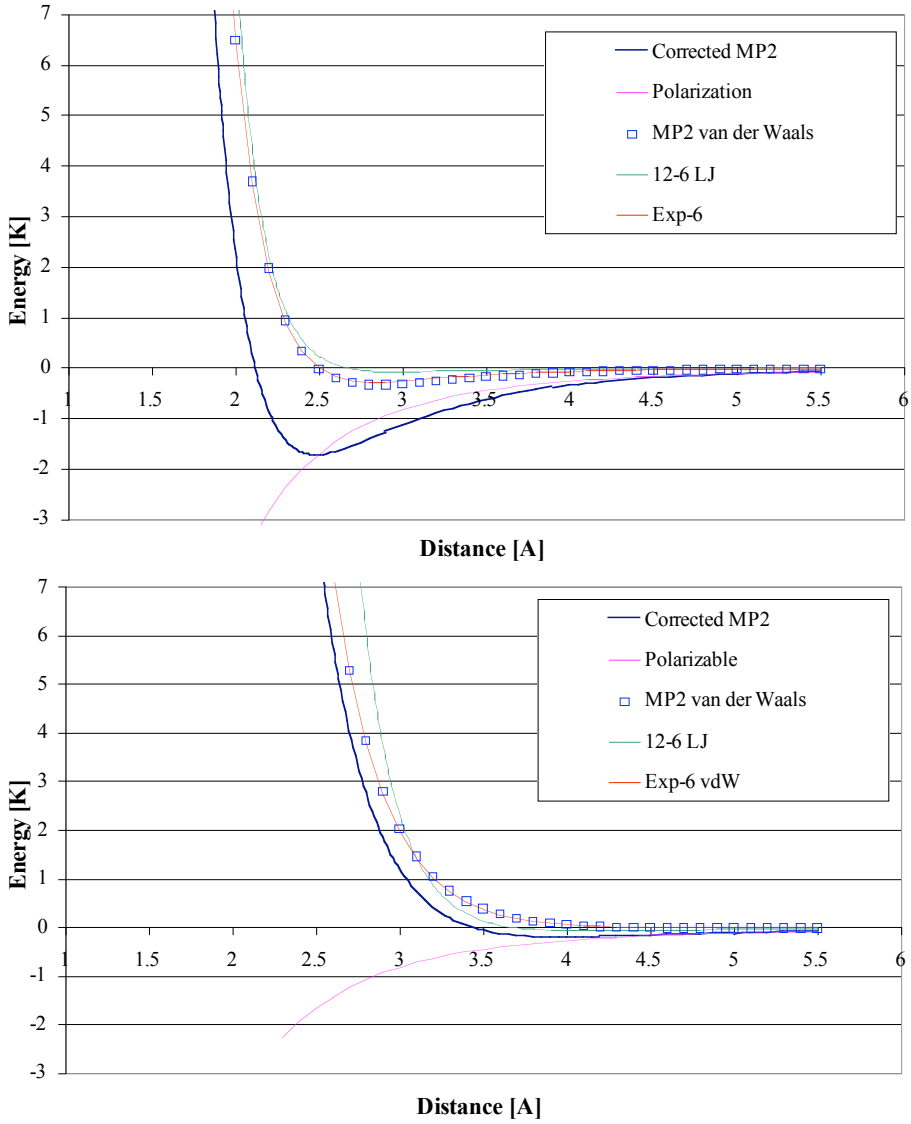


Figure B-1: *MP2 energy and its classical components for the dimer interaction between neon and the ions sodium (above) and chloride (below) The blue line is the corrected MP2 energy, The pink line is the polarization energy. The residue is the MP2 van der Waals energy (squares). The exponential six is fitted to the MP2 vdW (red line). For comparison, the LJ parameters from the pure species via mixing rules are green.*

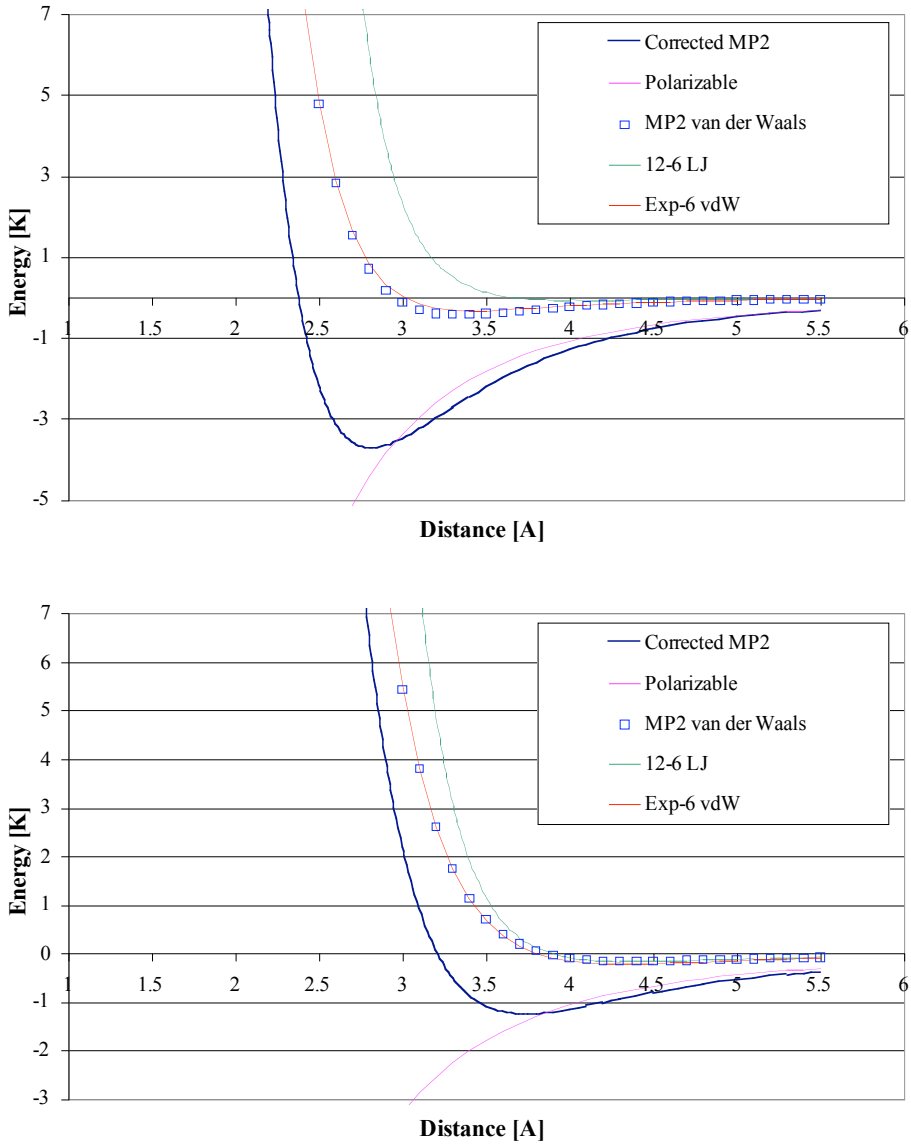
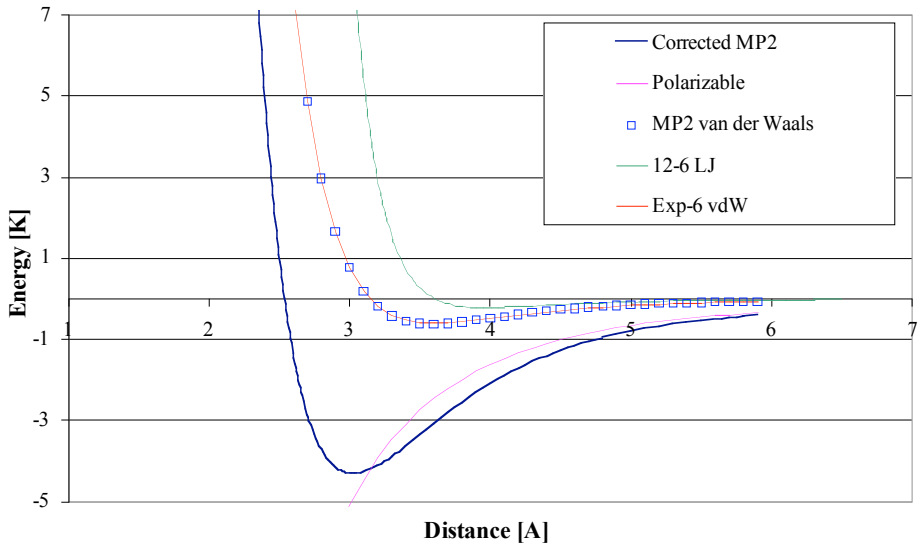


Figure B-2: *MP2 energy and its classical components for the dimer interaction between argon and the ions sodium (above) and chloride (below) The blue line is the corrected MP2 energy, The pink line is the polarization energy. The residue is the MP2 van der Waals energy (squares). The exponential six is fitted to the MP2 vdW (red line). For comparison, the LJ parameters from the pure species via mixing rules are green.*



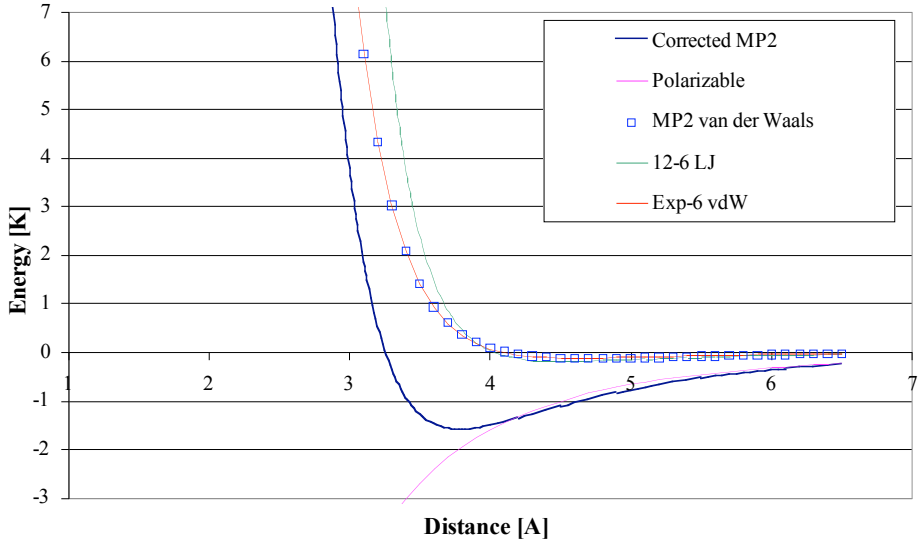


Figure B-3: *MP2 energy and its classical components for the dimer interaction between krypton and the ions sodium (above) and chloride (below) The Blue line is the corrected MP2 energy, The pink line is the polarization energy. The residue is the MP2 van der Waals energy (squares). The exponential six is fitted to the MP2 vdW (red line). For comparison, the LJ parameters from the pure species via mixing rules are green.*

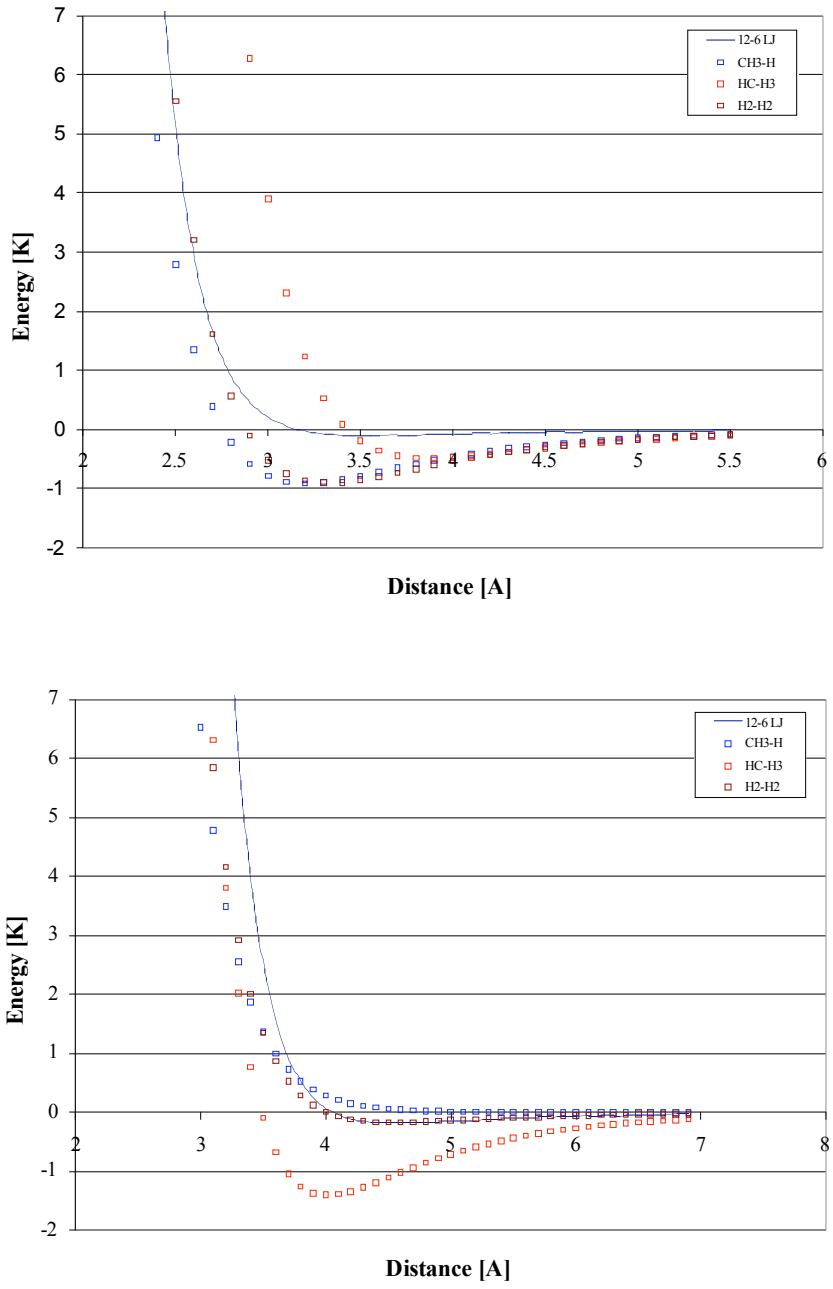


Figure B-4: *Fitted exponential six, short range interactions parameterized to the calculated MP2 vdW energy for the dimer interaction between methane and the ions sodium (above) and chloride (below) for three different configurations. Hydrogen towards the ion (Blue squares), ion in the void between three hydrogens (red squares) and ion in the void between two hydrogens (brown squares). For comparison, the LJ parameters from the pure species via mixing rules are green.*